1)

Given:

$$k_n = k_{n-1} + \mu e_n x_n$$

Find

$$\min \|k_{n+1} - k_n\|^2$$

$$\text{s.t.} \quad k_{n+1}^T x_n = d_n$$

results in $k_{n+1} = k_n + \frac{x_n e_n}{x_n^T x_n}$

Solution

Write out the Lagrangian

$$\mathcal{L}(k_{n+1}, \lambda) = \|k_{n+1} - k_n\|^2 - \lambda(k_{n+1}^T x_n - d_n)$$

$$= \|k_{n+1} - k_n\|^2 + \lambda(d_n - k_{n+1}^T x_n)$$

$$= (k_{n+1} - k_n)^2 + \lambda(d_n - k_{n+1}^T x_n)$$

$$= k_{n+1}^T k_n - 2k_{n+1}^T k_n + 2k_n^T k_n + \lambda(d_n - k_{n+1}^T x_n)$$

Take the partial derivative

$$\frac{\partial \mathcal{L}}{\partial k_{n+1}} = 2k_{n+1} - 2k_n - \lambda x_n = 0$$

$$\Rightarrow k_{n+1} = k_n + \frac{\lambda x_n}{2} \quad \text{, but what is } \lambda?$$

Using the constraint

$$k_{n+1}^T x_n = d_n = k_n^T x_n + \frac{\lambda x_n^T x_n}{2}$$

$$\frac{\lambda x_n^T x_n}{2} = d_n - k_n^T x_n \Rightarrow \lambda = 2\left(\frac{d_n - k_n^T x_n}{x_n^T x_n}\right)$$

Note that the numerator is just the error

Thus $\lambda = \frac{2e_n}{x_n^T x_n}$

Plugging λ back in we find

$$w_{n+1} = w_n + \left(\frac{2e_n}{x_n^T x_n}\right)\frac{x_n}{2} = w_n + \frac{x_n e_n}{x_n^T x_n}$$ //

2)

Given:

$c_k = [c_{k-1,k}\ c_{k-2,k}\ \cdots\ c_{1,k},\ 1]$

$A_k = [P_1\ P_2\ \cdots\ P_k]$ $\quad R_k = A^H A$

Find:

a) Show $\sigma_k^2 = e_k^H e_k = c_k^H \begin{bmatrix} R_{k-1} & h_k \\ h_k^H & r_{kk} \end{bmatrix} c_k$ (1)

Clarify what c_k is, explain what h_k and r_{kk} are.

b) Determine σ_k^2 by minimizing (1) with respect to c_k, subject to the constraint that the last element of $c_k = 1$. Show that this leads to

$$\begin{bmatrix} R_{k-1} & h_k \\ h_k^H & r_{kk} \end{bmatrix} c_k = \sigma_k^2 d \quad \text{where } d = [0\ 0\ \cdots\ 0\ 1]^T \quad (2)$$

c) Show (2) can be manipulated to become

$$\sigma_k^2 = r_{kk} - h_k^H R_{k-1}^{-1} h_k$$

Solution:

a) $e_k = A_k c_k$ $\quad \sigma_k^2 = e_k^H e_k = c_k^H A_k^H A_k c_k$

$$= c_k^H R_k c_k$$

$$R_k = A^H A = \left[\begin{array}{ccc|c} \langle P_1, P_1\rangle & \langle P_2, P_1\rangle & \cdots & \langle P_k, P_1\rangle \\ \langle P_1, P_2\rangle & \langle P_2, P_2\rangle & \cdots & \langle P_k, P_2\rangle \\ \vdots & \vdots & & \vdots \\ \hline \langle P_1, P_k\rangle & \langle P_2, P_k\rangle & \cdots & \langle P_k, P_k\rangle \end{array}\right] = \begin{bmatrix} R_{k-1} & h_k \\ h_k^H & r_{kk} \end{bmatrix}$$

where $r_{kk} = \langle p_k, p_k \rangle$

$$h_k^H = [\langle p_1, p_k \rangle \; \langle p_2, p_k \rangle \; \cdots \; \langle p_{k-1}, p_k \rangle]$$

- c_k is the set of coefficients
- h_k terms are the off-diagonal kth values of the matrix product $A_k^H A_k$. The cross correlation terms.
- r_{kk} is the k^{th} diagonal element or the autocorrelation term.

b)

$$\mathcal{L} = c_k^H R_k c_k - \lambda (c_k^H d - 1)$$

$$\frac{\partial \mathcal{L}}{\partial c_k} = 2 R_k c_k - \lambda d = 0$$

$$2 R_k c_k = \lambda d$$

If we let $\lambda = \sigma_k^2$ then

$$c_k^H R_k c_k \cancel{d}^{-1} = c_k^H R_k$$

c)

$$\begin{bmatrix} I & 0 \\ -h_k^H R_{k-1}^{-1} & I \end{bmatrix} \begin{bmatrix} R_{k-1} & h_k \\ h_k^H & r_{kk} \end{bmatrix} \begin{bmatrix} I & -R_{k-1}^{-1} h_k \\ 0 & I \end{bmatrix} = \begin{bmatrix} R_k & 0 \\ 0 & \sigma_k^2 / R_k \end{bmatrix}$$

Thus $\sigma_k^2 = r_{kk} - h_k^H R_{k-1}^{-1} h_k$ //

1/1

3.5-2)

Given:

$$(I - A(A^H A)^{-1} A^H)$$

Find:

Show that the above is a PSD matrix and hence the minimum error e_{min} has smaller norm than the original vector x.

Hint: consider $0 \le \|Bx\|^2$ where $B = I - A(A^H A)^{-1} A^H$

Solution

PSD matrices are defined as $x^T M x \ge 0 \; \forall \; x \in \mathbb{R}^n$

If we let

$$\|Bx\|^2 = \|(I - A(A^H A)^{-1} A^H) x\|^2 = \langle Bx, Bx \rangle \ge \forall x \in \mathbb{R}^n$$

Since we know this is PSD then

$$x^H (I - A(A^H A)^{-1} A^H) x = \|e_{min}\|^2 \ge 0$$

$$\|e_{min}\|^2 = x^H x - x^H A(A^H A) A^H x < \|x\|^2$$

3.8-4)

Given:

$$y_i \approx a_0 + a_1 x_i + a_2 x_i^2 \qquad (1)$$

$$y = Ac + e \qquad (2)$$

Find:

Formulate (1) as (2)

Solution

If $y_i \approx a_0 + a_1 x_i + a_2 x_i^2$

then $y_i = a_0 + a_1 x_i + a_2 x_i + e_i$

Let $y = [y_1 \ y_2 \ \cdots \ y_n]^T$

$e = [e_1 \ e_2 \ \cdots \ e_n]^T$

$c = [a_0 \ a_1 \ a_2]^T$

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & \vdots & \\ 1 & x_n & x_n^2 \end{bmatrix}$$

Thus, $y = Ac + e$ //

3.8-6)

Given:

$$y \approx ce^{ax} \qquad (1)$$

Find:

Formulate (1) as a linear regression problem with parameters c and a.

Solution

Take the natural log to remove the exponent

$$\ln(y) \approx ax + \ln(a)$$

as before we add the error to get equality

$$\ln(y_i) = a_1 x_i + \ln(a_2) + e$$

Let $y = [y_1 \; y_2 \; \dots \; y_n]^T$
$e = [e_1 \; e_2 \; \dots \; e_n]^T$
$c = [\, a_1 \;\; a_2 \,]^T$

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix}$$

Thus we get

$$y = Ac + e$$

//

3.8-10)

Given:

Define an inner product

$$\langle X, Y \rangle = tr(XY^H)$$

We want to approximate the matrix Y by the scalar linear combination of matrices $X_1, X_2, \ldots, X_m$ as

$$Y = c_1 X_1 + c_2 X_2 + \cdots + c_m X_m + E$$

Find:

Using the orthogonality principle, determine a set of normal equations that can be used to find $c_1, c_2, \ldots, c_m$ that minimize the induced norm of E

Solution:

$$Y - \underbrace{\sum_{i=1}^{m} c_i X_i}_{X^H C} = E$$

That is, we want to minimize the normed error

$$\|\underline{y} - \underline{x}^H C\| = \|E\|$$

Using the l_2 norm allows us to employ the induced norm and hence the orthogonality principle. So, we wish that

$$E \perp X$$

Using our defined inner product

$$\langle E, X \rangle = tr(EX^H) = tr(B) = \sum_i e_i^T A_{:,i} = 0$$

$$= \langle \underline{y} - \sum_{i=1}^{m} c_i X_i, X_j \rangle \quad \text{for} \quad j = 1, 2, \ldots, m$$

So just as before we get

$z = [\langle y, x_1 \rangle, \langle y, x_2 \rangle \cdots \langle y, x_n \rangle]^T$

$e = [e_1, e_2, \cdots, e_n]^T$

$c = [c_1, c_2, \cdots, c_m]^T$

$$A = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \cdots & \langle x_m, x_1 \rangle \\ \langle x_1, x_2 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_m, x_2 \rangle \\ & \vdots & & \\ \langle x_1, x_n \rangle & \langle x_2, x_n \rangle & \cdots & \langle x_m, x_n \rangle \end{bmatrix}$$

We get

$$z = Ac + e$$

Where minimization in the least squares sense results in

$$c = (A^H A)^{-1} A^H y \quad //$$

4)

Given:

a) $\dfrac{\partial Ax}{\partial x^T} = A$ b = $\dfrac{\partial x^T A}{\partial x} = A$ c= $\dfrac{\partial x^T x}{\partial x} = 2x$

d) $\dfrac{\partial x^T A x}{\partial x} = Ax + A^T x$ e) $\dfrac{\partial x^T A x}{\partial x} = 2Ax$ if A is symmetric

f) $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ $g\ \mathbb{R}^n \rightarrow \mathbb{R}^p$

$y = f(x)$ $z = g(y)$

$$(x_1, x_2, \dots, x_m) \longrightarrow \boxed{f} \xrightarrow{(y_1, \dots, y_m)} \boxed{g} \xrightarrow{(z_1, \dots, z_p)}$$

$$z_1 = g_1(y_1, \dots, y_n) = g_1(y_1(x_1, \dots, x_m), \dots, y_n(x_1 \dots, x_m))$$
$$z_2 = g_2(y_1, \dots, y_n) = g_1(y_1(x_1, \dots, x_m), \dots, y_n(x_1, \dots, x_m))$$
$$\vdots$$
$$z_p = g_p(y_1, y_n) = g_1(y_1(x_1, \dots, x_m), \dots y_n(x_1, \dots, x_m))$$

Where its derivative

$$\frac{\partial g}{\partial x^T} = \frac{\partial g}{\partial f^T} \frac{\partial f}{\partial x^T} \quad \text{or} \quad \frac{\partial z}{\partial x^T} = \frac{\partial z}{\partial y^T} \frac{\partial y}{\partial x^T}$$

Hint: A total derivative is $\dfrac{\partial z_1}{\partial x_1} = \dfrac{\partial z_1}{\partial y_1}\dfrac{\partial y_1}{\partial x_1} + \dfrac{\partial z_1}{\partial y_2}\dfrac{\partial y_2}{\partial x_1} + \dots + \dfrac{\partial z_1}{\partial y_n}\dfrac{\partial y_1}{\partial x_1}$

Find:

Derive the gradients (a) – (f)

Solution:

a)

$$\frac{\partial Ax}{\partial x^T} = A\frac{\partial x}{\partial x^T} = A\begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \cdots & \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_n}{\partial x_2} \\ & \vdots & & \end{bmatrix} = AI$$

b) Similarly to before

$$\left(\frac{\partial x^T A}{\partial x}\right)^T = \frac{\partial A x^T}{\partial x^T} = A$$

c)

$$\frac{\partial x^T x}{\partial x} = \frac{\partial}{\partial x}[x_1x_1 \quad x_2x_2 \quad \cdots \quad x_nx_n] = \frac{\partial}{\partial x}[x_1^2 \quad x_2^2 \quad \cdots \quad x_n^2]$$

$$\Rightarrow \frac{\partial x^T x}{\partial x} = 2x$$

d)

$$\frac{\partial x^T A x}{\partial x} = \sum_i \sum_j a_{ij} x_i x_j$$

$$= \sum_i a_{i1} x_i x_1 + \sum_j a_{1j} x_1 x_j + \sum_{i=2}\sum_{j=2} a_{ij} x_i x_j$$

$$\frac{\partial f}{\partial x_1} = \sum_i a_{i1} x_i + \sum_j a_{1j} x_j \quad \Rightarrow \quad \frac{\partial f}{\partial x} = \begin{bmatrix} \sum_i a_{i1} x_i \\ \vdots \\ \sum_i a_{in} x_i \end{bmatrix} + \begin{bmatrix} \sum_j a_{1j} x_j \\ \vdots \\ \sum_j a_{nj} x_j \end{bmatrix}$$

$$\vdots$$

Therefore $\frac{\partial x^T A x}{\partial x} = Ax + A^T x$

e) Similarly to before

$$\frac{\partial f}{\partial x_1} = \sum_i a_{1i} x_i + \sum_j a_{ij} x_j = 2\sum_i a_{1i} x_1 \quad \text{(because symmetry)}$$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 2\sum a_{1i} x_i \\ \vdots \\ 2\sum a_{ni} x_i \end{bmatrix} = 2Ax \quad //$$

f) By chain rule $\frac{df(g(x))}{dx} = \frac{df}{dg}\frac{dg}{dx}$

Thus for z_1

$$\frac{\partial z_1}{\partial x_1} = \frac{\partial z_1}{\partial y_1}\frac{\partial y_1}{\partial x_1} + \frac{\partial z_1}{\partial y_2}\frac{\partial y_2}{\partial x_1} + \dots + \frac{\partial z_1}{\partial y_n}\frac{\partial y_n}{\partial x_1}$$

Doing this for $\frac{\partial z_i}{\partial x_i}$ where $i = 1 \dots m$

we can stack the vectors

$$\frac{\partial z}{\partial x} = \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_p}{\partial y_n} \\ \frac{\partial z_1}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_p}{\partial y_n} \\ \vdots & & & \vdots \\ \frac{\partial z_1}{\partial y_p} & \frac{\partial z_2}{\partial y_2} & & \frac{\partial z_p}{\partial y_n} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_m} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_m} \\ \vdots & \vdots & & \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}$$

$$= \frac{\partial z}{\partial y^T}\frac{\partial y}{\partial x^T} \quad //$$

//

E.8-12)

Given:

a) $\frac{\partial z}{\partial \bar{z}} = 0$ b) $\frac{\partial z}{\partial z} = 1$ c) $\frac{\partial |z|^2}{\partial z} = \bar{z}$

d) $\frac{\partial |z|^2}{\partial \bar{z}} = z$ e) $\frac{\partial z^2}{\partial z} = 2z$ f) $\frac{\partial z^2}{\partial \bar{z}} = 0$

Find:

Using

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - j\frac{\partial}{\partial y}\right) \quad \text{and}$$

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + j\frac{\partial}{\partial y}\right)$$

Solution:

a) $\frac{1}{2}\left(\frac{\partial}{\partial x} - j\frac{\partial}{\partial y}\right)(x+jy) = \frac{1}{2}(1+0+0-1) = 0$ //

b)
$\frac{1}{2}\left(\frac{\partial}{\partial x} + j\frac{\partial}{\partial y}\right)(x+jy) = \frac{1}{2}(1+0+0+1) = 1$ //

c) $\frac{1}{2}\left(\frac{\partial}{\partial x} - j\frac{\partial}{\partial y}\right)(x^2+y^2) = \frac{1}{2}(2x - 2jy) = \bar{z}$ //

d) $\frac{1}{2}\left(\frac{\partial}{\partial x} + j\frac{\partial}{\partial y}\right)(x^2+y^2) = \frac{1}{2}(2x+2jy) = z$ //

e) $\frac{1}{2}\left(\frac{\partial}{\partial x} - j\frac{\partial}{\partial x}\right)(x^2+2jxy-y^2) = \frac{1}{2}(2x+2jy-j(2y+2jy))$

$= (x+jy) + (x+jy) = 2z$ //

f)

$$\frac{1}{2}\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right)(x^2+2ixy-y^2) = \frac{1}{2}(2x+2iy+i(2y+2iy))$$

$$=\frac{1}{2}\left[(2x+2iy)-(2x+2iy)\right]=0 \quad //$$

3-10-14)

Given

Zero-mean random vector $x = [x_1 \; x_2 \; x_3]$ with covariance

$$\mathrm{cov}(x) = E[xx^T] = \begin{bmatrix} 1 & 0.7 & 0.5 \\ 0.7 & 4 & 0.2 \\ 0.5 & 0.2 & 3 \end{bmatrix}$$

Find:

a) Determine optimal coefficients for the predictor of x_1

$$x_1 = c_1 x_2 + c_2 x_3$$

b) Determine the minimum mean-squared error

c) How is the estimator modified if the mean is

$$E_x = [1, 2, 3]^T$$

Solution

a)

$$R = \begin{bmatrix} 4 & 0.2 \\ 0.2 & 3 \end{bmatrix} \qquad p = \begin{bmatrix} 1 \\ 0.7 \end{bmatrix}$$

$$c = (R^H R)^{-1} R^H p = [0.1672 \quad 0.1555]^T \;//$$

b)

$$e = \sigma_x^2 - p' c = 0.8052 \;//$$

c) $\mathrm{cov}(Y_1, Y_2) = E(Y_1 Y_2) - \mu_1 \mu_2$

Because previously $\mu_1 = \mu_2 = 0$ to get new R_{zz} we calculate

$R_{zz}(i,j) - \mu_i \mu_j$ for $i,j = 1, 2, 3$

Therefore

$$R_{zz} = \begin{bmatrix} 1 & 0.7 & 0.5 \\ 0.7 & 4 & 0.2 \\ 0.5 & 0.2 & 3 \end{bmatrix} - E[x]^T E[x]$$

And repeating the calculations from before we find

$c = [\ 0.4973 \ \ 0.6034\ ]^T$ //

$e = 1.7599$ //

3.8-15)

Given:

A discrete-time radar signal is transmitted as

$$s[t] = A_0 e^{-j\omega_0 t}$$

The sampled recieved noisy sample is represented as

$$x[t] = A_1 e^{-j\omega_1 t} + v[t]$$

- ω_1 : recieved signal frequency, generally different than ω_0 because Doppler effect.
- $v[t]$: white noise signal with variance σ_n^2

Let a vector of samples be $x[t] = [\,x[0] \; x[1] \; \dots \; x[m-1]\,]^T$

Find:

a)

Show $E\left[x[t]x^H[t]\right] = \sigma_v^2 I + \sigma_1 S(\omega_1) S^H(\omega_1)$

Where

$S(\omega_1) = [1, e^{-j\omega_1}, e^{-j2\omega_1}, \dots, e^{-(m-1)j\omega_1}]^T$ and $\sigma_1^2 = E[|A_1|^2]$

b) $x[t]$ is applied to a FIR Wiener filter with m coefficients where

$$p = S(\omega_0)$$

Determine an expression for the tap-weight vector of the filter.

Solution:

a) I will be writing x_t in place of $x[t]$

$$E[x_t x_t^H] = E\begin{bmatrix} x_0 x_0 & x_0 x_1 & \cdots & x_0 x_{m-1} \\ x_1 x_0 & x_1 x_1 & \cdots & x_1 x_{m-1} \\ \vdots & & & \\ x_{m-1} x_0 & x_{m-1} x_1 & \cdots & x_{m-1} x_{m-1} \end{bmatrix}$$

1/2

2/2

There is no correlation of noise between variables; thus we get the term

$$\sigma_v^2 I .$$

From our matrix we can see that is simply

$$\sigma_1 S(\omega_1) S^H(\omega_1)$$

that arises from $E[x x^H]$.

b) Letting $P = S(\omega_0)$ and knowing

$$K = R^{-1} P$$

And in our case $R = \sigma_v^2 I + \sigma_1 S(\omega_0) S^H(\omega_0)$

$$K = (\sigma_v^2 I + \sigma_1 S(\omega_0) S(\omega_0)^H)^{-1} P$$

3.11-16)

Given:

Channel function with transfer function $H(z) = \frac{1}{1-0.2z^{-1}}$

w/ output $u[t]$ driven by an AR(1) signal

$$d[t] = 0.4d[t-1] + v(t)$$

- $v(t)$: zero mean and variance $\sigma_v^2 = 2$

The signal

$$f[t] = u[t] + n[t] \qquad \sigma_n^2 = 1.5$$

Find:

Design a second-order Wiener filter to minimize the average squared error between $f[t]$ and $d[t]$. What is the MSE?

Solution:

$$R_{ff} = R_{uu} + R_{nn} \qquad R_{nn} = \sigma_n^2 I \qquad R_{uu} = \begin{bmatrix} r_u(0) & r_u(1) \\ r_u(1) & r_u(0) \end{bmatrix}$$

$$r_u(0) = \left(\frac{1+a_2}{1-a_2}\right) \frac{\sigma_v^2}{(1+a_2)^2 - a_1^2}$$

$$r_v(1) = \frac{-a_1}{1+a_2} \sigma_u^2$$

To find these coefficients (a_i) we can write an equivalent model

$$H(z) = \frac{1}{(1-0.2z^{-1})(1+0.4z^{-1})} = \frac{1}{1 + \underbrace{0.2z^{-1}}_{a_1} \underbrace{- 0.08z^{-2}}_{a_2}}$$

$\Rightarrow r_u(0) = 2.1127 \quad r_v(1) = -0.4593$

2/2

$$R_{zz} = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix} + \begin{bmatrix} 2.1127 & -0.4593 \\ -0.4593 & 2.1127 \end{bmatrix} = \begin{bmatrix} 3.6127 & -0.4593 \\ -0.4593 & 3.6127 \end{bmatrix}$$

To find P, lets write $d[t]$ in terms of $u[t]$

$$d[t] = u[t] - 0.2u[t-1]$$

where it can be shown

$$P(k) = r_u(k) - 0.2r_u(k-1)$$

Thus,

$$P(0) = r_u(0) - 0.2r_u(-1) = 2.1127 - 0.2(-0.4593)$$
$$= 2.2046$$

$$P(1) = r_u(1) - 0.2(r_u(0)) = -0.4593 - 0.2(2.1127)$$
$$= -0.8818$$

$$w = R_{zz}^{-1}P = [0.5887, -0.1692]^T$$

MSE is found by $\sigma_d^2 = \dfrac{\sigma_v^2}{1-a_1^2} = \dfrac{2}{1-0.2^2} = 2.0833$

3.11-17)

Given:

$d[t] = x[t]$ be the desired value and

$$\hat{x}[t] = \sum_{i=1}^{m} w_{f,i}\, x[t-i]$$

is the predicted value of $x[t]$ using an mth order predictor.

$f_m[t] = x[t] - \hat{x}[t]$ is the forward predictor

and

$$f_m[t] = \sum_{i=0}^{m} a_{f,i}\, x[t-i] \qquad a_{f,0} = 1$$

$$a_{f,i} = -w_{f,i} \quad i = 1, 2, \ldots, M$$

Find:

We want to find an optimal set of coefficients $\{w_{f,i}\}$ to minimize $E[|f_M[t]|^2]$.

a) Using orthogonality principle write down the normal equations corresponding to the minimization problem. Use

$$r[j-l] = E[x[t-l]\,\bar{x}[t-j]]$$

to get Weiner-Hopf equations

b) Determine MMSE $P_m = \min E[|f_M[t]|^2]$

c) Show that the weights can be used to augment Wiener-Hopf as

$$\begin{bmatrix} r_0 & r^H \\ r & R \end{bmatrix} \begin{bmatrix} 1 \\ -w_f \end{bmatrix} = \begin{bmatrix} P_m \\ 0 \end{bmatrix}$$

d) $x[t]$ is an AR(m) with $H(z) = \dfrac{1}{1+\sum_{k=1}^{m} a_k z^{-k}}$

Show $w_{f,k} = -a_k$ and hence the coefficients of the prediction filter $f_m[t]$ are

$$a_{f,i} = a_i$$

e) Let the backward filter be

$$\hat{x}[t-m] = \sum_{i=1}^{m} w_{0,i}\, x[t-i+1]$$

and let the backward prediction error be

$$b_m[t] = x[t-m] - \hat{x}[t-m]$$

Show that WH equations for the optimal backward predictor is

$$R w_0 = \bar{r}^B$$

f) From the previous part show

$$R^H \bar{w}_0^B = r$$

Hence conclude that $\bar{w}_0^B = w_f$.

Solution:

a) We wish to minimize the inner product

$$\|e[t]\|^2 = E[e[t]\bar{e}[t]]$$

That is, we wish the data $\{x\}$ to be orthogonal to the error. Thus, we write

$$\langle d[t] - \sum_{l=1}^{\infty} h[l]\, x[t-l],\ x[t-i] \rangle = 0$$

$$\langle d[t], x[t-i] \rangle = \sum_{l=1}^{\infty} \underset{\text{not stochastic}}{h[l]} \langle x[t-l], x[t-i] \rangle$$

Let $\langle X, Y \rangle = E[X\bar{Y}]$ we get

$$\underbrace{E[\underset{d[t]=x[t]}{x[t]}\,\overline{x[t-i]}]}_{r(i)} = \sum_{l=1}^{\infty} h[l]\, \underbrace{E[x[t-l]\,\overline{x[t-i]}]}_{r(i-l)}$$

$$\Rightarrow r(i) = \sum_{l=1}^{\infty} h[l]\, r(i-l)$$

$h[l] = w_{f,l}$ and rather than ∞ go to m.

$$r(i) = \sum_{l=1}^{m} w_{f,l}\, r(i-l) = Rw_f = r$$

b) $$\|e\|^2_{min} = \sigma_x^2 - r^H w_f = \sigma_x^2 - r^H R^{-1} r \qquad (1)$$

$$\underbrace{\sigma_f^2}_{r_0} = \underbrace{\sigma_x^2 - \rho^H R^{-1} \rho}_{P_m}$$

c) From (1) we see the first row, the second row is from (a) if we write

$$r - Rw_f = 0$$

Thus we get

$$\begin{bmatrix} r[0] & r^H \\ r & R \end{bmatrix} \begin{bmatrix} 1 \\ -w_f \end{bmatrix} = \begin{bmatrix} P_m \\ 0 \end{bmatrix}$$

d) The Yule-Walker equations state

$$\underbrace{\begin{bmatrix} r_0 & r_1 & \cdots & r_{(p-1)} \\ \bar{r}_1 & r_0 & \cdots & r_{-(p-2)} \\ \vdots & & \ddots & \\ \bar{r}_{p-1} & \bar{r}_{p-2} & & r_0 \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}}_{a} = \underbrace{\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix}}_{\hat{r}}$$

We note $\hat{R} = R^H$, and because R is Toeplitz only the conjugated values are mirrored. Similarly for r and $\hat{r}$. Therefore we can equate $a = w_f$.

e) We still want to have orthogonality between the data and error a similar technique may be used

$$\langle b_m, x[t-j] \rangle = \langle x[t-m] - \sum_{i=1}^{m} W_{bi}\, x[t-i+1], x[t-j] \rangle$$

$$E[x[t-m]\, \overline{x[t-j]}] = \sum_{i=1}^{m} W_{bi}\, E[x[t-i+1]\, \overline{x[t-j]}]$$

As before then we can say

$r(i-m) = E[x[t-i+1]\, x[t-j]]$

$p(i) = E[x[t-m]\, \overline{x[t-j]}]$

From which we can write $R W_b = \overline{r^B}$ //

c) Recalling the Yule-Walker Equations and the toeplitz structure R^H results in the conjugated values to be mirrored over the diagonal. Thus the results will be the conjugate. So

$$R^H\, \overline{W_b^B} = r$$

$$\therefore\ \overline{W_b^B} = W_f.$$ //

3.11-19)

Given:

Suppose we have the random vector approximation problem

$$y = c_1 p_1 + c_2 p_2 + \cdots + c_m p_m + e$$

where we want to approximate y s.t. the norm of e is minimized.

Let

$$\langle x, y \rangle = tr(E[xy^H])$$

be our inner product.

Find:

a) Based on the inner product find a set of normal equations to find $c_1, \ldots, c_m$

b) Use the formula for the gradient of a trace to get to the same solution.

Solution:

a)

The form will be of

$$z = Ac + e \quad \text{where} \quad A = \begin{bmatrix} \langle p_1, p_1 \rangle & \langle p_2, p_1 \rangle & \cdots & \langle p_m, p_1 \rangle \\ \langle p_1, p_2 \rangle & \langle p_2, p_2 \rangle & \cdots & \langle p_m, p_2 \rangle \\ \langle p_1, p_n \rangle & \langle p_2, p_n \rangle & \cdots & \langle p_m, p_n \rangle \end{bmatrix}$$

$$e = [e_1, e_2, \ldots e_m]^T$$

$$c = [c_1, c_2, \ldots c_m]^T$$

$$z = [\langle x, p_1 \rangle, \langle x, p_2 \rangle \ldots]^T$$

Taking the expected value results in

$$A = \begin{bmatrix} E[\langle P_1, P_1\rangle] & E[\langle P_2, P_1\rangle] & \cdots & E[\langle P_m, P_1\rangle] \\ E[\langle P_1, P_2\rangle] & E[\langle P_2, P_2\rangle] & \cdots & E[\langle P_m, P_2\rangle] \\ \vdots & & & \\ E[\langle P_1, P_n\rangle] & E[\langle P_2, P_n\rangle] & \cdots & E[\langle P_m, P_n\rangle] \end{bmatrix}$$

$$z = [\, E[\langle x, P_1\rangle] \quad E[\langle X, P_2\rangle] \cdots \,]$$

Because $E[\cdot]$ is a linear operator, we can move it inside the $\langle \cdot, \cdot \rangle$ operator. Thus we get

$$\langle \cdot, \cdot \rangle = tr(E[ab])$$

b)

$$\text{Let } J = \|e\|^2 = \|x - \sum_{i=1}^{m} c_i P_i\|^2$$

$$= \langle x, x\rangle + \sum_{i=1}^{m}\sum_{j=1}^{m} c_i c_j \langle P_i, P_j\rangle$$

$$- 2\sum_{k=1}^{m} c_k \langle X, P_k\rangle$$

Note we are just doing an $m \times m$ for simplicity. Now lets take the derivatives

$$\frac{\partial J}{\partial c} = \begin{bmatrix} \langle P_1, P_1\rangle & \langle P_2, P_1\rangle & \cdots & \langle P_m, P_1\rangle \\ \langle P_1, P_2\rangle & \langle P_2, P_2\rangle & \cdots & \langle P_m, P_2\rangle \\ \vdots & & & \\ \langle P_1, P_m\rangle & \langle P_2, P_m\rangle & \cdots & \langle P_m, P_m\rangle \end{bmatrix} \underline{c} = \begin{bmatrix} \langle X, P_1\rangle \\ \langle X, P_2\rangle \\ \vdots \\ \langle X, P_m\rangle \end{bmatrix}$$

Then we can take the expectation of the trace and achieve the same result.

7)

b) The coefficients found were accurate to about a decimal. When I tried implementing forward/backward I could not get better results. I think my d vector is incorrect.

c) Looks similar but out of phase.

see P7.m for code

8)

b) It does a pretty good job, it seems that it is biased.

c) It does seem somewhat like the inverse

d) It seems to be performing better. Loss of a bias.

See Prim

9)

Given:

This problem introduces Gaus Quadrature

Find:

a) Show $\int_{-1}^{1} g(x)\,dx$ with substitution $t = \frac{1}{b-a}(2x-a-b)$ leads to $\int_{-1}^{1} f(t)\,dt$

b) Let $\{P_n(t), n=0,1,\dots\}$ be a set of orthogonal polynomials over $[-1,1]$ where $P_n(t)$ is a polynomial of degree n. Show

$$\langle P(t), P_m(t)\rangle = 0 \quad \forall \text{ degree} \le m-1$$

c) $f(t)$ is a polynomial of degree $2m-1$. Show $f(t)$ can be written as

$$f(t) = q(t)P_m(t) + r(t)$$

$-q(t)$ and $r(t)$ are of degree $\le m-1$

d) Show that

$$q(t) = \sum_{k=0}^{m-1} \alpha_k P_k(t) \quad \text{and} \quad \sum_{k=0}^{m-1} \beta_k P_k(t)$$

e) Show

$$\int_{-1}^{1} f(t)\,dt = \beta_0 \int_{-1}^{1} P_0(t)\,dt \qquad (1)$$

f) Let $t_1, t_2, \dots, t_m$ be the roots of $P_m(t)$. Show

$$\sum_{i=1}^{m} a_i f(t_i) = \sum_{k=0}^{m} \beta_k \sum_{i=1}^{m} a_i P_k(t_i) \qquad (2)$$

g) Show that if a_i are chosen as

$$\sum_{i=1}^{m} a_i P_k(t_i) = \begin{cases} \int_{-1}^{1} P_0(t)\,dt & k=0 \\ 0 & k=1,2,\dots,m-1 \end{cases} \qquad (3)$$

then (2) can be written as

$$\sum_{i=1}^{m} a_i f(t_i) = \beta_0 \int_{-1}^{1} P_0(t)\,dt$$

Solution:

$\frac{8}{5}$

a) $\int_{-1}^{1} g(x)\,dx$ and $t = \frac{1}{b-a}(2x - a - b)$

$$\Rightarrow x = \frac{a}{2} + \frac{b}{2} + t\,\frac{b-a}{2}$$

$$\Rightarrow \int_{-1}^{1} g\left(\frac{a}{2} + \frac{b}{2} + t\frac{b-a}{2}\right) dt$$

$$= \int_{-1}^{1} f(t)\,dt \;//$$

b) $P_n(t) = a_1 P_1 + a_2 P_2 + \cdots + a_n P_n + a_0$

$$P(t) = \beta_1 \bar{P}_1 + \beta_2 \bar{P}_2 + \cdots + \beta_{n-1} \bar{P}_{n-1} + \underbrace{0 \bar{P}_n}_{\text{note the } 0\, \bar{P}_n} + \beta_0$$

Therefore

$$\langle P_n(t), P(t) \rangle = a_1 \beta_1 P_1 \bar{P}_1 + a_2 \beta_2 P_2 \bar{P}_2 + \cdots + a_m \beta_{n-1} P_{n-1} \bar{P}_{n-1}$$
$$+ \underbrace{a_n 0}_{0}\, P_n \bar{P}_n + a_0 \beta_0$$

where the first n-1 terms are orthogonal so

$$\langle P_{n-1}(t), P(t) \rangle = 0$$

and the last term is 0, thus $\langle P_n(t), P(t) \rangle = 0$ //

c) Let $f(t)$ be a polynomial of degree $2m-1$

Let $P_m(t)$ be a polynomial of degreem m and $r(t)$ and $q(t)$ be polynomials of degree $\leq m-1$

If we divide $f(t)$ by a polynomial $g(t)$ we get

$$\frac{f(t)}{g(t)} = Q + \frac{R}{g(t)}$$ where Q is the quotient and R is the remainder

Q will be of degree $\deg(f) - \deg(g)$, assume $\deg(g) \leq m-1$

Now lets multiply by $g(t)$

$$f(t) = Q\,g(t) + R \quad ; \text{ let } \quad Q = P_m(t)$$
$$g(t) = q(t)$$
$$R = r(t)$$

Then

$$f(t) = q(t)\,P_m(t) + r(t) \quad //$$

d) Because $f(t) = q(t)\,P_m(t) + r(t)$ then the individual components must be a subset of the linear combinations. to see this more easily we write

$$f(t) = c_1 P_1 + c_2 P_2 + \cdots + c_{m-1} P_{m-1} = q(t)\,P_m(t) + r(t) \qquad (1)$$

$P_m(t)$ is obviously a subset, but we must also be able to write

$$q(t) = \sum_{k=1}^{m-1} \alpha_k P_k(t) \quad \text{and} \quad r(t) = \sum_{k=1}^{m} \beta_k P_k(t)$$

because the linear combinations must be equal the the RHS of (1).

e)

$$\int_{-1}^{1} f(t) = \underbrace{\int_{-1}^{1} q(t)\,P_m(t)\,dt}_{\langle q(t),\, P_m(t)\rangle = 0} + \int_{-1}^{1} r(t)\,dt$$

because they are orthogonal

$$\Rightarrow \int_{-1}^{1} f(t) = \int_{-1}^{1} r(t)\,dt = \int_{-1}^{1} \left(\sum_{k=0}^{m-1} \beta_k P_k(t)\right) dt$$

We can think of this as an inner product of

$$\int_{-1}^{1}\left(\sum_{k=0}^{m-1}\beta_k P_k(t)\right) 1\,dt = \int_{-1}^{1}\beta_0 P_0(t)\cdot 1\,dt + \int_{-1}^{1}\sum_{k=1}^{m-1}\beta_k P_k(t)\,dt$$

$P_0(t)$ is proportional to 1, thus the inner products of $\langle P_k, 1\rangle$ $k \geq 0$ will be 0! So we write

$$\int_{-1}^{1} f(t)\,dt) = \beta_0 \int_{-1}^{1} P_0(t)\,dt$$

f) We can say $\int_a^b f(t) \approx \sum_{i=1}^{m} a_i f(t_i)$

if t_i are the roots of $P_m(t)$ ($P_m(t_i)=0$). Using the result from (e) we say

$$\sum a_i f(t_i) = \sum_{i=1}^{m} a_i (q(t_i)\overset{\nearrow 0}{P_m(t_i)} + r(t_i))$$

$$= \sum_{i=1}^{m} a_i r(t_i) = \sum_{k=0}^{m-1}\beta_k \sum_{i=1}^{m} a_i P_k(t_i)$$

g) $$\sum_{k=0}^{m-1}\beta_k \sum_{i=1}^{m} a_i P_k(t_i) = \sum_{k=0}^{m-1}\sum_{i=1}^{m}\beta_k a_i P_k(t_i) = \underbrace{\beta_0 a_i P_0(t_i)}_{\text{Given the weights we have chosen}}$$

$$\Rightarrow \sum_{i=1}^{m} a_i f(t_i) = \beta_0 a_i P_0(t_i) = \beta_0 \int_{-1}^{1} P_0(t)\,dt$$

5/5

h) $[a_1 \; 0 \cdots] \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_m \end{bmatrix}$ s.t. $a_1 P_0(t) = \int_{-1}^{1} P_0(t)\,dt$

i) $\int_{-1}^{1} f(t)\,dt = \sum_{i=1}^{m} a_i f(t_i)$ where t_i are the roots

j) Let $\int_{-1}^{1} f(t)\,\omega(t)\,dt$ be the function of interest.

$W(t)$ is a weighting function. Let $F(t) := f(t)\,W(t)$. As before

$$\int_{-1}^{1} F(t)\,dt = \sum_{i=1}^{m} a_i F(t_i) = \sum_{i=1}^{m} a_i f(t_i) W(t_i)$$

10)

a) We can write $\|f(x,y) - \bar{f}(x,y)\|$ implies

$$\langle f - \bar{f}, \psi_j \rangle = \langle f - \sum_i \bar{x}_i \psi_i, \psi_j \rangle$$

from which we write the Grammian

$$\begin{bmatrix} \langle \psi_1 \psi_1 \rangle & \langle \psi_2 \psi_1 \rangle & \cdots & \langle \psi_N \psi_1 \rangle \\ \langle \psi_1 \psi_2 \rangle & \langle \psi_2 \psi_2 \rangle & \cdots & \langle \psi_N \psi_2 \rangle \\ & \vdots & & \\ \langle \psi_1, \psi_N \rangle & & \cdots & \langle \psi_N \psi_N \rangle \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \\ \bar{x}_N \end{bmatrix} = \begin{bmatrix} \langle f, \psi_1 \rangle \\ \langle f, \psi_2 \rangle \\ \vdots \\ \langle f, \psi_N \rangle \end{bmatrix}$$

Where we define $\langle \cdot ; \cdot \rangle = \iint \psi_i \psi_j = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$

Thus all non-diagonal terms are 0. We are then left with

$$d_i \bar{x}_i = \langle f, \psi_i \rangle \Rightarrow \bar{x}_i = \frac{\langle f, \psi_i \rangle}{d_i}$$

b)

$$y_i = \int \sum_j \bar{x}_j \psi_j \phi_i \, dx dy + v = \langle \sum_j \psi_j x_j, \phi_i \rangle + v$$

Much like the grammian we get

$$\underbrace{\begin{bmatrix} \langle \psi_1, \phi_1 \rangle & \langle \psi_2, \phi_1 \rangle & \cdots \\ \langle \psi_1, \phi_2 \rangle & \langle \psi_2, \phi_2 \rangle & \cdots \\ \vdots & & \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \end{bmatrix}}_{\bar{x}} + V = H \bar{x} + V$$

c) Analagous to before we write

$$\underbrace{\begin{bmatrix} \langle \phi_1, \phi_1 \rangle \langle \phi_2, \phi_1 \rangle \cdots \\ \langle \phi_1, \phi_2 \rangle \langle \phi_2, \phi_2 \rangle \cdots \\ \vdots \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \end{bmatrix}}_{\hat{x}} + V = H\hat{x} + V$$

d) The off diagonals represent a small element much like the pixels in fig 1a. So in other words, it is the correlation of two measurements over a region in the body.

In that light the diagonal terms are two beams being projected at the same direction, so the area of overlap is the entire beam.

e) The Gersngorin theorem can be thought of as if the off diagonal values in a square matrix over the complex numbers are small in norm, the eigenvalues cannot be far from the diagonal terms.

Looking at the statement in (d), the areas off the off-diagonal terms will be much smaller than the diagonal terms meaning that the diagonal values $\approx$ the eigenvalues.

To get eigenvalues the matrix must have a determinant which means the matrix must be invertible!

f) Having $y_i = \langle \phi_i, f \rangle$, we can jump right into the dual approximation by stating

$$\langle \phi_1, f \rangle = y_1 = \langle c_1\phi_1 + c_2\phi_2 + \cdots + c_m\phi_m, \phi_1 \rangle$$

$$\langle \phi_2, f \rangle = y_2 = \langle c_1\phi_1 + c_2\phi_2 + \cdots + c_m\phi_m, \phi_2 \rangle$$

$$\vdots$$

Which then gives rise to the matrix form

$$\begin{bmatrix} \langle \phi_1 \phi_1 \rangle \langle \phi_2 \phi_1 \rangle \cdots \langle \phi_m \phi_1 \rangle \\ \vdots \quad \vdots \quad \vdots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix}$$

much like we have seen up to now

which is identical to what we found in (c).

g) $V \sim N(0, Q) \quad Q = \sigma_v^2 I$

$$f_y(y|\hat{x}) = \frac{1}{(2\pi)^{d/2} |\det(Q)|^{1/2}} \exp\left[-\frac{1}{2}(y - H\hat{x})^H Q^{-1} (y - H\hat{x})\right]$$

$$\hat{x}_{ML} = \arg\max_{\hat{x}} f_y(y|\hat{x})$$

$$= y^H Q^{-1} y - y^H Q^{-1} H\hat{x} - \hat{x}^H H^H Q^{-1} y + \hat{x}^H H^H Q^{-1} H\hat{x}$$

$$= y^H Q^{-1} y - 2(y^H Q^{-1} H)\hat{x} + \hat{x}^H H^H Q^{-1} H\hat{x}$$

$$\frac{\partial}{\partial \hat{x}} = H^H Q^{-1} y + H^H Q^{-1} H\hat{x}$$

Given $Q = \sigma_n^2 I \Rightarrow \hat{x} = (H^H H)^{-1} H^H y$

h)

$$\arg\max_{\hat{x}} f_{\hat{x}}(\hat{x}) f_y(y|\hat{x}) = [\hat{x}^T R_x \hat{x}] + [(y - Hx)^T R_n (y - H\hat{x})]$$

Expanding we get

$$\tilde{A} = [\hat{x}^T R_x \hat{x}] + [y^T R y - 2yR_n H\hat{x} + \hat{x}^T H^T R_n H\hat{x}]$$

$$\frac{\partial \tilde{A}}{\partial \hat{x}} = \cancel{2}R_x\hat{x} + 0 - \cancel{2}yR_n H + \cancel{2}H^T R_n H\hat{x} = 0$$

$$R_x\hat{x} + H^T R_n H\hat{x} = yR_n H$$

$$\hat{x} = (R_x + H^T R_n H)^{-1} y R_n H$$

i) Suppose we want to approximate $f(x_0, y_0) = f_0$ we can say

$$\hat{f}(x,y) \approx f(x_0, y_0) = \sum_i x_i \phi_i = c^T x_i = f_0 \quad //$$

j)

$$\mathcal{L} = \|y - H\hat{x}\|^2 - \lambda(c^T\hat{x} - f_0) = \|y^2\| - 2\langle y, H\hat{x}\rangle + \|H\hat{x}\|^2$$

$$= y^T y - 2y^T H\hat{x} + \hat{x}^T H^T H\hat{x} - \lambda(c^T\hat{x} - f_0)$$

$$\frac{\partial \mathcal{L}}{\partial x} = -2y^T H - 2H^T H\hat{x} - \lambda c^T = 0$$

$$\hat{x} = -(H^T H)^{-1}\left(\frac{\lambda c^T}{2} + H^T y\right)$$

$$c^T\hat{x} - f_0 = -c^T (H^T H)^{-1}\left(\frac{\lambda}{2} c^T + H^T y\right)$$

$$\lambda = 2\left(\frac{f_0 (H^T H)}{c^T c} - \frac{H^T y}{c}\right)$$

$$\hat{x} = -(H^T H)^{-1}\left[\frac{2}{2}\left(\frac{f_0(H^T H)}{c^T c} - \frac{H^T y}{c}\right)c^T + H^T y\right]$$

$$= (H^T H)^{-1}\left(\frac{H^T y + f_0 - c^T (H^T H)^{-1} H^T y\, c}{c(H^T H)^{-1} c}\right) \quad //$$

11)

a)

$$y^2 = [W^H (s+n)][(s+n)^H w]$$

$$= W^H [ss^H + sn^H + ns^H + nn^H] W$$

$$E|y|^2 = W^H E[ss^H + \cancel{sn^H}^{\,0} + \cancel{ns^H}^{\,0} + nn^H] W$$

$$= W^H ss W + W^H \underbrace{nn^H}_{\text{cov}(n)} W$$

$$|Ey|^2 = (W^H s)(s^H W)$$

$$\Rightarrow \text{Var}(y) = E|y|^2 - |Ey|^2 = W^H R W$$

where $R = \text{cov}(n)$

b)

$$\text{SNR} = \frac{|Ey|^2}{\text{Var}(y)} = \frac{W^H s s^H W}{W R W}$$

c)

$$\frac{|\langle x, y\rangle|^2}{|\langle x, x\rangle|^2} \le |\langle y, y\rangle|^2 \qquad \text{Let } \langle a, b\rangle = a R^{-1} b$$

$$\langle Rw, Rw\rangle = \langle x, x\rangle = W^H \cancel{R^H} R^{-1} R W$$

$$\langle s, s\rangle = \langle y, y\rangle = s^H R^{-1} s$$

$$\langle Rw, s\rangle = s^H R^{-1} R w = s^H W$$

Thus $\text{SNR} \le s^H R^{-1} s$

Cauchy-Schwartz achieves equality when they are proportional.

$Rw \propto s \Rightarrow w \propto R^{-1}s$

So we can say $w = R^{-1}s$ //