Several Random Variables (Con't)

The joint central moment of order kq involving random variables x and y is defined as

$$\boxed{\begin{aligned}\mu_{kq} &= E[(x-m_x)^k(y-m_y)^q] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)^k(y-m_y)^q p(x,y)\,dx\,dy\end{aligned}}$$

where $m_x = E[x]$ and $m_y = E[y]$ are the means of x and y, as defined earlier. We note that

$$\boxed{\mu_{20} = E[(x-m_x)^2]} \quad \text{and} \quad \boxed{\mu_{02} = E[(y-m_y)^2]}$$

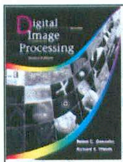
are the variances of x and y, respectively.

Several Random Variables (Con't)

The moment μ_{11}

$$\boxed{\begin{aligned}\mu_{11} &= E[(x-m_x)(y-m_y)] \\ C_{xy} = &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)(y-m_y)p(x,y)\,dx\,dy\end{aligned}}$$

is called the ***covariance*** of x and y. As in the case of correlation, the covariance is an important concept, usually given a special symbol such as C_{xy}.



Several Random Variables (Con't)

By direct expansion of the terms inside the expected value brackets, and recalling the $m_x = E[x]$ and $m_y = E[y]$, it is straightforward to show that

$$\boxed{\begin{aligned} C_{xy} &= E[xy] - m_y E[x] - m_x E[y] + m_x m_y \\ &= E[xy] - E[x]E[y] \\ &= R_{xy} - E[x]E[y]. \end{aligned}}$$

From our discussion on correlation, we see that the covariance is zero if the random variables are either uncorrelated *or* statistically independent. This is an important result worth remembering.

$$\begin{aligned} C_{xy} &= E[(x-m_x)(y-m_y)] = E[xy - m_x y - m_y x + m_x m_y] \\ &= E[xy] - m_x E[y] - m_y E[x] + m_x m_y \end{aligned}$$

If $R_{xy} = E[x]E[y]$ The variables are uncorrelated.

and $C_{xy} = R_{xy} - E[x]E[y] = 0$

Several Random Variables (Con't)

If we divide the covariance by the square root of the product of the variances we obtain

$$\boxed{\begin{aligned}\gamma &= \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}} \\ &= \frac{C_{xy}}{\sigma_x \sigma_y} \\ &= E\left[\frac{(x-m_x)}{\sigma_x}\frac{(y-m_y)}{\sigma_y}\right].\end{aligned}}$$

The quantity γ is called the *correlation coefficient* of random variables x and y. It can be shown that γ is in the range $-1 \le \gamma \le 1$ (see Problem 12.5). As discussed in Section 12.2.1, the correlation coefficient is used in image processing for matching.

The Multivariate Gaussian Density

As an illustration of a probability density function of more than one random variable, we consider the ***multivariate Gaussian probability density function***, defined as

$$\boxed{p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right]}}$$

where n is the ***dimensionality*** (number of components) of the random vector $\mathbf{x}$, $\mathbf{C}$ is the ***covariance matrix*** (to be defined below), $|\mathbf{C}|$ is the determinant of matrix $\mathbf{C}$, $\mathbf{m}$ is the ***mean vector*** (also to be defined below) and T indicates transposition (see the review of matrices and vectors).

The Multivariate Gaussian Density (Con't)

The *mean vector* is defined as

$$\boxed{\mathbf{m} = E[\mathbf{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix}}$$

and the *covariance matrix* is defined as

$$\boxed{\mathbf{C} = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T].}$$

The Multivariate Gaussian Density (Con't)

The element of **C** are the covariances of the elements of **x**, such that

$$\boxed{c_{ij} = C_{x_i x_j} = E[(x_i - m_i)(x_j - m_j)]}$$

where, for example, x_i is the *i*th component of **x** and m_i is the *i*th component of **m**.

The Multivariate Gaussian Density (Con't)

Covariance matrices are ***real*** and ***symmetric*** (see the review of matrices and vectors). The elements along the main diagonal of **C** are the variances of the elements **x**, such that $c_{ii} = \sigma_{x_i}^2$. When all the elements of **x** are uncorrelated or statistically independent, $c_{ij} = 0$, and the covariance matrix becomes a ***diagonal matrix***. If all the variances are equal, then the covariance matrix becomes proportional to the ***identity matrix***, with the constant of proportionality being the variance of the elements of **x**.

$$\begin{bmatrix} .3 & -.5 & -.2 \\ -.5 & .2 & 0 \\ -.2 & 0 & .6 \end{bmatrix}$$

element are real
symmetric about the diagonal.

the diagonal elements are the variances of the variables.

$$\begin{bmatrix} .4 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .2 \end{bmatrix}$$

in this case the variables are statistically independent

The Multivariate Gaussian Density (Con't)

Example: Consider the following ***bivariate*** ($n = 2$) Gaussian probability density function

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right]}$$

with

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

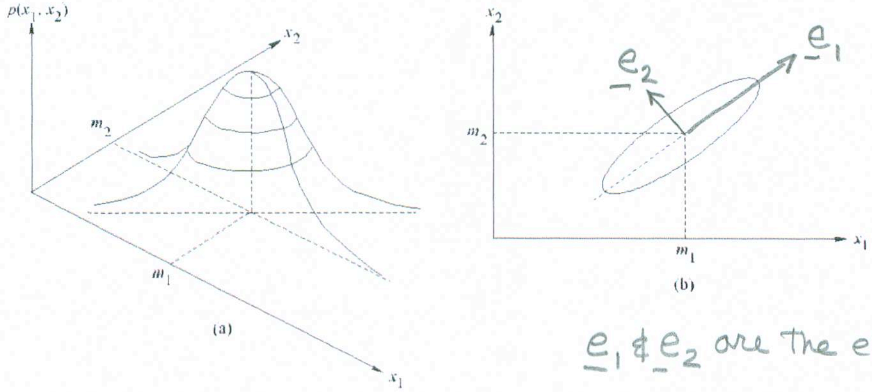
and

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

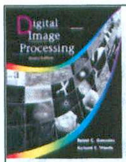
The Multivariate Gaussian Density (Con't)

where, because **C** is known to be symmetric, $c_{12} = c_{21}$. A schematic diagram of this density is shown in Part (a) of the following figure. Part (b) is a horizontal slice of Part (a). From the review of vectors and matrices, we know that the main directions of data spread are in the directions of the eigenvectors of **C**. Furthermore, if the variables are uncorrelated or statistically independent, the covariance matrix will be diagonal and the eigenvectors will be in the same direction as the coordinate axes x_1 and x_2 (and the ellipse shown would be oriented along the x_1 - and x_2-axis). If, the variances along the main diagonal are equal, the density would be symmetrical in all directions (in the form of a bell) and Part (b) would be a circle. Note in Parts (a) and (b) that the density is centered at the mean values (m_1,m_2).

The Multivariate Gaussian Density (Con't)



$\underline{e}_1$ & $\underline{e}_2$ are the eigenvectors of $\underline{C}$



Linear Transformations of Random Variables

As discussed in the *Review of Matrices and Vectors*, a linear transformation of a vector **x** to produce a vector **y** is of the form **y** = **Ax**. Of particular importance in our work is the case when the rows of **A** are the eigenvectors of the covariance matrix. Because **C** is real and symmetric, we know from the discussion in the *Review of Matrices and Vectors* that it is always possible to find *n* orthonormal eigenvectors from which to form **A**. The implications of this are discussed in considerable detail at the end of the *Review of Matrices and Vectors*, which we recommend should be read again as a conclusion to the present discussion.

Example: estimating points on a line

Y	4	1	0	1	4	9
X	−2	−1	0	1	2	3
ζ	⚀	⚁	⚂	⚃	⚄	⚅

Fig. 8.1.1. Definition of a random variable X.

← compute m_Y, σ_Y^2
← compute m_X, σ_X^2

- Estimate the value of X given Y by points on a straight line

$$\hat{X} = aY + b$$

- Write the mean square error as

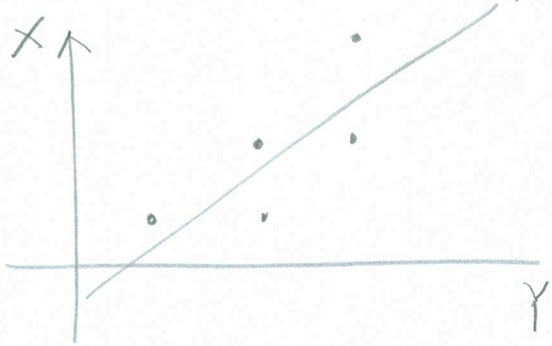
$$E(e^2) = E\left\{\left[X - \hat{X}\right]^2\right\} = E\left\{\left[X - (aY + b)\right]^2\right\}$$

- Set partial derivative of mean square error wrt b equal to zero to get b

$$\frac{\partial}{\partial b} E(e^2) = E\{2[X - aY - b](-1)\} = -2E(X) + 2aE(Y) + 2b = 0$$

$$b = E(X) - aE(Y) = m_X - am_Y$$

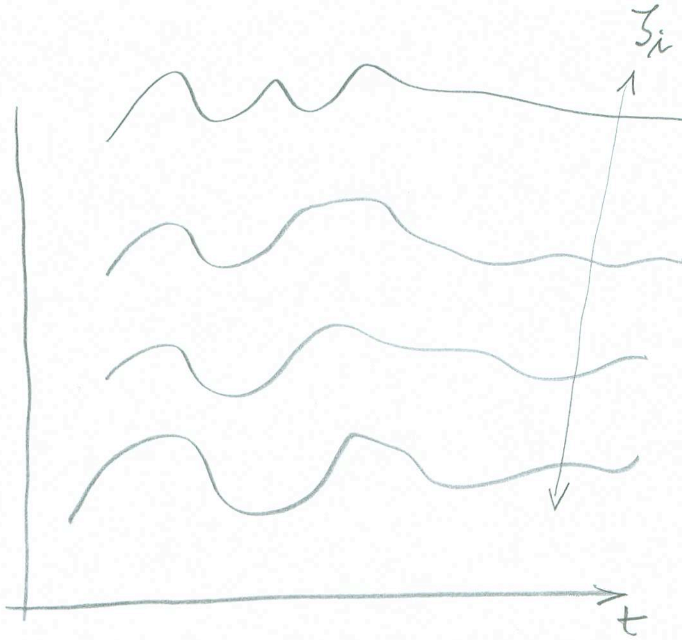
let this be $\hat{X} = aY + b$



estimating points on a line is an example of a minimum mean square estimator

Stochastic processes and ensembles

- A stochastic process produces an output waveform rather than just a number
- A specific output waveform is denoted by $X(t,\zeta_i)$
- A collection of time functions $X(t,\zeta_i)$ is called an ensemble
- Mix, Fig.6.1.1 illustrates an ensemble



This is an ensemble.

Mean Square Estimation

- Let $e = X - \hat{X}$ where e is the error between the random variable X and our estimate $\hat{X}$
- The mean squared error is:

$$E\left(e^2\right) = E\left[\left(X - \hat{X}\right)^2\right]$$

- The value of $\hat{X}$ which minimizes $E\left(e^2\right)$ is the minimum mean-square estimate of X

This is what we used in the Wiener filter

Example: estimating points on a line

substituting for b

- The mean square error is then

$$E(e^2) = E\left\{\left[X - aY - m_X + am_Y\right]^2\right\} = E\left\{\left[(X - m_X) - a(Y - m_Y)\right]^2\right\}$$

covariance variance

$$E(e^2) = E\left\{(X - m_X)^2 - \overset{2}{a}(X - m_X)(Y - m_Y) + a^2(Y - m_Y)^2\right\} = \sigma_X^2 - 2a\mu_{11} + a^2\sigma_Y^2$$

- Take the derivative wrt a and set equal to zero to get $a = \frac{\mu_{11}}{\sigma_Y^2}$

- We can calculate the means and variances of the data to get

$$a = \frac{\mu_{11}}{\sigma_Y^2} = 0.319 \quad b = m_X - am_Y = 0.5 - (0.319)(3.17) = -0.5$$

- or $\hat{X} = 0.319Y - 0.5$

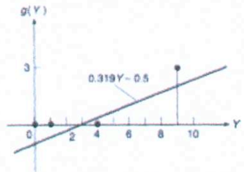


Fig. 8.1.2 Estimates for X.

$a = \frac{\mu_{11}}{\sigma_Y^2}$ where $\mu_{11} = E\left[(x - m_x)(y - m_y)\right]$

the covariance of x and y

$b = m_X - am_Y$

Continuous Waveform Calculations

- The inner product $\langle v_1(t) \mid v_2(t) \rangle = \int_{-\infty}^{\infty} v_1(t) v_2(t)\, dt$

- The norm or length $\| v(t) \| = \sqrt{\int_{-\infty}^{\infty} v^2(t)\, dt}$

- Distance metric $d(v_1, v_2) = \sqrt{\int_{-\infty}^{\infty} [v_1(t) - v_2(t)]^2\, dt}$ the distance between two waveforms.

We want to treat waveforms like vectors and matrices

Discrete Waveform Calculations

- The inner product $\langle v_1(t) \mid v_2(t) \rangle = \sum_{n=-\infty}^{\infty} v_1(n) v_2(n)$
- The norm or length $\|v(t)\| = \sqrt{\sum_{n=-\infty}^{\infty} v^2(n)}$
- Distance metric $d(v_1, v_2) = \sqrt{\sum_{n=-\infty}^{\infty} [v_1(n) - v_2(n)]^2}$

Random Variable Calculations

- The inner product $\langle X \mid Y \rangle = E(XY)$
- The norm or length $E(X) = \sqrt{E(X^2)}$
- Distance metric $d(X,Y) = \|X - Y\| = \sqrt{E\left((X-Y)^2\right)}$
- Orthogonal requires $\langle X \mid Y \rangle = E(XY) = 0$

Expand idea of functions as vectors to random variables

Note the inclusion of E( ) everywhere.

See slide #13

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)\,p(x)\,dx \qquad \text{continuous}$$

$$E[g(x)] = \sum_{i=1}^{N} g(x_i)\,P(x_i) \qquad \text{discrete}$$

Linear estimator

$$\hat{d} = h_0 x(n) + h_1 x(n-1) + h_2 x(n-2) + h_3 x(n-3) + \ldots + h_p x(n-p)$$

where x(i) is the data, the h_i's are constants, and $\hat{d}$ is the estimate of the output d

In general $x(n) = s(n) + w(n)$ where s is the actual signal and w is white noise

Extrapolation: $d(n) = s(n+k)$ estimate a future value

Interpolation: $d(n) = s(n-k)$ estimate a previous value

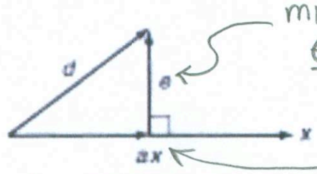
Smoothing: $d(n) = s(n)$ estimate the current value

we want to make an estimate $\hat{d}$ of d which is a linear function of a number p+1 of known, previous NOISY inputs

The noise is assumed to be white noise.

Orthogonality Principle

minimize this by making $\underline{e} \perp \underline{x}$



this is the estimate of d from $\underline{a} \cdot \underline{X}$

- The error $\underline{e}$ is the difference between the estimate $\underline{a}\underline{x}$ and the actual parameter $\underline{d}$ to be estimated
- The length of the error vector $\underline{e}$ is minimized when the error is orthogonal to the data x.

For the Kalman filter (and others) we minimize $\underline{e}$ by making $\underline{e} \perp \underline{x}$ where $\hat{d} = \underline{a} \cdot \underline{X}$, a linear estimator

Single Observation

what does an estimate mean?

- Given one observation x(n) and we want to estimate s(n)

observation $x(n) = s(n) + w(n)$ $\qquad d(n) = s(n)$ $\hat{d}$ is an estimate of s(n)

- Require the error $e(n) = d(n) - \hat{d}(n)$ to be orthogonal to the data x(n)

$$E\{e(n)x(n)\} = E\{(d(n) - \hat{d}(n))x(n)\} = 0$$

- Using the estimate $\hat{d}(n) = h_0 x(n)$ gives

if you have only one value the estimate is simple.

$$E\{(d(n) - h_0 x(n))x(n)\} = E\{d(n)x(n)\} - h_0 E\{x(n)x(n)\} = 0$$

- This can be re-arranged to give

$$E\{d(n)x(n)\} - h_0 E\{x(n)x(n)\} = R_{SX}(0) - h_0 R_{XX}(0) = 0$$

- Which says the optimum estimator is given when

$$h_0 = \frac{R_{SX}(0)}{R_{XX}(0)}$$

See slide #29

$$R_{xy} = E[xy] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,y\,p(x,y)\,dx\,dy$$

This is called the correlation of x and y
also called the cross-correlation function
and the moment η_{11}

NOTE:

$$E[d(n)x(n)] = E[s(n)x(n)] = R_{SX}(0)$$

this is always the difference between the times of the two waveforms.

typically the second minus the first

Multiple Observations

- Given two observation x(n) and x(n-1) and we want to estimate s(n)

$$x(n) = s(n) + w(n) \qquad d(n) = s(n)$$

- Require the error $e(n) = d(n) - \hat{d}(n)$ to be orthogonal to the data x(n)

$$E\{e(n)x(n)\} = E\left\{\left(d(n) - \hat{d}(n)\right)x(n)\right\} = 0$$

- Using the estimate $\hat{d}(n) = h_0 x(n) + h_1 x(n-1)$ now gives two equations

$$E\left\{\left(d(n) - h_0 x(n) - h_1 x(n-1)\right)x(n)\right\} = E\{d(n)x(n)\} - h_0 E\{x(n)x(n)\} - h_1 E\{x(n-1)x(n)\} = 0$$

$$E\left\{\left(d(n) - h_0 x(n) - h_1 x(n-1)\right)x(n-1)\right\} = E\{d(n)x(n-1)\} - h_0 E\{x(n)x(n-1)\} - h_1 E\{x(n-1)x(n-1)\} = 0$$

Two observations is much more interesting mathematically.

Multiple Observations

- Rewriting these equations in terms of autocorrelation functions

$$E\{d(n)x(n)\} - h_0 E\{x(n)x(n)\} - h_1 E\{x(n-1)x(n)\} = R_{DX}(0) - h_0 R_{XX}(0) - h_0 R_{XX}(-1) = 0$$

$$E\{d(n)x(n-1)\} - h_0 E\{x(n)x(n-1)\} - h_1 E\{x(n-1)x(n-1)\} = R_{DX}(1) - h_0 R_{XX}(1) - h_1 R_{XX}(0) = 0$$

- And putting them in matrix form gives

$$\begin{bmatrix} R_{XX}(0) & R_{XX}(-1) \\ R_{XX}(1) & R_{XX}(0) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} R_{DX}(0) \\ R_{DX}(1) \end{bmatrix} \quad \begin{bmatrix} R_{SX}(0) \\ R_{SX}(1) \end{bmatrix}$$

- Which can be solved for h_0 and h_1.

remember $d(x) = s(x)$

Single Observation Example

- Find the optimum h_0 and mean-square error in estimating s(n) if the data is x(n)=s(n)+w(n). The noise w(n) is white Gaussian noise with zero mean and unit variance. The signal, which is also zero mean and is independent of the noise, has an autocorrelation function given by $R_{SS}(n)=0.9^{|n|}$

sometimes these are known analytically or a general model is known.

- The solution requires that we compute both $R_{XX}(0)$ and $R_{SX}(0)$.
- Computing $R_{XX}(0)$

$$R_{XX}(0)=E\{x(n)x(n)\}=E\{(s(n)+w(n))(s(n)+w(n))\}$$
$$R_{XX}(0)=E\{s(n)s(n)\}+E\{s(n)w(n)\}+E\{w(n)s(n)\}+E\{w(n)w(n)\}$$
$$R_{XX}(0)=R_{SS}(0)+R_{SW}(0)+R_{WS}(0)+R_{WW}(0)$$

- Both cross-correlations are zero since the signal is independent of the noise and for white noise $R_{WW}(n)=\delta(n)$ giving.

$$R_{XX}(0)=R_{SS}(0)+R_{SW}(0)+R_{WS}(0)+R_{WW}(0)=0.9^0+0+0+\delta(0)=1+1=2$$

cross-correlations

autocorrelation of noise is $\delta(0)$

autocorrelation of signal

Single Observation Example (cont.)

- Computing $R_{SX}(0)$. $R_{SX}(0) = E\{s(n)x(n)\} = E\{s(n)(s(n)+w(n))\}$
  $R_{SX}(0) = E\{s(n)s(n)\} + E\{s(n)w(n)\}$
  $R_{SX}(0) = R_{SS}(0) + R_{SW}(0) = 0.9^0 + 0 = 1$
- We can evaluate the optimum estimator coefficient as
  $$h_0 = \frac{R_{SX}(0)}{R_{XX}(0)} = \frac{1}{2}$$
- The mean squared error is given by
  $$E(e^2) = E\big((s(n)-\hat{s}(n))(s(n)-\hat{s}(n))\big) = E\big((s(n)-h_0x(n))(s(n)-h_0x(n))\big)$$
  $$E(e^2) = E(s(n)s(n)) - h_0E(s(n)x(n)) - h_0E(x(n)s(n)) + h_0^2E(x(n)x(n))$$
  $$E(e^2) = R_{SS}(0) - h_0R_{SX}(0) - h_0R_{XS}(0) + h_0^2R_{XX}(0) = 0.9^0 - \left(\frac{1}{2}\right)1 - \left(\frac{1}{2}\right)1 + \left(\frac{1}{2}\right)^2 2 = \frac{1}{2}$$

$$E(e^2) = E\Big[\underbrace{[s(n)-\hat{s}(n)]}_{\text{simply the difference between the noise-free value and the estimate from the noisy values}}[s(n)-\hat{s}(n)]\Big]$$

Two Observation Example

- Expand the previous example to two observations, i.e., find the optimum h_0 and h_1 in estimating s(n) if the data is x(n)=s(n)+w(n). The noise w(n) is white Gaussian noise with zero mean and unit variance. The signal, which is also zero mean and is independent of the noise, has an autocorrelation function given by $R_{SS}(n)=0.9^{|n|}$
- The solution requires that we compute evaluate the matrix

$$\begin{bmatrix} R_{XX}(0) & R_{XX}(-1) \\ R_{XX}(1) & R_{XX}(0) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} R_{SX}(0) \\ R_{SX}(1) \end{bmatrix}$$

$\delta(1)=0$

- The new quantities to be evaluated are $R_{XX}(1)$, $R_{XX}(-1)$, and $R_{SX}(1)$.

$$R_{XX}(1)=R_{SS}(1)+R_{SW}(1)+R_{WS}(1)+R_{WW}(1)=0.9^1+0+0+\delta(1)=0.9$$

$$R_{XX}(-1)=R_{SS}(-1)+R_{SW}(-1)+R_{WS}(-1)+R_{WW}(-1)=0.9^{|-1|}+0+0+\delta(1)=0.9$$

$$R_{SX}(1)=E\{s(n)x(n+1)\}=E\{s(n)(s(n+1)+w(n+1))\}$$

$$R_{SX}(1)=E\{s(n)s(n+1)\}+E\{s(n)w(n+1)\}$$

$$R_{SX}(1)=R_{SS}(1)+R_{SW}(1)=0.9^1+\underset{0}{\cancel{\delta(1)}}=0.9$$

Not that different from the previous example except we <u>MUST</u> use a matrix approach.

Two Observation Example

- Evaluating the matrices gives

$$\begin{bmatrix} 2 & 0.9 \\ 0.9 & 2 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.9 \end{bmatrix}$$

- Which can be solved to give h_0=0.3730 and h_1=0.2821. The mean-square error is calculated as

$$E(e^2) = E\left(\left(s(n) - \hat{s}(n)\right)\left(s(n) - \hat{s}(n)\right)\right) = E\left(\left(s(n) - h_0 x(n) - h_1 x(n-1)\right)\left(s(n) - h_0 x(n) - h_1 x(n-1)\right)\right)$$

$$E(e^2) = E(s(n)s(n)) - h_0 E(s(n)x(n)) - h_1 E(s(n)x(n-1)) - h_0 E(x(n)s(n)) + h_0^2 E(x(n)x(n))$$
$$+ h_0 h_1 E(x(n)x(n-1)) - h_1 E(x(n-1)s(n)) + h_0 h_1 E(x(n-1)x(n)) + h_1^2 E(x(n-1)x(n-1))$$

$$E(e^2) = R_{SS}(0) - h_0 R_{SX}(0) - h_1 R_{SX}(1) - h_0 R_{XS}(0) + h_0^2 R_{XX}(0) + h_0 h_1 R_{XX}(1)$$
$$- h_1 R_{XS}(-1) + h_0 h_1 R_{XX}(-1) + h_1^2 R_{XX}(0)$$

$$E(e^2) = 0.9^0 - (0.373)(1) - (0.2821)(0.9) - (0.373)(1) + (0.373)^2 (2) + (0.373)(0.2821)(0.9)$$
$$- (0.2821)(0.9) + (0.373)(0.2821)(0.9) + (0.2821)^2 (2) = 0.373$$

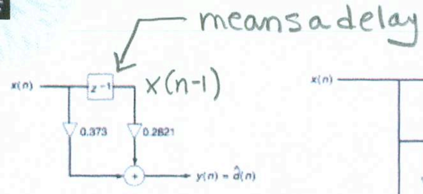
$$R_{SX}(-1) = R_{SS}(-1) + R_{SW}(-1) = 0.9^{|-1|} + \delta(1) = 0.9$$

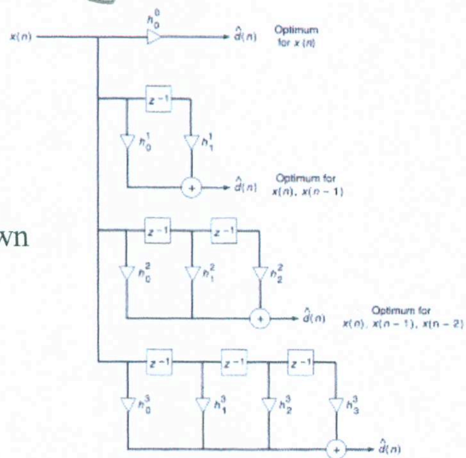
$$E(e^2) = E\left[\left(s(n) - \hat{s}(n)\right)\left(s(n) - \hat{s}(n)\right)\right]$$

the estimate of the noisy signal

The noise - free actual value

Estimator Filter Architecture





The two observation estimator example drawn as a filter. And expanded to more observations.

We can generalize this process to many more observations.

Kalman Filter

- What is the optimum estimator filter for n samples of a signal which is evolving over time?
- Kalman (1960) proposed a signal model which can be used to recursively estimate a signal evolving over time.

Optimum Filtering

- Kalman filters are often used to provide accurate estimates of position and velocity
- A Kalman filter is an efficient recursive filter which estimates the state of a dynamical system from a series of incomplete and noisy measurements
- Estimates can be
  - past time (interpolation or smoothing)
  - present time (filtering)
  - future time (prediction)

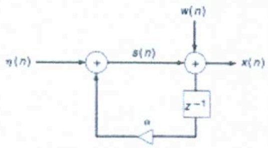
Design a Kalman Filter for a simple system

Simple system can be defined in multiple ways:

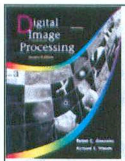
- Impulse function — $h(n) = \alpha^n u(n)$
- Transfer function — $H(z) = \dfrac{1}{1 - \alpha z^{-1}}$
- Difference equation — $s(n) = \alpha s(n-1) + \eta(n)$ ← this is the normal approach

Kalman signal model

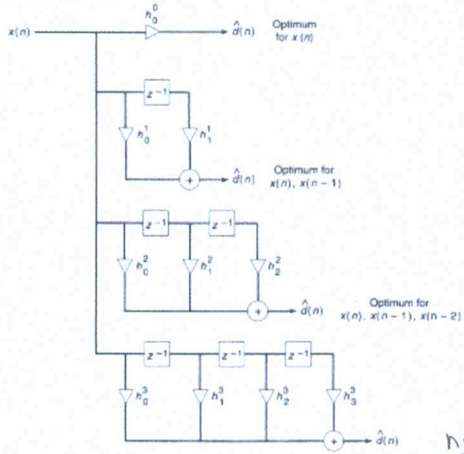
The Kalman signal model for this system is



- s(n) is the output signal
- w(n) is white noise in the observations
- x(n) is the actual observed output (s + n)
- η(n) is the white noise which drives the system



Recursive Estimation?



In our previous architecture we need to compute a new coefficient and add a delay (processing block)

Instead, can we recursively do a mean square estimate of the signal using the previous estimate and the new signal observation? If so, it would have to behave like

$$\underset{\text{new estimate}}{\hat{d}(n)} = \underbrace{A_n\hat{d}(n-1)}_{\text{previous estimate}} + \underbrace{K_n x(n)}_{\text{new signal observation}}$$

Can we use a fixed length (size) architecture to recursively update the h_i^n?

The Kalman Filter

- Assume that we can write $A_n = (1 - K_n)\alpha$
- Then the optimum estimator can be written

this is insight by Kalman

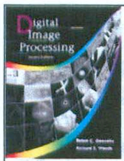
$$\hat{d}(n) = A_n\hat{d}(n-1) + K_n x(n) = (1 - K_n)\alpha\hat{d}(n-1) + K_n x(n)$$

- The normal form for this is

$$\hat{d}(n) = \alpha\hat{d}(n-1) + K_n\left[x(n) - \alpha\hat{d}(n-1)\right]$$

- Where the first term is called the forward prediction term and the second is called the residual or correction term

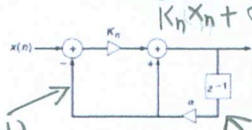
the purpose is to get a specific architecture for implementation



The Kalman Filter

- For the specified system, the Kalman filter uses a time varying gain K_n as shown below to set the error orthogonal to the signal

$K_n x_n + \alpha \hat{d}(n-1)$ this is $\hat{d}(n)$



$\alpha \hat{d}(n-1)$

$\hat{d}(n-1)$

The key idea is that everything but K_n is constant
K_n is adjusted over time to keep $\underline{e}(n)$ orthogonal
to $\underline{x}(n)$

Basic Kalman theory

- The mean square error is
$$\varepsilon(n) = E\left[e^2(n)\right] = E\left\{\left[d(n) - \hat{d}(n)\right]^2\right\} = E\left\{\left[d(n) - A_n\hat{d}(n-1) + K_n x(n)\right]^2\right\}$$
- Since we are using a linear estimator the error is also given by
$$\varepsilon(n) = E\left[e(n)d(n)\right]$$
- Solving these (without proof) requires
$$K_n = \frac{\varepsilon(n)}{E\left[\left(w(n) - m_w\right)^2\right]} = \frac{\varepsilon(n)}{\sigma_W^2}$$
this is the formula for K_n
- Where
$$\varepsilon(n) = \left[\frac{\sigma_\eta^2 + \alpha^2\varepsilon(n-1)}{\sigma_\eta^2 + \sigma_W^2 + \alpha^2\varepsilon(n-1)}\right]\sigma_W^2$$
- and
$$\varepsilon(0) = \frac{\sigma_S^2\sigma_W^2}{\sigma_S^2 + \sigma_W^2}$$

$\varepsilon(n)$ is the error over time (this changes).
σ_W^2 is the variance of the noise (and doesn't chang)

Kalman Filter (algorithm)

The signal has an exponential autocorrelation function. The parameters α and σ_η^2 must be known. The additive noise w(n) is white with known variance σ_W^2. Then

Step 1. Set n=0 and calculate the initial mean square error $\varepsilon(0)=\dfrac{\sigma_S^2\sigma_W^2}{\sigma_S^2+\sigma_W^2}$

Step 2. Calculate the Kalman gain $K_n = \varepsilon(n) / \sigma_W^2$

Step 3. Input the data x(n) and calculate the estimate.

new

$$\hat{s}(n)=\alpha\hat{s}(n-1)+K_n\left[x(n)-\alpha\hat{s}(n-1)\right]$$

For n=0 assume $\hat{s}(0)=0$ so that $\hat{s}(0)=K_n x(0)$

Step 4. Let n=n+1

Step 5. Update the error $\varepsilon(n)=\left[\dfrac{\sigma_\eta^2+\alpha^2\varepsilon(n-1)}{\sigma_\eta^2+\sigma_W^2+\alpha^2\varepsilon(n-1)}\right]\sigma_W^2$

where $\sigma_\eta^2=(1-\alpha^2)\sigma_S^2$

Step 6. Go to Step 2.

Example of Kalman Filtering

- Consider a particle moving in the plane at <u>constant velocity</u> subject to random perturbations in its trajectory. The new position (x_1, x_2) is the old position plus the velocity (Δx_1, Δx_2) plus noise w

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \Delta x_1(t) \\ \Delta x_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \Delta x_1(t) \\ \Delta x_2(t) \end{bmatrix} + \begin{bmatrix} w_{x1} \\ w_{x2} \\ w_{\Delta x1} \\ w_{\Delta x2} \end{bmatrix}$$

- We assume we only observe the position of the particle

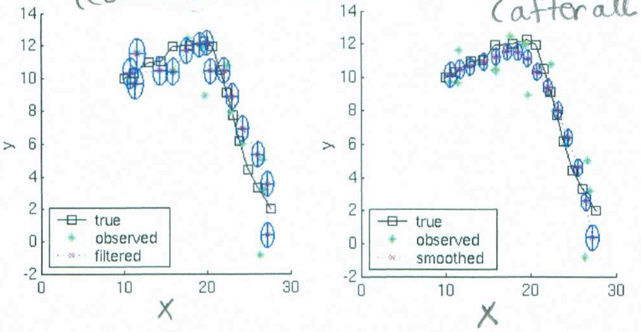
$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \Delta x_1(t) \\ \Delta x_2(t) \end{bmatrix} + \begin{bmatrix} v_{x1} \\ v_{x2} \end{bmatrix}$$

new position (x, y)   velocity (V_x, V_y)

Example of Kalman Filtering

filtered
(estimate next value)

smoothed
(after all data – curve fit)



Suppose we start out at position (10,10) moving to the right with velocity (1,0). We sampled a random trajectory of length 15. The figures show the filtered and smoothed trajectories. The mean squared error of the filtered estimate is 4.9; for the smoothed estimate it is 3.2. Not only is the smoothed estimate better, but we know that it is better, as illustrated by the smaller uncertainty ellipses initial state

$$\hat{s}_1 = A s_0$$

prediction of next state

$$\hat{P}_1 = A P_0 A^T + Q$$ update covariance matrix for predicted state

look for new location of feature and measure it

This is m_1

Compute the Kalman gain matrix using m_1

$$K_1 = \hat{P}_1 H^T (H \hat{P}_1 H^T + R)^{-1}$$ where $m_1 = H s_0$

measurement — initial state (previous measurement)

$$P_1 = \hat{P}_1 + K_1 H \hat{P}_1$$ update covariance

this is the new covariance.

Kalman gain

$$\hat{s}_2 = \hat{s}_1 + K_1(m_1 - H\hat{s}_1)$$

previous estimate

this is the new prediction for the next state

Repeat process.