AR Spectral Estimation In Theory: Connections to Linear Prediction

AR Spectral Estimation in Theory

Assume $x[n]$ is an AR(p) process and that σ_w^2, p and $\gamma[k]$ are <u>known</u>.

The parameters $a[1], \ldots, a[p]$ are the solution of the following linear system of equations:

$$\begin{bmatrix} \gamma[0] & \gamma[-1] & \cdots & \gamma[-(p-1)] \\ \gamma[1] & \gamma[0] & \cdots & \gamma[-(p-2)] \\ \vdots & \vdots & \ddots & \vdots \\ \gamma[p-1] & \gamma[p-2] & & \gamma[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = - \begin{bmatrix} \gamma[1] \\ \gamma[2] \\ \vdots \\ \gamma[p] \end{bmatrix}$$

Yule-Walker equations.

In theory, we just solve this system of eqns.

In practice, we need to estimate γ first.

$\boxed{\text{next lecture}}$

Linear Prediction

Problem Statement:

Given p consecutive observations of a WSS random process

$$x[n-p], x[n-p+1], \ldots, x[n-2], x[n-1]$$

predict the next observation $x[n]$ using the linear predictor

$$\hat{x}[n] = -\sum_{k=1}^{p} a_p[k]\, x[n-k]$$

that minimizes the power of the prediction error $e[n] = x[n] - \hat{x}[n]$

$$\rho_p = E\{|e[n]|^2\} = E\{|x[n] - \hat{x}[n]|^2\}$$

As we will see, if x is AR(p), the solution to the linear prediction problem is given by the AR coefficients. More generally (when x is not AR(p)), the linear prediction problem is given by the Yule-Walker equations.

$\boxed{\text{Assume everything is real-valued}}$

$$\rho_p = E\left\{ \left| x[n] - \hat{x}[n] \right|^2 \right\} \qquad \underline{a}_p := \begin{bmatrix} a_p[1] \\ \vdots \\ a_p[p] \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x[n-1] \\ \vdots \\ x[n-p] \end{bmatrix}$$

$$= E\left\{ \left| x[n] + \underline{a}_p^T \underline{x} \right|^2 \right\}$$

$$= E\left\{ |x[n]|^2 + 2 a_p^T \underline{x}\, x[n] + \underline{a}_p^T \underline{x}\, \underline{x}^T \underline{a}_p \right\}$$

$$= \gamma_0 + 2 a_p^T \underline{\gamma}_p + \underline{a}_p \Gamma_p \underline{a}_p \qquad (*)$$

where $\gamma_k := \gamma[k]$

$$\underline{\gamma}_p = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} \qquad \Gamma_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{-(p-1)} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{-(p-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{-(p-2)} & \cdots & \gamma_0 \end{bmatrix}$$

We need to find $\underline{a}_p$ that minimizes $(*)$

Consider the more general problem of minimizing (w.r.t. $\underline{a}$)

$$\underline{a}^T C \underline{a} - 2\underline{a}^T \underline{b} + \underline{d}$$

where C is a nonnegative definite matrix.

Solution one

$$\frac{\partial}{\partial \underline{a}}(\cdot) = 2C \cdot \underline{a} - 2\underline{b} = 0$$

$$\Longrightarrow \underline{a} = C^{-1}\underline{b}$$

Solution two

Complete the square:

$$\underline{a}^T C \underline{a} - 2\underline{a}^T b + d$$

$$= (\underline{a} - C^{-1}\underline{b})^T C (\underline{a} - C^{-1}\underline{b}) - \underline{b}^T C^{-1}\underline{b} + d$$

Since C is nonneg. def, this expression is minimized when $\underline{a} = C^{-1}\underline{b}$

In our case we have

$$C = \Gamma_p \,, \quad \underline{b} = -\underline{\gamma}_p \,, \quad d = \gamma_0$$

$$\Longrightarrow \quad \underline{a}_p = -\Gamma_p^{-1} \cdot \underline{\gamma}_p$$

or, equivalently,

$$\boxed{-\Gamma_p \cdot \underline{a}_p = \underline{\gamma}_p}$$

Look familiar? They're the Yule-Walker equations. In the context of linear prediction they are also called the <u>Weiner-Hopf</u> equations or <u>normal</u> equations.

The corresponding prediction error power is

$$\rho_p = E\{|e[n]|^2\} =$$

$\boxed{\text{The Orthogonality Principle}}$

"The predictor values $x[n-p], \ldots, x[n-1]$ are <u>orthogonal</u> to the error." That is,

$$E\left\{ \left(x[n] - \hat{x}[n] \right) \cdot x[n-k] \right\} = 0$$

for $k = 1, 2, \ldots, p$.

Proof:
$$E\left\{ \left(x[n] - \left(\Gamma_p^{-1} \cdot \underline{\gamma}_p \right)^T \underline{x} \right) x[n-k] \right\}$$

$$= \gamma[k] - \underline{\gamma}_p^T \Gamma_p^{-1} \underbrace{E\left\{ \underline{x} \cdot x[n-k] \right\}}_{k^{th} \text{ column of } \Gamma_p}$$

$$= \gamma[k] - \gamma_p^T \cdot \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow k^{th} \text{ position}$$

$$= \gamma[k] - \gamma[k]$$

$$= 0.$$

Using the orthogonality principle,

$$\rho_p = E\left\{ \left(x[n] - \hat{x}[n]\right)\left(x[n] - \hat{x}[n]\right) \right\}$$

$$= E\left\{ \left(x[n] - \hat{x}[n]\right) x[n] \right\}$$

$$= E\left\{ \left(x[n] + \left(\Gamma_p^{-1} \underline{\gamma}_p\right)^T \underline{x} \right) \cdot x[n] \right\}$$

$$= \gamma_0 + \underline{\gamma}_p^T \Gamma_p^{-1} \underline{\gamma}_p$$

The Prediction Error Filter

$$e[n] = x[n] + \sum_{k=1}^{p} a_p[k]\, x[n-k]$$

$$\Big\downarrow \text{ z-transform}$$

$$E(z) = X(z)\underbrace{\left[1 + \sum_{k=1}^{p} a_p[k]\, z^{-k}\right]}_{= A(z)}$$

<u>The prediction error filter</u>

$$x[n] \longrightarrow \boxed{A(z)} \longrightarrow e[n]$$

Theorem If $x[n]$ is $AR(p)$ with driving noise power σ^2, then $e[n]$ is white and $\rho_p = \sigma^2$.

Proof: Suppose

$$x[n] = -\sum a_p[k]\, x[n-k] + w[n].$$

The YW eqns. tell us $-\Gamma_p \underline{a}_p = \underline{\gamma}_p$.

The WH eqns tell us the optimal predictors satisfy $-\Gamma_p \underline{a}_p = \underline{\gamma}_p$. Hence

$$\hat{x}[n] = -\sum a_p[k]\, x[n-k]$$

$$\Rightarrow \quad e[n] = x[n] - \hat{x}[n]$$

$$= w[n]$$

And $\rho_p = E\{|e[n]|^2\} = E\{|w[n]|^2\} = \sigma^2$.

$$\underbrace{w[n] \longrightarrow \boxed{\frac{1}{A(z)}} \longrightarrow}_{\text{SYNTHESIS}} \underbrace{x[n] \longrightarrow \underset{\substack{\uparrow \\ \text{PEF}}}{\boxed{A(z)}} \longrightarrow}_{\text{ANALYSIS}} w[n] = e[n]$$

Whitening

If $x[n]$ is $AR(p)$, then the p-th order PEF <u>whitens</u> the data, extracting all correlation info

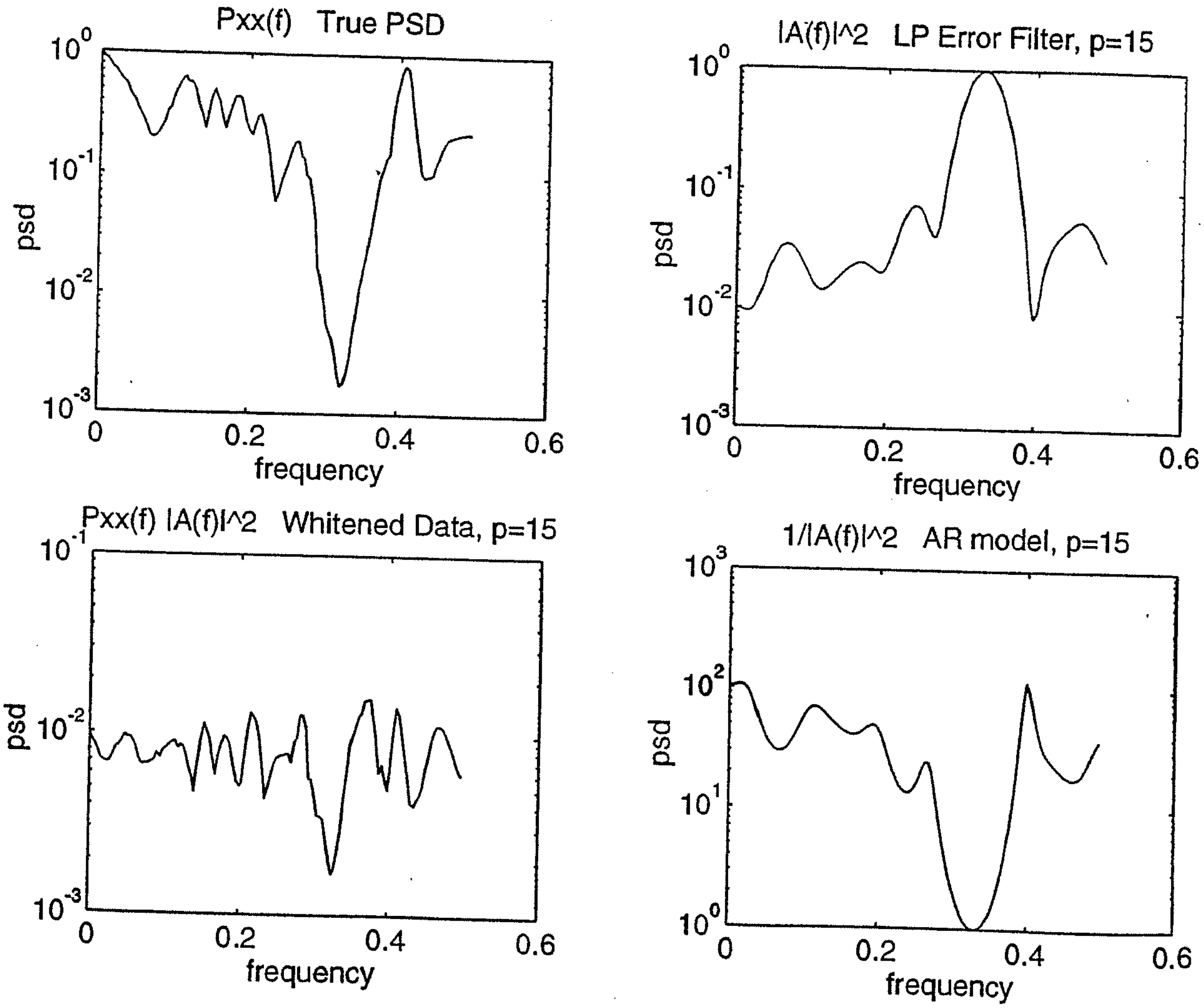
If $x[n]$ is not $AR(p)$, then the whitening is only <u>approximate</u> (the error $e[n]$ is not white but exhibits less correlation than $x[n]$).

If we allow non-white <u>driving noise</u>, every RP has an AR-like representation:

$$e[n] \longrightarrow \boxed{\frac{1}{A(z)}} \longrightarrow x[n]$$

<u>Innovations sequence</u> (not white!) supplies not only the random element of $x[n]$, but also provides any structure in $x[n]$ that cannot come out of $\frac{1}{A(z)}$

$x[n]$ is not AR(15)



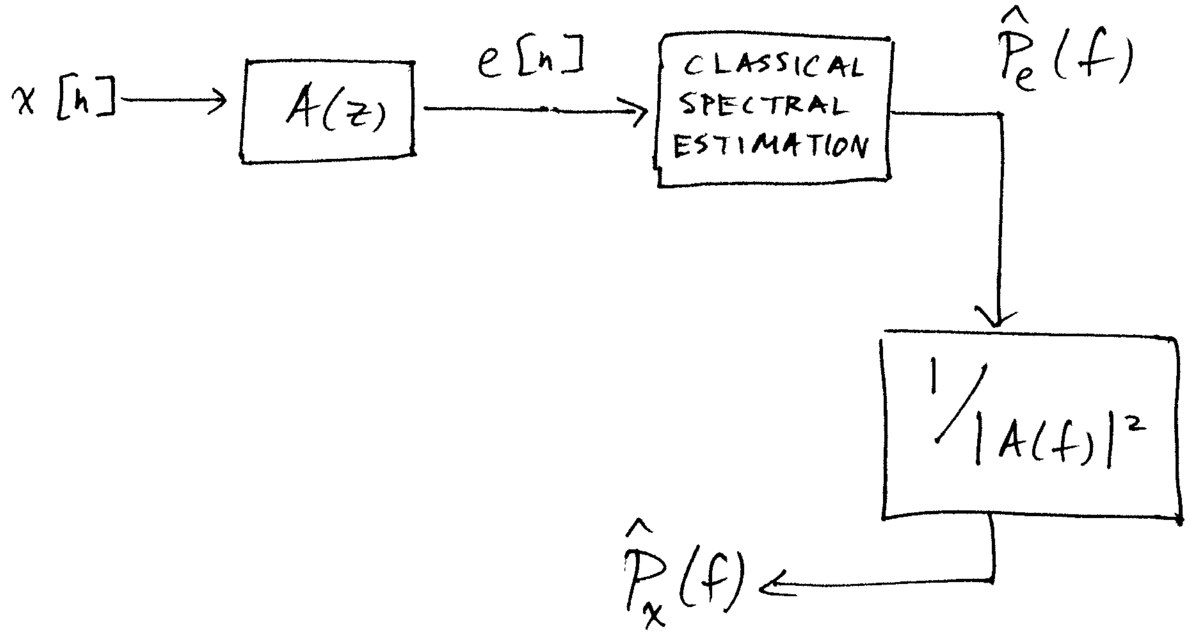
flatter than $P_{xx}(f)$

<u>Theorem</u> If $p = \infty$, then the PEF makes <u>any</u> RP perfectly white.

<u>Proof</u>: By orthogonality principle, $e[n] \perp$ all past data.

Since $e[n-1] = x[n-1] - \hat{x}[n-1]$ is a linear combo of these past values, $e[n] \perp e[n-1]$. Similar argument applies to all other lags > 1

Pre-Whitening Filters



Why? Recall that for white noise, classical methods are <u>unbiased</u>. For approximately white processes, the bias is also lessened, even if shorter windows are used $\Rightarrow$ can decrease window size to <u>decrease variance</u>.

Effictive and widely used technique in practice.

The PEF is Minimum Phase

Recall Any rational spectrum $P_{xx}(z) = \sigma^2 \left| \frac{B(z) B^*(1/z^*)}{A(z) A^*(1/z^*)} \right|^2$

can be represented by a filter $H(z)$ that is both $\underbrace{\text{causal / stable}}_{|\text{poles}| < 1}$ and $\underbrace{\text{minimum phase}}_{|\text{zeros}| \leq 1}$

since poles and zeros of $P_{xx}(z)$ come in pairs that are reciprocals.

Question We know the Yule-Walker equations provide <u>an</u> $AR(p)$ model that matches a given ACV function. But is the corresponding filter $\sigma^2/A(z)$ causal & stable? In other words, is $A(z)$ minimum phase?

Answer Yes. To prove this, we rely on the connection to linear prediction once again.

<u>Theorem</u> If the autocovariance of a random process $x[n]$ (arbitrary; not necessarily AR)

$$\Gamma_p = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots & \gamma_{-(p-1)} \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots & \gamma_{-(p-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix}$$

is positive definite, then the prediction error filter $A(z) = \left(1 + \sum a_p[k] z^{-k}\right)$ is minimum phase.

$\boxed{\text{solution to YW/WH eqns.}}$

<u>Proof:</u> Consider the error sequence $e[n] = x[n] - \hat{x}[n]$.

Recall $\underline{a}_p$ was choosen such that

$$\rho = E\left\{ |e[n]|^2 \right\} = \gamma_{ee}[0]$$

is minimized.

The PSD of $e[n]$ is

$$P_{ee}(f) =$$

Recall that $\rho = \gamma_{ee}[0]$ is given by

$$\gamma_{ee}[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ee}(f)\,df$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 P_{xx}(f)\,df$$

So $\underline{a}_p$ minimizes this expression

<u>Claim</u>: Write $A(z) = (1-\alpha_1 z^{-1}) \cdots (1-\alpha_p z^{-1})$

If $|\alpha_i| > 1$ for some i, then the prediction power ρ can be decreased by replacing α_i by $1/\alpha_i^*$.

To see this, suppose $|\alpha_i| > 1$. Write

$$A(z) = (1-\alpha_i z^{-1}) \cdot \tilde{A}(z)$$

the remaining poles.

We have

$$\rho = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left|1-\alpha_i e^{-2\pi i f}\right|^2 \left|\tilde{A}(f)\right|^2 P_{XX}(f)\, dx$$

Now observe

$$\left|1-\alpha_i e^{-2\pi i f}\right|^2 = |\alpha_i|^2 \cdot \left|\frac{1}{\alpha_i} - e^{-2\pi i f}\right|^2$$

$$= |\alpha_i|^2 \cdot \left| e^{2\pi i f} - \frac{1}{\alpha_i^*}\right|^2$$

$$\boxed{\begin{array}{l} |-z| = |z| \\ |z^*| = |z| \end{array}}$$

$$= |\alpha_i|^2 \cdot \left|1 - \frac{1}{\alpha_i^*} e^{-2\pi i f}\right|^2$$

$$\boxed{\left|e^{2\pi i f}\right| = 1}$$

$$> \left|1 - \frac{1}{\alpha_i^*} e^{-2\pi i f}\right|^2$$

$$\boxed{\text{since } |\alpha_i| > 1 \text{ by assumption}}$$

Hence, the error prediction power is decreased by swapping a pole outside the unit circle with it conjugate-inverse inside the unit circle.

The Levinson Algorithm

Given a $p \times p$ positive definite, Toeplitz matrix Γ_p and a $p \times 1$ vector $\underline{\gamma}_p$, the Levinson algorithm is an $O(p^2)$ algorithm for computing

$$\underline{a}_p = -\Gamma_p^{-1} \underline{\gamma}_p .$$

The algorithm is also called the Levinson-Durbin algorithm (Durbin refined Levinson's initial solution).

The algorithm is recursive: It computes

$$\underline{a}_p = \begin{bmatrix} \underline{a}_{p-1} & \underline{d}_{p-1} \\ 0 & k_p \end{bmatrix}$$

where $\underline{d}_{p-1}, k_p$ are the "update coefficients."

Write

$$\Gamma_p = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \cdots & \gamma_{-(p-1)} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{-(p-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}, \qquad \underline{\gamma}_p = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix}$$

Durbin's modification takes advantage of the relationship between Γ_p and $\underline{\gamma}_p$.

<u>Base of recursion</u>:

$$-\left[\gamma_0\right] \cdot \left[\underline{a}_1\right] = \left[\gamma_1\right]$$

$$\Longrightarrow \quad a_1 = -\frac{\gamma_1}{\gamma_0}$$

<u>General recursion</u>:

We will capitalize on the recursive structure of Γ_p, $\underline{\gamma}_p$.

By the YW/WH/normal equations,

$$-\Gamma_p \cdot \underline{a}_p = \underline{\gamma}_p .$$

Now observe

$$\Gamma_p = \begin{bmatrix} \Gamma_{p-1} & \underline{\tilde{\gamma}}_{p-1} \\ \underline{\tilde{\gamma}}_{p-1}^T & \gamma_0 \end{bmatrix}, \quad \underline{\tilde{\gamma}}_p = \underline{\gamma}_p \text{ upside-down}$$

and

$$\underline{\gamma}_p = \begin{bmatrix} \underline{\gamma}_{p-1} \\ \gamma_p \end{bmatrix}$$

Therefore,

$$-\begin{bmatrix} \Gamma_{p-1} & \underline{\tilde{\gamma}}_{p-1} \\ \underline{\tilde{\gamma}}_{p-1}^T & \gamma_0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} \underline{a}_{p-1} \\ 0 \end{bmatrix} + \begin{bmatrix} \underline{d}_{p-1} \\ k_p \end{bmatrix} \right\} = \begin{bmatrix} \underline{\gamma}_{p-1} \\ \gamma_p \end{bmatrix}$$

From this, we get two sets of equations:

$$\Gamma_{p-1}\,\underline{a}_{p-1} + \Gamma_{p-1}\,\underline{d}_{p-1} + \tilde{\underline{\gamma}}_{p-1}\,k_p = -\underline{\gamma}_{p-1} \qquad \boxed{\text{vector equation}}$$

$$\tilde{\underline{\gamma}}_{p-1}^{T}\,\underline{a}_{p-1} + \tilde{\underline{\gamma}}_{p-1}^{T}\,\underline{d}_{p-1} + r_0\,k_p = -\gamma_p \qquad \boxed{\text{scalar equation}}$$

<u>Goal</u>: solve this system of equations for $\underline{d}_{p-1}$, k_p.

How can we simplify the first equation?

From the normal equations of order $p-1$, we know

$$\Gamma_{p-1} \, \underline{a}_{p-1} = - \underline{\gamma}_{p-1}$$

$$\Rightarrow \quad \Gamma_{p-1} \, \underline{d}_{p-1} + \underline{\tilde{\gamma}}_{p-1} \, k_p = 0$$

$$\Rightarrow \quad \underline{d}_{p-1} = -k_p \cdot \Gamma_{p-1}^{-1} \cdot \underline{\tilde{\gamma}}_{p-1}$$

<u>Claim</u>: $\quad \Gamma_{p-1}^{-1} \, \underline{\tilde{\gamma}}_{p-1} = - \underline{\tilde{a}}_p$

<u>Proof</u>: Since Γ_{p-1} is Toeplitz, if

$$\Gamma_{p-1} \, \underline{a}_{p-1} = - \underline{\gamma}_{p-1}$$

then

$$\Gamma_{p-1} \, \underline{\tilde{a}}_{p-1} = - \underline{\tilde{\gamma}}_{p-1}$$

<u>Exercise</u>: verify for $p-1 = 3$.

$\boxed{\text{Assume everything is real}}$

$$k_p = -\frac{\gamma_p + \tilde{\underline{\gamma}}_{p-1}^T \underline{a}_{p-1}}{\gamma_0 + \tilde{\underline{\gamma}}_{p-1}^T \cdot \tilde{\underline{a}}_{p-1}}$$

$$= -\frac{\gamma_p + \tilde{\underline{\gamma}}_{p-1}^T \underline{a}_{p-1}}{\gamma_0 + \underline{\gamma}_{p-1}^T \underline{a}_{p-1}}$$

$$= \boxed{-\frac{\gamma[p] + \sum_{k=1}^{p-1} a_p[k]\gamma[p-k]}{\gamma[0] + \sum_{k=1}^{p-1} a_p[k]\gamma[k]}}$$

The coefficients k_p are called <u>reflection coefficients</u>.

<u>Useful fact</u>: The following are equivalent:

1. $A_p(z)$ is minimum phase and $\rho_{MIN} > 0$
2. Γ_{p+1} is positive definite.
3. $\gamma[0] > 0$ and $|k_i| < 1$, $1 \le i \le p$.

<u>Another fact</u>: The prediction error power after i steps given by

$$\rho_i = (1 - |k_i|^2)\, \rho_{i-1}$$

and therefore

$$\rho_{min} = \gamma[0] \cdot \prod_{i=1}^{p} (1 - |k_i|^2)$$

Read / Skim Kay Chapter 6 to learn about

- other properties of reflection coeffs.
- connections to backward error prediction and lattice filters
- other problems that are equivalent to AR spectral estimation
  - maximum entropy
  - maximum spectral flatness.