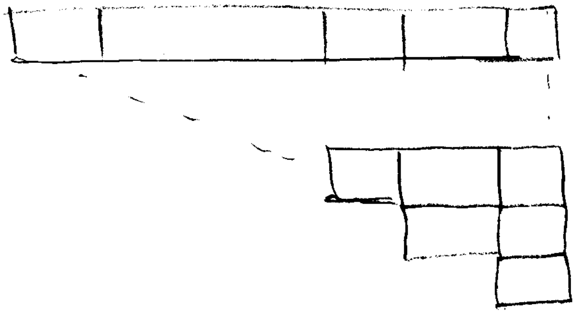
4.4 Orthogonality, Gram Schmidt and Cholesky | Haykin 6.7

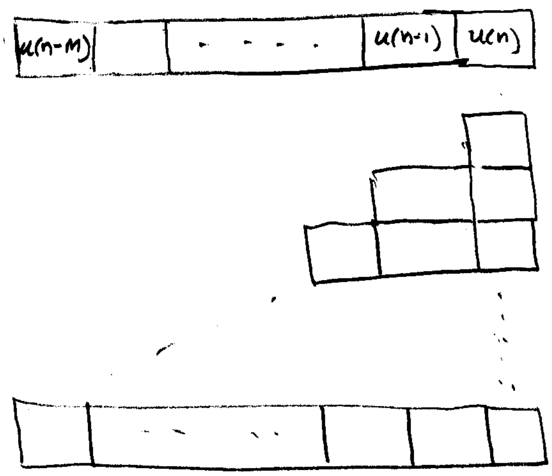


$\underline{a}_M \longrightarrow f_M(n) = u(n) - \hat{u}(n)$

$\underline{a}_2 \longrightarrow f_2(n) = u(n) - \hat{u}(n)$

$\underline{a}_1 \longrightarrow f_1(n) = u(n) - \hat{u}(n)$

$\underline{a}_0 \longrightarrow f_0(n) = u(n) - \hat{u}(n)$



$\tilde{\underline{a}}_0^* \longrightarrow b_0(n) = u(n) - \hat{u}(n)$

$\tilde{\underline{a}}_1^* \longrightarrow b_1(n) = u(n-1) - \hat{u}(n-1)$

$\tilde{\underline{a}}_2^* \longrightarrow b_2(n) = u(n-2) - \hat{u}(n-2)$

$\tilde{\underline{a}}_M^* \longrightarrow b_M(n) = u(n-M) - \hat{u}(n-M)$

By their Gram-Schmidt construction, the backward errors $b_0(n), \dots, b_M(n)$ at time n are orthogonal (uncorrelated), but the forward don't seem to be. This is an artifact of our notation conventions – set it up differently, and

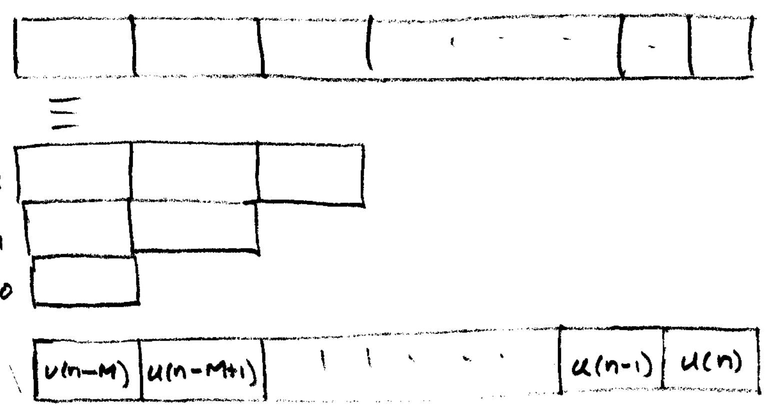
$f_M(n) \longleftarrow \underline{a}_M$

$\vdots \qquad \equiv$

$f_2(n-M+2) \longleftarrow \underline{a}_2$

$f_1(n-M+1) \longleftarrow \underline{a}_1$

$f_0(n-M) \longleftarrow \underline{a}_0$



So the forward errors are also orthogonal when read at different times

- Summarize the orthogonal prediction errors from last page:

$$f_M(n) = u(n) + a^*_{M,1}\, u(n-1) + a^*_{M,2}\, u(n-2) + \cdots + a^*_{M,M}\, u(n-M)$$

$$f_{M-1}(n-1) = u(n-1) + a^*_{M-1,1}\, u(n-2) + \cdots + a^*_{M-1,M-1}\, u(n-M)$$

$$\vdots$$

$$f_0(n-M) = u(n-M)$$

and

$$b_0(n) = u(n)$$

$$b_1(n) = a_{1,1}\, u(n) + u(n-1)$$

$$b_2(n) = a_{2,2}\, u(n) + a_{2,1}\, u(n-1) + u(n-2)$$

$$\vdots$$

$$b_M(n) = a_{MM}\, u(n) + a_{M,M-1}\, u(n-1) + a_{M,M-2}\, u(n-2) + \cdots u(n-M)$$

Both sets can be written in matrix form, but it's more common to use backword errors:

$$\underline{b}(n) = L\, \underline{u}(n)$$

where

$$L = \begin{bmatrix} \underline{a}_0 \\ \hline \underline{a}_1 \\ \hline \vdots \\ \hline \underline{a}_M \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ a_{11} & 1 & & 0 & \\ a_{22} & a_{21} & 1 & & \\ \vdots & & & & \\ a_{MM} & a_{M,M-1} & \cdots & a_{M,1} & 1 \end{bmatrix}$$

"1"s on the main diagonal
reflection coeffs $\kappa_1, \cdots \kappa_M$ in first column

note $|L| = 1$, all eigenvals $= 1$

Also a vector of time-slipped forword errors can be written

$$\underline{f}'(n) = L^+\, \underline{u}(n) \quad \text{where} \quad \underline{f}'(n) = \begin{bmatrix} f_M(n) \\ f_{M-1}(n-1) \\ \vdots \\ f_0(n-M) \end{bmatrix}$$

- Factorization of R, R^{-1}

If $\underline{b}(n) = L\,\underline{u}(n)$, then its correlation matrix is

$$E[\underline{b}(n)\,\underline{b}^{+}(n)] = E[L\,\underline{u}(n)\,\underline{u}^{+}(n)\,L^{+}] = L\,E[\underline{u}(n)\,\underline{u}^{+}(n)]\,L^{+}$$

or

$$D = L R L^{+} \qquad \text{where } D = \mathrm{diag}[P_0, P_1, \cdots P_M] = \begin{bmatrix} P_0 & & 0 \\ & P_1 & \\ 0 & & \ddots P_M \end{bmatrix}$$

and we have diagonalized the correlation matrix a different way (i.e. not eigen decomposition)

Reconstruction: $R = L^{-1} D L^{-+}$ (L^{-1} is also lower triangular)

$$= L^{-1} D^{\frac{1}{2}} D^{1/2} L^{-+} = \left(L^{-1} D^{\frac{1}{2}}\right)\left(L^{-1} D^{1/2}\right)^{+}$$

this is an L-U Cholesky factorization

Also useful to factorize R^{-1}:

$$R^{-1} = \left(L^{-1} D L^{-+}\right)^{-1} = L^{+} D^{-1} L = \left(D^{-1/2} L\right)^{+}\left(D^{-1/2} L\right)$$

a U-L Cholesky factorization

→ Good for statistical problems, e.g.

$$p(\underline{u}) = \frac{1}{(2\pi)^{M+1}|\frac{1}{2}R|} \exp\left(-\underline{u}^{+} R^{-1} \underline{u}\right) \sim \exp\left(-\underline{u}^{+}(D^{-1/2}L)^{+}(D^{-1/2}L)\underline{u}\right)$$

$$= \exp\left(-\underline{e}_b^{+}\underline{e}_b\right) \quad \text{where } \underline{e}_b = D^{-1/2} L \underline{u} = D^{-1/2}\underline{b}$$

normalized backward errors

or with time-slipped forward errors $\underline{f}'(n) = L^{+}\underline{u}(n)$,

$$\tilde{D} = L^{+} R L \qquad R = L^{-+}\tilde{D} L^{-1} \quad \text{where } \tilde{D} = \begin{bmatrix} P_M & & 0 \\ & \ddots P_1 & \\ 0 & & P_0 \end{bmatrix}$$

$$R^{-1} = L \tilde{D}^{-1} L^{+}$$

$$= \left(L \tilde{D}^{-1/2}\right)\left(L \tilde{D}^{-1/2}\right)^{+}$$

$$p(\underline{u}) \sim \exp\left(-u^{+} L \tilde{D}^{-1/2}\left(L \tilde{D}^{-1/2}\right)^{+}\underline{u}\right) = \exp\left(-\underline{e}_f'^{+}\underline{e}_f'\right)$$

where $\underline{e}_f' = \tilde{D}^{-1/2} L^{+} \underline{u}$ normalized timeslipped forward

- Relate this decorrelation (diagonalization of R) to eigenanalysis (Karhunen-Loève):
  - One relation: $R = Q\Lambda Q^+ \quad Q = [\underline{q}_0 | \cdots | \underline{q}_M]$

    so $D = LRL^+ = LQ\Lambda Q^+L^+ = (LQ)\Lambda(LQ)^+$

    and the two diagonal matrices

    $$D = \mathrm{diag}[P_0, \cdots, P_M] \qquad \Lambda = \mathrm{diag}[\lambda_0, \cdots, \lambda_M]$$

    are related simply by

    $$D = F\Lambda F^+ \quad \text{where} \quad F = LQ \quad \left(\begin{matrix}\text{analysis of}\\ \text{eigen vecs}\end{matrix}\right)$$

  - Another relation (Yule-Walker) $\boxed{\text{Haykin 6.4}}$

    $$R_{M+1}\,\underline{a}_M = \begin{bmatrix} P_M \\ 0 \\ \vdots \\ 0 \end{bmatrix} = P_M\,\underline{i}_{M+1} \qquad \underline{i}_{M+1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

    $$\underline{a}_M = P_M R_M^{-1}\,\underline{i}_{M+1} = P_M Q\Lambda^{-1}Q^+\,\underline{i}_{M+1}$$

    $$= P_M Q \begin{bmatrix} \lambda_0^{-1}\underline{q}_0^+ \\ \lambda_1^{-1}\underline{q}_1^+ \\ \vdots \\ \lambda_M \underline{q}_M^+ \end{bmatrix} \underline{i}_{M+1} = P_M Q \begin{bmatrix} q_{00}^*/\lambda_0 \\ q_{10}^*/\lambda_1 \\ \vdots \\ q_{M0}^*/\lambda_M \end{bmatrix}$$

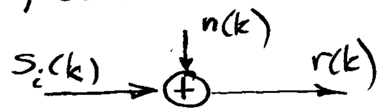
    so

    $$\underline{a}_M = P_M \sum_{i=0}^{M} \frac{q_{i0}^*}{\lambda_i}\,\underline{q}_i \quad \text{and since } a_{M,0} = 1,$$

    $$P_M = \left(\sum_{i=0}^{M} \frac{|q_{i0}|^2}{\lambda_i}\right)^{-1}$$

Hence eigenanalysis gives G-S analysis.

- Why do we care about this part?
  - orthogonality benefits will become clear in joint process estimation.
  - factorization benefits are seen in this example:



receive one of N possible signals, $i = 1,.\,N$, in coloured noise.

Take M+1 samples and form vector

$$\underline{r} = \underline{s}_i + \underline{n}$$

ML detection says max $pr(\underline{r}|\underline{s}_i) \sim \exp\left(-(\underline{r}-\underline{s}_i)^+ R_n^{-1}(\underline{r}-\underline{s}_i)\right)$

or minimize $(\underline{r}-\underline{s}_i)^+ R_n^{-1}(\underline{r}-\underline{s}_i)$ use time-slipped forward

$$= \underline{d}_i^+ (L\tilde{D}^{-\frac{1}{2}})(L\tilde{D}^{-\frac{1}{2}})^+ \underline{d}_i \qquad \text{where } \underline{d}_i = \underline{r} - \underline{s}_i$$

$$= \underline{e}_i^+ \underline{e}_i$$

where $\underline{e}_i = \tilde{D}^{-1/2} L^+ \underline{d}_i$

$$= \underbrace{\begin{bmatrix} \frac{1}{\sqrt{P_M}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{P_0}} \end{bmatrix}}_{\text{normalization}} \underbrace{\begin{bmatrix} 1 & a_{M,1}^* & a_{M,2}^* & \cdots & a_{MM}^* \\ 0 & 1 & a_{M-1,1}^* & \cdots & a_{M-1,M-1}^* \\ & & 1 & & \\ & & & 1 & a_{11}^* \\ & & & 0 & 1 \end{bmatrix}}_{\text{series of whitening filters}} \underline{d}_i$$

and if noise is finite order AR (eg order 1)

$$\underline{e}_i = \begin{bmatrix} 1/\sqrt{P_1} & \frac{a_{11}^*}{\sqrt{P_1}} & 0 & 0 & & 0 \\ 0 & \frac{1}{\sqrt{P_1}} & \frac{a_{11}^*}{\sqrt{P_1}} & 0 & & \\ & & \ddots & & & 0 \\ & 0 & & & \frac{1}{\sqrt{P_1}} & \frac{a_{11}^*}{\sqrt{P_1}} \\ & & & & 0 & \frac{1}{\sqrt{P_0}} \end{bmatrix} \underline{d}_i$$