

# Geometric interpretation of signals: applications

*David G. Messerschmitt*



Electrical Engineering and Computer Sciences  
University of California at Berkeley

Technical Report No. UCB/EECS-2006-92

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-92.html>

June 27, 2006

Copyright © 2006, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

# Geometric interpretation of signals: applications

by

David G. Messerschmitt  
University of California at Berkeley

In this tutorial we give some examples of applications of the geometric framework developed in "Geometric interpretation of signals: background" [\[pdf\]](#). By *signal*, we usually mean a finite or infinite sequence of complex-valued samples. Both deterministic and random signals can be modeled w.r.t. their geometric properties.

## Stationary signals

A particularly important model that arises in signal processing is the *stationary signal*.

Given a Hilbert space  $\mathbb{H}$ , suppose we are given an infinite sequence of vectors

$\vec{X}(k), \vec{Y}(k) \in \mathbb{H}, -\infty < k < \infty$  where the inner products have the special property

$$\begin{aligned} r_x(l-m) &= \langle \vec{X}(l) | \vec{X}(m) \rangle \text{ or } r_x(m) = \langle \vec{X}(k) | \vec{X}(k-m) \rangle \\ r_{xy}(l-m) &= \langle \vec{X}(l) | \vec{Y}(m) \rangle \text{ or } r_{xy}(m) = \langle \vec{X}(k) | \vec{Y}(k-m) \rangle \end{aligned}$$

In this model,  $k$  is usually interpreted as ‘time’, and the inner product of two vectors is a function of the *difference* in time indices, not the absolute time. Putting it another way, the geometric properties of these signals do not depend on the absolute time index. In the special case of a wide-sense stationary random process (below), this is equivalent to two such signals being *jointly wide-sense stationary*. In analogy to the random process case, we call  $r_x(m)$  the *autocorrelation function*.

**Example.** (Deterministic continuous-time signals) In Hilbert space  $\mathbb{L}_2$  let

$$h(t) \leftrightarrow \vec{H} \in \mathbb{L}_2$$

and define an infinite sequence of vectors which are (in the time domain) time-translates of  $h(t)$  by times  $kT$ ,

$$h(t-kT), -\infty < t < \infty \leftrightarrow \vec{H}(k), \text{ where } \vec{H}(0) = \vec{H}.$$

In this case,

$$r_h(m) = \int_{-\infty}^{\infty} h(t) \cdot h^*(t+mT) \cdot dt.$$

This situation arises in sampling (see later) and in digital communications (see the homework).

**Example.** (Deterministic discrete-time signals) In Hilbert space  $\mathbb{H}_2$  let

$$h(k), -\infty < k < \infty \leftrightarrow \vec{H} \in \mathbb{H}_2$$

and define an infinite sequence of vectors which are (in the time domain) time-translates of  $h(k)$  by time  $m$ ,

$$h(k-m) \leftrightarrow \vec{H}(m), \text{ where } \vec{H}(0) = \vec{H}.$$

In this case,

$$r_h(m) = \sum_{k=-\infty}^{\infty} h(k) \cdot h^*(k+m).$$

This is the discrete-time parallel to the last example.

**Example.** (Random discrete-time signals) In Hilbert space  $\mathbb{H}_2(\Omega)$  let  $x(k)$ ,  $-\infty < k < \infty$ , be a wide-sense stationary (WSS) random process, and define

$$x(k) \leftrightarrow \vec{X}(k) \in \mathbb{H}_2(\Omega), -\infty < k < \infty$$

In this case,

$$r_x(m) = E[x(k) \cdot x^*(k-m)].$$

This situation arises in the modeling of a WSS discrete-time random process (see later).

Note the essential difference among these examples: In the two deterministic signal cases (continuous- and discrete-time), a vector is associated with an entire time signal, and in the random process case a vector is associated with just one sample of a time signal. This is the distinction between time and ensemble averaging—two complementary but distinct modeling approaches (which are connected through the ergodic theory). In a Hilbert space framework all these cases can be treated uniformly w.r.t. geometric properties.

The properties of an inner product impute upon the autocorrelation function some special properties. These properties are very familiar for a WSS random process, but carry over without change to the larger class of stationary signals. The autocorrelation is a Hermitian function, possessing complex-conjugate symmetry,

$$r_{xy}(m) = r_{yx}^*(-m)$$

$$r_x(m) = r_x^*(-m)$$

This in turn implies that  $r_x(0)$  is real-valued, and further,

$$0 \leq r_x(m) \leq r_x(0).$$

The left inequality follows from

$$r_x(0) = \|\vec{X}(k)\|^2 \geq 0$$

and the right inequality follows from the Schwarz inequality

$$|r_x(m)| = \left| \langle \vec{X}(k+m) | \vec{X}(k) \rangle \right| \leq \|\vec{X}(k+m)\| \cdot \|\vec{X}(k)\| = r_x(0)$$

Further, for any set of scalars  $d_1, d_2, \dots, d_n$

$$\left\| \sum_{l=1}^n d_l \cdot \vec{X}(k+l) \right\|^2 = \sum_{l=1}^n \sum_{m=1}^n d_l \cdot d_m^* \cdot r_x(l-m) \geq 0.$$

Defining

$$\mathbf{d} = [d_1 \quad d_2 \quad \dots \quad d_n] \text{ and}$$

$$\mathbf{R}_n = \begin{bmatrix} r_x(0) & r_x(-1) & \dots & r_x(1-n) \\ r_x(1) & r_x(0) & \dots & r_x(-2) \\ \dots & \dots & \dots & r_x(-1) \\ r_x(n-1) & r_x(n-2) & \dots & r_x(0) \end{bmatrix} = \begin{bmatrix} r_x(0) & r_x^*(1) & \dots & r_x^*(n-1) \\ r_x(1) & r_x(0) & \dots & r_x^*(2) \\ \dots & \dots & \dots & r_x^*(1) \\ r_x(n-1) & r_x(n-2) & \dots & r_x(0) \end{bmatrix}$$

This establishes that the *autocorrelation matrix*  $\mathbf{R}_n$  is non-negative definite (where  $\mathbf{d}^H$  is the conjugate transpose of  $\mathbf{d}$ ,

$$\mathbf{d} \mathbf{R}_n \mathbf{d}^H \geq 0.$$

To summarize, any autocorrelation matrix is:

- Toeplitz ( $r_{ij} = r_{i-j}$ )
- Hermitian ( $\mathbf{R}_n = \mathbf{R}_n^H$ )
- Non-negative definite

Since the autocorrelation matrix is Hermitian, its eigenvalues are real-valued, and since it is non-negative definite these real-valued eigenvalues are non-negative [1, page 42].

Further, there exists an orthonormal set of eigenvectors  $\mathbf{v}_i$ ,  $1 \leq i \leq n$ , and  $\mathbf{R}_n$  can be expressed in terms of these eigenvectors and its eigenvalues  $\lambda_i$ ,  $1 \leq i \leq n$  as [1, p44]

$$\mathbf{R}_n = \sum_{i=1}^n \lambda_i \cdot \mathbf{v}_i \mathbf{v}_i^H.$$

The *power spectrum* is defined as the Z-transform of the autocorrelation function,

$$P_x(z) = \sum_{k=-\infty}^{\infty} r_x(k) \cdot z^{-k}.$$

(Again, “power spectrum” is a terminology that arises in the wide-sense stationary random process case, but we are adopting this terminology for a larger class of signals.) From the Hermitian property of the autocorrelation, it follows that

$$P_x(z) = P_x^*\left(\frac{1}{z^*}\right) \text{ and hence } P_x(e^{j\omega}) \text{ is real-valued.}$$

The positive-definite property of the autocorrelation matrix implies further that

$$P_x(e^{j\omega}) \geq 0 \quad [1, \text{page 95}].$$

If the power spectrum is continuous on the unit circle, then it admits a *spectral factorization*

$$P_x(z) = \sigma^2 \cdot S_x(z) \cdot S_x^*\left(\frac{1}{z^*}\right) \text{ or } r_x(m) = \sigma^2 \cdot s_x(m) \otimes s_x^*(-m)$$

where “ $\otimes$ ” denotes convolution,

$$x(k) \otimes y(k) = \sum_m x(m) \cdot y(k-m) \text{ or } r_x(m) = \sum_n s_x(n) \cdot s_x^*(n-m)$$

and  $s_x(m)$  is *monic* ( $s_x(0) = 1$ ), *causal* ( $s_x(m) = 0$  for  $m < 0$ ), and *minimum-phase* (the region of convergence of both  $S_x(z)$  and  $\log S_x(z)$  include  $|z| \geq 1$ ).

The two factorizations lead to two useful ways to express a Hermitian form in the autocorrelation matrix. Given two  $1 \times n$  (row) matrices  $\mathbf{x}$  and  $\mathbf{y}$ ,

$$\begin{aligned} \mathbf{xR}_n\mathbf{y}^H &= \sum_{i=1}^n \lambda_i \cdot \mathbf{x}^H \mathbf{v}_i \mathbf{v}_i^H \mathbf{y} = \sum_{i=1}^n \lambda_i \cdot (\mathbf{xv}_i) \cdot (\mathbf{yv}_i)^* \\ \mathbf{xR}_n\mathbf{y}^H &= \sum_{m=1}^n \sum_{k=1}^n x_m \cdot r(m-k) \cdot y_k^* = \sum_{l=-\infty}^{\infty} \left( \sum_{m=1}^n x_m \cdot s_x(l+m) \right) \cdot \left( \sum_{k=1}^n y_k \cdot s_x(l+k) \right)^* \end{aligned}$$

## Linear predictors

Assume we are given a stationary signal in vector space (sequence of vectors  $\vec{X}(k)$  (elements of  $\mathbb{L}_2$  or  $\mathbb{l}_2$  or  $\mathbb{L}_2(\Omega)$ ) with autocorrelation function  $r_x(m)$ ). An optimum (in the sense of minimum error norm) infinite-order linear predictor of  $\vec{X}(k)$  based upon its entire past has error at time  $k$  equal to

$$\vec{E}(k) = \vec{X}(k) - P\left(\vec{X}(k) : \left\{ \vec{X}(i), -\infty < i < k \right\}\right).$$

(Recall the notation:  $\left\{ \vec{Y}_1, \vec{Y}_2, \dots, \vec{Y}_n \right\}$  is the subspace spanned by the vectors enclosed in brackets.) This signal is also called the *innovations*, since it embodies what is new about  $\vec{X}(k)$ ; that is, what cannot be easily inferred (based on a linear estimate) from its past. According to the projection theorem, this error is orthogonal to the subspace  $\left\{ \vec{X}(k), -\infty < i < k \right\}$ . But also note that all past innovations are vectors of this subspace,

$$E(i) \in \left\{ \vec{X}(i), -\infty < i < k \right\} \text{ for } -\infty < l < k.$$

Thus, we immediately conclude that the innovations process is *white*; that is,

$\left\langle \vec{E}(l) \middle| \vec{E}(k) \right\rangle = 0$  for  $l < k$  (and hence for  $l \neq k$ , due to the Hermitian symmetry in the error's autocorrelation function). Equivalently,  $r_e(m) = r_e(0) \cdot \delta_m$  and the power spectrum of the error is 'white',  $P_e(z) = r_e(0)$ . Intuitively, if the innovations were not white, there

would be some residual information in it about the past that could be inferred, thus allowing the error to be further reduced.

Define the transfer function from  $\vec{X}(k)$  to  $\vec{E}(k)$  as  $A_\infty(z)$  (the subscript  $\infty$  denotes an infinite-order predictor, in anticipation of finite-order predictors later),

$$A_\infty(z) = \sum_{i=0}^{\infty} a_i \cdot z^{-i}, \quad a_0 = 1.$$

This is a monic causal transfer function. In particular,

$$\vec{E}(k) = \sum_{i=0}^{\infty} a_i \cdot \vec{X}(k-i).$$

Then the white innovations implies that

$$P_x(z) \cdot A_\infty(z) \cdot A_\infty^*\left(\frac{1}{z^*}\right) = r_e(0)$$

Writing this in the form

$$A_\infty(z) \cdot A_\infty^*\left(\frac{1}{z^*}\right) = \frac{r_e(0)}{P_x(z)}$$

we see that the linear predictor transfer function is the spectral factorization of the inverse of the power spectrum of the signal (where the prediction error norm normalizes the spectral factorization to be monic). The inverse of the power spectrum possesses such a spectral factorization since its region of convergence includes the unit circle and is presumed to be non-zero on the unit circle. (Actually rational zeros on the unit circle are acceptable, although that case has to be treated with care.)

## Sampling

The idea behind sampling is to approximate a continuous-time complex valued signal  $x(t)$  by a discrete-time signal  $x(k)$  (called the 'sample values') through the relationship

$$x(t) \cong \sum_{k=-\infty}^{\infty} x(k) \cdot h(t-kT)$$

for some given 'pulse shape'  $h(t)$ . We expect that this approximation can become equality if  $x(t)$  is bandlimited to less than half the sampling rate and  $h(t)$  is chosen appropriately, but we are interested in more general cases where  $h(t)$  is perhaps any continuous-time signal and perhaps  $x(t)$  is not appropriately bandlimited in relation to the sampling rate. This approximation can be moved to signal space and interpreted geometrically if we associate

$$\begin{aligned} x(t), -\infty < t < \infty &\leftrightarrow \vec{X} \in \mathbb{L}_2 \\ h(t-kT), -\infty < t < \infty &\leftrightarrow \vec{H}_k \in \mathbb{L}_2 \end{aligned}$$

Then it is natural to seek the samples that minimize the  $\mathbb{L}_2$  norm of the error (this minimizes the energy of the error between the sampled approximation and the signal),

$$\min_{x(k), -\infty < k < \infty} \left\| \vec{X} - \sum_{k=-\infty}^{\infty} x(k) \cdot \vec{H}_k \right\|.$$

(Note the mixture of models inherent here: while  $x(k)$  is a discrete-time signal, we choose to consider it instead as a sequence of scalar coefficients!). This has a superficial similarity to the prediction problem, except that the approximation is in terms of future as well as past signals. From the projection theorem, since the sampled representation is an element of subspace  $\{\vec{H}_k, -\infty < k < \infty\}$

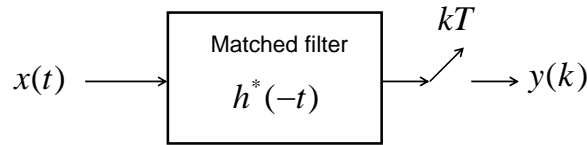
$$\left\langle \vec{X} - \sum_{k=-\infty}^{\infty} x(k) \cdot \vec{H}_k \middle| \vec{H}_i \right\rangle = 0, \quad -\infty < i < \infty.$$

The first thing we notice is that the signal space approach tells us we should first calculate the quantity

$$y(i) = \left\langle \vec{X} \middle| \vec{H}_i \right\rangle = \int_{-\infty}^{\infty} x(\tau) \cdot h^*(\tau - iT) \cdot d\tau.$$

This is known as a *sampled matched filter*, because it can be obtained by passing the signal through a filter with impulse response  $h^*(-t)$  (a matched filter) and sampling the output at times  $iT$ , as shown in the future below.

$$y(i) = \int_{-\infty}^{\infty} x(\tau) \cdot h^*(-(t - \tau)) \cdot d\tau \Big|_{t=iT}.$$



Why is this called a matched filter? Signal space methods give us a strong hint: Suppose a finite-energy signal  $x(t)$  is input to a matched filter  $h^*(-t)$  and the output is sampled at time  $t = 0$ . If the signal is constrained to have unit energy ( $\|\vec{X}\|^2 = 1$ ), what signal maximizes the modulus of the output? We can write the output as

$$\int x(\tau) \cdot h^*(-(t - \tau)) \cdot d\tau \Big|_{t=0} = \int x(\tau) \cdot h^*(\tau) \cdot d\tau = \left\langle \vec{X} \middle| \vec{H} \right\rangle.$$

The modulus is thus bounded by (Schwarz inequality)



$$\left| \langle \vec{X} | \vec{H} \rangle \right| \leq \|\vec{X}\| \cdot \|\vec{H}\| = \|\vec{H}\|$$

with equality iff  $\vec{X} = \alpha \cdot \vec{H}$  (or if the constraint is met,  $|\alpha| = 1/\|\vec{H}\|$ ). The maximum

response (in modulus) occurs when the input signal is the conjugate time-reversal of the impulse response.

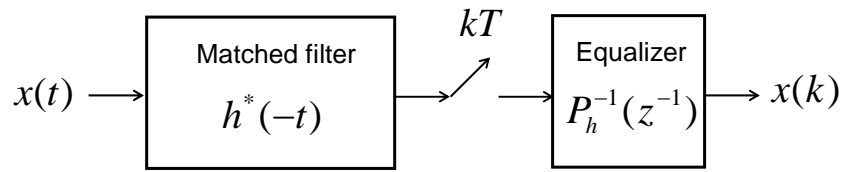
Once the signal  $x(t)$  has been filtered (by a matched filter) and sampled, the orthogonality principle yields a set of linear equations to solve for the samples  $x(k)$ :

$$y(i) - \sum_{k=-\infty}^{\infty} r_h(k-i) \cdot x(k) = 0, \quad -\infty \leq i \leq \infty.$$

Taking the Z-transform allows us to solve for the distance-minimizing set of samples  $x(k)$ ,

$$X(z) = P_h^{-1}(z^{-1}) \cdot Y(z).$$

These are the sample values that minimize the distance between the sampled approximation and the original continuous-time signal, as shown in the following figure. The equalizer has a two-sided unit-sample response; thus, it is not causal. This is generally true of the matched filter. This is, of course, because we did not include any causality constraint in the original formulation of the problem, so the solution is taking maximum advantage of the information available.



### Lattice filters

If the goal is *computation* of a predictor (as opposed to analytical results), then we must be satisfied with a finite-order predictor. This motivates us to define a optimum  $n$ -th order *forward prediction error* at time  $k$  as

$$\vec{E}_n(k) = \vec{X}(k) - P(\vec{X}(k) : \{\vec{X}(k-i), 1 \leq i \leq n\}), \quad \vec{E}_0(k) \equiv \vec{X}(k).$$

The projection theorem gives us a set of linear equations that can be solved for the coefficients of the predictor. If the prediction filter of order  $n$  (from  $\vec{X}_k$  to  $\vec{E}_n(k)$ ) is

$$A_n(z) = \sum_{i=0}^n a_n(i) \cdot z^{-i}, \quad a_n(0) = 1$$

then

$$0 = \left\langle \sum_{i=0}^n a_n(i) \cdot \vec{X}(k-i) \left| \vec{X}(k-j) \right. \right\rangle = \sum_{i=0}^n a_n(i) \cdot r_x(j-i), \quad 1 \leq j \leq n.$$

Unfortunately, this finite-order quasi-innovations process is not white in general because of the finite order.

It is useful to define a *backward prediction error* at time  $k$  of order  $n$  as

$$\vec{E}_n^R(k) = \vec{X}(k) - P\left(\vec{X}(k) : \left\{ \vec{X}(k+i), 1 \leq i \leq n \right\}\right), \quad \vec{E}_0^R(k) \equiv \vec{X}(k).$$

This is the error of a predictor of  $\bar{X}(k)$  in terms of  $n$  samples in the *future*. The superscript ‘R’ can be read as ‘reverse’ (to maintain consistency of notation with [1]), although ‘backward’ is a more common term in the literature. The transfer function of the backward predictor (from  $\bar{X}(k)$  to  $\bar{E}_n^R(k)$ ) is

$$\sum_{i=0}^n a_n^*(i) \cdot z^i = A_n^* \left( \frac{1}{z^*} \right).$$

(This is easily verified from the projection theorem and also follows from the Hermitian property of the autocorrelation function.) Although it is non-causal, and hence non-realizable, in the sequel we will always formulate this backward prediction error with sufficient delay so as to render the cascade to be causal.

One reason the forward and backward predictors are of both interest is that they give us two ways to apply a Gram-Schmidt orthogonalization to a subspace spanned by a finite continuous sequence of samples. Consider the following figure, which shows two subspaces of interest and two Gram-Schmidt orthogonalizations.

$$\begin{aligned} \left\{ \vec{X}(k-n+1), \vec{X}(k-n+2), \dots, \vec{X}(k) \right\} &= \\ \left\{ \vec{E}_0(k-n+1), \vec{E}_1(k-n+2), \dots, \vec{E}_{n-1}(k) \right\} \\ &\quad \underbrace{\hspace{10em}} \\ &\quad \underbrace{\hspace{10em}} \\ \left\{ \vec{X}(k-n), \vec{X}(k-n+1), \dots, \vec{X}(k-1) \right\} &= \\ \left\{ \vec{E}_0^R(k-1), \vec{E}_1^R(k-2), \dots, \vec{E}_{n-1}^R(k-n) \right\} \end{aligned}$$

In the bottom case, the subspace shown is the one from which  $\vec{E}_n(k)$ , a *forward* predictor of order  $n$  at time  $k$ , is drawn, and we have chosen an orthogonal basis consisting of *backward* prediction errors of different orders. This is precisely the Gram-Schmidt orthogonalization procedure applied to the samples starting at  $k-1, k-2$ , through  $k-n$ , in that order (moving backward in time). In the top case, the subspace shown is the one from which  $\vec{E}_n^R(k-n)$ , a *backward* predictor of order  $n$  at time  $k-n$ , is drawn, and we have chosen an orthogonal basis consisting of *forward* prediction errors of different orders. This is precisely the Gram-Schmidt orthogonalization procedure applied to the samples starting at  $k-n+1, k-n+2$ , through  $k$ , in that order (moving forward in time).

It is convenient to express the forward and backward predictors in terms of the orthogonal basis, because (1) the coefficients are particularly easy to determine and (2) this conveniently expresses prediction errors of order  $n$  to lower-order prediction errors. Applying the orthogonality principle to the forward and backward predictors,

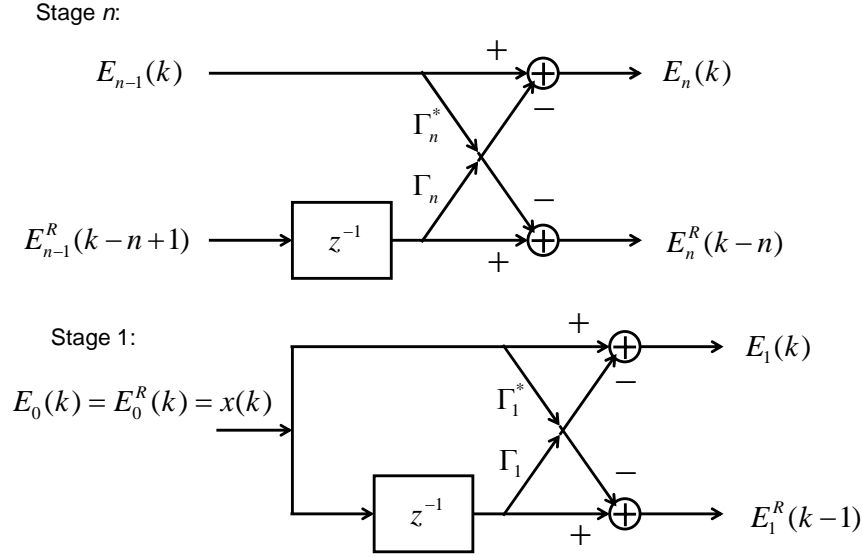
$$\begin{aligned}\vec{E}_n(k) : \left\langle \vec{X}(k) - \sum_{i=1}^n \Gamma_i \cdot \vec{E}_{i-1}^R(k-i) \middle| \vec{E}_{j-1}^R(k-j) \right\rangle &= 0, \quad 1 \leq j \leq n \\ \vec{E}_n^R(k-n) : \left\langle \vec{X}(k-n) - \sum_{i=1}^n \Lambda_i \cdot \vec{E}_{i-1}(k-n+i) \middle| \vec{E}_{j-1}(k-n+j) \right\rangle &= 0, \quad 1 \leq j \leq n\end{aligned}$$

The scalars  $\Gamma_i$  are called *reflection coefficients*, and are not a function of  $k$  due to the WSS assumption. Similarly for the coefficients  $\Lambda_i$ , although we will see shortly that they are actually redundant (this is not surprising, since a linear predictor of order  $n$  is expected to have  $n$  free parameters); that is,  $\Lambda_i = \Gamma_i^*$  as well be shown later.

Due to WSS, the prediction error can be expressed order-recursively as

$$\begin{aligned}\vec{E}_n(k) &= \vec{X}(k) - \sum_{i=1}^n \Gamma_i \cdot \vec{E}_{i-1}^R(k-i) = \vec{E}_{n-1}(k) - \Gamma_n \cdot \vec{E}_{n-1}^R(k-n) \\ \vec{E}_n^R(k-n) &= \vec{X}(k-n) - \sum_{i=1}^n \Lambda_i \cdot \vec{E}_{i-1}(k-n+i) = \vec{E}_{n-1}^R(k-n) - \Lambda_n \cdot \vec{E}_{n-1}(k)\end{aligned}$$

As shown in the following figure (in the time domain rather than signal space), these recursions together define both the forward and backward prediction errors of an arbitrary (but finite) number of orders, starting with the first, then second, etc. This structure is called a *lattice filter*.



It is useful to review the technique we used. First we observed that the predictor can be expressed in terms of an orthogonal basis obtained using the Gram-Schmidt orthogonalization procedure. This procedure turns out to define lower-order predictors. There are two natural options for orthogonalization—moving forward in time, and moving backward. We exploited both by defining both a forward predictor (our primary interest) and a backward predictor (an intermediate step). The natural orthogonalization is to move backward in time for the forward predictor and forward in time for the backward predictor. In both cases, this choice results in an order-recursive representation of the predictors—the  $n$ -th order predictor can be obtained incrementally from an  $(n-1)$ -th order predictor.

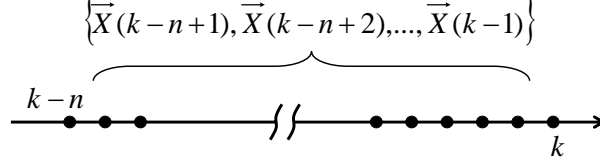
It remains to determine the reflection coefficients, and derive some of their interesting properties. The orthogonality relationship easily allows us to solve for the coefficients (exploiting the orthogonality of the Gram-Schmidt basis) as (considering, without loss of generality, only  $j = n$ ),

$$\Gamma_n = \frac{\langle \vec{X}(k) | \vec{E}_{n-1}^R(k-n) \rangle}{\|\vec{E}_{n-1}^R(k-n)\|^2},$$

$$\Lambda_n = \frac{\langle \vec{X}(k-n) | \vec{E}_{n-1}(k) \rangle}{\|\vec{E}_{n-1}(k)\|^2}.$$

These relations can be more conveniently expressed in terms of prediction errors (rather than signal values). The structure of the subspaces reflected in these errors is illustrated in the figure below. From the orthogonality principle, both the errors  $\vec{E}_{n-1}(k)$  and

$\vec{E}_{n-1}^R(k-n)$  are orthogonal to the subspace spanned by  $\{\vec{X}(k-n+1), \vec{X}(k-n+2), \dots, \vec{X}(k-1)\}$ , and both forward and backward predictors are vectors in that subspace.



Hence,

$$\begin{aligned} \langle \vec{X}(k) | \vec{E}_{n-1}^R(k-n) \rangle &= \langle \vec{E}_{n-1}(k) | \vec{E}_{n-1}^R(k-n) \rangle \\ \langle \vec{X}(k-n) | \vec{E}_{n-1}(k) \rangle &= \langle \vec{E}_{n-1}^R(k-n) | \vec{E}_{n-1}(k) \rangle \end{aligned}$$

Further, the forward and backward prediction errors of the same order will have the same norm, even if displaced in time,

$$\|\vec{E}_{n-1}^R(k-n)\| = \|\vec{E}_{n-1}(k)\|.$$

Combining these two observations, the reflection coefficients can be expressed as a normalized inner product of forward and backward errors,

$$\Gamma_i = \Lambda_i^* = \frac{\langle \vec{E}_{n-1}(k) | \vec{E}_{n-1}^R(k-n) \rangle}{\|\vec{E}_{n-1}(k)\| \cdot \|\vec{E}_{n-1}^R(k-n)\|}.$$

The Schwartz inequality tells us therefore that the modulus of the reflection coefficients are less than or equal to unity,

$$|\Gamma_n| \leq 1$$

Finally, we can determine the impact on the prediction error norm as the order of the predictor is increased by one,

$$\begin{aligned} \vec{E}_n(k) &= \vec{E}_{n-1}(k) - \Gamma_n \cdot \vec{E}_{n-1}^R(k-n) \\ \|\vec{E}_n(k)\|^2 &= \|\vec{E}_{n-1}(k)\|^2 + |\Gamma_n|^2 \cdot \|\vec{E}_{n-1}^R(k-n)\|^2 - \Gamma_n^* \cdot \langle \vec{E}_{n-1}(k) | \vec{E}_{n-1}^R(k-n) \rangle - \Gamma_n \cdot \langle \vec{E}_{n-1}^R(k-n) | \vec{E}_{n-1}(k) \rangle \\ &\Rightarrow \|\vec{E}_n(k)\|^2 = (1 - |\Gamma_n|^2) \cdot \|\vec{E}_{n-1}(k)\|^2 \end{aligned}$$

Thus, the closer the reflection coefficient modulus  $|\Gamma_n|$  is to unity, the more the norm of the prediction error is reduced in increasing the predictor order from  $n-1$  to  $n$ . Alternatively, this reflects a large inner product (alignment) of the forward and backward errors of order  $n-1$ .

The backward predictor cascaded with a delay (from  $\overrightarrow{X_k}$  to  $\overrightarrow{E_n^R}(k-n)$ ) has transfer function

$$A_n^R(z) = z^{-n} \cdot A_n^*\left(\frac{1}{z^*}\right).$$

The lattice filter is then represented by the order-iterative relation

$$\begin{bmatrix} A_n(z) \\ A_n^R(z) \end{bmatrix} = \begin{bmatrix} 1 & z^{-1} \cdot \Gamma_n \\ \Gamma_n^* & z^{-1} \end{bmatrix} \cdot \begin{bmatrix} A_{n-1}(z) \\ A_{n-1}^R(z) \end{bmatrix}.$$

This expression is easily inverted, assuming that  $|\Gamma_n| < 1$ , as

$$\begin{bmatrix} A_{n-1}(z) \\ A_{n-1}^R(z) \end{bmatrix} = \frac{1}{z^{-1}(1-|\Gamma_n|^2)} \begin{bmatrix} z^{-1} & -z^{-1} \cdot \Gamma_n \\ -\Gamma_n^* & 1 \end{bmatrix} \cdot \begin{bmatrix} A_n(z) \\ A_n^R(z) \end{bmatrix}$$

Armed with these relations, in the following we develop conversions among the three equivalent sets of parameters: autocorrelation coefficients  $r_x(m)$ , predictor filter coefficients  $a_N(k)$ , and reflection coefficients  $\Gamma_n$ .

### Minimum-phase property

As seen before, for a valid autocorrelation function we will have  $|\Gamma_n| < 1$ . It is also true that the finite predictor filter  $A_n(z)$  is strictly minimum phase (zeros all interior to the unit circle) if the reflection coefficients  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  are all less than unity in magnitude.

To show this, assume that  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  all have magnitude less than unity. We will show that  $A_n(z)$  is minimum phase. Proceeding by induction,  $A_1(z) = 1 + \Gamma_1 \cdot z^{-1}$  has one pole (at  $z = 0$ ) and one zero (at  $z = \Gamma_1$ ) inside the unit circle; it is minimum phase. Assume that  $A_{n-1}(z)$  is minimum phase, and thus has  $n-1$  poles (at  $z = 0$ ) and  $n-1$  zeros inside the unit circle. Starting with the order-recursive relation

$$A_n(z) = A_{n-1}(z) + \Gamma_n \cdot z^{-1} A_{n-1}^R(z),$$

dividing through by  $A_{n-1}(z)$  and replacing  $A_{n-1}^R(z)$  by  $z^{1-n} \cdot A_{n-1}^*\left(\frac{1}{z^*}\right)$ ,

$$\frac{A_n(z)}{A_{n-1}(z)} = 1 + \Gamma_n \cdot z^{-n} \frac{A_{n-1}^*\left(\frac{1}{z^*}\right)}{A_{n-1}(z)}.$$

The term  $z^{-n} \frac{A_{n-1}^*\left(\frac{1}{z^*}\right)}{A_{n-1}(z)}$  is an allpass filter, so as  $z$  traverses around the unit circle its mapping to the  $Z$ -plane also traverses the unit circle. Thus, as  $z$  traverses around the unit circle,  $\frac{A_n(z)}{A_{n-1}(z)}$  fails to traverse the unit circle since  $|\Gamma_n| < 1$ . This implies that  $\frac{A_n(z)}{A_{n-1}(z)}$  has the same number of poles and zeros inside the unit circle, and we already know that this is true of the denominator  $A_{n-1}(z)$ . Knowing that  $A_n(z)$  has  $n$  poles (at  $z = 0$ ) inside the unit circle, it must also have  $n$  zeros inside the unit circle; it is minimum phase.

### Levinson-Durbin recursion

The preceding gives us a way of turning an autocorrelation function into reflection coefficients and a modeling error:

$$\varepsilon_0 = r_x(0), \quad a_0(0) = 1$$

$$\Gamma_n^* = \frac{\langle \vec{X}(k-n) | \vec{E}_{n-1}(k) \rangle}{\varepsilon_{n-1}} = \frac{\langle \vec{X}(k-n) | \sum_{i=0}^{n-1} a_{n-1}(i) \cdot X(k-i) \rangle}{\varepsilon_{n-1}} = \frac{\sum_{i=0}^{n-1} a_{n-1}^*(i) \cdot r_x(i-n)}{\varepsilon_{n-1}}$$

$$\varepsilon_n = \varepsilon_{n-1} \cdot (1 - |\Gamma_n|^2)$$

$$\begin{bmatrix} a_n(1) \\ a_n(2) \\ \dots \\ a_n(n-1) \\ a_n(n) \end{bmatrix} = \begin{bmatrix} a_{n-1}(1) \\ a_{n-1}(2) \\ \dots \\ a_{n-1}(n-1) \\ 0 \end{bmatrix} + \Gamma_n \cdot \begin{bmatrix} a_{n-1}^*(n-1) \\ a_{n-1}^*(n-2) \\ \dots \\ a_{n-1}^*(1) \\ 1 \end{bmatrix}$$

### Step-up recursion

Given a set of reflection coefficients, the step-up recursion determines the predictor coefficients for predictors of all orders, beginning with the first order:

$$\begin{bmatrix} a_n(1) \\ a_n(2) \\ \dots \\ a_n(n-1) \\ a_n(n) \end{bmatrix} = \begin{bmatrix} a_{n-1}(1) \\ a_{n-1}(2) \\ \dots \\ a_{n-1}(n-1) \\ 0 \end{bmatrix} + \Gamma_n \cdot \begin{bmatrix} a_{n-1}^*(n-1) \\ a_{n-1}^*(n-2) \\ \dots \\ a_{n-1}^*(1) \\ 1 \end{bmatrix}$$

### Step-down recursion

Given the predictor coefficients of all orders, the step-down recursion determines the reflection coefficients, beginning with the last:

$$\Gamma_n = a_n(n)$$

$$\begin{bmatrix} a_{n-1}(1) \\ a_{n-1}(2) \\ \dots \\ a_{n-1}(n-2) \\ a_{n-1}(n-1) \end{bmatrix} = \frac{1}{1 - |\Gamma_n|^2} \cdot \left( \begin{bmatrix} a_n(1) \\ a_n(2) \\ \dots \\ a_n(n-2) \\ a_n(n-1) \end{bmatrix} - \Gamma_n \cdot \begin{bmatrix} a_n^*(n-1) \\ a_n^*(n-2) \\ \dots \\ a_n^*(2) \\ a_n^*(1) \end{bmatrix} \right)$$

### Reference

[1] Monson H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, 1996.