COM521500
Math. Methods for SP I
Lecture 10: Toeplitz Matrices and Linear Prediction

Toeplitz Matrices

A matrix of the form

$$\mathbf{R} = \begin{bmatrix} r_0 & r_{-1} & \dots & \dots & r_{-K} \\ r_1 & r_0 & r_{-1} & & \vdots \\ \vdots & r_1 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & r_{-1} \\ r_K & \dots & \dots & r_1 & r_0 \end{bmatrix}$$

is called a Toeplitz matrix.

Linear Prediction

Signal model for an AR process:

$$x[n] = \sum_{i=1}^{K} h_i x[n-i] + w[n]$$

where $\{h_i\}$ is the set of AR coefficients, and $w[n]$ is a time-uncorrelated WSS process.

We assume that the signals are real valued.

Minimum-mean-squared-error (MMSE) estimation

$$\mathbf{h}_{opt} = \arg \min_{\mathbf{h} \in \mathbb{R}^K} \mathrm{E}\left\{ \left(x[n] - \sum_{i=1}^{K} h_i x[n-i] \right)^2 \right\}$$

Let $f(\mathbf{h}) = \mathrm{E}\left\{ \left(x[n] - \sum_{i=1}^{K} h_i x[n-i] \right)^2 \right\}$. Then,

$$f(\mathbf{h}) = r_0 - 2\sum_{i=1}^{K} h_i r_{-i} + \sum_{i=1}^{K} \sum_{k=1}^{K} h_i h_k r_{k-i}$$

where $r_i = \mathrm{E}\{x[n+i]x[n]\}$ is the auto-correlation sequence of $x[n]$.

Using the symmetry of auto-correlation sequences (i.e., $r_i = r_{-i}$), the objective function f can be written as

$$f(\mathbf{h}) = r_0 - 2\mathbf{r}_p^T\mathbf{h} + \mathbf{h}^T\mathbf{R}\mathbf{h}$$

where

$$\mathbf{r}_p = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_K \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} r_0 & r_{-1} & \dots & r_{-K+1} \\ r_1 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{K-1} & & r_1 & r_0 \end{bmatrix}.$$

As a correlation matrix, $\mathbf{R}$ is symmetric and PSD.

Assume that $\mathbf{R}$ is also PD, for simplicity.

By finding the gradient of f, the solution $\mathbf{h}_{opt}$ is obtained as

$$\mathbf{h}_{opt} = \mathbf{R}^{-1}\mathbf{r}_p \qquad (*)$$

$(*)$ is known as the **Wiener-Hopf equations.**

LS: an alternative to AR coefficient estimation

By letting $\mathbf{x}[n] = [\ x[n-1], x[n-2], \ldots, x[n-K]\ ]^T$, and

$$\mathbf{x}_p = \begin{bmatrix} x[K] \\ \vdots \\ x[N] \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}^T[K] \\ \vdots \\ \mathbf{x}^T[N] \end{bmatrix},$$

where N is the data length, the AR signal model can be written in a matrix form

$$\mathbf{x}_p = \mathbf{X}\mathbf{h} + \mathbf{w}$$

LS solution to $\mathbf{h}$:

$$\mathbf{h}_{LS} = \arg\min_{\mathbf{h}\in\mathbb{R}^K} \|\mathbf{x}_p - \mathbf{X}\mathbf{h}\|_2^2$$

Using the LS concepts, we obtain

$$\mathbf{h}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}_p$$

Note that a mild assumption of nonsingular $\mathbf{X}^T\mathbf{X}$ has been made.

Relationship of MMSE estimation and LS

By noting that

$$\mathbf{X} = \begin{bmatrix} x[0] & x[1] & \dots & x[K-1] \\ x[1] & x[2] & \dots & x[K] \\ \vdots & & & \vdots \\ x[N-K] & \dots & \dots & x[N-1] \end{bmatrix},$$

the following properties are obtained:

$$\lim_{N\to\infty} \mathbf{X}^T\mathbf{X} = \mathbf{R}, \qquad \lim_{N\to\infty} \mathbf{X}^T\mathbf{x}_p = \mathbf{r}_p$$

Hence, the LS method approaches the MMSE estimation method when the data length N is sufficiently large.

Forward/Backward Prediction-Error Eqns.

We focus on the MMSE estimation framework, instead of LS.

The MMSE is

$$\begin{aligned} \sigma^2 &= f(\mathbf{h}_{opt}) \\ &= r_0 - \mathbf{r}_p^T \mathbf{h}_{opt} \\ &= [\ r_0, r_1, \dots, r_K\ ] \begin{bmatrix} 1 \\ -\mathbf{h}_{opt} \end{bmatrix} \end{aligned}$$

On the other hand, the Wiener-Hopf eqns. can be reorganized as

$$[\ \mathbf{r}_p\ \mathbf{R}\ ] \begin{bmatrix} 1 \\ -\mathbf{h}_{opt} \end{bmatrix} = \mathbf{0}$$

By augmenting the MMSE to the above eqn., we obtain

$$\underbrace{\begin{bmatrix} r_0 & r_{-1} & \dots & r_{-K+1} \\ r_1 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{K-1} & & r_1 & r_0 \end{bmatrix}}_{\mathbf{R}_K} \begin{bmatrix} 1 \\ -h_1 \\ \vdots \\ -h_K \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

These are called the forward prediction error eqns.

Exploiting the symmetric and Toeplitz properties of $\mathbf{R}_K$, we obtain, from the forward prediction eqns.,

$$\begin{bmatrix} r_0 & r_{-1} & \dots & r_{-K+1} \\ r_1 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{K-1} & & r_1 & r_0 \end{bmatrix} \begin{bmatrix} -h_K \\ \vdots \\ -h_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma^2 \end{bmatrix}$$

These are called the backward prediction error eqns.

Levinson-Durbin Recursion (LDR)

Suppose that we have a $(m-1)$th order AR solution

$$\begin{bmatrix} r_0 & \dots & r_{-m+1} \\ \vdots & \ddots & \vdots \\ r_{m-1} & \dots & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{h}^{(m-1)} \end{bmatrix} = \begin{bmatrix} \sigma^2_{(m-1)} \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{h}^{(m-1)} \in \mathbb{R}^{(m-1)}$.

We seek to find ways of obtaining mth order AR solution from the $(m-1)$th.

LDR is shown by construction. We suppose that

$$\begin{bmatrix} 1 \\ -h_1^{(m)} \\ \vdots \\ -h_{m-1}^{(m)} \\ -h_m^{(m)} \end{bmatrix} = \begin{bmatrix} 1 \\ -h_1^{(m-1)} \\ \vdots \\ -h_{m-1}^{(m-1)} \\ 0 \end{bmatrix} + \rho_m \begin{bmatrix} 0 \\ -h_{m-1}^{(m-1)} \\ \vdots \\ -h_1^{(m-1)} \\ 1 \end{bmatrix}$$

for some coefficient ρ_m.

The coefficients ρ_m are known as the **partial correlation coefficients** (PARCOR).

Forward prediction:

$$\mathbf{R}_m \begin{bmatrix} 1 \\ -h_1^{(m-1)} \\ \vdots \\ -h_{m-1}^{(m-1)} \\ 0 \end{bmatrix} = \begin{bmatrix} & & & r_{-m} \\ & \mathbf{R}_{m-1} & & \vdots \\ & & & r_{-1} \\ r_m & \ldots & r_1 & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ -h_1^{(m-1)} \\ \vdots \\ -h_{m-1}^{(m-1)} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2_{(m-1)} \\ \mathbf{0} \\ \Delta^{(m-1)} \end{bmatrix}$$

where $\Delta^{(m-1)} = r_m - \sum_{i=m-1}^{1} r_i h_{m-i}^{(m-1)}$.

Backward prediction:

$$\mathbf{R}_m \begin{bmatrix} 0 \\ -h_m^{(m-1)} \\ \vdots \\ -h_1^{(m-1)} \\ 1 \end{bmatrix} = \begin{bmatrix} r_0 & r_{-1} & \ldots & r_{-m} \\ r_1 & & & \\ \vdots & & \mathbf{R}_{m-1} & \\ r_m & & & \end{bmatrix} \begin{bmatrix} 0 \\ -h_m^{(m-1)} \\ \vdots \\ -h_1^{(m-1)} \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \Delta^{(m-1)} \\ \mathbf{0} \\ \sigma^2_{(m-1)} \end{bmatrix}$$

Combining the equations of forward and backward predictions,

$$\begin{bmatrix} \sigma^2_{(m-1)} \\ \mathbf{0} \\ \Delta^{(m-1)} \end{bmatrix} + \rho_m \begin{bmatrix} \Delta^{(m-1)} \\ \mathbf{0} \\ \sigma^2_{(m-1)} \end{bmatrix} = \begin{bmatrix} \sigma^2_{(m)} \\ \mathbf{0} \\ 0 \end{bmatrix}$$

Hence, the solution to ρ_m is

$$\rho_m = -\Delta^{(m-1)}/\sigma^2_{(m-1)}$$

and

$$\sigma^2_{(m)} = \sigma^2_{(m-1)}(1 - \rho_m^2)$$

LDR algorithm:

1. Set $m = 1$, $\Delta^{(0)} = r_1$, $\mathbf{h}^{(0)} = \{\}$ and $\sigma^2_{(0)} = r_0$.
2. From $\Delta^{(m-1)}$, $\sigma^2_{(m-1)}$ compute ρ_m, $\mathbf{h}^{(m)}$, $\sigma^2_{(m)}$, and $\Delta^{(m)}$.
3. If $m < K$, increment m and repeat Step 2.

The FLOPS of LDR is K^2, which is lower than the $K^3/3$ FLOPS required by using Cholesky decomposition to find $\mathbf{h}_{opt}$.

Toeplitz Factorizations

Note that we still assume $\mathbf{R}_K$ to be symmetric PD.

From the forward prediction-error eqns.,

$$\mathbf{R}_m \begin{bmatrix} 1 \\ -\mathbf{h}^{(m)} \end{bmatrix} = \begin{bmatrix} \sigma^2_{(m)} \\ \mathbf{0}_m \end{bmatrix}$$

Hence,

$$\mathbf{R}_K \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -\mathbf{h}^{(m)} \end{bmatrix} = \begin{bmatrix} \times \\ \vdots \\ \times \\ \sigma^2_{(m)} \\ \mathbf{0}_m \end{bmatrix}$$

Subsequently

$$\mathbf{R}_K \underbrace{\begin{bmatrix} 1 & & & \\ -h_1^{(K)} & 1 & & \\ -h_2^{(K)} & -h_1^{(K-1)} & & \\ \vdots & \vdots & \ddots & \\ -h_K^{(K)} & -h_{K-1}^{(K-1)} & & 1 \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} \sigma_{(K)}^2 & \times & \cdots & \times \\ & \sigma_{(K-1)}^2 & & \times \\ & & \ddots & \vdots \\ & & & \sigma_{(0)}^2 \end{bmatrix}}_{\mathbf{U}}$$

Consider $\mathbf{A}^T\mathbf{R}_K\mathbf{A}$.

$\mathbf{A}^T\mathbf{R}_K\mathbf{A}$ is symmetric as long as $\mathbf{R}_K$ is symmetric.

On the other hand, $\mathbf{A}^T\mathbf{R}_K\mathbf{A} = \mathbf{A}^T\mathbf{U}$ is a multiplication of two upper triangular matrices, which is upper triangular.

As a consequence, $\mathbf{A}^T\mathbf{R}_K\mathbf{A}$ is diagonal:

$$\mathbf{A}^T\mathbf{R}_K\mathbf{A} = \mathbf{D}$$

where

$$\mathbf{D} = \mathrm{Diag}(\sigma_{(K)}^2, \ldots, \sigma_{(0)}^2).$$

Let $\mathbf{B} = \mathbf{A}\mathbf{D}^{-1/2}$. We have that

$$\mathbf{B}^T \mathbf{R}_K \mathbf{B} = \mathbf{I}$$

or

$$\mathbf{R}_K = \mathbf{B}^{-T}\mathbf{B}^{-1}$$

which means $\mathbf{B}^T$ is **the inverse Cholesky factor** of $\mathbf{R}_K$.

Determinants and PARCOR

From the Toeplitz factorization, $\mathbf{R}_K = \mathbf{A}^{-T}\mathbf{D}\mathbf{A}^{-1}$. $\mathbf{A}$ is lower triangular with unit diagonals, and hence $\mathbf{A}^{-1}$ shares the same property. Subsequently,

$$\begin{aligned} \det(\mathbf{R}_K) &= \det(\mathbf{A}^{-T})\det(\mathbf{D})\det(\mathbf{A}) \\ &= \det(\mathbf{D}) = \prod_{m=1}^{K} \sigma^2_{(m)} \end{aligned}$$

where we recall $\sigma^2_{(m)} = r_0 \prod_{i=1}^{m}(1-\rho_m^2)$.

Since $\mathbf{R}_K$ is PD, it must be true that $|\rho_m| < 1$ for all m.

Lattice Filter

Let

$$H^{(m)}(z) = [\ 1\ \ -h_1^{(m)} \ldots - h_m^{(m)}\ ] \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-m} \end{bmatrix}$$

and

$$G^{(m)}(z) = [\ -h_m^{(m)} \ldots - h_1^{(m)}\ \ 1\ ] \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-m} \end{bmatrix}$$

Our objective is to implement $H^{(K)}(z)$, the linear prediction filter. From

$$\begin{bmatrix} 1 \\ -h_1^{(m)} \\ \vdots \\ -h_{m-1}^{(m)} \\ -h_{m-1}^{(m)} \end{bmatrix} = \begin{bmatrix} 1 \\ -h_1^{(m-1)} \\ \vdots \\ -h_{m-1}^{(m-1)} \\ 0 \end{bmatrix} + \rho_m \begin{bmatrix} 0 \\ -h_m^{(m-1)} \\ \vdots \\ -h_1^{(m-1)} \\ 1 \end{bmatrix},$$

we obtain

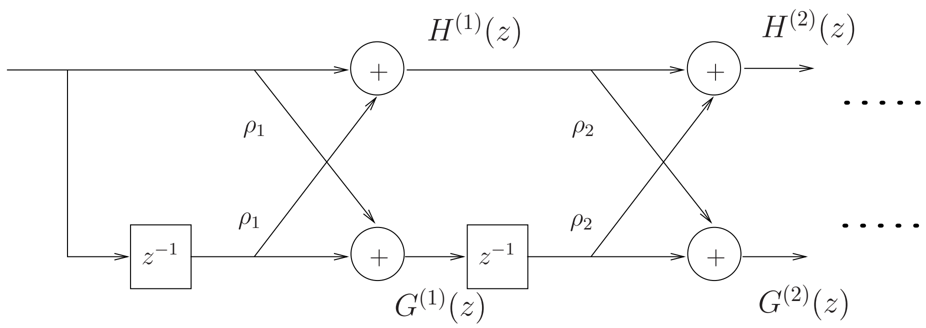
$$H^{(m)}(z) = H^{(m-1)}(z) + \rho_m z^{-1} G^{(m-1)}(z)$$

Consider the backward version of the previous prediction eqns.:

$$\begin{bmatrix} -h_m^{(m)} \\ \vdots \\ -h_2^{(m)} \\ -h_m^{(1)} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -h_{m-1}^{(m-1)} \\ \vdots \\ -h_1^{(m-1)} \\ 1 \end{bmatrix} + \rho_m \begin{bmatrix} 1 \\ -h_1^{(m-1)} \\ \vdots \\ -h_{m-1}^{(m-1)} \\ 0 \end{bmatrix},$$

The following eqn. is obtained

$$G^{(m)}(z) = z^{-1}G^{(m-1)}(z) + \rho_m H^{(m-1)}(z)$$



Lattice filter architecture.