Statistical Signal Processing - Definitions

1 Basic Definitions

Correlation: The correlation of two random variables, x and y, is defined as $r_{xy} = E(xy^*)$. This is a second-order joint moment of x and y.

Covariance: The covariance of two random variables, x and y, is $c_{xy} = \text{Cov}(x, y) = E[(x - m_x)(y - m_y)^*) = E(xy^*) - m_x m_y^* = r_{xy} - m_x m_y^*$.

Correlation Coefficient: The correlation coefficient is a normalized covariance, given by $\rho_{xy} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$. where $\sigma_x = \text{Var}(x)$.

Statistically Independent: Two random variables, x and y, are said to be statistically independent if $f_{x,y}(\alpha, \beta) = f_x(\alpha) f_y(\beta)$.

Uncorrelated: Two random variables x and y are said to be uncorrelated if their covariance, c_{xy}, is zero. This is equivalent to $r_{xy} = E(x)E(y^*) = m_x m_y^*$.

Discrete-Time Random Process: A discrete-time random process is an index sequence of r.v.'s.

Ensemble Averages: Ensemble averages are averages over the probability space of the random process. Note that these averages may vary with time (i.e. vary with n).

Mean of a Random Process: The mean of a random process is the (deterministic) sequence $m_x(n) = E(x(n))$.

Variance of a Random Process: The variance of a random process, $x(n)$, is $\sigma_x(n) = E(|x(n) - m_x(n)|^2)$.

Autocovariance: The autocovariance of a random process, $x(n)$, is $c_{xx}(k, l) = E((x(k) - m_x(k))(x(l) - m_x(l))^*)$.

Autocorrelation: The autocorrelation of a random process, $x(n)$, is $r_{xx}(k, l) = E(x(k)x^*(l))$.

Remark 1: The mean and variance are referred to as the first order statistics, while the autocovariance and autocorrelation are referred to as the second order statistics of a random process.

Remark 2: As in the case case of r.v.'s, the autocorrelation and autocovariance functions provide information about the degree of linear dependence between two random variables. For

example, if $c_{xx}(k,l) = 0$ for $k \neq l$, then the r.v.'s $x(k)$ and $x(l)$ are uncorrelated and knowledge of one does not help in the estimation of the other using a linear estimator.

Cross-Covariance: The cross-covariance of two random processes, $x(n)$ and $y(n)$ is $c_{xy}(k,l) = E[(x(k) - m_x(k))(y(l) - m_y(l))^*] = r_{xy}(k,l) - m_x(k)m_y^*(k)$.

Cross-Correlation: The cross-correlation of two random processes, $x(n)$ and $y(n)$ is $r_{xy}(k,l) = E[x(k)y^*(l)]$.

Uncorrelated Random Processes: Two random processes, $x(n)$ and $y(n)$, are uncorrelated if $c_{xy}(k,l) = \delta(k-l)$.

Orthogonal Random Processes: Two random processes, $x(n)$ and $y(n)$, are orthogonal if $r_{xy}(k,l) = m_x(k)m_y^*(l)$.

Proposition: If two random processes $x(n)$ and $y(n)$ are uncorrelated, then the autocorrelation of the sum $z(n) = x(n) + y(n)$ is equal to the sum of the autocorrelations of $x(n)$ and $y(n)$, i.e., $r_{zz}(k,l) = r_{xx}(k,l) + r_{yy}(k,l)$.

First-Order Stationary: A random process is first-order statationary if the first-order density function of $x(n)$ is independent of time, i.e., $f_{x(n)}(\alpha) = f_{x(n+k)}(\alpha)$ for all k.

Second-Order Stationary: A random process, $x(n)$, is second-order statationary if the second-order density function $f_{x(n_1),x(n_2)}(\alpha_1,\alpha_2)$ depends only on the difference $n_2 - n_1$. Equivalently, if for all k, $x(n)$ and $x(n+k)$ have the same second-order joint density function, i.e. $f_{x(n_1),x(n_2)}(\alpha_1,\alpha_2) = f_{x(n_1+k),x(n_2+k)}(\alpha_1,\alpha_2)$. If a process is second-order stationary, then it will be first-order stationary.

Lag: Suppose $x(n)$ is second-order stationary. Then $r_{xx}(k,l) = r_{xx}(k-l,0) = r_{xx}(k-l)$, where $k-l$ is called the lag.

Stationary in the Strict Sense (SSS): A random process is stationary in the strict sense, if $x(n)$ is stationary for all orders.

Wide Sense Stationary (WSS): A random process is said to be WSS if: (1) $m_x(n) = m_x$, (2) $r_{xx}(k,l) = r_{xx}(k-l)$, and (3) $c_{xx}(0) = \mathrm{Var}(x(k)) < \infty$.

Remark 3: If $x(n)$ is a WSS Gaussian process, then it is SSS.

Jointly Wide Sense Stationary: If two processes, $x(n)$ and $y(n)$, are each WSS and $r_{xy}(k,l) = r_{xy}(k-l)$, then they are jointly WSS.

Properties of WSS Processes: If $x(n)$ is WSS, then (1) $r_{xx}(k) = r_{xx}^*(-k)$, (2) $r_{xx}(0) = E[|x(n)|^2] \geq 0$, and (3) $r_{xx}(0) \geq |r_{xx}(k)|$ for all k. If, in addition, $r_{xx}(k_0) = r_{xx}(0)$ for some k_0, then $r_{xx}(k)$ is periodic with period k_0 and $E[|x(n) - x(n-k_0)|^2] = 0$, here, $x(n)$ is said to be mean-square periodic.

Autocovariance and Autocorrelation Matrices: Then $\mathbf{R}_{xx} = E(\mathbf{x}\mathbf{x}^H)$ and $\mathbf{C}_{xx} = \mathbf{R}_{xx} - \mathbf{m}_x\mathbf{m}_x^H$ for $\mathbf{x} = (x(0), x(1), \ldots, x(p))^T$. Note that $\mathbf{R}_{xx}$ is Hermetian ($\mathbf{R}_{xx} = \mathbf{R}_{xx}^H$),

toeplitz matrix and thus $\mathbf{R}_{xx} = toep(r_{xx}(0), r_{xx}(1), \ldots, r_{xx}(p))$.

Properties of a WSS Autocorrelation Matrix: If $x(n)$ is a WSS process, then $\mathbf{R}_{xx}$ is nonnegative definite, $\mathbf{R}_{xx} > 0$. Since $\mathbf{R}_{xx}$ is Hermetian and nonnegative definite, the eigenvalues are all real and nonnegative.

Sample Mean, Variance, and Autocorrelation: The sample mean, variance, and autocorrelation are given by

$$\widehat{m}_x = \frac{1}{N}\sum_{n=0}^{N-1} x[n], \quad \widehat{\sigma}_x^2 = \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\widehat{m}_x)^2, \quad \widehat{r}_{xx}[m] = \frac{1}{N}\sum_{n=0}^{N-|m|-1} x[n]x[n+|m|].$$

Ergodicity: Ergodicity is an averaging over time rather than over the probability space, e.g., $\hat{m}_x = \frac{1}{N}\sum_{n=1}^{N} x_1(n)$, for a realization, $x_1(n)$, of a r.v. $x(n)$. Note that this only makes sense if $m_x(n) = m_x$.

Ergodic in the Mean: If the sample mean $\hat{m}_x(N) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)$ of a WSS process converges to m_x in the square-mean sense (i.e. convergence in L^2), then the process is said to be ergodic in the mean and we write $\lim_{N\to\infty} \hat{m}_x(N) = m_x$. Note that $\hat{m}_x(N)$ is a random variable.

Proposition: A WSS process is ergodic in the mean iff $\lim_{N\to\infty} E(\hat{m}_x(N)) = m_x$ and $\lim_{N\to\infty} Var(\hat{m}_x(N)) = 0$.

Mean Ergodic Theorem 1: Let $x(n)$ be a WSS random process with autocovariance sequence $c_{xx}(k)$. Then $x(n)$ is ergodic in mean if $\lim_{k\to\infty} c_{xx}(k) = 0$.

Autocorrelation Ergodic: Let $x(n)$ be a WSS random process. Then $r_{xx}(k) = E(x(n)x^*(n-k))$. Let $y_k(n) = x(n)x^*(n-k)$. Then

$$\hat{r}_{xx}(k,N) = \frac{1}{N}\sum_{n=0}^{N-1} y_k(n) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)x^*(n-k).$$

If $\hat{r}_{xx}(k,N) \to^{L^2} r_{xx}(k)$, then $x(n)$ is Autocorrelation ergodic. Since $\hat{r}_{xx}(k,N)$ is the sample mean of $y_k(n)$, it follows that $x(n)$ will be autocorrelation ergodic if $y_k(n)$ is ergodic in mean.

Autocorrelation Ergodic Theorem: Let $x(n)$ be a WSS random process. Then $x(n)$ is Autocorrelation Ergodic iff $\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} c_{xx}^2(k) = 0$.

White Noise: A WSS process, $v(n)$, is said to be white noise if $c_{vv}(k) = \sigma_v^2\delta(k)$. Thus, white noise is a sequence of uncorrelated r.v.'s, each having variance σ_v^2.

Power Spectrum: The power spectrum of a WSS or SSS random process is the DTFT of the autocorrelation sequence, i.e. $P_{xx}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_{xx}(k)e^{-jk\omega}$. Then using the inverse DTFT we have $r_{xx}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xx}(e^{j\omega})e^{jk\omega}\,d\omega$. We may also define the z-transform of P_{xx} by $P_{xx}(z) = \sum_{k=-\infty}^{\infty} r_{xx}(k)z^{-k}$.

Properties of the Power Spectrum of a WSS Process: The power spectrum of a WSS process, $x(n)$, is (1) *real valued*: $P_{xx}(e^{j\omega}) = P_{xx}^*(e^{j\omega})$, $P_{xx}(z) = P_{xx}^*(1/z^*)$ and if, in addition, $x(n)$ is real, then the power spectrum is even, i.e. $P_{xx}(e^{j\omega}) = P_{xx}(e^{-j\omega})$ and $P_{xx}(z) = P_{xx}(z^*)$,

(2) *nonnegative*: $P_{xx}(e^{j\omega}) \geq 0$, (3) *total power* of a mean zero WSS process is $E(|x(n)|^2) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xx}(e^{j\omega})\,d\omega$, and (4) *eigenvalue extremal property*: the eigenvalues (denoted by λ_i) of the $n \times n$ autocorrelation matrix of a mean zero WSS process satisfy

$$\min_{\omega} P_{xx}(e^{j\omega}) \leq \lambda_i \leq \max_{\omega} P_{xx}(e^{j\omega}).$$

Calculating the Power Spectrum Egotically: Let $x(n)$ be a WSS process. Then

$$P_{xx}(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{2N+1} E\left[\left|\sum_{n=-N}^{N} x(n)e^{-jn\omega}\right|^2\right] = \lim_{N\to\infty} \frac{1}{2N+1} E|X_N(e^{j\omega})|^2.$$

Filtering Random Processes: Let $x(n)$ be a WSS process, $h(n)$ be a stable linear shift invariant filter, and $y(n) = x(n) * h(n)$. If $\sigma_y^2 < \infty$, then $y(n)$ is a WSS process and (1) $r_{yx}(k) = r_{xx}(k) * h(k)$, (2) $r_{yy}(k) = r_{xx}(k) * h(k) * h^*(-k)$, (3) $\sigma_y^2 = E(|y(n)|^2) = \mathbf{h}^H \mathbf{R}_{xx} \mathbf{h}$, (4) $P_{yy}(e^{j\omega}) = P_{xx}(e^{j\omega})|H(e^{j\omega})|^2$, and (5) $P_{yy}(z) = P_{xx}(z)H(z)H^*(1/z^*)$.

Remark: Suppose that $x(n) = w(n) * h(n)$, where $w(n)$ is white noise with variance σ_w^2 and $h(n)$ is real. Then $P_{xx}(z) = \sigma_w^2 H(z)H(1/z)$. Thus if we are given $P_{xx}(z)$ and we filter $x(n)$ with unit variance white noise, then $P_{xx}(z) = H(z)H(1/z)$.

Physical Description of the Power Spectrum: The power spectrum describes how the signal power is distributed as a function of frequency.

Spectral Factorization: Let $x(n)$ be a WSS process such that $P_{xx}(e^{j\omega})$ is continuous and $\ln(P_{xx}(z))$ is analytic in $\rho < |z| < 1/\rho$ for $\rho < 1$. Then the exists $c(k)$ such that

$$\ln(P_{xx}(z)) = \sum_{k=-\infty}^{\infty} c(k)z^{-k}, \quad \text{where } c(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(P_{xx}(e^{j\omega}))e^{j\omega}\,d\omega.$$

And since $P_{xx}(e^{j\omega})$ is real,

$$P_{xx}(z) = \exp(c(0)) \exp\left(\sum_{k=1}^{\infty} c(k)z^{-k}\right) \exp\left(\sum_{k=-\infty}^{-1} c(k)z^{-k}\right) = \exp(c(0))Q(z)Q^*(1/z^*).$$

Regular Process: The WSS process, $x(n)$, is a regular process if the power spectrum can be factored $P_{xx}(z) = \sigma_0^2 Q(z)Q^*(1/z^*)$ as above.

Properties of a Regular Process: 1) Any regular process may be realized as the output of a causal and stable filter that is driven by white noise having a variance σ_0^2. This is known as innovations representation. (2) The inverse filter $1/H(z)$ is a whitening filter. That is, if $x(n)$ is filtered with $1/H(z)$, then the output is white noise with a variance of σ_0^2. The formation of this white noise process is called innovations process. (3) Since $v(n)$ (the white noise) and $x(n)$ are related by an invertible transformation, either process may be derived from the other. Therefore they contain the same information.

2 Signal Modeling

General Difference Equation: A general difference equation is given by

$$y[n] + \sum_{k=1}^{p} a_p(k) y[n-k] = \sum_{k=0}^{q} b_q(k) x[n-k]$$

and thus

$$H(z) \equiv \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{q} b_q(k) z^{-k}}{1 + \sum_{k=1}^{p} a_p(k) z^{-k}}.$$

Autoregressive Moving Average Process (ARMA): A process with power spectrum

$$P_{xx}(e^{j\omega}) = \sigma_v^2 \frac{|B_q(e^{j\omega})|^2}{|A_p(e^{j\omega})|^2},$$

where

$$B_q(z) = \sum_{k=0}^{q} b_q(k) z^{-k} \quad \text{and} \quad A_p(z) = 1 + \sum_{k=1}^{p} a_p(k) z^{-k}$$

is an autoregressive moving average process of order (p,q) or simply ARMA(p,q). It is the result of filtering white noise, $v(n)$, with a causal linear shift-invariant filter with system function $H(z) = B_q(z)/A_p(z)$. Notice that $P_{xx}(z)$ has $2p$ poles and $2q$ zeros.

Yule Walker Equations (Property of an ARMA Process): Let $x(n)$ be the process generated by filtering $v(n)$ by $h(n)$, where $H(z)$ is given above. Then $x(n)$ and $v(n)$ are related by the difference equation

$$x(n) + \sum_{l=1}^{p} a_p(l) x(n-l) = \sum_{l=0}^{q} b_q(l) v(n-l)$$

and thus

$$\begin{aligned} r_{xx}(k) + \sum_{l=1}^{p} a_p(l) r_{xx}(k-l) &= \sum_{l=0}^{q} b_q(l) r_{vx}(k-l) = \sigma_v^2 \sum_{l=0}^{q} b_q(l) h^*(l-k) \\ &= \sigma_v^2 \sum_{l=0}^{q-k} b_q(l+k) h^*(l) = \begin{cases} \sigma_v^2 c_q(k), & 0 \le k \le q \\ 0, & k > q. \end{cases} \end{aligned}$$

This is known as the Yule-Walker equations.

Autoregressive Process (AR): A process with power spectrum

$$P_{xx}(z) = \sigma_v^2 \frac{|b(0)|^2}{A_p(z) A_p^*(1/z^*)}, \quad P_{xx}(e^{j\omega}) = \sigma_v^2 \frac{|b(0)|^2}{|A_p(e^{j\omega})|^2},$$

where $A_p(z)$ is given above is an autoregressive process of order p or AR(p). Note that AR(p) = ARMA(p,0). This process is the result of filtering white noise ($v(n)$) with variance σ_v^2 with

a causal linear shift-invariant filter with system function $H(z) = b(0)/A_p(z)$. The Yule-Walker equations are

$$r_{xx}(k) + \sum_{l=1}^{p} a_p(l) r_{xx}(k-l) = \sigma_v^2 |b(0)|^2 \delta(k), \quad k \geq 0.$$

Since this equation is linear in the coefficients $a_p(k)$ one can find $r_{xx}(k)$ explicitly.

Moving Average Process (MA): A process with power spectrum

$$P_{xx}(z) = \sigma_v^2 B_q(z) B_q^*(1/z^*), \quad P_{xx}(e^{j\omega}) = \sigma_v^2 |B_q(e^{j\omega})|^2,$$

where $B_q(z)$ is given above in the definition of an ARMA process is a moving average process of order q. Note that MA(q) = ARMA(0,q). This process is the result of filtering white noise with variance σ_v^2 with a filter with system function $H(z) = B_q(z)$. The Yule-Walker equations are

$$r_{xx}(k) = \sigma_v^2 b_q(k) * b_q^*(-k) = \sigma_v^2 \sum_{l=0}^{q-|k|} b_q(l+|k|) b_q^*(l).$$

3 Algorithms for Signal Modeling

Note: Each of the following algorithms use difference definitions of the autocorrelation sequence r_{xx} and thus different autocorrelation matrices. We also note that $\varepsilon_p = \sigma_p^2 = b(0)$.

Prony's Method: In Prony's Method we attempt to approximate a signal $x(n)$ by $h(n) = h(n) * \delta(n)$ with transfer function $H(z) = B_q(z)/A_p(z)$ with $B_q(z)$ and $A_p(z)$ as given above. We find $h[n]$ by minimizing the error $E(z) = A_p(z)X(z) - B_q(z)$. Now let

$$r_{xx} = \sum_{n=q+1}^{\infty} x(n-l) x^*(n-k).$$

To find $a_p(k)$ we solve

$$\mathbf{R}_{xx} \mathbf{a}_p = \varepsilon_{p,q} \mathbf{u}_1,$$

where $(\mathbf{R}_{xx})_{m,n} = r_{xx}(m,n)$ for $0 \leq m, n \leq p$, $\mathbf{a}_p = (1, a_p(1), \ldots, a_p(p))^T$, and $\mathbf{u}_1 = (1, 0, \ldots, 0)^T$. We determine $b_q(k)$ by setting the error, $e[n]$ to zero for $n = 0, 1, \ldots, q$, i.e.

$$b_q(n) = x(n) + \sum_{k=1}^{p} a_p(k) x(n-k), \quad n = 0, 1, \ldots, q.$$

Note that $\mathbf{R}_{xx} = \mathbf{X}_q^H \mathbf{X}_q$, where $\mathbf{X}_q \in \mathbb{R}^{\infty \times p}$ such that

$$\mathbf{X}_q = \begin{bmatrix} x(q) & x(q-1) & \cdots & x(q-p+1) \\ x(q+1) & x(q) & \cdots & x(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}.$$

We also note that one can solve for $a_p(k)$ by $\bar{\mathbf{R}}_{xx}\bar{\mathbf{a}}_p = -\mathbf{r}_{xx}$, where $(\bar{\mathbf{R}}_{xx})_{m,n} = r_{xx}(m,n)$ for $1 \le m,n \le p$, $\bar{\mathbf{a}}_p = (a_p(1),\ldots,a_p(p))^T$, and $\mathbf{r}_{xx} = (r_{xx}(1,0),\ldots,r_{xx}(p,0))^T$.

Autocorrelation Method: Suppose that $x[n]$ is known for $n = 0,\ldots,N$. We attempt to approximate $x[n]$ with an all-pole model with a variant of Prony's Method. We wish to minimize the error: $\varepsilon_p \equiv \sum_{n=0}^{\infty}|e[n]|^2$, where $e[n] = x[n] + \sum_{k=1}^{p} a_p(k)x(n-k)$. To minimize ε_p we assume $x[n] = 0$ for $n < 0$ and $n > N$, i.e., we use Prony's Method on the signal $\widetilde{x}(n) = x(n)w_R(n)$, where $w_R(n) = 1$ for $n = 0,1,\ldots,N$ and zero otherwise (rectangular window). Let $r_{xx} = \sum_{n=k}^{N} x(n)x^*(n-k)$ for $k \ge 0$. We find $a_p(k)$ by the least squares solution of

$$\mathbf{X}_p\bar{\mathbf{a}}_p = -\mathbf{x}_1$$

which is

$$\bar{\mathbf{a}}_p = -(\mathbf{X}_p^H\mathbf{X}_\mathbf{p})^{-1}\mathbf{X}_p^H\mathbf{x}_1,$$

where $\bar{\mathbf{a}}_p = (a_p(1),\ldots,a_p(p))^T$, $\mathbf{x}_1 = (x(1),\ldots,x(N),0,\ldots,0)^T \in \mathbb{R}^{N+p}$ and $\mathbf{X}_p \in \mathbb{R}^{(N+p)\times p}$ such that

$$\mathbf{X}_p = \begin{bmatrix} x(0) & 0 & 0 & \cdots & 0 \\ x(1) & x(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(p-1) & x(p-2) & x(p-3) & \cdots & x(0) \\ x(p) & x(p-1) & x(p-2) & \cdots & x(1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(N-1) & x(N-2) & x(N-3) & \cdots & x(N-p-1) \\ x(N) & x(N-1) & x(N-2) & \cdots & x(N-p) \\ 0 & x(N) & x(N-1) & \cdots & x(N-p+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x(N) \end{bmatrix}.$$

Covariance Method: Suppose that $x[n]$ is known for $n = 0,\ldots,N$. We attempt to approximate $x[n]$ with an all-pole model with a variant of Prony's Method. We wish to minimize the error: $\varepsilon_p^N \equiv \sum_{n=p}^{\infty}|e[n]|^2$, where $e[n] = x[n] + \sum_{k=1}^{p} a_p(k)x(n-k)$. If we let $r_{xx}(k,l) = \sum_{n=p}^{N} x(n-l)x^*(n-k)$, then to find $a_p(k)$ we simply use Prony's Method (with $b_q(k) = 0$), i.e. we solve $\bar{\mathbf{R}}_{xx}\bar{\mathbf{a}}_p = -\mathbf{r}_{xx}$.

Levinson-Durbin Recursion: The Levinson-Durbin Recursion is an algorithm to solve equations of the form: $\mathbf{R}_{xx}\mathbf{a} = \mathbf{b}$. The algorithm is recursive on the order of the model (i.e. size of the matrix $\mathbf{R}_{xx}$). Consider the special case $\mathbf{R}_{xx}\mathbf{a}_p = \varepsilon_p\mathbf{u}_1$. Then the algorithm is as follows:

1. Initialize the recursion
   (a) $a_0(0) = 1$
   (b) $\varepsilon_0 = r_{xx}(0)$
2. For $j = 0,1,\cdots p-1$
   (a) $\gamma_j = r_{xx}(j+1) + \sum_{i=1}^{j} a_j(i)r_{xx}(j-i+1)$
   (b) $\Gamma_{j+1} = -\gamma_j/\varepsilon_j$

(c) For $i = 1, 2, \ldots, j$

$$a_{j+1}(i) = a_j(i) + \Gamma_{j+1} a_j^*(j - i + 1)$$

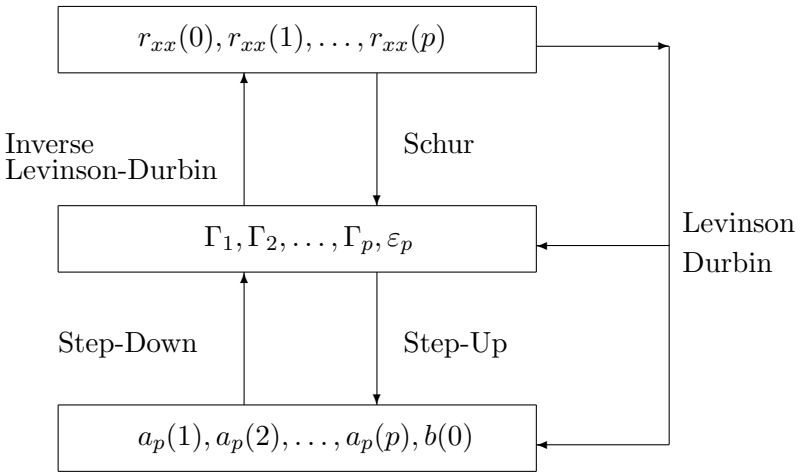
(d) $a_{j+1}(j+1) = \Gamma_{j+1}$

(e) $\varepsilon_{j+1} = \varepsilon_j[1 - |\Gamma_{j+1}|^2]$

3. $b(0) = \sqrt{\varepsilon_p}$

The term Γ_{j+1} is referred to as the $(j+1)$st reflection coefficient.

Mapping Between the Autocorrelation Sequence, the All-Pole Model, and the Reflection Coefficients: Below is a diagram describing how the autocorrelation sequence, the all-Pole model, and the reflection coefficients are related by according algorithm.



Linear Prediction: Linear prediction is concerned with the estimation of $x(n+1)$ in terms of a linear combination of the current and previous values of $x(n)$. Thus an FIR linear predictor of order $p-1$ has the form

$$\widehat{x}(n+1) = \sum_{k=0}^{p-1} w(k) x(n-k),$$

where $w(k)$ for $k = 0, \ldots, p-1$ are the coefficients of the prediction filter. The Wiener-Hopf equations for the optimum linear predictor are

$$\begin{bmatrix} r_{xx}(0) & r_{xx}^*(1) & \cdots & r_{xx}^*(p-1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}^*(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ \vdots \\ r_{xx}(p) \end{bmatrix}.$$

Selecting an Appropriate Model Order: If a model order is chosen too small, then the resulting spectrum will be smoothed and will have poor resolution. If, on the other hand, the model order is too large, then the spectrum may contain spurious peaks and may lead to spectral line splitting. Below are a few suggested technique to choose the order length. Note that these are only educated guesses, the correct model order is often found experimentally and is subjective choice. In the following we let ε_p be the model error, N is the data record length, and p is the

order of the model. In the following, we choose the p that minimizes each expression.

$$\begin{aligned}
\text{Akaike Information Criterion:} \quad & AIC(p) &&= N \log \varepsilon_p + 2p \\
\text{Minimum Description Length:} \quad & MDL(p) &&= N \log \varepsilon_p + p \log N \\
\text{Akaike's Final Prediction Error:} \quad & FPE(p) &&= \varepsilon \frac{N+p+1}{N-p-1} \\
\text{Parzen's Criterion Autoregressive Tranfer Function:} \quad & CAT(p) &&= \frac{1}{N} \sum_{j=1}^{p} \frac{N-j}{N\varepsilon_j} - \frac{N-p}{N\varepsilon_p}
\end{aligned}$$

4 Non-parametric Spectrum Estimation

Periodogram: The periodogram is a method for approximating the power spectrum of a wss process $x[n]$. The periodogram is of $x[n]$ for $n = 0, \ldots, L-1$ is

$$I(e^{j\omega}) = P_{per}(e^{j\omega}) = \frac{1}{L}|X(e^{j\omega})|^2,$$

where $X[k]$ is the DFT of $x[n]$.

Modified Periodogram: Let $w[n]$ be a window of size L such that $U = \|w\|_{l^2}^2 = \frac{1}{L}\sum_{n=0}^{L-1}|w[n]|^2$ and $x[n]$ be a wss process. Then the modified periodogram of $x[n]$ is

$$I(e^{j\omega}) = P_M(e^{j\omega}) = \frac{1}{LU}\left|\sum_{n=-\infty}^{\infty} x[n]w[n]e^{-jn\omega}\right|^2.$$

Bartlett's Method: Let $x_i[n] = x[n+iL]$ for $n = 0, 1, \ldots, L-1$ and $i = 0, 1, \ldots, K-1$ and let

$$P_{per}^{(i)}(e^{j\omega}) = \frac{1}{L}\left|\sum_{n=0}^{L-1} x_i[n]e^{-jn\omega}\right|^2.$$

Then Bartlett's Method approximates the PSD by

$$P_B(e^{j\omega}) = \frac{1}{K}\sum_{i=0}^{K-1} P_{per}^{(i)}(e^{j\omega}) = \frac{1}{KL}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} x[n+iL]e^{-j\omega n}\right|^2.$$

Welch's Method: Let $x[n]$ be a wss process of length N. Let $w[n]$ be a window of size L such that $U = \frac{1}{L}\|w\|_{l^2}^2$. Let $1 \le D \le L$ and $N = L + D(K-1)$. Let $x_i[n] = x[n+iD]$, $n = 0, 1, \ldots, L-1$. Then Welch's method approximates the PSD by

$$P_W(e^{j\omega}) = \frac{1}{K}\sum_{i=0}^{K-1} P_M^{(i)}(e^{j\omega}) = \frac{1}{KLU}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} w[n]x[n+iD]e^{-jn\omega}\right|^2.$$

Note that the amount of overlap between $x_i[n]$ and $x_{i+1}[n]$ is $L-D$ points. Thus for 50 percent overlap we have $D = L/2$ and $K = 2\frac{N}{L} - 1$.

Properties of the Above Spectrum Estimators: Let w_B be the rectangular window. Welch's method estimates assume 50 percent overlap and Bartlett window. Variance estimates are approximate.

Estimator	Bias	Resolution	Variance
Periodogram	$\frac{1}{2\pi}P_{xx}(e^{j\omega}) * W_B(e^{j\omega})$	$\Delta\omega = 0.89\frac{2\pi}{N}$	$P_{xx}^2(e^{j\omega})$
Modified Periodogram	$\frac{1}{2\pi LU}P_{xx}(e^{j\omega}) * \lvert W(e^{j\omega})\rvert^2$	window dependent	$P_{xx}^2(e^{j\omega})$
Bartlett's Method	$\frac{1}{2\pi}P_{xx}e^{j\omega} * W_B(e^{j\omega})$	$0.89K\frac{2\pi}{N}$	$\frac{1}{K}P_{xx}^2(e^{j\omega})$
Welch's Method	$\frac{1}{2\pi LU}P_{xx}(e^{j\omega}) * \lvert W(e^{j\omega})\rvert^2$	window dependent	$\frac{9}{16}\frac{L}{N}P_{xx}^2(e^{j\omega})$

Minimum Variance Method: Let $\mathbf{e} = (1, e^{j\omega}, \ldots, e^{jp\omega})^T$, $\mathbf{R}_{xx}$ be the (approximate) autocorrelation matrix, and p be the order of the method. Then the minimum variance estimates the spectrum by

$$P_{MV}(e^{j\omega}) = \frac{p+1}{\mathbf{e}^H\mathbf{R}_{xx}^{-1}\mathbf{e}}.$$

Maximum Entropy Method: Let $\mathbf{e} = (1, e^{j\omega}, \ldots, e^{jp\omega})^T$. The maximum entropy method estimates the spectrum by

$$P_{mem}(e^{j\omega}) = \frac{\varepsilon_p}{|\mathbf{e}^H\mathbf{a}_p|^2},$$

where ε_p and $a_p(k)$ satisfy

$$\begin{bmatrix} r_{xx}(0) & r_{xx}^*(1) & \cdots & r_{xx}^*(p) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}^*(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(p) & r_{xx}(p-1) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ \vdots \\ a_p(p) \end{bmatrix} = \varepsilon_p \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$