Fundamentals of Adaptive Filtering

Paulo S. R. Diniz - *diniz@smt.ufrj.br*



SIGNAL PROCESSING LABORATORY

COPPE/Poli - Federal University of Rio de Janeiro

October, 2012

slidec24ed.tex

Fundamentals of Adaptive Filtering



Contents

- A brief review of deterministic and random signal representations.
- The properties of the correlation matrix of the input signal vector
- The Wiener solution: the minimum mean-square error (MSE) solution
- Properties of the MSE surface
- The classical methods of optimization: The Newton and steepest descent algorithms

Signal Representation



Deterministic Signals

A deterministic signal has a defined mathematical function. For example

$$x(k) = e^{-\alpha\, k} \cos(\omega k) + u(k)$$

where $u(k)$ is the unit step sequence.

Signal Representation



The response of a linear time-invariant filter is

$$y(k) = x(k) * h(k) = \sum_{n=-\infty}^{\infty} x(n)h(k-n) \quad (1)$$

$$= \sum_{n=-\infty}^{\infty} h(n)x(k-n) = h(k) * x(k) \quad (2)$$

where $h(k)$ is the impulse response of the filter.

The $\mathcal{Z}$ transform of $x(k)$ is

$$\mathcal{Z}\{x(k)\} = X(z) = \sum_{k=-\infty}^{\infty} x(k)z^{-k}$$

Signal Representation



The convolution operation can be replaced by

$$Y(z) = H(z)\, X(z)$$

For finite-energy waveforms, the Fourier transform is

$$\begin{aligned} \mathcal{F}\{x(k)\} &= X(e^{j\omega}) \\ &= \sum_{k=-\infty}^{\infty} x(k) e^{-j\omega k} \end{aligned} \tag{3}$$

Signal Representation



Random Signals

A random variable $\mathrm{X}(\varrho)$ is a function that assigns a number to every outcome denoted by ϱ of a given experiment.

A stochastic process is a rule to describe the time evolution depending on ϱ, therefore it is a function of two variables $\mathrm{X}(k, \varrho)$.

The set of all experimental outcomes, i.e. the ensemble, is the domain of ϱ.

We denote $x(k)$ as a sample of the given process with ϱ fixed, where in this case if k is also fixed, $x(k)$ is a number.

Signal Representation



The expected value, or mean-value, of the process is defined by

$$m_x(k) = E[x(k)] \tag{4}$$

The integral definition of the expected value is expressed as

$$E[x(k)] = \int_{-\infty}^{\infty} y p_{x(k)}(y) dy \tag{5}$$

where $p_{x(k)}(y)$ is the probability density function (pdf) of $x(k)$.

Signal Representation



To interpret the pdf, define the distribution function of a random variable as

$$P_{x(k)}(y) = \ prob.\ of\ x(k)\ be\ smaller\ or\ equal\ y$$

or

$$P_{x(k)}(y) = \int_{-\infty}^{y} p_{x(k)}(z)dz \tag{6}$$

The derivative of the distribution function is the pdf

$$p_{x(k)}(y) = \frac{dP_{x(k)}(y)}{dy} \tag{7}$$

Signal Representation



The autocorrelation function of the process $x(k)$ is defined by

$$r_x(k,l) = E[x(k)x(l)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yz\, p_{x(k),x(l)}(y,z)\,dy\,dz \tag{8}$$

where $p_{x(k),x(l)}(y,z)$ is the joint probability density of $x(k)$ and $x(l)$ defined as

$$p_{x(k),x(l)}(y,z) = \frac{\partial^2 P_{x(k),x(l)}(y,z)}{\partial y \partial z} \tag{9}$$

where $P_{x(k),x(l)}(y,z) = probability\ of\ \{x(k) \leq y\ and\ x(l) \leq z\}$.

Signal Representation



The autocovariance function is defined as

$$\begin{aligned} \sigma_x^2(k,l) &= E[(x(k) - m_x(k))(x(l) - m_x(l))] \\ &= r_x(k,l) - m_x(k)m_x(l) \end{aligned} \tag{10}$$

For $k = l$, $\sigma_x^2(k,l) = \sigma_x^2(k)$ which is the variance of $x(k)$.

The Gaussian pdf is defined by

$$p_{x(k)}(y) = \frac{1}{\sqrt{2\pi\sigma_x^2(k)}} e^{-\frac{(y-m_x(k))^2}{2\sigma_x^2(k)}} \tag{11}$$

where $m_x(k)$ and $\sigma_x^2(k)$ are the mean and variance of $x(k)$.

Signal Representation



Given a random variable x composed by the sum of n independent random variables x_i as follows

$$x = \sum_{i=1}^{n} x_i \tag{12}$$

the central limit theorem states that under certain general conditions, the pdf of x approaches a Gaussian density function for large n.

Signal Representation



The mean and variance of y are given respectively by

$$m_x = \sum_{i=1}^{n} m_{x_i} \tag{13}$$

$$\sigma_x^2 = \sum_{i=1}^{n} \sigma_{x_i}^2 \tag{14}$$

Signal Representation



The mean and variance of y can grow, so

$$x^{'} = \frac{x - m_x}{\sigma_x} \tag{15}$$

In this case, for $n \to \infty$ it follows that

$$p_{x'}(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \tag{16}$$

Signal Representation



Conditional distribution:

$$P_{x(k)}(y|B) = \frac{P(\{x(k) \leq y\} \cap B)}{P(B)}$$

$\triangleq$ *probability of* $x(k) \leq y$ *assuming* B *has occurred*

Consists of all outcomes $\varrho \in B$ such that $\mathrm{X}(\varrho) \leq y$.

The integral definition

$$m_{x|B}(k) = E[x(k)|B] = \int_{-\infty}^{\infty} y p_{x(k)}(y|B) dy \qquad (17)$$

where $p_{x(k)}(y|B)$ is the pdf of $x(k)$ conditioned on B.

Signal Representation



The conditional variance is

$$\begin{aligned}\sigma^2_{x|B}(k) &= E[(x(k) - m_{x|B}(k))^2|B] \\ &= \int_{-\infty}^{\infty}(y - m_{x|B}(k))^2 p_{x(k)}(y|B)dy \qquad (18)\end{aligned}$$

There are processes such that the mean and autocorrelation functions are time invariant, i.e.

$$m_x(k-i) = m_x(k) = E[x(k)] = m_x; \qquad (19)$$

Signal Representation



$$r_x(k,i) = E[x(k-j)x(i-j)] = r_x(k-i) = r_x(l) \tag{20}$$

and as a consequence

$$\sigma_x^2(l) = r_x(l) - m_x^2 \tag{21}$$

These processes are said to be stationary in the wide-sense (WSS).

If the nth-order statistics of a process is shift invariant, the process is said to be nth-order stationary.

Signal Representation



If the process is nth-order stationary for any value of n, it is stationary in strict sense.

Two processes are considered jointly WSS if and only if any linear combination of them is also WSS. For example, given a WSS process

$$y(k) = k_1\, x_1(k) + k_2\, x_2(k) \tag{22}$$

where k_1 and k_2 are arbitrary constants, then $x_1(k)$ and $x_2(k)$ have shift invariant means, autocorrelations and cross-correlations.

Signal Representation



For complex signals, $x(k) = x_r(k) + jx_i(k)$ and $y = y_r + y_i$, the integral definition of the expected value is

$$E[x(k)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y p_{x_r(k),x_i(k)}(y_r, y_i) dy_r dy_i \tag{23}$$

where $p_{x_r(k),x_i(k)}(y_r, y_i)$ is the joint (pdf) of $x_r(k)$ and $x_i(k)$.

Signal Representation



The autocorrelation function of $x(k)$ is

$$\begin{aligned} & r_x(k,l) = E[x(k)x^*(l)] \\ = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \\ & yz^* p_{x_r(k),x_i(k),x_r(l),x_i(l)}(y_r, y_i, z_r, z_i)dy_r dy_i dz_r dz_i \end{aligned} \tag{24}$$

where $p_{x_r(k),x_i(k),x_r(l),x_i(l)}(y_r, y_i, z_r, z_i)$ is the joint pdf of $x(k)$ and $x(l)$.

Signal Representation



For complex signals the autocovariance is

$$\begin{aligned} \sigma_x^2(k,l) &= E[(x(k) - m_x(k))(x(l) - m_x(l))^*] \\ &= r_x(k,l) - m_x(k)m_x^*(l) \end{aligned} \tag{25}$$

Signal Representation



Autoregressive Moving Average Process

If the process resulting from the output of a system described by

$$y(k) = \sum_{j=0}^{M} b_j x(k-j) + \sum_{i=1}^{N} a_i y(k-i) \tag{26}$$

where $x(k)$ is a white noise, is called autoregressive moving average (ARMA). For $b_j = 0$ for $j = 0, 1, \cdots, M$, the process is called autoregressive (AR) process. For $a_i = 0$ for $i = 1, 2, \cdots, N$, the process is moving average (MA).

A stochastic process is called a Markov process if its past has no influence on the future if the present is specified.

Signal Representation



Wold Decomposition

$$x(k) = x_r(k) + x_p(k) \tag{27}$$

where $x_r(k)$ is a regular process that is equivalent to the response of a stable linear, time-invariant and causal filter to a white noise, and $x_p(k)$ is a perfectly predictable (deterministic or singular) process.

Signal Representation



Markov Process

- A stochastic process is called a Markov process if its past has no influence in the future when the present is specified.
- The present behavior of the process depends only on the most recent past, i.e., all behavior previous to the most recent past is not required.
- Take the sequence

$$y(k) = ay(k-1) + n(k) \tag{28}$$

where $n(k)$ is a white noise process.

Signal Representation



Markov Process

- The process represented by $y(k)$ is determined by $y(k-1)$ and $n(k)$, and no information before the instant $k-1$ is required.
- An mth-order Markov process satisfies the following condition: for all $k \geq 0$, and for a fixed m:

$$\begin{aligned} & P_{x(k)}\left(y|x(k-1), x(k-2), \ldots, x(0)\right) \\ & = P_{x(k)}\left(y|x(k-1), x(k-2), \ldots, x(k-m)\right) \end{aligned} \tag{29}$$

Signal Representation



Power Spectral Density

The Fourier transf. of the autocorr. function of a stat. process is the power spectral density,

$$R_x(e^{j\omega}) = \sum_{l=-\infty}^{\infty} r_x(l)e^{-j\omega l} = \mathcal{F}[r_x(l)] \tag{30}$$

where $r_x(l)$ is the autocorrelation of $x(k)$.

Signal Representation



The inverse Fourier transform is

$$r_x(l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R_x(e^{j\omega})e^{j\omega\, l}d\omega = \mathcal{F}^{-1}[R_x(e^{j\omega})] \qquad (31)$$

$R_x(e^{j\omega})$ is a determ. function of ω. The MSE of $x(k)$ is

$$r_x(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R_x(e^{j\omega})d\omega \qquad (32)$$

Signal Representation



For random signal applied as input to a LTI filter,

$$y(k) = \sum_{n=-\infty}^{\infty} x(n)h(k-n) = x(k) * h(k) \quad (33)$$

$$r_y(l) = r_x(l) * r_h(l) \quad (34)$$

$$R_y(e^{j\omega}) = R_x(e^{j\omega})|H(e^{j\omega})|^2 \quad (35)$$

$$r_{yx}(l) = r_x(l) * h(l) = E[x^*(l)y(l)] \quad (36)$$

$$R_{yx}(e^{j\omega}) = R_x(e^{j\omega})H(e^{j\omega}) \quad (37)$$

where $r_h(l) = h(l) * h(-l)$.

Signal Representation



Ergodicity

The time average $x(k)$ is

$$\hat{m}_{x_N} = \frac{1}{2N+1} \sum_{k=-N}^{N} x(k) \tag{38}$$

If

$$\sigma^2_{\hat{m}_{x_N}} = \lim_{N \to \infty} E\{|\hat{m}_{x_N} - m_x|^2\} = 0$$

the process is said to be mean-ergodic in the mean square sense.

Signal Representation



$\hat{m}_{x_N}$ is an unbiased estimate of m_x since

$$E[\hat{m}_{x_N}] = \frac{1}{2N+1} \sum_{k=-N}^{N} E[x(k)] = m_x \tag{39}$$

Signal Representation



The process will be considered ergodic if the variance of $\hat{m}_{x_N} \to 0$ ($\sigma^2_{\hat{m}_{x_N}} \to 0$) when $N \to \infty$.

The variance $\sigma^2_{\hat{m}_{x_N}}$ can be expressed as

$$\sigma^2_{\hat{m}_{x_N}} = \frac{1}{2N+1} \sum_{l=-2N}^{2N} \sigma_x^2(k+l,k) \left(1 - \frac{|l|}{2N+1}\right) \qquad (40)$$

where $\sigma_x^2(k+l,k)$ is the autocovariance of $x(k)$. The variance of $\hat{m}_{x_N}$ tends to zero if and only if

Signal Representation



$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N} \sigma_x^2(k+l,k) \to 0$$

For second-order statistics we can define

$$x_l(k) = x(k+l)x^*(k)$$

The mean-ergodicity of $x_l(k)$ implies the mean-square ergodicity of the autocorrelation of $x(k)$.

Signal Representation



The time average of $x_l(k)$ is given by

$$\hat{m}_{x_{l,N}} = \frac{1}{2N+1} \sum_{k=-N}^{N} x_l(k) \tag{41}$$

If the variance of $\hat{m}_{x_{l,N}}$ tends to zero as N tends to infinity the process $x(k)$ is said to be mean-square ergodic of the autocorrelation, i.e.

$$\lim_{N \to \infty} E\{|\hat{m}_{x_{l,N}} - r_x(l)|^2\} = 0$$

Signal Representation



The condition above is satisfied if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N} E\{x(k+l)x^*(k)x(k+l+i)x^*(k+i)\}$$

$$-r_x^2(l) = 0 \tag{42}$$

The Correlation Matrix



For a given input vector the correlation matrix is

$$\mathbf{R} =$$

$$\begin{bmatrix} E[|x_0(k)|^2] & E[x_0(k)x_1^*(k)] & \cdots & E[x_0(k)x_N^*(k)] \\ E[x_1(k)x_0^*(k)] & E[|x_1(k)|^2] & \cdots & E[x_1(k)x_N^*(k)] \\ \vdots & & \cdots & \\ E[x_N(k)x_0^*(k)] & E[x_N(k)x_1^*(k)] & \cdots & E[|x_N(k)|^2] \end{bmatrix}$$

$$= E[\mathbf{x}(k)\mathbf{x}^H(k)] \qquad (43)$$

Signal Representation



The main properties of the $\mathbf{R}$ matrix are listed below:

1. The matrix $\mathbf{R}$ is positive semidefinite.

   Usually the matrix $\mathbf{R}$ is positive definite, unless the signals that compose the input vector are linearly dependent.

2. For WSS processes $x_i(k)$, the matrix $\mathbf{R}$ is Hermitian, i.e.

$$\mathbf{R} = \mathbf{R}^H \tag{44}$$

3. When the input signal vector is composed of delayed versions of the same signal taken from a WSS process, the matrix $\mathbf{R}$ is Toeplitz.

Signal Representation



If $\mathbf{R}$ is nonsingular for a given N, the input signal is said to be persistently exciting of order $N+1$. This means $R_x(e^{j\omega})$ is different from zero at least at $N+1$ points of $0 < \omega \leq 2\pi$. It also means that an Nth-order FIR filter with at least one nonzero coefficient can not filter $x(k)$ to zero.

λ is an eigenvalue of $\mathbf{R}$, with a corresponding nonzero eigenvector $\mathbf{q}$, if and only if

$$det(\mathbf{R} - \lambda\mathbf{I}) = 0 \tag{45}$$

and

$$\mathbf{Rq} = \lambda\mathbf{q} \tag{46}$$

Signal Representation



Properties of the eigenvalues and eigenvectors of $\mathbf{R}$:

- The eigenvalues of $\mathbf{R}^m$ are λ_i^m, for $i = 0, 1, 2, ...N$.
- Suppose $\mathbf{R}$ has $N+1$ l.i. eigenvectors $\mathbf{q}_i$, then if we form $\mathbf{Q}$ with columns consisting of the $\mathbf{q}_i$'s, it follows that:

Signal Representation



$$\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q} = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & & \vdots \\ \vdots & 0 & \ldots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & & \lambda_N \end{bmatrix} = \mathbf{\Lambda} \qquad (47)$$

Signal Representation



- The nonzero eigenvectors $\mathbf{q}_0$, $\mathbf{q}_1$,...$\mathbf{q}_N$ that correspond to different eigenvalues, are l. i.

  A matrix of order $(N+1)$ is diagonalizable if it possesses $(N+1)$ eigenvectors l.i.. A matrix with repeated eigenvalues can be diagonalized or not, depending on the linear dependency of the eigenvectors. A nondiagonalizable matrix is called defective.

Signal Representation



- Since $\mathbf{R}$ is a Hermitian matrix and representing a corretation matrix, i.e., $\mathbf{R}^H = \mathbf{R}$, its eigenvalues are real and greater or equal than zero.

  Note that $\mathbf{Q}$ is not unique since each $\mathbf{q}_i$ can be multiplied by an arbitrary nonzero constant. We consider only normalized eigenvectors, that is

$$\mathbf{q}_i^H \mathbf{q}_i = 1 \quad \text{for } i = 0, 1, ...N \tag{48}$$

- If $\mathbf{R}$ is a Hermitian matrix with different eigenvalues, the eigenvector are orthogonal to one another. As a consequence there is a diagonalizing matrix $\mathbf{Q}$ that is unitary, i.e. $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$.

Signal Representation



An important result is that any Hermitian matrix $\mathbf{R}$ can be diagonalized by a suitable unitary matrix $\mathbf{Q}$, even if the eigenvalues of $\mathbf{R}$ are not distinct. Therefore, for Hermitian matrices with repeated eigenvalues it is always possible to find a complete set of orthonormal eigenvectors.

A useful form to decompose a Hermitian matrix is

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \sum_{i=0}^{N} \lambda_i \mathbf{q}_i \mathbf{q}_i^H \tag{49}$$

that is known as *spectral decomposition*.

Signal Representation



From this decomposition we have

$$\mathbf{w}^H\mathbf{R}\mathbf{w} = \sum_{i=0}^{N} \lambda_i \mathbf{w}^H \mathbf{q}_i \mathbf{q}_i^H \mathbf{w} = \sum_{i=0}^{N} \lambda_i |\mathbf{w}^H \mathbf{q}_i|^2 \tag{50}$$

Since $\mathbf{q}_i = \lambda_i \mathbf{R}^{-1}\mathbf{q}_i$, the eigenvectors of a matrix and its inverse coincide, whereas the eigenvalues are reciprocals of each other. As a consequence

$$\mathbf{R}^{-1} = \sum_{i=0}^{N} \frac{1}{\lambda_i} \mathbf{q}_i \mathbf{q}_i^H \tag{51}$$

Signal Representation



Any Hermitian matrix can be written in the form

$$\mathbf{R} = \begin{bmatrix} \sqrt{\lambda_0}\mathbf{q}_0 \; \sqrt{\lambda_1}\mathbf{q}_1 \ldots \sqrt{\lambda_N}\mathbf{q}_N \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_0}\mathbf{q}_0^H \\ \sqrt{\lambda_1}\mathbf{q}_1^H \\ \vdots \\ \sqrt{\lambda_N}\mathbf{q}_N^H \end{bmatrix}$$

$$= \mathbf{L}\mathbf{L}^H \qquad (52)$$

Signal Representation



- The sum of the eigenvalues of $\mathbf{R}$ is equal to the trace of $\mathbf{R}$, and the product of eigenvalues of $\mathbf{R}$ is equal to the determinant of $\mathbf{R}$.
- The Rayleigh's quotient defined as

$$\mathcal{R} = \frac{\mathbf{w}^H \mathbf{R} \mathbf{w}}{\mathbf{w}^H \mathbf{w}} \tag{53}$$

of a Hermitian matrix is bounded by

$$\lambda_{min} \leq \mathcal{R} \leq \lambda_{max} \tag{54}$$

where the min. and max. are reached when the vector $\mathbf{w}$ is chosen to be the eigenvectors with minimum and maximum eigenvalues.

Signal Representation



The norm $\mathbf{R}$ is defined by

$$\|\mathbf{R}\|^2 = \max_{\mathbf{w}\neq 0} \frac{\|\mathbf{R}\mathbf{w}\|^2}{\|\mathbf{w}\|^2} = \max_{\mathbf{w}\neq 0} \frac{\mathbf{w}^H \mathbf{R}^H \mathbf{R}\mathbf{w}}{\mathbf{w}^H \mathbf{w}} \tag{55}$$

It is a measure of how $\mathbf{w}$ grows in magnitude when it is multiplied by $\mathbf{R}$.

When the matrix $\mathbf{R}$ is Hermitian

$$\|\mathbf{R}\| = \lambda_{max}$$

Signal Representation



A common problem we encounter in adaptive filtering is the solution of

$$\mathbf{R}\mathbf{w} = \mathbf{p} \tag{56}$$

In case there is an error in the vector $\mathbf{p}$,

$$\frac{\|\Delta\mathbf{w}\|}{\|\mathbf{w}\|} \leq \frac{\lambda_{max}}{\lambda_{min}} \frac{\|\Delta\mathbf{p}\|}{\|\mathbf{p}\|} \tag{57}$$

where the ratio $\lambda_{max}/\lambda_{min}$ is called condition number of a matrix, that is

$$C = \frac{\lambda_{max}}{\lambda_{min}} = \|\mathbf{R}\|\|\mathbf{R}^{-1}\| \tag{58}$$

Example



Example 2.1

Input signal vector:

$$\mathbf{x}(k) = [x(k)x(k-1)\ldots x(k-N)]^T$$

Given the following input signals:

(a)

$$x(k) = n(k)$$

(b)

$$x(k) = a\cos\omega_0 k + n(k)$$

Example



(c)

$$x(k) = \sum_{i=0}^{M} b_i n(k-i)$$

(d)

$$x(k) = -a_1 x(k-1) + n(k)$$

(e)

$$x(k) = a\mathrm{e}^{\jmath(\omega_0 k + n(k))}$$

where $n(k)$ is a white noise with zero mean and variance σ_n^2; in case (e) $n(k)$ is uniformly distributed in the range $-\pi$ to π.

Calculate the autocorrelation matrix $\mathbf{R}$ for $N = 3$.

Example



Solution:

(a)

$$\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^T(k)] = \sigma_n^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Example



(b)

$$
\begin{aligned}
r(l) &= E[a^2 \cos(\omega_0 k)\cos(\omega_0 k - \omega_0 l) + n(k)n(k-l)] \\
&= a^2 E[\cos(\omega_0 k)\cos(\omega_0 k - \omega_0 l)] + \sigma_n^2 \delta(l) \\
&= \frac{a^2}{2}[\cos(\omega_0 l) - \cos(2\omega_0 k - \omega_0 l)] + \sigma_n^2 \delta(l)
\end{aligned}
$$

Example



(c) Since $n(k)$ is a white noise

$$
\begin{aligned}
r(l) &= E[x(k)x(k-l)] \\
&= E[\sum_{j=0}^{M-l}\sum_{i=0}^{M} b_i b_j n(k-i)n(k-l-j)] \\
&= \sum_{j=0}^{M-l} b_j b_{l+j} E[n^2(k-l-j)] \\
&= \sigma_n^2 \sum_{j=0}^{M} b_j b_{l+j} \\
& \quad 0 \leq l+j \leq M
\end{aligned}
$$

Example



For $M = 3$, the correlation matrix has the following form

$$\mathbf{R} = \sigma_n^2$$

$$\begin{bmatrix} \sum_{i=0}^{3} b_i^2 & \sum_{i=0}^{2} b_i b_{i+1} & \sum_{i=0}^{1} b_i b_{i+2} & b_0 b_3 \\ \sum_{i=0}^{2} b_i b_{i+1} & \sum_{i=0}^{3} b_i^2 & \sum_{i=0}^{2} b_i b_{i+1} & \sum_{i=0}^{1} b_i b_{i+2} \\ \sum_{i=0}^{1} b_i b_{i+2} & \sum_{i=0}^{2} b_i b_{i+1} & \sum_{i=0}^{3} b_i^2 & \sum_{i=0}^{2} b_i b_{i+1} \\ b_0 b_3 & \sum_{i=0}^{1} b_i b_{i+2} & \sum_{i=0}^{2} b_i b_{i+1} & \sum_{i=0}^{3} b_i^2 \end{bmatrix}$$

Example



(d) By solving the difference equation, we have

$$x(k) = (-a_1)^l x(k-l) + \sum_{j=0}^{l-1} (-a_1)^j n(k-j)$$

Multiplying $x(k-l)$ on both sides of the equation above and taking the expected value, we obtain

$$E[x(k)x(k-l)] = (-a_1)^l E[x^2(k-l)]$$

since $x(k-l)$ is independent of $n(k-j)$ for $j \leq l-1$. For $l = 0$

$$E[x^2(k)] = a_1^2 E[x^2(k-1)] + E[n^2(k)]$$

Example



therefore,

$$E[x^2(k)] = \frac{\sigma_n^2}{1 - a_1^2}$$

Example



$$r(l) = \frac{(-a_1)^{|l|}}{1 - a_1^2}\sigma_n^2 \tag{59}$$

and the 3×3 autocorrelation matrix becomes

$$\mathbf{R} = \frac{\sigma_n^2}{1 - a_1^2}\begin{bmatrix} 1 & -a_1 & a_1^2 \\ -a_1 & 1 & -a_1 \\ a_1^2 & -a_1 & 1 \end{bmatrix}$$

Example



(e) In this case, we are interested in calculating the autocorrelation of a complex sequence, that is

$$\begin{aligned} r(l) &= E[x(k)x^*(k-l)] \\ &= a^2 E[\mathrm{e}^{-\jmath(-\omega_0 l - n(k) + n(k-l))}] \end{aligned} \tag{60}$$

Example



By recalling the definition of expected value, for $l \neq 0$,

$$r(l) = a^2 \mathrm{e}^{\jmath\omega_0 l} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-\jmath(-n_0+n_1)} p_{n(k),n(k-l)}(n_0, n_1) dn_0 dn_1$$

$$= a^2 \mathrm{e}^{\jmath\omega_0 l} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-\jmath(-n_0+n_1)} p_{n(k)}(n_0) p_{n(k-l)}(n_1) dn_0 dn_1$$

Example



$$= a^2 \mathrm{e}^{\jmath\omega_0 l} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-\jmath(-n_0+n_1)} \frac{1}{2\pi} \frac{1}{2\pi} dn_0 dn_1$$

$$= a^2 \mathrm{e}^{\jmath\omega_0 l} \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-\jmath(-n_0+n_1)} dn_0 dn_1$$

$$= a^2 \mathrm{e}^{\jmath\omega_0 l} \frac{1}{4\pi^2} \left[\int_{-\pi}^{\pi} \mathrm{e}^{\jmath n_0} dn_0 \right] \left[\int_{-\pi}^{\pi} \mathrm{e}^{-\jmath n_1} dn_1 \right]$$

Example



$$= a^2 e^{\jmath\omega_0 l} \frac{1}{4\pi^2} \left[\frac{e^{\jmath\pi} - e^{-\jmath\pi}}{\jmath}\right] \left[\frac{-e^{-\jmath\pi} + e^{\jmath\pi}}{\jmath}\right]$$

$$= -a^2 e^{\jmath\omega_0 l} \frac{1}{\pi^2} (\sin\pi)(\sin\pi) = 0 \qquad (61)$$

where in the fifth equality it is used the fact that $n(k)$ and $n(k-l)$, for $l \neq 0$, are independent.

Example



For $l = 0$

$$r(0) = E[x(k)x^*(k)] = a^2 \mathrm{e}^{\jmath(\omega_0 0)} = a^2$$

Therefore,

$$r(l) = E[x(k)x^*(k-l)] = a^2 \mathrm{e}^{\jmath(\omega_0 l)} \delta(l)$$

where in the 3×3 case

Example



$$\mathbf{R} = \begin{bmatrix} a^2 & 0 & 0 \\ 0 & a^2 & 0 \\ 0 & 0 & a^2 \end{bmatrix}$$

In the end it was verified the fact that when we have two exponential functions with uniformly distributed white noise in the range of $-k\pi$ to $k\pi$ as exponents, these exponentials are nonorthogonal only if $l = 0$, where k is a positive integer.

Wiener Filter



The MSE is defined as

$$\begin{aligned} F(e(k)) &= \xi(k) = E[e^2(k)] \\ &= E[d^2(k) - 2d(k)y(k) + y^2(k)] \end{aligned}$$

The FIR adaptive filter consists of a linear combiner, i.e.

$$y(k) = \sum_{i=0}^{N} w_i(k)x_i(k) = \mathbf{w}^T(k)\mathbf{x}(k)$$

where $\mathbf{x}(k) = [x_0(k)\, x_1(k) \ldots x_N(k)]^T$ and
$\mathbf{w}(k) = [w_0(k)\, w_1(k) \ldots w_N(k)]^T$ are the input signal and the adaptive coefficient vectors.

Example



Input Vector: Tapped delay line

$$y(k) = \sum_{i=0}^{N} w_i(k)x(k-i) = \mathbf{w}^T(k)\mathbf{x}(k)$$

where

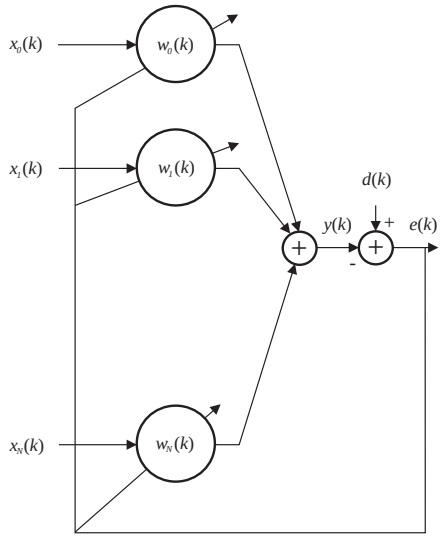
$$\mathbf{x}(k) = [x(k)\, x(k-1) \ldots x(k-N)]^T$$

$$\mathbf{w}(k) = [w_0(k)\, w_1(k) \ldots w_N(k)]^T$$

are the input and tap-weight vectors.

Wiener Filter

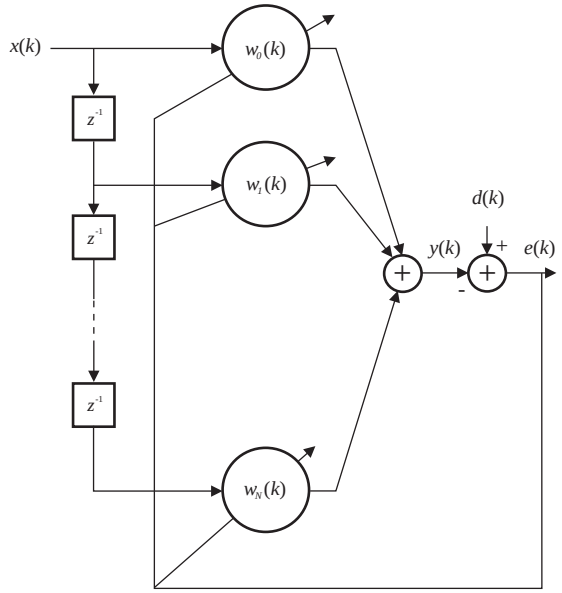




(a)

Wiener Filter





(b)

Figure 1: (a) Linear combiner; (b) Adaptive FIR filter.

Wiener Filter



The objective function is then

$$\begin{aligned} E[e^2(k)] &= \xi(k) \\ &= E[d^2(k)] - 2E[d(k)\mathbf{w}^T(k)\mathbf{x}(k)] \\ &+ E[\mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)] \end{aligned}$$

Wiener Filter



For fixed coefficients the MSE is given by

$$\begin{aligned}\xi &= E[d^2(k)] - 2\mathbf{w}^T E[d(k)\mathbf{x}(k)] \\ &+ \mathbf{w}^T E[\mathbf{x}(k)\mathbf{x}^T(k)]\mathbf{w} \\ &= E[d^2(k)] - 2\mathbf{w}^T\mathbf{p} + \mathbf{w}^T\mathbf{R}\mathbf{w}\end{aligned}$$

where $\mathbf{p} = E[d(k)\mathbf{x}(k)]$ is the cross-correlation vector between the desired and input signals, $\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^T(k)]$ is the input signal correlation matrix.

Wiener Filter



Gradient vector of the MSE function

$$\begin{aligned} \mathbf{g_w} &= \frac{\partial \xi}{\partial \mathbf{w}} \\ &= \left[\frac{\partial \xi}{\partial w_0} \; \frac{\partial \xi}{\partial w_1} \cdots \frac{\partial \xi}{\partial w_N}\right]^T \\ &= -2\mathbf{p} + 2\mathbf{R}\mathbf{w} \end{aligned} \tag{62}$$

Optimal values for the tap-weight coefficients

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$$

This solution is called the Wiener solution.

Wiener Filter



Precise estimations of $\mathbf{R}$ and $\mathbf{p}$ are not available.

The minimum MSE provided by the Wiener solution is

$$\begin{aligned} \xi_{min} &= E[d^2(k)] - 2\mathbf{w}_o{}^T\mathbf{p} + \mathbf{w}_o{}^T\mathbf{R}\mathbf{R}^{-1}\mathbf{p} \\ &= E[d^2(k)] - \mathbf{w}_o{}^T\mathbf{p} \end{aligned} \tag{63}$$

Wiener Filter



The gradient vector can be expressed as follows

$$\begin{aligned} \mathbf{g_w} &= \frac{\partial E[e^2(k)]}{\partial \mathbf{w}} \\ &= E[2e(k)\frac{\partial e(k)}{\partial \mathbf{w}}] = E[2e(k)\mathbf{x}(k)] \end{aligned} \tag{64}$$

At Wiener solution, the gradient vector is zero implying that

$$E[e(k)\mathbf{x}(k)] = \mathbf{0} \tag{65}$$

or

Wiener Filter



$$E[e(k)x(k-i)] = 0 \tag{66}$$

for $i = 0, 1, \ldots, N$. This means the error signal is orthogonal to the elements of the input signal vector.

The orthogonality principle also applies to

$$E[e(k)\mathbf{w}_o^T\mathbf{x}(k)] = E[e(k)y(k)] = 0 \tag{67}$$

Wiener Filter



Example

The input signal of a first-order adaptive filter is

$$x(k) = \alpha_1 x_1(k) + \alpha_2 x_2(k)$$

where $x_1(k)$ and $x_2(k)$ are first-order AR procs. and mutually uncorrelated both with unit variance. Generated by applying distinct white noises to first-order filters whose poles are placed at $-s_1$ and $-s_2$.

(a) Calculate the autocorrelation matrix of the input signal.

(b) If the desired signal is $x_2(k)$, calculate the Wiener solution.

Wiener Filter



Solution:

(a)

The models for the signals involved are described by

$$x_i(k) = -s_i x_i(k-1) + \kappa_i n_i(k)$$

for $i = 1, 2$.

According to equation (59) the autocorrelation of either $x_i(k)$ is given by

$$E[x_i(k)x_i(k-l)] = \kappa_i^2 \frac{(-s_i)^{|l|}}{1-s_i^2}\sigma_{n,i}^2 \tag{68}$$

where $\sigma_{n,i}^2$ is the variance of $n_i(k)$.

Wiener Filter



Since each signal $x_i(k)$ has unit variance, then by applying $l = 0$ to the equation above

$$\kappa_i^2 = \frac{1 - s_i^2}{\sigma_{n,i}^2} \tag{69}$$

Now by utilizing the fact that $x_1(k)$ and $x_2(k)$ are uncorrelated, the autocorrelation of the input signal is

$$\mathbf{R} = \begin{bmatrix} \alpha_1^2 + \alpha_2^2 & -\alpha_1^2 s_1 - \alpha_2^2 s_2 \\ -\alpha_1^2 s_1 - \alpha_2^2 s_2 & \alpha_1^2 + \alpha_2^2 \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} \alpha_2 \\ -\alpha_2 s_2 \end{bmatrix}$$

Wiener Filter



(b)

The Wiener solution can then be expressed as

$$\begin{aligned}\mathbf{w}_o &= \mathbf{R}^{-1}\mathbf{p} \\ &= \frac{1}{(\alpha_1^2+\alpha_2^2)^2-(\alpha_1^2 s_1+\alpha_2^2 s_2)^2}\begin{bmatrix}\alpha_1^2+\alpha_2^2 & \alpha_1^2 s_1+\alpha_2^2 s_2\\ \alpha_1^2 s_1+\alpha_2^2 s_2 & \alpha_1^2+\alpha_2^2\end{bmatrix} \\ &\quad \times \begin{bmatrix}\alpha_2 \\ -\alpha_2 s_2\end{bmatrix}\end{aligned}$$

Wiener Filter



$$\mathbf{w}_o = \frac{1}{(1+\frac{\alpha_2^2}{\alpha_1^2})^2 - (s_1+\frac{\alpha_2^2}{\alpha_1^2}s_2)^2} \begin{bmatrix} 1+\frac{\alpha_2^2}{\alpha_1^2} & s_1+\frac{\alpha_2^2}{\alpha_1^2}s_2 \\ s_1+\frac{\alpha_2^2}{\alpha_1^2}s_2 & 1+\frac{\alpha_2^2}{\alpha_1^2} \end{bmatrix} \times \begin{bmatrix} \frac{\alpha_2}{\alpha_1^2} \\ -\frac{\alpha_2}{\alpha_1^2}s_2 \end{bmatrix}$$

Wiener Filter



$$\mathbf{w}_o = \alpha_2 \begin{bmatrix} \frac{1}{\alpha_1^2+\alpha_2^2-s_1\alpha_1^2-s_2\alpha_2^2} & 0 \\ 0 & \frac{1}{\alpha_1^2+\alpha_2^2+s_1\alpha_1^2+s_2\alpha_2^2} \end{bmatrix} \begin{bmatrix} \frac{1-s_2}{2} \\ -\frac{1+s_2}{2} \end{bmatrix}$$

Let's assume that in this example our task was to detect the presence of $x_2(k)$ in the input signal. For a fixed input-signal power, from this solution it is possible to observe that lower signal to interference at the input, that is lower $\frac{\alpha_2^2}{\alpha_1^2}$, leads to a Wiener solution vector with lower norm. This result reflects the fact that the Wiener solution tries to detect the desired signal at the same time it avoids enhancing the undesired signal, i.e., the interference $x_1(k)$.

Wiener Filter



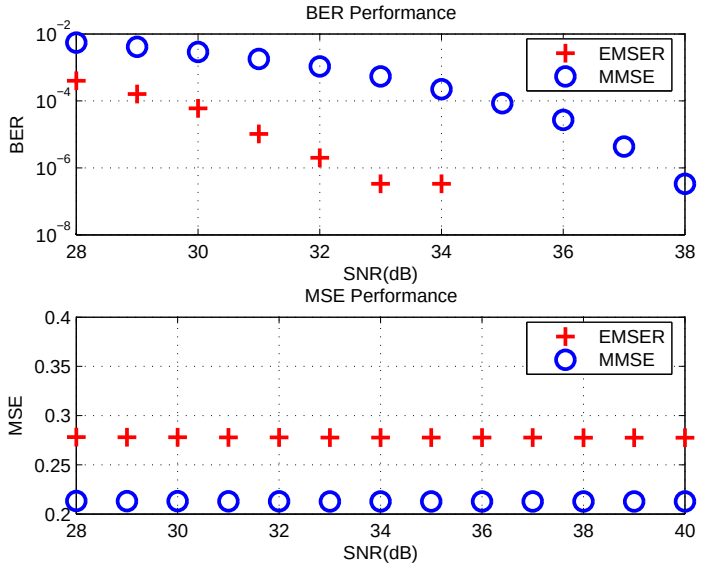


Figure 2: SER × BER.

Linearly-Constrained Wiener Filter



In a number of applications, it is required to impose some linear constraints on the filter coefficients such that the optimal solution is the one that achieves the minimum MSE, provided the constraints are met. Typical constraints are: unity norm of the parameter vector; linear phase of the adaptive filter; prescribed gains at given frequencies.

Consider the a narrowband beamformer required to pass without distortion all signals arriving at 90° with respect to the array of antennas.

Linearly-Constrained Wiener Filter



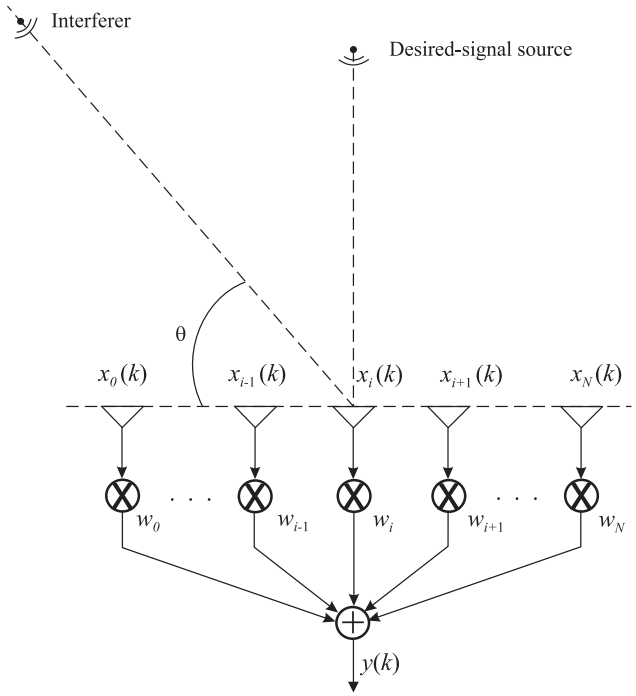


Figure 3: Narrowband beamformer.

Linearly-Constrained Wiener Filter



Taking the antenna with input signal $x_0(k)$ as a time reference t_0, the wavefront will reach the ith antenna at

$$t_i = t_0 + i\frac{d\cos\theta}{c}$$

where θ is the angle between the antenna array and the interferer direction of arrival, d is the distance between neighboring antennas, and c is the speed of propagation of the wave (3×10^8 m/s).

Linearly-Constrained Wiener Filter



The LCMV (Linearly-Constrained Minimun Variance) filter is the one that minimizes the array output signal energy

$$\xi = E[y^2(k)] = E[\mathbf{w}^T\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}]$$

$$\text{subject to :} \qquad \sum_{j=0}^{N} c_j w_j = f \tag{70}$$

where $\mathbf{w} = [w_0 \ \cdots \ w_N]^T$,$\mathbf{x}(k) = [x_0(k) \ \cdots \ x_N(k)]^T$, and $\mathbf{c} = [1 \ 1 \ \cdots \ 1]^T$ is the constraint vector, since $\theta = 90°$. The desired gain is usually $f = 1$.

Linearly-Constrained Wiener Filter



If the desired signal impinges the array at an angle θ with respect to the array, the incoming signal reaches the ith antenna delayed by $i\frac{d\cos\theta}{c}$ with respect to the 0th antenna.

Consider the case of a narrowband array such that all antennas detect the impinging signal with the same amplitude when measured taking into consideration their relative delays, which are multiples of $\frac{d\cos\theta}{c}$.

Linearly-Constrained Wiener Filter



In such a case the optimal receiver coefficients would be

$$w_i = \frac{\mathrm{e}^{\jmath\omega\tau_i}}{N+1} \tag{71}$$

for $i = 0, 1, \ldots, N$, in order to add coherently the delays of the desired incoming signal at a given direction θ. The impinging signal appear at the ith antenna multiplied by $\mathrm{e}^{-\jmath\omega\tau_i}$, for the case of array configuration of Fig. 3. In this uniform linear array, the antenna locations are

$$p_i = id$$

for $i = 0, 1, \ldots, N$.

Linearly-Constrained Wiener Filter



Using the 0th antenna as reference, the signal will reach the array according to the following pattern

$$\begin{aligned}\tilde{\mathbf{c}} &= \mathrm{e}^{\jmath\omega t}\left[1\ \mathrm{e}^{-\jmath\omega\frac{d\cos\theta}{c}}\ \mathrm{e}^{-\jmath\omega\frac{2d\cos\theta}{c}}\ \ldots \mathrm{e}^{-\jmath\omega\frac{Nd\cos\theta}{c}}\right]^T \\ &= \mathrm{e}^{\jmath\omega t}\left[1\ \mathrm{e}^{-\jmath\frac{2\pi}{\lambda}d\cos\theta}\ \mathrm{e}^{-\jmath\frac{2\pi}{\lambda}2d\cos\theta}\ \ldots \mathrm{e}^{-\jmath\frac{2\pi}{\lambda}Nd\cos\theta}\right]^T \end{aligned} \quad (72)$$

where the equality $\frac{\omega}{c} = \frac{2\pi}{\lambda}$ was employed, with λ being the wavelength corresponding to the frequency ω.

Linearly-Constrained Wiener Filter



By defining $\psi(\omega,\theta) = \frac{2\pi}{\lambda} d \cos\theta$, the output signal of the beamformer is

$$\begin{aligned} y &= \mathrm{e}^{\jmath\omega t} \sum_{i=0}^{N} w_i \mathrm{e}^{-\jmath\psi(\omega,\theta)i} \\ &= \mathrm{e}^{\jmath\omega t} H(\omega,\theta) \end{aligned} \tag{73}$$

where $H(\omega,\theta)$ modifies the amplitude and phase of transmitted signal at a given frequency ω and depends on the impinging angle.

Linearly-Constrained Wiener Filter



If the antenna separation is $d = \frac{\lambda}{2}$, $\theta = 60°$, and N is odd, then the constraint vector would be

$$\begin{aligned} \mathbf{c} &= \left[1 \quad \mathrm{e}^{-\jmath\frac{\pi}{2}} \quad \mathrm{e}^{-\jmath\pi} \ldots \mathrm{e}^{-\jmath\frac{N\pi}{2}}\right]^T \\ &= \left[1 \quad -\jmath \quad -1 \ldots \mathrm{e}^{-\jmath\frac{N\pi}{2}}\right]^T \end{aligned} \tag{74}$$

Linearly-Constrained Wiener Filter



Using the method of Lagrange multipliers, we can rewrite the constrained minimization problem described in equation (70) as

$$\xi_c = E[\mathbf{w}^T\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}] + \lambda(\mathbf{c}^T\mathbf{w} - f) \tag{75}$$

The gradient of ξ_c with respect to $\mathbf{w}$ is equal to

$$\mathbf{g_w} = 2\mathbf{R}\mathbf{w} + \lambda\mathbf{c} \tag{76}$$

where $\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^T(k)]$. For a positive definite matrix $\mathbf{R}$, the value of $\mathbf{w}$ that satisfies $\mathbf{g_w} = \mathbf{0}$ is unique and minimizes ξ_c.

Linearly-Constrained Wiener Filter



Denoting $\mathbf{w}_o$ as the optimal solution

$$2\mathbf{R}\mathbf{w}_o + \lambda\mathbf{c} = \mathbf{0}$$
$$2\mathbf{c}^T\mathbf{w}_o + \lambda\mathbf{c}^T\mathbf{R}^{-1}\mathbf{c} = \mathbf{0}$$
$$2f + \lambda\mathbf{c}^T\mathbf{R}^{-1}\mathbf{c} = \mathbf{0}$$

to obtain the second equality, we premultiply the first equation by $\mathbf{c}^T\mathbf{R}^{-1}$. Then,

$$\lambda = -2(\mathbf{c}^T\mathbf{R}^{-1}\mathbf{c})^{-1}f$$

and the LCMV filter is

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{c}(\mathbf{c}^T\mathbf{R}^{-1}\mathbf{c})^{-1}f \qquad (77)$$

Linearly-Constrained Wiener Filter



If more constraints need to be satisfied by the filter, these can be easily incorporated in a constraint matrix and in a gain vector, such that

$$\mathbf{C}^T\mathbf{w} = \mathbf{f} \tag{78}$$

In this case

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C})^{-1}\mathbf{f} \tag{79}$$

If there is a desired signal, we use MSE. The optimal solution is

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} + \mathbf{R}^{-1}\mathbf{C}(\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C})^{-1}(\mathbf{f} - \mathbf{C}^T\mathbf{R}^{-1}\mathbf{p}) \tag{80}$$

where $\mathbf{p} = E[d(k)\ \mathbf{x}(k)]$.

The Generalized Sidelobe Canceller



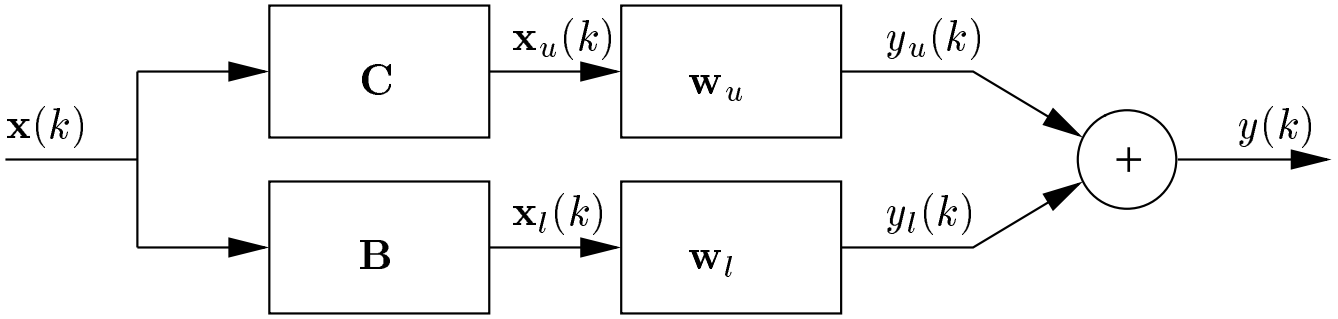


Figure 4: The generalized sidelobe canceller.

The Generalized Sidelobe Canceller



For this structure

$$\mathbf{T} = [\mathbf{C}\ \mathbf{B}] \tag{81}$$

where $\mathbf{C}$ is the constraint matrix and $\mathbf{B}$ is a *blocking matrix* that spans the null space of $\mathbf{C}$, i.e., matrix $\mathbf{B}$ satisfies

$$\mathbf{B}^T\mathbf{C} = \mathbf{0} \tag{82}$$

The Generalized Sidelobe Canceller



The output signal shown in Fig. 4 is formed as

$$\begin{aligned} y(k) &= \mathbf{w}_u^T \mathbf{C}^T \mathbf{x}(k) + \mathbf{w}_l^T \mathbf{B}^T \mathbf{x}(k) \\ &= (\mathbf{C}\mathbf{w}_u + \mathbf{B}\mathbf{w}_l)^T \mathbf{x}(k) \\ &= (\mathbf{T}\mathbf{w})^T \mathbf{x}(k) \\ &= \bar{\mathbf{w}}^T \mathbf{x}(k) \end{aligned} \tag{83}$$

where $\mathbf{w} = [\mathbf{w}_u^T \ \mathbf{w}_l^T]^T$ and $\bar{\mathbf{w}} = \mathbf{T}\mathbf{w}$.

The Generalized Sidelobe Canceller



The linear constraints are satisfied if $\mathbf{C}^T\bar{\mathbf{w}} = \mathbf{f}$. But as $\mathbf{C}^T\mathbf{B} = \mathbf{0}$, then the condition to be satisfied becomes

$$\mathbf{C}^T\bar{\mathbf{w}} = \mathbf{C}^T\mathbf{C}\mathbf{w}_u = \mathbf{f} \tag{84}$$

Therefore, for the GSC structure there is a necessary condition that the upper part of the coefficient vector, $\mathbf{w}_u$, should be initialized as

$$\mathbf{w}_u = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f} \tag{85}$$

The Generalized Sidelobe Canceller



Minimization of the output energy is achieved with a proper choice of $\mathbf{w}_l$.

$$\begin{aligned}\min_{\mathbf{w}_l} E[y^2(k)] &= \min_{\mathbf{w}_l} E\{[y_u(k) + \mathbf{w}_l^T \mathbf{x}_l(k)]^2\} \\ &= \mathbf{w}_{l,o} \\ &= -\mathbf{R}_l^{-1}\mathbf{p}_l \end{aligned} \tag{86}$$

where

$$\begin{aligned}\mathbf{R}_l &= E[\mathbf{x}_l(k)\mathbf{x}_l^T(k)] \\ &= E[\mathbf{B}^T\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{B}] \\ &= \mathbf{B}^T[\mathbf{x}(k)\mathbf{x}^T(k)]\mathbf{B} \\ &= \mathbf{B}^T\mathbf{R}\mathbf{B} \end{aligned} \tag{87}$$

The Generalized Sidelobe Canceller



and

$$\begin{aligned}
\mathbf{p}_l &= E[y_u(k)\ \mathbf{x}_l(k)] = E[\mathbf{x}_l(k)\ y_u(k)] \\
&= E[\mathbf{B}^T\mathbf{x}(k)\ \mathbf{w}_u^T\mathbf{C}^T\mathbf{x}(k)] \\
&= E[\mathbf{B}^T\mathbf{x}(k)\ \mathbf{x}^T(k)\mathbf{C}\mathbf{w}_u] \\
&= \mathbf{B}^T E[\mathbf{x}(k)\ \mathbf{x}^T(k)]\mathbf{C}\mathbf{w}_u \\
&= \mathbf{B}^T\mathbf{R}\mathbf{C}\mathbf{w}_u \\
&= \mathbf{B}^T\mathbf{R}\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f} \qquad (88)
\end{aligned}$$

where we utilized the results and definitions from equations (83) and (85).

The Generalized Sidelobe Canceller



Using the equations (87), (88), and (85), it is possible to show that

$$\mathbf{w}_{l,o} = -(\mathbf{B}^T\mathbf{R}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{R}\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f} \tag{89}$$

Given that $\mathbf{w}_{l,o}$ is the solution to an unconstrained minimization problem of transformed quantities, any unconstrained adaptive filter can be used to estimate recursively this optimal solution. The drawback in the implementation of the GSC structure comes from the transformation of the input signal vector via a constraint matrix and a blocking matrix due to computational complexity.

Example



Consider a linearly-constrained Wiener filter whose main beam should point to an impinging angle of 140 degrees. Design arrays with 5 and 40 elements respectively.

Example



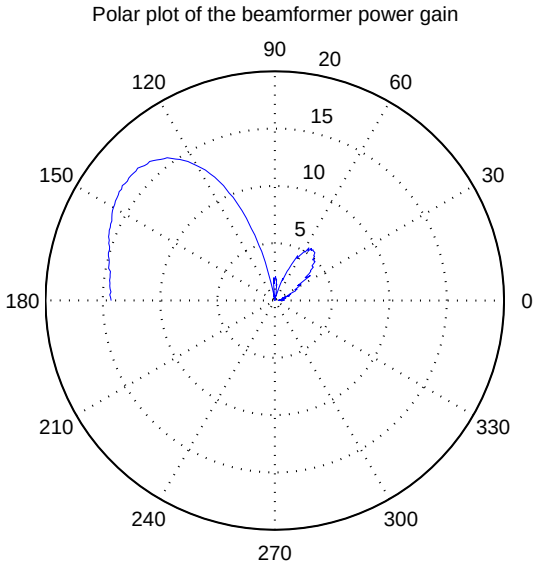


Figure 5: 5 elements array.

Example



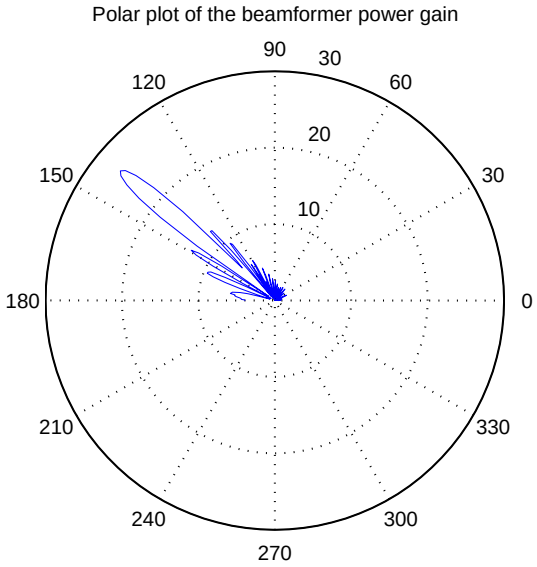


Figure 6: 40 elements array.

Mean-Square Error Surface



The MSE is a quadratic function of the parameters $\mathbf{w}$. For a fixed $\mathbf{w}$, the MSE is

$$\xi = \sigma_d^2 - 2\mathbf{w}^T\mathbf{p} + \mathbf{w}^T\mathbf{R}\mathbf{w} \tag{90}$$

Translating the coefficient vector

$$\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}_o \tag{91}$$

Wiener Filter



the MSE can be expressed as

$$\begin{aligned}
\xi &= \sigma_d^2 - \mathbf{w}_o^T\mathbf{p} + \mathbf{w}_o^T\mathbf{p} - 2\mathbf{w}^T\mathbf{p} + \mathbf{w}^T\mathbf{R}\mathbf{w} \\
&= \xi_{min} - \Delta\mathbf{w}^T\mathbf{p} - \mathbf{w}^T\mathbf{R}\mathbf{w}_o + \mathbf{w}^T\mathbf{R}\mathbf{w} \\
&= \xi_{min} - \Delta\mathbf{w}^T\mathbf{p} + \mathbf{w}^T\mathbf{R}\Delta\mathbf{w} \\
&= \xi_{min} - \mathbf{w}_o^T\mathbf{R}\Delta\mathbf{w} + \mathbf{w}^T\mathbf{R}\Delta\mathbf{w} \\
&= \xi_{min} + \Delta\mathbf{w}^T\mathbf{R}\Delta\mathbf{w} \qquad (92)
\end{aligned}$$

Mean-Square Error Surface



By diagonalizing $\mathbf{R}$ follows

$$\begin{aligned} \xi &= \xi_{min} + \Delta\mathbf{w}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\Delta\mathbf{w} \\ &= \xi_{min} + \mathbf{v}^T\mathbf{\Lambda}\mathbf{v} \\ &= \xi_{min} + \sum_{i=0}^{N}\lambda_i v_i^2 \end{aligned} \tag{93}$$

where $\mathbf{v} = \mathbf{Q}^T\Delta\mathbf{w}$. The MSE surface is in an uncoupled form

$$\mathbf{g_v} = [2\lambda_0 v_0 \;\; 2\lambda_1 v_1 \; ... \; 2\lambda_N v_N]^T$$

This property means the rotated parameters represent the principal axes of the hyperellipse of constant MSE.

Mean-Square Error Surface



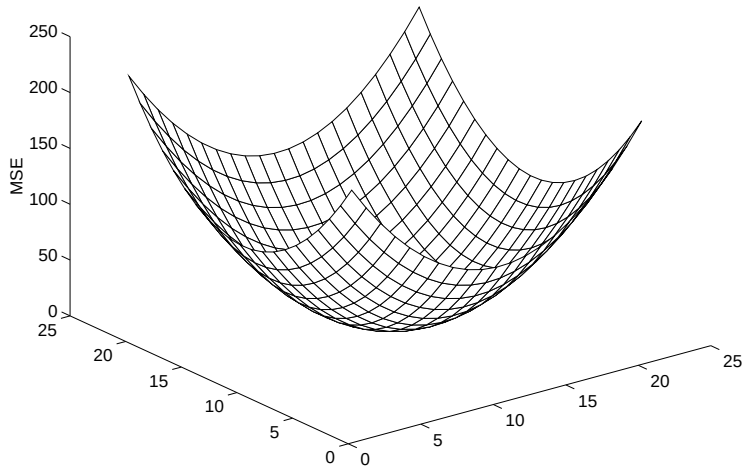


Figure 7: Mean-square error surface.

Mean-Square Error Surface



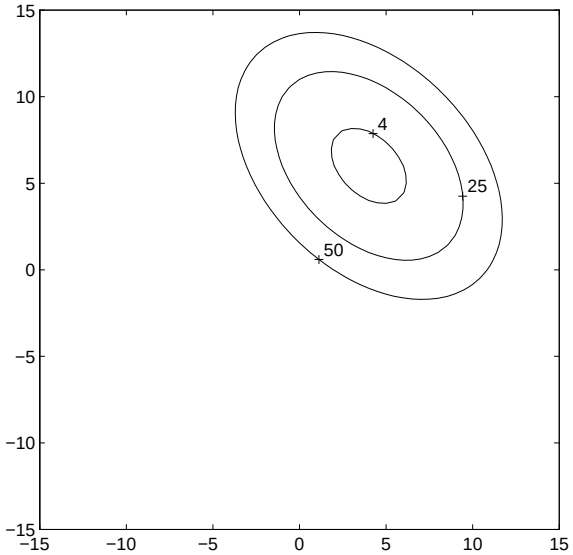


Figure 8: Contours of the MSE surface.

Mean-Square Error Surface



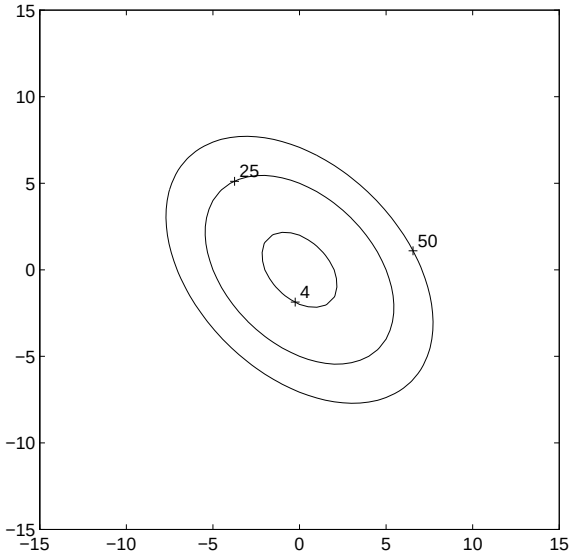


Figure 9: Translated contours of the MSE surface.

Mean-Square Error Surface



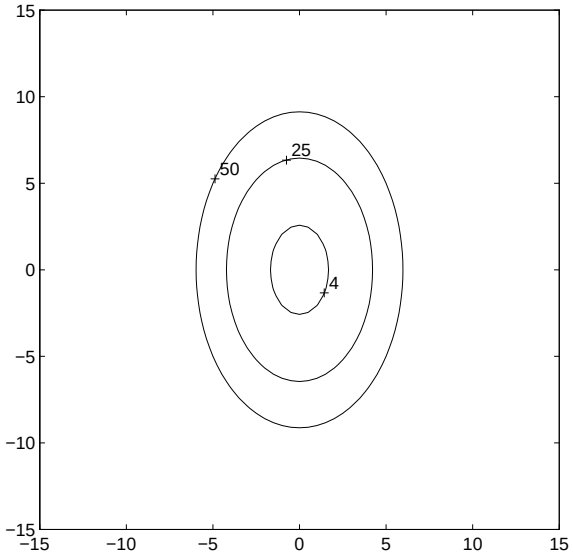


Figure 10: Rotated contours of the MSE surface.

Newton Algorithm



Coefficient updating using the Newton method

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \mathbf{R}^{-1} \mathbf{g}_{\mathbf{W}}(k)$$

the Wiener solution is reached in one step.

Newton like search algorithm updating equation

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \hat{\mathbf{R}}^{-1}(k) \hat{\mathbf{g}}_{\mathbf{W}}(k)$$

where $\hat{\mathbf{R}}^{-1}(k)$ is an estimate of $\mathbf{R}^{-1}$ and $\hat{\mathbf{g}}_{\mathbf{W}}(k)$ is an estimate of $\mathbf{g}_{\mathbf{W}}$, and μ is the convergence factor.

Steepest Descent Algorithm



Steepest descent updating equation

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \mathbf{g}_{\mathbf{w}}(k)$$

Alternative gradient-based algorithms replace $\mathbf{g}_{\mathbf{w}}(k)$ by an estimate $\hat{\mathbf{g}}_{\mathbf{w}}(k)$.

Note that

$$\mathbf{w}(k+1) = \mathbf{w}(k) - 2\mu \mathbf{R}\mathbf{w}(k) + 2\mu \mathbf{p} \tag{94}$$

An analysis is required to determine the influence of μ in the convergence of the algorithm.

Steepest Descent Algorithm



The steepest descent algorithm can be as

$$\begin{aligned}\Delta\mathbf{w}(k+1) &= \Delta\mathbf{w}(k) - 2\mu(\mathbf{R}\mathbf{w}(k) - \mathbf{R}\mathbf{w}_o) \\ &= \Delta\mathbf{w}(k) - 2\mu\mathbf{R}\Delta\mathbf{w}(k) \\ &= (\mathbf{I} - 2\mu\mathbf{R})\,\Delta\mathbf{w}(k) \qquad (95)\end{aligned}$$

where $\mathbf{p} = \mathbf{R}\mathbf{w}_o$ was employed. It can be shown that

$$\Delta\mathbf{w}(k+1) = (\mathbf{I} - 2\mu\mathbf{R})^{k+1}\Delta\mathbf{w}(0) \qquad (96)$$

Steepest Descent Algorithm



or

$$\mathbf{w}(k+1) = \mathbf{w}_o + (\mathbf{I} - 2\mu\mathbf{R})^{k+1}(\mathbf{w}(0) - \mathbf{w}_o) \qquad (97)$$

Premultiplying by $\mathbf{Q}^T$, yields

Steepest Descent Algorithm



$$\mathbf{Q}^T \Delta\mathbf{w}(k+1) = (\mathbf{I} - 2\mu\mathbf{Q}^T\mathbf{R}\mathbf{Q})\mathbf{Q}^T\Delta\mathbf{w}(k)$$

$$= \quad \mathbf{v}(k+1)$$

$$= \quad (\mathbf{I} - 2\mu\mathbf{\Lambda})\mathbf{v}(k) =$$

$$\begin{bmatrix} 1-2\mu\lambda_0 & 0 & \cdots & 0 \\ 0 & 1-2\mu\lambda_1 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & 1-2\mu\lambda_N \end{bmatrix} \times \quad \mathbf{v}(k)$$

where $\mathbf{v}(k+1) = \mathbf{Q}^T\Delta\mathbf{w}(k+1)$ is the rotated coef. vector error.

Steepest Descent Algorithm



Using induction

$$\mathbf{v}(k+1) = (\mathbf{I} - 2\mu\mathbf{\Lambda})^{k+1}\mathbf{v}(0)$$

$$\mathbf{v}(k+1) =$$

$$\begin{bmatrix} (1-2\mu\lambda_0)^{k+1} & 0 & \cdots & 0 \\ 0 & (1-2\mu\lambda_1)^{k+1} & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & (1-2\mu\lambda_N)^{k+1} \end{bmatrix}$$

$$\times \quad \mathbf{v}(0) \tag{98}$$

Steepest Descent Algorithm



To guarantee the convergence of the coefficients, each element $1 - 2\mu\lambda_i$ must be less than one, therefore

$$0 < \mu < \frac{1}{\lambda_{max}} \tag{99}$$

The μ value in this range guarantees that the coefficient vector approaches $\mathbf{w}_o$.

Steepest Descent Algorithm



The MSE transient behavior

$$\xi(k) = \xi_{min} + \Delta\mathbf{w}^T(k)\mathbf{R}\Delta\mathbf{w}(k) \tag{100}$$

the MSE can then be simplified as follows

$$\begin{aligned}
\xi(k) &= \xi_{min} + \Delta\mathbf{w}^T(k)\mathbf{Q}\mathbf{\Lambda}\ \mathbf{Q}^T\Delta\mathbf{w}(k) \\
&= \xi_{min} + \mathbf{v}^T(k)\mathbf{\Lambda}\ \mathbf{v}(k) \\
&= \xi_{min} + \sum_{i=0}^{N}\lambda_i v_i^2(k) \qquad (101)
\end{aligned}$$

where $\mathbf{v}(k) = \mathbf{Q}^T\Delta\mathbf{w}(k)$.

Steepest Descent Algorithm



It can be shown that

$$\xi(k) = \xi_{min} + \mathbf{v}^T(k-1)(\mathbf{I}-2\mu\mathbf{\Lambda})\mathbf{\Lambda}\ (\mathbf{I}-2\mu\mathbf{\Lambda})\mathbf{v}(k-1)$$

$$= \xi_{min} + \sum_{i=0}^{N} \lambda_i (1-2\mu\lambda_i)^{2k} v_i^2(0) \qquad (102)$$

Steepest Descent Algorithm



The convergence will follow $(N+1)$ geometric decaying curves with ratios $r_{wi} = (1 - 2\mu\lambda_i)$. These curves can be fitted to an exponential envelope with time constant τ_i

$$r_{wi} = \mathrm{e}^{\frac{-1}{\tau_{wi}}} = 1 - \frac{1}{\tau_{wi}} + \frac{1}{2!\tau_{wi}^2} + \cdots \tag{103}$$

In general, r_{wi} is slightly smaller than one. Therefore,

$$r_{wi} = (1 - 2\mu\lambda_i) \approx 1 - \frac{1}{\tau_{wi}} \tag{104}$$

then

$$\tau_{wi} \approx \frac{1}{2\mu\lambda_i}$$

for $i = 0, 1, \cdots, N$.

Steepest Descent Algorithm



For the convergence of the MSE, the range of values for μ is the same. The geometric decaying curves have ratios given by $r_{ei} = (1 - 4\mu\lambda_i)$ with exponential envelopes with

$$\tau_{ei} \approx \frac{1}{4\mu\lambda_i} \tag{105}$$

for $i = 0, 1, \cdots, N$.

Example



Example 2.2

R and **p** are known:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.4045 \\ 0.4045 & 1 \end{bmatrix}$$

$$\mathbf{p} = [0 \;\; 0.2939]^T$$

Example



$$E[d^2(k)] = 0.5$$

(a) Deduce the equation for the MSE.

(b) Choose a small value for μ, and starting the parameters at $[0 \quad -2]^T$ plot the convergence path of the steepest-descent algorithm in the MSE surface.

(c) Repeat the previous item for the Newton algorithm.

Example



Solution:

(a) The MSE function is given by

$$\begin{aligned}
\xi &= E[d^2(k)] - 2\mathbf{w}^T\mathbf{p} + \mathbf{w}^T\mathbf{R}\mathbf{w} \\
&= \sigma_d^2 - 2[w_1 \;\; w_2]\begin{bmatrix} 0. \\ 0.2939 \end{bmatrix} \\
&\quad +[w_1 \;\; w_2]\begin{bmatrix} 1 & 0.4045 \\ 0.4045 & 1 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}
\end{aligned}$$

Example



Result

$$\xi = 0.5 + w_1^2 + w_2^2 + 0.8090 w_1 w_2 - 0.5878 w_2$$

(b) Using $\mu = 0.1/\lambda_{max}$, where $\lambda_{max} = 1.4045$.

Example



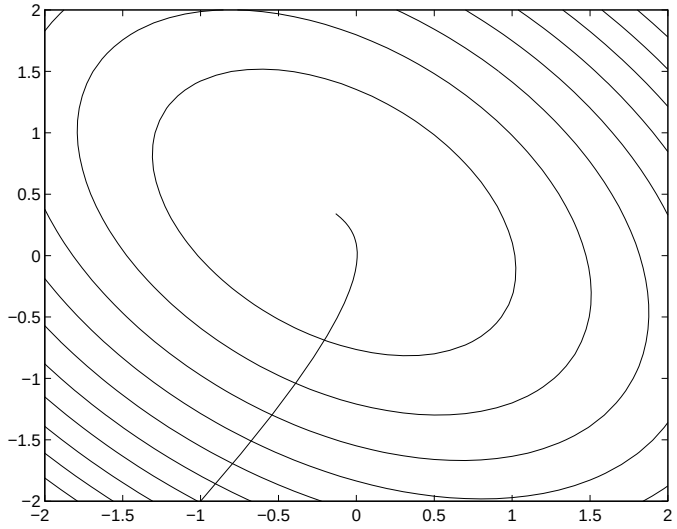


Figure 11: Convergence path of the steepest-descent algorithm.

Example



(c) Using a convergence factor $\mu = 0.1/\lambda_{max}$.

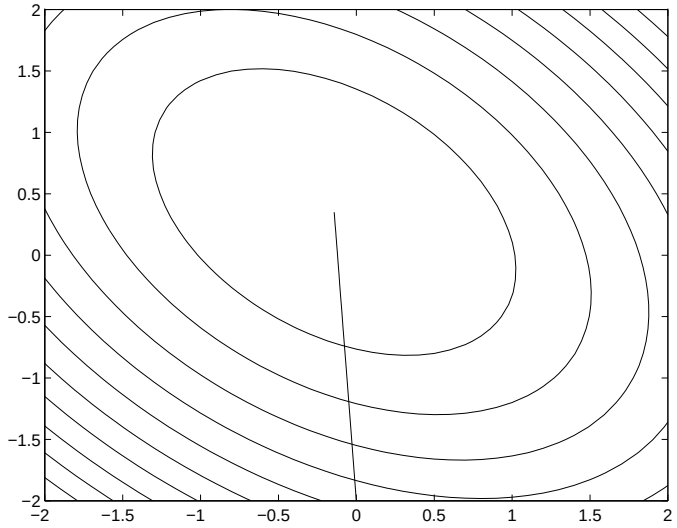


Figure 12: Convergence path of the Newton algorithm.

Adaptive Filtering Algorithms



FIR

- Gradient Based
  - LMS
  - Signs
  - Frequency domain
- Gauss Newton

} **MSE**

Adaptive Filtering Algorithms



- RLS
  - Conventional
  - Fast Conventional
  - QR Decomposition
  - Fast QR
  - Lattice
  - QR Lattice
  - Hybrid Lattice/Conv. QR

} **LS**

Adaptive Filtering Algorithms



IIR

- Output Error Based $\begin{cases} MSE \\ LS \\ OTHERS \end{cases}$
- SM Method $\{Bias\ Problems$
- Eq. Error Method $\{Bias\ Problems$
- Composite Error $\begin{cases} Moderate\ Bias \\ Convergence\ Problems \end{cases}$

Applications Revisited



- System Identification
- Signal Enhancement
- Signal Prediction
- Channel Equalization

Applications Revisited



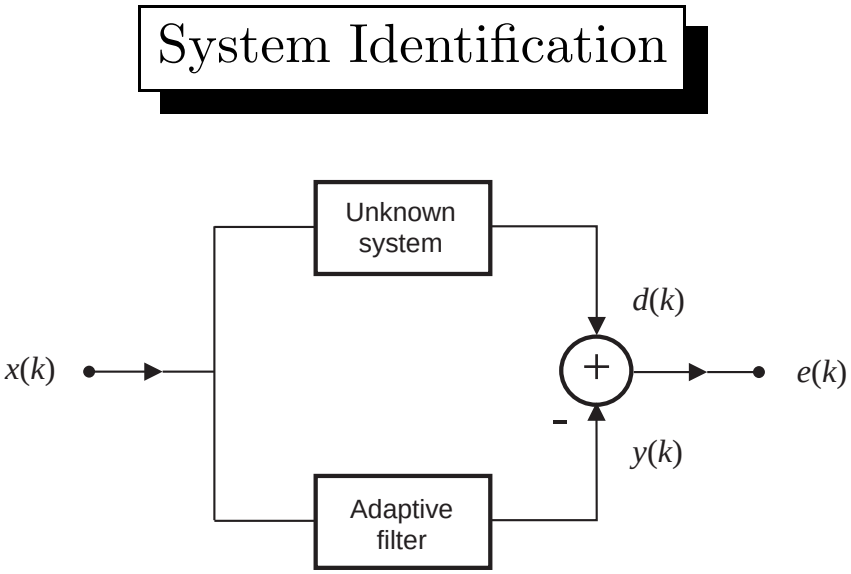


Figure 13: System identification.

Example



Assume the unknown system has impulse response given by $h(k)$, for $k = 0,\ 1,\ 2,\ 3, ..., \infty$. The error signal is then given by

$$\begin{aligned} e(n) &= d(k) - y(k) \\ &= \sum_{l=0}^{\infty} h(l)x(k-l) - \sum_{i=0}^{N} w_i(k)x(k-i) \end{aligned} \tag{106}$$

$w_i(k)$ are the coefficients of the adaptive filter.

Applications Revisited



Assuming that $x(k)$ is a white noise, the MSE for a fixed $\mathbf{w}$ is given by

$$\begin{aligned}
\xi &= E[(\mathbf{h}^T\mathbf{x}_\infty(k) - \mathbf{w}^T\mathbf{x}_{N+1}(k))^2] \\
&= E[\mathbf{h}^T\mathbf{x}_\infty(k)\mathbf{x}_\infty^T(k)\mathbf{h} - 2\mathbf{h}^T\mathbf{x}_\infty(k)\mathbf{x}_{N+1}^T(k)\mathbf{w} \\
&+ \mathbf{w}^T\mathbf{x}_{N+1}(k)\mathbf{x}_{N+1}^T(k)\mathbf{w}] \\
&= \sigma_x^2\sum_{i=0}^{\infty} h^2(i) - 2\sigma_x^2\mathbf{h}^T\begin{bmatrix}\mathbf{I}_{N+1}\\ \mathbf{0}\end{bmatrix}\mathbf{w} + \mathbf{w}^T\mathbf{R}_{N+1}\mathbf{w} \quad (107)
\end{aligned}$$

where $\mathbf{x}_\infty(k)$ and $\mathbf{x}_{N+1}(k)$ are the input signal vector with infinite and finite lengths, respectively.

Applications Revisited



The derivative of ξ with respect to the coefficients is

$$\mathbf{w}_o = \mathbf{h}_{N+1} \tag{108}$$

where

$$\mathbf{h}_{N+1}^T = \mathbf{h}^T \begin{bmatrix} \mathbf{I}_{N+1} \\ \mathbf{0} \end{bmatrix} \tag{109}$$

If the input is white, the best model is a system whose impulse response coincides with the $N+1$ first samples of the unknown system impulse response.

Applications Revisited



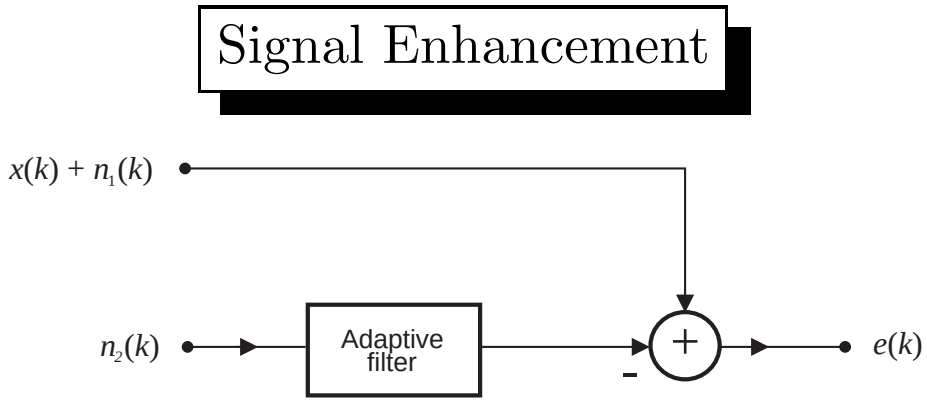


Figure 14: Signal enhancement ($n_1(k)$ and $n_2(k)$ are noise signals correlated to each other).

Applications Revisited



The reference signal consists of a desired signal $x(k)$ that is corrupted by an additive noise $n_1(k)$. The input signal is a noise signal $n_2(k)$ that is correlated with the interference signal $n_1(k)$, but uncorrelated with $x(k)$.

In this application, the error signal is given by

$$\begin{aligned} e(k) &= x(k) + n_1(k) - \sum_{l=0}^{N} w_l n_2(k-l) \\ &= x(k) + n_1(k) - y(k) \end{aligned} \tag{110}$$

Applications Revisited



The resulting MSE is then given by

$$E[e^2(k)] = E[x^2(k)] + E[(n_1(k) - y(k))^2] \tag{111}$$

with $x(k)$ uncorrelated with $n_1(k)$ and $n_2(k)$.

The equation above shows that if the adaptive filter, having $n_2(k)$ as input, is able to perfectly predict $n_1(k)$, the minimum MSE is given by

Applications Revisited



$$\xi_{min} = E[x^2(k)] \tag{112}$$

where the error signal, in this situation, is the desired signal $x(k)$.

The effectiveness of the signal enhancement depends on the high correlation between $n_1(k)$ and $n_2(k)$.

In some applications, it is useful to include a delay of L samples in the reference signal or in the input signal, such that their relative delay yields a maximum crosscorrelation between $y(k)$ and $n_1(k)$, reducing the MSE.

Applications Revisited



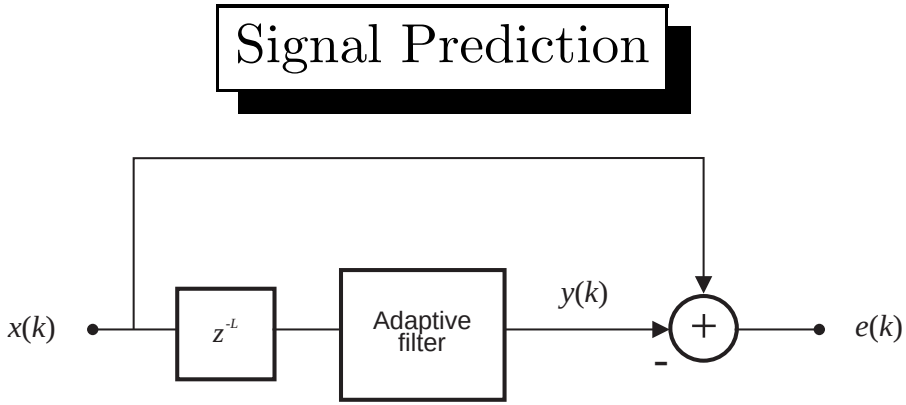


Figure 15: Signal prediction.

Applications Revisited



The input consists of a delayed version of the desired signal. The MSE is given by

$$\xi = E[(x(k) - \mathbf{w}^T \mathbf{x}(k-L))^2] \tag{113}$$

The predictor is able to predict the present sample of the input signal using as information old samples such as
$x(k-L)$, $x(k-L-1)$, $x(k-L-N)$.

Applications Revisited



The minimum MSE is given by

$$\xi_{min} = r(0) - \mathbf{w}_o^T \begin{bmatrix} r(L) \\ r(L+1) \\ . \\ . \\ . \\ r(L+N) \end{bmatrix} \tag{114}$$

where $\mathbf{w}_o$ is the optimum coefficient vector, and $r(l) = E[x(k)x(k-l)]$ for a stationary process.

Applications Revisited



Main application is in linear prediction coding of speech signals, where the predictor's task is to estimate the speech parameters. These parameters $\mathbf{w}$ are part of the coding information that is transmitted or stored along with other informations inherent to the speech characteristics, such as pitch period.

The adaptive signal predictor is used for adaptive line enhancement (ALE), where the input signal is a narrowband signal (predictable), added to a wideband signal. The predictor output will be an enhanced version of the narrowband signal.

Applications Revisited



Another application of the signal predictor is the suppression of narrowband interference in a wideband signal. The input signal has the same general characteristics of the ALE, however we are now interested in getting rid of the narrowband interferer. For such application, the output signal of interest is the error signal.

Applications Revisited



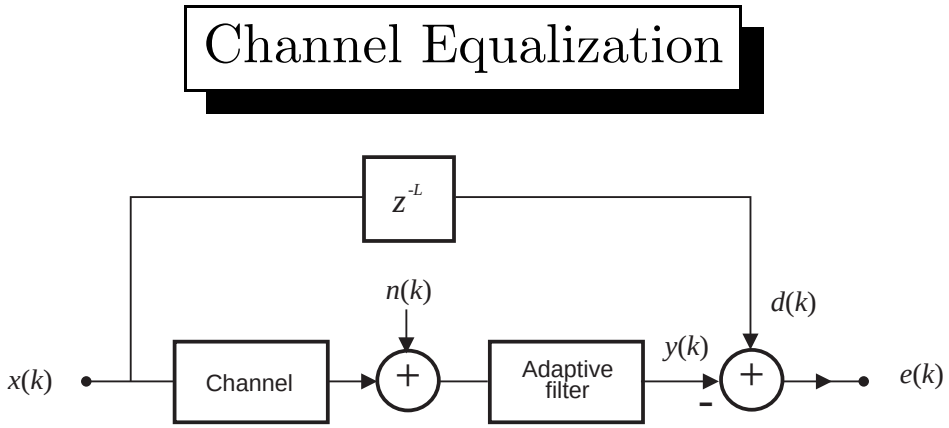


Figure 16: Channel equalization.

Applications Revisited



The channel equalization or inverse filtering consists of estimating a transfer function to compensate for the linear distortion caused by the channel.

In the ideal situation, where $n(k) = 0$ and the equalizer has sufficient order, the error signal is zero if

$$W(z)H(z) = z^{-L} \tag{115}$$

$W(z)$ and $H(z)$ are the equalizer and unknown system transfer functions.

Applications Revisited



Therefore

$$W(z) = \frac{z^{-L}}{H(z)} \tag{116}$$

From the equation above we can conclude that if $H(z)$ is an IIR transfer function with nontrivial numerator and denominator polynomials, $W(z)$ will also be IIR. If $H(z)$ is an allpole model, $W(z)$ is FIR. If $H(z)$ is an allzero model, $W(z)$ is an allpole transfer function.

Applications Revisited



The delay in the reference signal plays an important role in the equalization process. Without the delay, the desired signal is $x(k)$, whereas the signal $y(k)$ will be mainly influenced by old samples of the input signal, since the unknown system is usually causal. The equalizer performs the task of predicting $x(k)$ simultaneously with the task of equalizing the channel. The introduction of a delay alleviates the prediction task, leaving the equalizer free to invert the channel response.

A rule of thumb is that the best delay should be close to half the time span of the equalizer. We should in practice try different delays.

Applications Revisited



If the unknown system is not of minimum phase, the optimum equalizer is either stable and noncausal, or unstable and causal. Both solutions are unacceptable. The noncausal stable solution could be better approximated by a causal FIR filter when the delay is included in the desired signal. The delay forces a time shift in the ideal impulse response of the equalizer, allowing the time span where most of the energy is concentrated to be in the *causal* region.

If the channel noise signal is present and is uncorrelated with the channel's input signal, the error signal and $y(k)$ will be noisier.

Applications Revisited



The channel impulse response is a time spread sequence described by $h(k)$ with the received signal being given by

$$re(k+J) = x(k)h(J) + \sum_{l=-\infty,\, l\neq k}^{k+J} x(l)h(k+J-l) + n(k+J) \qquad (117)$$

where J denotes the channel time delay (including the sampler phase).

Applications Revisited



The first term of corresponds to the desired information, the second term is the interference of the symbols sent before and after $x(k)$. The third term accounts for the channel noise.

The equalizer coefficients are updated in two distinct circumstances by employing different reference signals. During the equalizer training period, a previously chosen training signal is transmitted through the channel and a properly delayed version of this signal, that is prestored in the receiver end, is used as reference signal.

Applications Revisited



After convergence the error between the adaptive filter output and the decision device output is utilized to update the coefficients.

The resulting scheme is the so called decision-directed adaptive equalizer.

A more general equalizer scheme is the so called decision-feedback equalizer (DFE). The basic idea is to feed back, via a second FIR filter, the decisions made by the decision device that is applied to the equalized signal. Assuming the decisions were correct, we are actually feeding back the symbols $x(l)$, for $l < k$, of equation (117).

Applications Revisited



The DFE is able to cancel the post-cursor ISI, leaving more freedom for the feedforward section to take care of the remaining terms of the ISI.

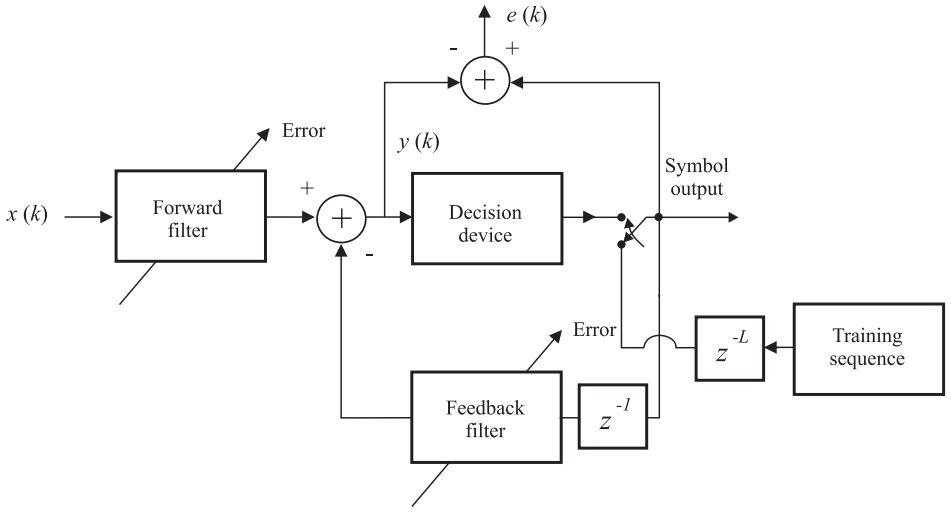


Figure 17: Decision-feedback equalizer.

Applications Revisited



Main characteristics

- The signals that are fed back are symbols, being noise free and allowing computational savings.
- The noise enhancement is reduced if compared with the feedforward only equalizer.
- Short time recovery when incorrect decisions are made.
- Reduced sensitivity to sampling phase.

Applications Revisited



Digital Communication System

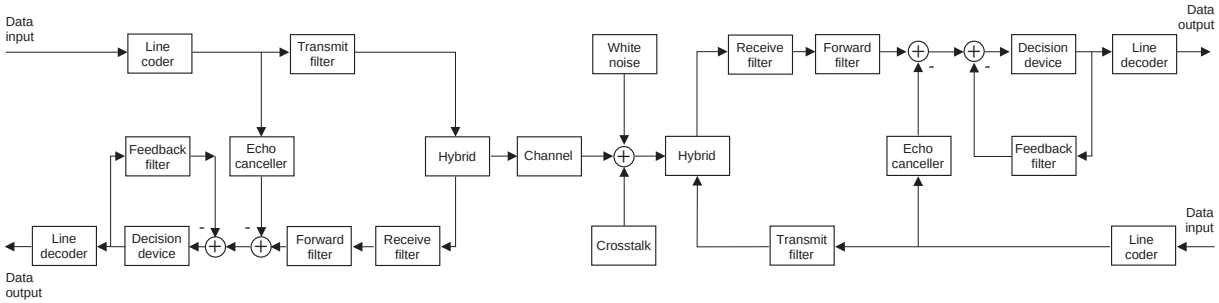


Figure 18: General digital communication transceiver.

Examples



Example 2.4

To verify the effectiveness of the Wiener solution in environments related to the applications of noise cancellation, prediction, equalization, and identification.

(a) In a noise cancellation environment a sinusoid is corrupted by noise as follows

$$d(k) = \cos \omega_0 k + n_1(k)$$

Examples



with

$$n_1(k) = -an_1(k-1) + n(k)$$

$|a| < 1$ and $n(k)$ is a zero-mean white noise with variance $\sigma_n^2 = 1$.
The input signal is described by

$$n_2(k) = -bn_2(k-1) + n(k)$$

where $|b| < 1$.

Examples



(b) In a prediction case the input signal is modeled as

$$x(k) = -ax(k-1) + n(k)$$

with $n(k)$ being a white noise with unit variance and $|a| < 1$.

Examples



(c) In an equalization problem a zero-mean white noise signal $s(k)$ with variance c is transmitted through a channel with an AR model described by

$$\hat{x}(k) = -a\hat{x}(k-1) + s(k)$$

with $|a| < 1$ and the received signal given by

$$x(k) = \hat{x}(k) + n(k)$$

whereas $n(k)$ is a zero-mean white noise with variance d and uncorrelated with $s(k)$.

Examples



(d) In a system identification problem a zero-mean white noise signal $x(k)$ with variance c is employed as the input signal to identify an AR system whose model is described by

$$v(k) = -av(k-1) + x(k)$$

where $|a| < 1$ and the desired signal is given by

$$d(k) = v(k) + n(k)$$

Examples



Repeat the problem if the system to be identified is an MA whose model is described by

$$v(k) = -ax(k-1) + x(k)$$

For all these cases describe the Wiener solution with two coefficients.

Examples



Solution:

A 2×2 matrix inversion is performed as

$$\mathbf{R}^{-1} = \frac{1}{r_{11}r_{22} - r_{12}r_{21}} \begin{bmatrix} r_{22} & -r_{12} \\ -r_{21} & r_{11} \end{bmatrix}$$

where r_{ij} is the element of row i and column j of the matrix $\mathbf{R}$.

Examples



For two first-order AR modeled signals $x(k)$ and $v(k)$, whose poles are respectively placed at $-a$ and $-b$ with the same white noise input with unit variance, their cross-correlations are given by

$$E[x(k)v(k-l)] = \frac{(-a)^l}{1-ab}$$

for $l > 0$, and

$$E[x(k)v(k-l)] = \frac{(-b)^l}{1-ab}$$

for $l < 0$, are frequently required in the following solutions.

Examples



(a)

The input signal in this case is given by $n_2(k)$, whereas the desired signal is given by $d(k)$. The elements of the correlation matrix are computed as

$$E[n_2(k)n_2(k-l)] = \frac{(-b)^{|l|}}{1-b^2}$$

Examples



The expression for the cross-correlation vector is given by

$$\mathbf{p} = \begin{bmatrix} E[(\cos\omega_0 k + n_1(k))n_2(k)] \\ E[(\cos\omega_0 k + n_1(k))n_2(k-1)] \end{bmatrix}$$

$$= \begin{bmatrix} E[n_1(k)n_2(k)] \\ E[n_1(k)n_2(k-1)] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{1-ab}\sigma_n^2 \\ -\frac{a}{1-ab}\sigma_n^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{1-ab} \\ -\frac{a}{1-ab} \end{bmatrix}$$

where we used $\sigma_n^2 = 1$.

Examples



The Wiener solution is given by

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{1-ab} \\ -\frac{a}{1-ab} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{b-a}{1-ab} \end{bmatrix}$$

Examples



The special case where $a = 0$

$$\mathbf{w}_o = \begin{bmatrix} 1 \\ b \end{bmatrix}$$

such that the error signal is given by

Examples



$$
\begin{aligned}
e(k) = d(k) - y(k) &= \cos\omega_0 k + n(k) - \mathbf{w}_o^T \begin{bmatrix} n_2(k) \\ n_2(k-1) \end{bmatrix} \\
&= \cos\omega_0 k + n(k) - n_2(k) - b n_2(k-1) \\
&= \cos\omega_0 k + n(k) + b n_2(k-1) - n(k) - b n_2(k-1) = \cos\omega_0 k
\end{aligned}
$$

As can be observed the cosine signal is fully recovered since the Wiener filter was able to restore $n(k)$ and remove it from the desired signal.

Examples



(b)

In the prediction case the desired signal is $x(k+L)$. Since

$$E[x(k)x(k-L)] = \frac{(-a)^{|L|}}{1-a^2}$$

the input signal correlation matrix is

$$\mathbf{R} = \begin{bmatrix} E[x^2(k)] & E[x(k)x(k-1)] \\ E[x(k)x(k-1)] & E[x^2(k-1)] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{1-a^2} & -\frac{a}{1-a^2} \\ -\frac{a}{1-a^2} & \frac{1}{1-a^2} \end{bmatrix}$$

Examples



Vector $\mathbf{p}$ is described by

$$\mathbf{p} = \begin{bmatrix} E[x(k+L)x(k)] \\ E[x(k+L)x(k-1)] \end{bmatrix} = \begin{bmatrix} \frac{(-a)^{|L|}}{1-a^2} \\ \frac{(-a)^{|L+1|}}{1-a^2} \end{bmatrix}$$

Examples



The expression for the optimal coefficient vector is easily derived.

$$
\begin{aligned}
\mathbf{w}_o &= \mathbf{R}^{-1}\mathbf{p} \\
&= (1-a^2)\begin{bmatrix} \frac{1}{1-a^2} & \frac{a}{1-a^2} \\ \frac{a}{1-a^2} & \frac{1}{1-a^2} \end{bmatrix}\begin{bmatrix} \frac{(-a)^L}{1-a^2} \\ \frac{(-a)^{L+1}}{1-a^2} \end{bmatrix} \\
&= \begin{bmatrix} (-a)^L \\ 0 \end{bmatrix}
\end{aligned}
$$

where L is considered positive.

Examples



The predictor result tells us that an estimate $\hat{x}(k+L)$ of $x(k+L)$ can be obtained as

$$\hat{x}(k+L) = (-a)^L x(k)$$

The actual value of $x(k+L)$ is

$$x(k+L) = (-a)^L x(k) + \sum_{i=0}^{L-1} (-a)^i n(k-i)$$

Examples



If $x(k)$ is an observed data at a given instant of time, the best estimate of $x(k+L)$ in terms of $x(k)$ is to average out the noise as follows

$$\hat{x}(k+L) = (-a)^L x(k) + E[\sum_{i=0}^{L-1}(-a)^i n(k-i)] = (-a)^L x(k)$$

since $E[n(k-i)] = 0$.

Examples



(c)

In this equalization problem, matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \begin{bmatrix} E[x^2(k)] & E[x(k)x(k-1)] \\ E[x(k)x(k-1)] & E[x^2(k-1)] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{1-a^2}c + d & -\frac{a}{1-a^2}c \\ -\frac{a}{1-a^2}c & \frac{1}{1-a^2}c + d \end{bmatrix}$$

Examples



By utilizing as desired signal $s(k-L)$ and recalling that it is a white noise and uncorrelated with the other signals involved in the experiment, the cross-correlation vector between the input and desired signals has the following expression

$$\mathbf{p} = \begin{bmatrix} E[x(k)s(k-L)] \\ E[x(k-1)s(k-L)] \end{bmatrix} = \begin{bmatrix} (-1)^L a^L c \\ (-1)^{L-1} a^{L-1} c \end{bmatrix}$$

Examples



The coefficients of the underlying Wiener solution are given by

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$$

$$= \frac{1}{\frac{c^2}{1-a^2}+2\frac{dc}{1-a^2}+d^2} \begin{bmatrix} \frac{1}{1-a^2}c+d & \frac{a}{1-a^2}c \\ \frac{a}{1-a^2}c & \frac{1}{1-a^2}c+d \end{bmatrix} \begin{bmatrix} (-1)^L a^L c \\ (-1)^{L-1} a^{L-1} c \end{bmatrix}$$

Examples



$$\mathbf{w}_o = \frac{(-1)^L a^L c}{\frac{c^2}{1-a^2} + 2\frac{cd}{1-a^2} + d^2} \begin{bmatrix} \frac{c}{1-a^2} + d - \frac{c}{1-a^2} \\ \frac{ac}{1-a^2} - a^{-1}d - \frac{a^{-1}c}{1-a^2} \end{bmatrix}$$

$$= \frac{(-1)^L a^L c}{\frac{c^2}{1-a^2} + 2\frac{cd}{1-a^2} + d^2} \begin{bmatrix} d \\ -a^{-1}d - a^{-1}c \end{bmatrix}$$

Examples



If there is no additional noise, i.e. $d = 0$, the result above becomes

$$\mathbf{w}_o = \begin{bmatrix} 0 \\ (-1)^{L-1} a^{L-1} (1 - a^2) \end{bmatrix}$$

The Wiener solution is correcting the gain of $x(k-1)$, while not using its most recent component $x(k)$. Because $s(k-L)$ at instant k has a defined correlation with any previously received symbol. If the signal $s(k)$ is colored the Wiener filter would have a nonzero first coefficient in a noiseless environment. In case there is environ. noise, the solution finds a perfect balance between the desired signal modeling and the noise amplification.

Examples



(d)

In the system identification example the input signal correlation matrix is given by

$$\mathbf{R} = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$$

with the desired signal $d(k)$ the cross-correlation vector is described as

$$\mathbf{p} = \begin{bmatrix} E[x(k)d(k)] \\ E[x(k-1)d(k)] \end{bmatrix} = \begin{bmatrix} c \\ -ca \end{bmatrix}$$

Examples



The coefficients of the underlying Wiener solution are given by

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} = \begin{bmatrix} \frac{1}{c} & 0 \\ 0 & \frac{1}{c} \end{bmatrix} \begin{bmatrix} c \\ -ca \end{bmatrix} = \begin{bmatrix} 1 \\ -a \end{bmatrix}$$

Examples



Note that this solution represents the best way a first-order FIR model can approximate an IIR model, since

$$W_o(z) = 1 - az^{-1}$$

and

$$\frac{1}{1 + az^{-1}} = 1 - az^{-1} + a^2 z^{-2} + \ldots$$

On the other hand, if the unknown model is the described FIR model such as $v(k) = -ax(k-1) + x(k)$, the Wiener solution remains the same and corresponds exactly to the unknown system model.