Optimal Filtering
IV: Wiener Filters and Linear Prediction

ECE416 Adaptive Algorithms

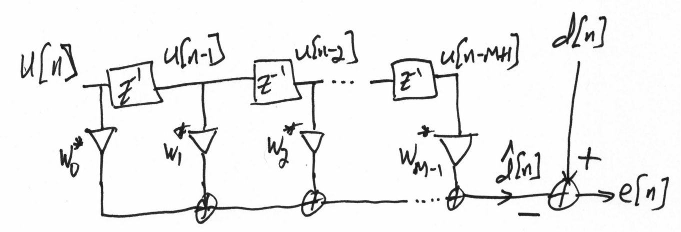
Prof. Fred L. Fontaine

Department of Electrical Engineering
The Cooper Union

Spring 2022

Wiener Filters

Let $u[n], d[n]$ be jointly WSS. A *Wiener filter* is a causal FIR that produces the optimal estimate of present d from a (finite) block of past and present u, in the mean-square sense.



Wiener Filter

$$\hat{d}[n] = \sum_{m=0}^{M-1} w_m^* u[n-m] = \mathbf{w}^H \mathbf{u}_M[n]$$
$$e[n] = d[n] - \hat{d}[n]$$

With $R_M = E\left(\mathbf{u}_M[m]\,\mathbf{u}_M^H[n]\right)$ and $\mathbf{p} = E\left(\mathbf{u}_M[n]\,d^*[n]\right)$, we know the optimal filter is:

$$\mathbf{w}_0 = R_M^{-1}\mathbf{p}$$

Wiener Filter

This is optimal in the *minimum mean-square error (MMSE) sense:*

$$\min E\left(|e[n]|^2\right)$$

With $J(\mathbf{w}) = E\left(|e[n]|^2\right)$, we have from previous study of linear regression problem that:

$$\begin{aligned} J(\mathbf{w}) &= J_{\min} + (\mathbf{w} - \mathbf{w}_0)^H R (\mathbf{w} - \mathbf{w}_0) \\ J_{\min} &= \sigma_d^2 - \mathbf{w}_0^H R \mathbf{w}_0 \end{aligned}$$

Wiener Filter

We can interpret the Wiener filter as computing the projection of $d[n]$ onto:

$$\mathcal{U}_M[n] = span\{u[n], u[n-1], \cdots, u[n-M+1]\}$$

Linear Prediction

We can formally define *linear prediction* as the problem of projecting $u[n+1]$ onto $\mathcal{U}_M[n]$, i.e., it is essentially a special case of a Wiener filter with $d[n] = u[n+1]$. The problem is when we write:

$$e[n] = u[n+1] - \sum_{m=0}^{M-1} w_m^* u[n-m]$$

this has the form of a *noncausal* filter.

Therefore, we will modify the statement of the problem slightly, and instead project $u[n]$ onto $\mathcal{U}_M[n-1]$, i.e., use $\{u[n-1], \cdots, u[n-M]\}$ to estimate $u[n]$.

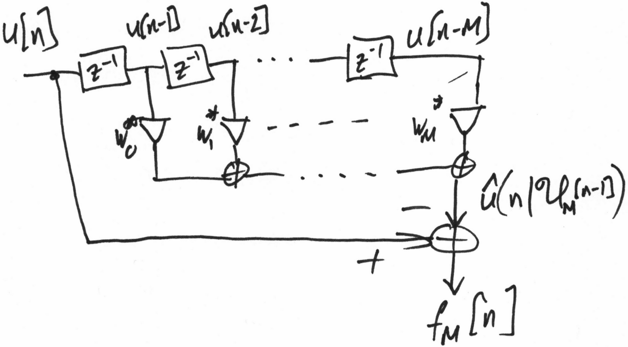
This is called *forward linear prediction.*

Forward Linear Prediction

$$
\begin{aligned}
f_M[n] &= u[n] - \hat{u}\left(n \middle| \mathcal{U}_M[n-1]\right) \\
&= u[n] - \mathbf{w}_M^H \mathbf{u}_M[n-1]
\end{aligned}
$$

The subscript $\mathbf{w}_M$ reminds us of the parameter M (we will eventually consider multiple values simultaneously).

We call $f_M[n]$ the *forward prediction error* signal, and the mapping $u \longrightarrow f_M$, which is a causal FIR filter of *order* M (recall the order is the number of delays; the length is $M+1$), is called the *forward prediction error filter (FPEF)* of order M.



u[n]
u[n-1]
u[n-2]
u[n-M]
z^-1
z^-1
z^-1
w_0^*
w_1^*
w_M^*
û(n|𝒰_M[n-1])
−
+
f_M[n]

Forward Linear Prediction

It will be convenient to replace $\mathbf{w}_M \in \mathbb{C}^M$ with an augement vector $\mathbf{a}_M \in \mathbb{C}^{M+1}$:

$$\mathbf{a}_M = \begin{array}{c} \left[\begin{array}{c} 1 \\ -\mathbf{w}_M \end{array} \right] \end{array} \begin{array}{l} \updownarrow 1 \\ \updownarrow M \end{array}$$

Then:

$$f_M[n] = \mathbf{a}_M^H \mathbf{u}_{M+1}[n]$$

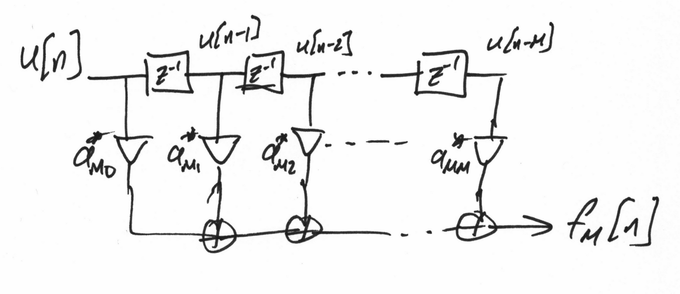
Forward Linear Prediction

Let us denote the elements of $\mathbf{a}_M$ as follows:

$$\mathbf{a}_M = \begin{bmatrix} a_{M,0} \\ a_{M,1} \\ \vdots \\ a_{M,M} \end{bmatrix}$$

and note that $a_{M,0} = 1$ *always.*

FPEF



Forward Linear Prediction

Define:

$$\mathbf{r}_M = E\left(\mathbf{u}_M\left[n-1\right]u^*\left[n\right]\right) = \begin{bmatrix} r_{-1} \\ r_{-2} \\ \vdots \\ r_{-M} \end{bmatrix}$$

where we use subscripts for convenience. Then the FPEF is given by:

$$\mathbf{w}_M = R_M^{-1}\mathbf{r}_M$$

Forward Linear Prediction

The *power* of the prediction error is:

$$P_M = E\left(|f_M[n]|^2\right) = \mathbf{a}_M^H R_{M+1} \mathbf{a}_M$$

As a special case, for 0^{th} order linear prediction, we use 0 to estimate $u[n]$, and this gives $\mathbf{a}_0 = a_{0,0} = 1, f_0[n] = u[n]$, and:

$$P_0 = r[0]$$

where $r[m]$ is the correlation function of u.

Forward Linear Prediction

Given the M^{th} order FPEF, $\mathbf{a}_M$, the $(M+1)^{st}$ order FPEF allows us to use not only $u[n-1], \cdots, u[n-M]$ to estimate $u[n]$, but also $u[n-M-1]$. If we just set $a_{M+1,m} = a_{M,m}$, $0 \leq m \leq M$, and $a_{M+1,M+1} = 0$, we get the same predicted value and the power would be P_M. We could perhaps do better than that, but the point is:

$$P_{M+1} \leq P_M$$

On the other hand, all $P_M \geq 0$ (in fact, if R_{M+1} is not singular, $P_M > 0$).

Forward Linear Prediction

We have a sequence:

$$r[0] = P_0 \geq P_1 \geq \cdots \geq P_{M+1} \geq P_M \geq \cdots \geq 0$$

Therefore the sequence converges!

$$P_\infty = \lim_{M \to \infty} P_M \geq 0$$

This alone is not enough to ensure the FPEFs converge: in what sense would we mean convergence, anyway?

Forward Linear Prediction

Theorem: u is a regular process iff $P_\infty > 0$, in which case the FPEFs converge to the whitening filter.

Informally, for large enough M, $f_M[n]$ is "approximately" white, and approximates the *innovations* signal of u. **We will discuss why $f_M[n]$ is "approximately white" later!**

In fact, we get:

$$P_\infty = e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S(\omega) d\omega}$$

Forward Linear Prediction

If $f_M[n]$ has PSD $S_{fM}(\omega)$, and:

$$A_M(z) = \sum_{m=0}^{M} a_{M,m}^* z^{-m}$$

is the transfer function of the FPEF, then the PSD of u is:

$$S(\omega) = \frac{S_{fM}(\omega)}{|A_M(\omega)|^2}$$

Forward Linear Prediction

Note that:

$$P_M = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{fM}(\omega)\, d\omega$$

If we treat f_M as (approximately) white, then $S_{fM}(\omega) \approx P_M$ and we have:

$$S(\omega) \approx \frac{P_M}{|A_M(\omega)|^2}$$

In other words, we can interpret the process of obtaining the FPEF as a *parametric* estimate of the PSD of u (the parameters being $\mathbf{a}_M$ and P_M).

Forward Linear Prediction

Indeed, if $f_M[n]$ is *exactly* white, then it is the innovations of u and the FPEF is the whitening filter, which means u is AR(M). Therefore, this is sometimes called the *AR power spectral estimate*, S_{AR}.

Although we will not prove it, this is equivalent to the solution of another optimization problem. obtained via what is called the *maximum entropy method (MEM).*

Maximum Entropy Spectral Estimation

Find PSD $S(\omega)$ that maximizes:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S(\omega)\, d\omega$$

subject to the constraints:

$$r[m] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)\, e^{j\omega m} d\omega$$

for prescribed $r[m]$, $0 \leq m \leq M$, such that the Hermitian Toeplitz matrix formed from them is pd. The solution is denoted S_{MEM}, and it turns out:

$$S_{MEM} = S_{AR}$$

Backward Linear Prediction

Forward linear prediction is to esimate $u[n]$ using $\mathcal{U}_M[n-1] = span\{u[n-1], \cdots, u[n-M]\}$. The idea of *backward* linear prediction would be to estimate $u[n]$ using future values, but to articulate it in a causal form:

Estimate $u[n-M]$ **using**

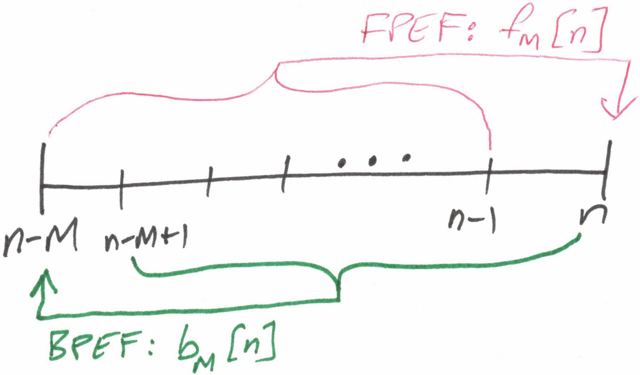
$$\mathcal{U}_M[n] = span\{u[n], u[n-1], \cdots, u[n-M]\}$$

The *backward prediction error filter (BPEF)* computes:

$$\begin{aligned} b_M[n] &= u[n-M] - \hat{u}(n-M|\mathcal{U}_M[n]) \\ &= \sum_{m=0}^{M} c^*_{Mm} u[n-m] \end{aligned}$$

where $c_{MM} = 1$, and $\mathbf{c}_M$ is the vector of coefficients of the BPEF. Note the FPEF and BPEF are both FIR filters.

Forward and Backward Linear Prediction



Setting Up the Normal Equations

The *normal equations* establish the orthogonality conditions of the FPEF/BPEF.

The *augmented normal equations* add one extra condition, which is the form we will be focusing on.

We first review the Toeplitz struction of the correlation matrix.

Toeplitz Correlation Matrices

Let R_M denote the $M \times M$ correlation matrix of $u_M[n]$. Using $r_m = r[m]$, for example for $M = 5$ we get:

$$R_5 = \begin{bmatrix} r_0 & r_1 & r_2 & r_3 & r_4 \\ r_{-1} & r_0 & r_1 & r_2 & r_3 \\ r_{-2} & r_{-1} & r_0 & r_1 & r_2 \\ r_{-3} & r_{-2} & r_{-1} & r_0 & r_1 \\ r_{-4} & r_{-3} & r_{-2} & r_{-1} & r_0 \end{bmatrix}$$

The matrix has *Toeplitz symmetry:* it is constant on the diagonals. Since $r_m = r^*_{-m}$, it is Hermitian as well. It is also pd (unless there is a degeneracy).

Toeplitz Correlation Matrices

The Toeplitz symmetry leads to interesting "embeddings." For example, R_4 can be seen inside R_5.

Let us define the $\mathbf{r}_M$ vector as:

$$\mathbf{r}_M = \begin{bmatrix} r_{-1} \\ r_{-2} \\ \vdots \\ r_{-M} \end{bmatrix} = \begin{bmatrix} r_1^* \\ r_2^* \\ r_3^* \\ r_4^* \end{bmatrix}$$

Also, for a column vector $\mathbf{x}$, define $\mathbf{x}^B$ by flipping its components upside-down:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \longrightarrow \mathbf{x}^B = \begin{bmatrix} x_N \\ \vdots \\ x_2 \\ x_1 \end{bmatrix}$$

Finally $\mathbf{x}^*$ means just conjugating the entries (no transpose).

Toeplitz Correlation Matrices

$$R_{M+1} = \begin{bmatrix} R_M & \mathbf{r}_M^{B*} \\ \mathbf{r}_M^{BT} & r_0 \end{bmatrix} = \begin{bmatrix} r_0 & \mathbf{r}_M^{H} \\ \mathbf{r}_M & R_M \end{bmatrix}$$

In the first case, we have pulled out the last row and column, in the second case the first row and column.

Normal Equations

The Wiener filter equation is $R\mathbf{w} = \mathbf{p}$.

For the M^{th} order FPEF, $\mathbf{p} = E\left(u_M\left[n-1\right]u^*\left[n\right]\right) = \mathbf{r}_M$.

For the M^{th} order BPEF, $\mathbf{p} = E\left(u_M\left[n\right]u^*\left[n-M\right]\right) = \mathbf{r}_M^{B*}$. Then the *normal equations* are:

$$\begin{aligned} R_M\mathbf{w}_{FPEF} &= \mathbf{r}_M \\ R_M\mathbf{w}_{BPEF} &= \mathbf{r}_M^{B*} \end{aligned}$$

Rewriting the Normal Equations

The coefficient vectors for the FPEF and BPEF are:

$$\mathbf{a}_M = \begin{bmatrix} 1 \\ -\mathbf{w}_{FPEF} \end{bmatrix} \begin{matrix} \updownarrow 1 \\ \updownarrow M \end{matrix} \qquad \mathbf{c}_M = \begin{bmatrix} -\mathbf{w}_{BPEF} \\ 1 \end{bmatrix} \begin{matrix} \updownarrow M \\ \updownarrow 1 \end{matrix}$$

We can rewrite the normal equations for the FPEF as:

$$\underset{\substack{\leftrightarrow \quad \leftrightarrow \\ 1 \quad\;\; M}}{\begin{bmatrix} \mathbf{r}_M & R_M \end{bmatrix}} \mathbf{a}_M = \mathbf{0} \;\; \updownarrow M$$

We recognize the matrix as the bottom M rows of R_{M+1}.

Rewriting the Normal Equations

We get:

$$R_{M+1}\mathbf{a}_M = \begin{bmatrix} ? \\ \mathbf{0} \end{bmatrix} \begin{array}{l} \updownarrow 1 \\ \updownarrow M \end{array}$$

But since the top element of $\mathbf{a}_M$ is $a_{M0} = 1$, we get:

$$? = \mathbf{a}_M^H R_{M+1}\mathbf{a}_M = E\left(\left|f_M\left[n\right]\right|^2\right) = P_M$$

Augmented Normal Equations for FPEF

$$R_{M+1}\mathbf{a}_M = \begin{bmatrix} P_M \\ \mathbf{0} \end{bmatrix} \begin{array}{l} \updownarrow 1 \\ \updownarrow M \end{array}$$

These are not articulated in the usual form. Usually, a matrix equation such as $Ax = y$ would prescribe y and we solve for x, which is otherwise unconstrainted.

Here: we want to find a vector $\mathbf{a}_M$ with 1 on top such that on the right we get M zeros at the bottom, and the top element in unconstrained.

Once $\mathbf{a}_M$ is found, we interpret that top element as the forward prediction error power $P_M = E\left(|f_M[n]|^2\right)$. Note that from $\mathbf{a}_M^H R_{M+1}\mathbf{a}_M = P_M$, we see $P_M \geq 0$ and in fact $P_M > 0$ unless R_{M+1} is degenerate!

Augmented Normal Equations for BPEF

A similar analysis yields:

$$R_{M+1}\mathbf{c}_M = \left[\begin{array}{c} \mathbf{0} \\ P_M \end{array}\right] \begin{array}{l} \updownarrow M \\ \updownarrow 1 \end{array}$$

Specifically, we can view this as solving for a vector $\mathbf{c}_M$ with 1 at the *bottom*, yielding the M zeros at the top (the orthogonality conditions). The bottom element is unconstrained, and turns out to be:

$$P_M = \mathbf{c}_M^H R_{M+1}\mathbf{c}_M = E\left(\left|b_M\left[n\right]\right|^2\right)$$

What is not obvious is this is the same as before, i.e.:

$$P_M = E\left(\left|f_M\left[n\right]\right|^2\right) = E\left(\left|b_M\left[n\right]\right|^2\right)$$

FPEF and BPEF Relation

Theorem:

$$\mathbf{a}_M = \mathbf{c}_M^{B*}$$

In words, the coefficients of the BPEF are the conjugated and reversed coefficients of the FPEF.

To see this, consider the equation $Ax = y$. If we flip the equations in reverse order (flip the rows of A), and flip the variables in reverse order (flip the columns of A), we get:

$$(\text{flip rows and flip columns of } A)\, x^B = y^B$$

FPEF and BPEF Relation

Because of the Hermitian Toeplitz symmetry of R_{M+1}, we get:

$$R_{M+1} = conj\,(\text{flip rows and flip columns of } \mathbf{R}_{M+1})$$

So:

$$R_{M+1}\mathbf{a}_M = \begin{bmatrix} P_M \\ \mathbf{0} \end{bmatrix} \longrightarrow R_{M+1}\mathbf{a}_M^{B*} = \begin{bmatrix} \mathbf{0} \\ P_M \end{bmatrix}$$

where we have used the fact that P_M is real, i.e., $P_M = P_M^*$. Also, $\mathbf{a}_M^{B*}$ has bottom element 1, so we recognize $\mathbf{a}_M^{B*} = \mathbf{c}_M$.

As a corrolary, we see that the forward and backward prediction error powers are each P_M.

Levinson-Durbin Recursion

It is possible to determine the M^{th} order FPEF/BPEF from the $(M-1)^{st}$ order FPEF/BPEF, via the *Levinson-Durbin recursion*.

Note that the 0^{th} order FPEF/BPEF is simply $a_{00} = 1$, so:

$$f_0[n] = b_0[n] = u[n]$$

and this gives:

$$P_0 = r[0]$$

Levinson-Durbin Recursion

Now take $m \geq 1$, and assume we have $\mathbf{a}_{m-1}, P_{m-1}$ for the $(m-1)^{st}$ order FPEF. Note that these are derived from $r[k]$ for $0 \leq k \leq m-1$. We use the decompositions of R_{m+1} to observe that:

$$R_{m+1} \begin{bmatrix} \mathbf{a}_{M-1} \\ 0 \end{bmatrix} = \begin{bmatrix} R_m & \mathbf{r}_m^{B*} \\ \mathbf{r}_m^{BT} & r_0 \end{bmatrix} \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} = \begin{bmatrix} P_{m-1} \\ \mathbf{0} \\ \Delta_{m-1} \end{bmatrix}$$

where:

$$\Delta_{m-1} = \mathbf{r}_m^{BT} \mathbf{a}_{m-1} = r[m] + \sum_{k=1}^{m-1} r[m-k]\, \mathbf{a}_{m-1,k}$$

Levinson-Durbin Recursion

We can do this similarly for the BPEF and we get:

$$R_{m+1} \begin{bmatrix} \mathbf{a}_{m-1} & 0 \\ 0 & \mathbf{a}_{m-1}^{B*} \end{bmatrix} = \begin{bmatrix} P_{m-1} & \Delta_{m-1}^{*} \\ \mathbf{0} & \mathbf{0} \\ \Delta_{m-1} & P_{m-1} \end{bmatrix}$$

Define:

$$\kappa_m = -\frac{\Delta_{m-1}}{P_{m-1}}$$

and the *Levinson-Durbin recursion* is:

$$\mathbf{a}_m = \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} + \kappa_m \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix}$$

Note that $a_{m0} = a_{m-1,0} + \kappa_m 0 = a_{m-1,0} = 1$, so that condition still holds!

Levinson-Durbin Recursion

Then:

$$R_{m+1}\mathbf{a}_m = \begin{bmatrix} \xi \\ \mathbf{0} \\ 0 \end{bmatrix}$$

In other words, κ_m is chosen to zero out the bottom coefficient! The value we get on top, labelled ξ above, is actually then P_m and we get:

$$\begin{aligned} P_m &= P_{m-1} - \kappa_m \Delta_{m-1}^* = P_{m-1} - \kappa_m \kappa_m^* P_{m-1} \\ &= \left(1 - |\kappa_m|^2\right) P_{m-1} \end{aligned}$$

Reflection Coefficients

κ_m is called the *reflection coefficient*. Note that $|\kappa_m| \leq 1$, and if $|\kappa_m| = 1$, then we get $P_m = 0$ (exactly) and R_{M+1} is degenerate. The process $u[n]$ is then *predictable* (not regular), i.e., we can express $u[n]$ exactly as a linear combination of $\{u[n-1], \cdots, u[n-M]\}$.

Otherwise, $|\kappa_m| < 1$, and indeed:

$$P_M = \prod_{m=1}^{M} \left(1 - |\kappa_m|^2\right) P_0$$

Levinson-Durbin Recursion

We write out the Levinson-Durbin recursion elementwise, for both the FPEF and BPEF vectors. For simplicity, so we don't need to separate out the boundary points $k = 0, m$, we adopt the convention that:

$$a_{m,k} = 0 \text{ for } k < 0 \text{ and } k > m$$

Then:

$$\begin{aligned} a_{m,k} &= a_{m-1,k} + \kappa_m a^*_{m-1,m-k} \\ a^*_{m,k} &= \kappa^*_m a_{m-1,k} + a^*_{m-1,m-k} \end{aligned}$$

Levinson-Durbin Recursion

The transfer functions of the FPEF and BPEF are:

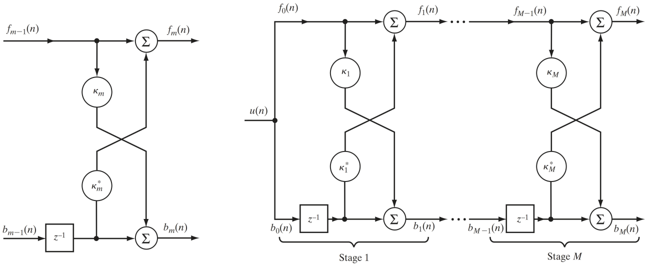
$$A_m(z) = \sum_{k=0}^{m} a_{mk}^* z^{-k} \quad \hat{A}_m(z) = \sum_{k=0}^{m} a_{m,m-k} z^{-k}$$

We can then rewrite the Levinson-Durbin recursion as:

$$\begin{aligned} A_m(z) &= A_{m-1}(z) + \kappa_m z^{-1} \hat{A}_{m-1}(z) \\ \hat{A}_m(z) &= \kappa_m^* A_{m-1}(z) + z^{-1} \hat{A}_{m-1}(z) \end{aligned}$$

This corresponds to the *lattice filter* realization of the FPEF/BPEF.

Lattice Filter



FPEF is Minimum-Phase

Theorem: The FPEF is minimum-phase.

Recall a minimum-phase fitler is stable and causal, and has a stable, causal inverse. Since the FPEF is FIR, this statement basically is equivalent to saying $A_m(z)$ has all its zeros in $|z| < 1$.

To prove this, we need to transform the usual form of the z-transform to one where we can apply principles of complex analysis. First recall the *paraconjugate:*

$$\tilde{H}(z) = H^*(1/z^*)$$

When $|z| = 1, 1/z^* = z$, and thus $\left|\tilde{H}(z)\right| = |H(z)|$ *on the unit circle.*

FPEF is Minimum-Phase

Introduce $B_m, \hat{B}_m$ as *polynomial* functions (with positive powers of z):

$$
\begin{aligned}
B_m(z) &= \tilde{A}_m(z) = \sum_{k=0}^{m} a_{m,k} z^m \\
\hat{B}_m(z) &= z^m \tilde{B}_m(z) = \sum_{k=0}^{m} a^*_{m,m-k} z^m
\end{aligned}
$$

Note that when $|z| = 1$, $|B_m(z)| = \left|\hat{B}_m(z)\right|$. The Levinson-Durbin recursions become recursions in the polynomial pair $\left\{B_m, \hat{B}_m\right\}$:

$$
B_m(z) = B_{m-1}(z) + \kappa_m z \hat{B}_{m-1}(z)
$$

Rouche's Theorem

Rouche's Theorem: If f, g are analytic inside and on a simple closed curve C, and $|f| > |g|$ for all points on C, then $f + g$ has the same number of zeros (counting multiplicity) as f inside C.

Here, since $B_{m-1}, \hat{B}_{m-1}$ are polynomials, they are certainly analytic.

Also, with $|\kappa_m| < 1$, we get $|B_{m-1}(z)| > \left|\kappa_m z \hat{B}_{m-1}(z)\right|$ on $|z| = 1$.

Conclusion: FPEF is Minimum-Phase

We conclude that B_m, B_{m-1} have the same number of zeros in $\{|z| < 1\}$.

Thus, B_m for *all* m have the same number of zeros inside the unit circle (and none on the unit circle).

But $B_0 = 1$ (constant), so $B_m(z)$ never has zeros on $\{|z| \leq 1\}$.

Therefore, $A_m(z)$ has no zeros in $\{|z| \geq 1\}$, and the FPEF is minimum-phase!

Correlation Function and Reflection Coefficients

Given a sequence $r[0], r[1], r[2], \cdots$, it represents the correlation of a WSS process iff the Hermitian Toeplitz matrices R_m of all orders are positive semi-definite.

From this sequence, we can derive $\kappa_1, \kappa_2, \cdots$. The condition $R_m \geq 0$ yields $|\kappa_m| \leq 1$, with strict pd iff strict inequality for the reflection coefficients.

We can run Levinson-Durbin backwards: given P_0 and κ_m, $m \geq 1$, observe $r[0] = P_0$.

Assume we have already recovered $r[k]$, $0 \leq k \leq m-1$.

Note κ_k for $1 \leq k \leq m-1$ gives us the FPEF $\mathbf{a}_{k-1}$.

From P_{m-1} and κ_m we find Δ_{m-1}, and then:

$$r[m] = \Delta_{m-1} - \sum_{k=1}^{m-1} r[m-k]\, a_{m-1,k}$$

Correlation Function and Reflection Coefficients

Up to a normalization $r[m] \longrightarrow r[m]/r[0]$, i.e., take $r[0] = P_0 = 1$, the sequence of valid correlations $\{r[m]\}_{0 \leq m < \infty}$, meaning Hermitian Toeplitz matrix R_m are pd for all orders, equivalently the DTFT $S(\omega)$ is ≥ 0 for all frequencies, has a one-to-one relationship with a sequence $\{\kappa_m\}_{1 \leq m < \infty}$ with $|\kappa_m| < 1$.

If any $|\kappa_m| = 1$, then $P_m = 0$ and the $\{r[m]\}$ sequence is degenerate (R_m is not invertible, the process is predictable).

FPEF and AR(M) Processes

Theorem: $f_M[n]$ is "approximately white". This is more accurate for large M, in general. It is *exactly* white iff the process is AR(M) and this occurs iff $\kappa_{M+k} = 0$ for all $k \geq 1$, and then $f_M[n]$ is the innovations signal with power P_M.

Justification: Take $n_0 \geq 1$. In general, $f_M[n - n_0] \in span\{u[n-k], n_0 \leq k \leq n_0 + M\}$, and $f_M[n] \perp span\{u[n-k], 1 \leq k \leq M\}$. For $n_0 < M$, these intervals partially overlap. The orthogonality conditions give $f_M[n]$ orthogonal to the "recent" past of $u[n]$, and in general we would expect weak correlation with the "old" past. As M gets larger, this statement gets stronger, and we don't get perfect result $f_M[n] \perp f_M[n-n_0]$ only for long time lags.

Case of AR(M) Process

For $\kappa_{M=k} = 0$, all $k \geq 1$, the Levinson-Durbin recursion yields that $\mathbf{a}_M$ is the FPEF for all order $M+k, k \geq 0$. This means $f_M[n] \perp u[n-k]$ for *all* $k \geq 1$. But since $f_M[n-n_0] \in span\{u[n-k], k \geq 1\}$, we get $f_M[n] \perp f_M[n-n_0]$ *exactly* and thus it is exactly white. Therefore:

$$f_M[n] = u[n] + \sum_{k=1}^{M} a_{m,k}^{*} u[n-k]$$

is the whitening filter, and u is AR(M), where $f_M[n]$ is the innovations with $E\left(|f_M[n]|^2\right) = P_M$. Note that technically for this to be the whitening filter, we need not just that f_M is white but also that this filter is minimum-phase, but indeed it is!

Case of AR(M) Processes

Note that we could have several $\kappa_m = 0$, but later values are not 0. Recall *all* possible sequences with $|\kappa_m| < 1$ are valid. For example, the sequence $0.2, 0.3, 0, 0, 0, 0, 0.5, 0.3, 0, -0.8, \cdots$ is possible, and does not correspond to an AR(M) process.

Orthogonality of Backward Prediction Error

Recall that $b_0[n] = u[n]$.

Theorem: For each fixed n, $\left\{ \frac{1}{\sqrt{P_m}} b_m[n] \right\}_{0 \leq m \leq M}$ is an orthonormal basis for *span* $\{u[n-m], 0 \leq m \leq M\}$.
Indeed, the mapping from
$\{u[n], u[n-1], \cdots, u[n-M]\} \longrightarrow \{b_0[n], b_1[n], \cdots, b_M[n]\}$ via Levinson-Durbin is a Gram-Schmidth orthogonalization without the normalization (i.e., b_m's are orthogonal, not orthonormal; dividing by $\sqrt{P_m}$ would make them orthonormal).

Orthogonality of Backward Prediction Error

Proof: $b_m[n] \in span\{u[n], u[n-1], \cdots, u[n-m]\}$. For $M > m$, $b_M[n] \perp span\{u[n], u[n-1], \cdots, u[n-M+1]\}$ and $b_m[n]$ is in this span!

Whereas the whiteness of the forward prediction errors is only approximate, the orthogonality of the backward prediction errors of all orders is *exact*.

In fact, we see that the state variables in the lattice filter are the backward prediction errors of different orders, and represents an exact orthogonalization of the data that would otherwise be in a transversal filter!

Entropy and Cholesky Factor

Define the upper triangular matrix:

$$\mathbf{A}_M = \begin{bmatrix} a_{00} = 1 & a_{11} & a_{22} & \cdots & a_{mm} \\ & a_{10} = 1 & a_{21} & \cdots & a_{m,m-1} \\ & & a_{20} = 1 & & \\ & \mathbf{0} & & \ddots & \vdots \\ & & & & a_{m,0} = 1 \end{bmatrix}$$

Entropy and Cholesky Factor

Then:

$$R_{M+1}\mathbf{A}_M = \begin{bmatrix} P_0 & 0 & 0 & \cdots & 0 \\ & P_1 & 0 & \cdots & 0 \\ & & P_2 & & \\ & \mathbf{X} & & \ddots & \vdots \\ & & & & P_M \end{bmatrix}$$

where the $\mathbf{X}$ denotes "don't care". Note that $\det \mathbf{A}_M = 1$, and clearly $\mathbf{A}_M$ is invertible.

Entropy and Cholesky Factor

This gives:

$$\det R_M = \prod_{m=0}^{M-1} P_m$$

Then:

$$\log\left(\det R_M\right)^{1/M} = \frac{1}{M}\sum_{m=0}^{M-1} \log P_m$$

Entropy and Cholesky Factor

$\det R_M = \prod_{m=0}^{M-1} \lambda_{Mm}$ where λ_{Mm} denote the eigenvalues of R_M. We get:

$$\frac{1}{M} \sum \log \lambda_{Mm} = \frac{1}{M} \sum \log P_m$$

Since $P_m \longrightarrow P_\infty$, and we are taking the average on the right, we get:

$$\log P_\infty = \lim_{M \longrightarrow \infty} \frac{1}{M} \sum \log \lambda_{Mm}$$

Then we apply *Szego's Theorem* to convert the average value of $g\left(\lambda_{Mm}\right)$ for large M to the average of $g\left(S\left(\omega\right)\right)$, the PSD:

$$\log P_\infty = \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} \log S\left(\omega\right) d\omega$$

Entropy and Cholesky Factor

Also, we get:

$$\mathbf{A}_M^H \mathbf{R}_{M+1} \mathbf{A}_M = diag\left\{P_m\right\}_{m=0}^{M}$$

Note that $L_M = \mathbf{A}_M^{-1}$ exists, has 1 on the diagonal, and is also upper triangular. We get:

$$\mathbf{R}_{M+1} = L_M^H diag\left\{P_m\right\}_{m=0}^{M} L_M$$

which we see is lower-diagonal-upper triangular. Indeed, the *Cholesky factor* of R_{M+1} is $L_M^H diag\left\{\sqrt{P_m}\right\}_{m=0}^{M}$.

The Gram-Schmidth orthogonalization achieved by the computation of the BPEFs via Levinson-Durbin is equivalent to the BPEF filter coefficients underlying the Cholesky factor of R.

OPTIMAL FILTERING
IV: WIENER FILTERS AND LINEAR PREDICTION

ECE416 ADAPTIVE ALGORITHMS