University of BRISTOL

Computer Science Year 2

Algorithms & Data

Estimation, Regression, Classification

Prof Alin Achim

Algorithms & Data Mathematical Preliminaries

Matrix Methods & Numerical Linear Algebra

- Vectors and Matrices:
  vector representation, linear equations, special matrix forms
- Dot products, Vector norms and Projections
- Solutions to linear equations
  square matrices, matrix inverse, Gaussian elimination
- Quadratic forms

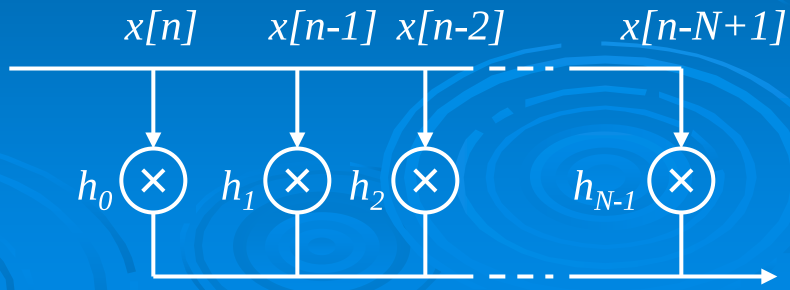
Vectors

- A vector is an array of real or complex-valued numbers. Vectors are usually denoted by lower case letters
- e.g..

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix}$$

- A vector of N elements is said to be a N-dimensional vector
- A vector is useful for representing the values of the discrete (time or space) sampled elements of a signal. e.g. FIR filter input

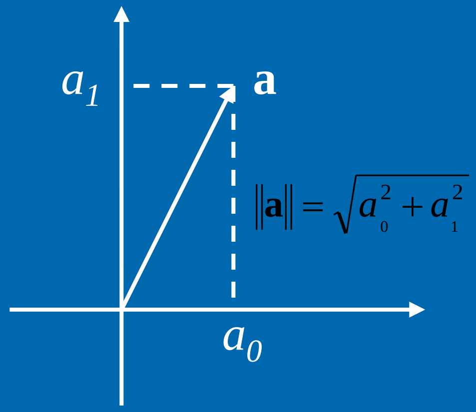
$$\mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-N+1] \end{bmatrix}$$



Length of a vector (norm)

- The length of a vector is given by: (known as the Euclidean L_2 norm)

$$\|\mathbf{a}\| = \left\{ \sum_{i=0}^{N-1} a_i^2 \right\}^{1/2}$$



- For example, the case of a 2-dimensional vector. Can be generalised to N-dimensions

- A vector can be normalised to be unit norm by dividing by its norm. e.g. assuming $||\mathbf{a}|| \neq 0$, then

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{\|\mathbf{a}\|}$$

where the norm 1 vector lies in the same direction as **a**

Dot product (inner product)

- The dot product between two real vectors is defined:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b} = \sum_{i=0}^{N-1} a_i b_i$$

- dot product between complex vectors is defined:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^H \mathbf{b} = \sum_{i=0}^{N-1} a_i^* b_i$$

- The result of a dot product is always a scalar value (real or complex number).

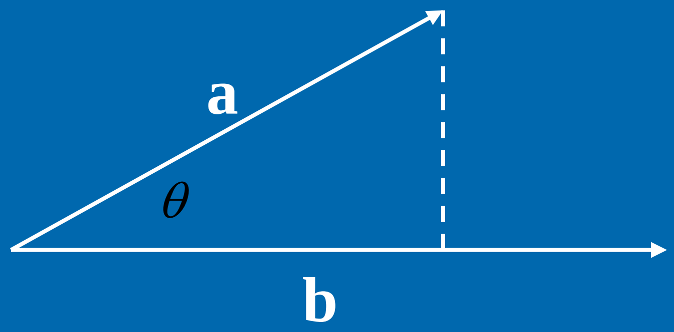
- Note that for real vectors: $\langle \mathbf{b}, \mathbf{a} \rangle = \mathbf{b}^T \mathbf{a} = \sum_{i=0}^{N-1} b_i a_i = \mathbf{a}^T \mathbf{b}$

- Note that for complex vectors:

$$\langle \mathbf{b}, \mathbf{a} \rangle = \mathbf{b}^H \mathbf{a} = \sum_{i=0}^{N-1} b_i^* a_i = \left[\sum_{i=0}^{N-1} a_i^* b_i \right]^* = \left[\mathbf{a}^H \mathbf{b} \right]^*$$

Dot product (inner product)

- The dot product between two vectors also defines the geometrical relationship between the two vectors through the relationship: $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos\theta$
-



If $\|\mathbf{b}\| = 1$ then $\mathbf{a}^T \mathbf{b} = \|\mathbf{a}\| \cos\theta$ which is the projection of **a** onto the direction of **b**.

If $\mathbf{a}^T \mathbf{b} = \|\mathbf{a}\| \cos\theta = 0$

then the projection of **a** onto **b** is zero. This is only possible if the two vectors a and b are *orthogonal*.

Dot product (inner product)

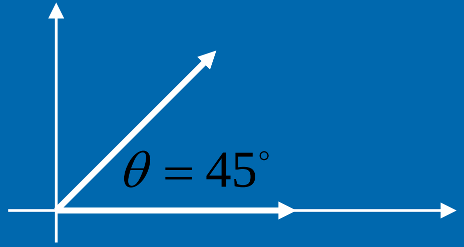
example (1), if:

$$\mathbf{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{b} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

are two unit length vectors, then the inner product:

$$\mathbf{a}^T\mathbf{b} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} = \cos\theta$$

$$\Rightarrow \theta = 45^\circ$$



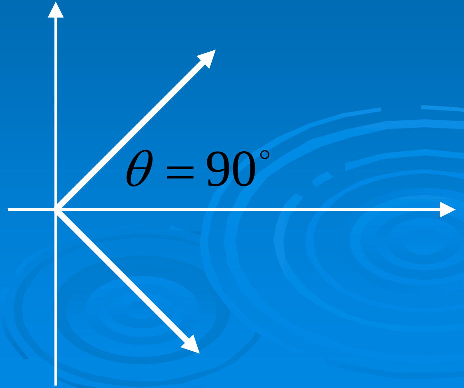
example (2), if:

$$\mathbf{a} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \mathbf{b} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

are two unit length vectors, then the inner product:

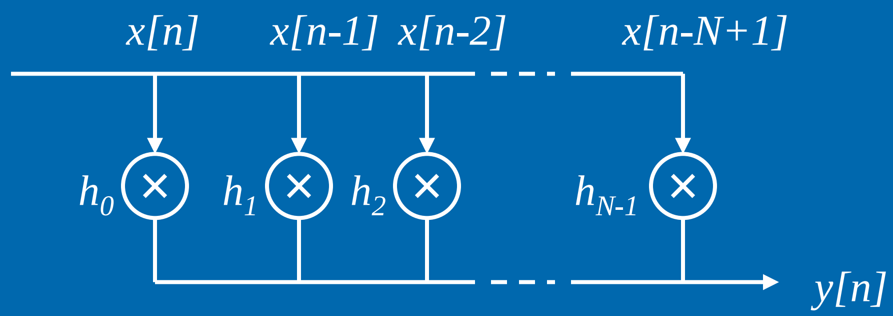
$$\mathbf{a}^T\mathbf{b} = \frac{1}{2} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 = \cos\theta$$

$$\Rightarrow \theta = 90^\circ$$



Dot product and its uses

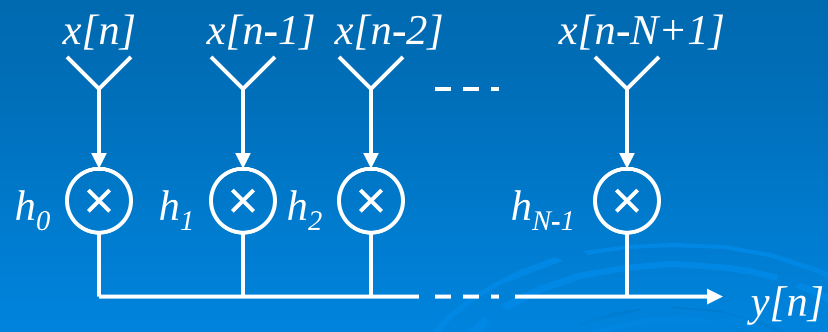
(1) The dot product can concisely describe the output of a time-invariant finite impulse response filter:



real case:

$$y[n] = \sum_{i=0}^{N-1} h_i x[n-i] = \mathbf{h}^T \mathbf{x}[n]$$

(2) The dot product can concisely describe the output of a time-invariant multi-element antenna:



real case:

$$y[n] = \sum_{i=0}^{N-1} h_i x[n-i] = \mathbf{h}^T \mathbf{x}[n]$$

The outer product

- The outer product of a real N dimensional vector is defined:

$$\mathbf{a}\mathbf{a}^T = \begin{bmatrix} a_1 \\ a_2 \\ \\ a_N \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \dots & a_N \end{bmatrix} = \begin{bmatrix} a_1 a_1 & a_1 a_2 & \dots & a_1 a_N \\ a_2 a_1 & a_2 a_2 & \dots & a_2 a_N \\ & & \dots & \\ a_N a_1 & a_N a_2 & \dots & a_N a_N \end{bmatrix} = \mathbf{A}$$

- The result of a real outer product is an N by N *square symmetric* matrix.

$$\Rightarrow \mathbf{A}^T = \mathbf{A}$$

- The outer product of a complex N dimensional vector is defined:

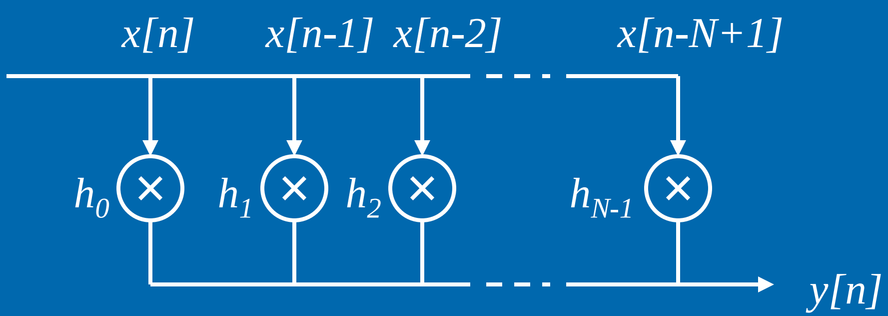
$$\mathbf{a}\mathbf{a}^H = \begin{bmatrix} a_1 \\ a_2 \\ \\ a_N \end{bmatrix} \begin{bmatrix} a_1^* & a_2^* & \dots & a_N^* \end{bmatrix} = \begin{bmatrix} a_1 a_1^* & a_1 a_2^* & \dots & a_1 a_N^* \\ a_2 a_1^* & a_2 a_2^* & \dots & a_2 a_N^* \\ & & \dots & \\ a_N a_1^* & a_N a_2^* & \dots & a_N a_N^* \end{bmatrix}$$

- The result of a complex outer product is an *Hermitian* matrix.

$$\Rightarrow \mathbf{A}^H = \mathbf{A}$$

The outer product and its uses

- The outer product can concisely describe the output power of a time-invariant finite impulse response filter:



real case:

$$y[n] = \sum_{i=0}^{N-1} h_i x[n-i] = \mathbf{h}^T \mathbf{x}[n]$$

The instantaneous output power is:

$$|y[n]|^2 = y[n]y[n]^T = (\mathbf{h}^T \mathbf{x}[n])(\mathbf{h}^T \mathbf{x}[n])^T = \mathbf{h}^T \mathbf{x}[n]\mathbf{x}[n]^T \mathbf{h} = \mathbf{h}^T \mathbf{X}[n]\mathbf{h}$$

where the matrix **X**[n] is the outer product:

$$\mathbf{X}[n] \hat{=} \mathbf{x}[n]\mathbf{x}[n]^T$$

Other Special Matrix Forms

- Toeplitz

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 1 & 3 & 5 \\ 4 & 2 & 1 & 3 \\ 6 & 4 & 2 & 1 \end{bmatrix}$$

- Hankel

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 5 & 7 & 4 \\ 5 & 7 & 4 & 2 \\ 7 & 4 & 2 & 1 \end{bmatrix}$$

Linear set of equations :

Many of the problems encountered in optimum signal processing e.g. Wiener filtering, spectral estimation and DoA estimation, require the analysis or solution of a set of linear equations of the form:

$$\mathbf{Ax} = \mathbf{b}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & \mathrm{M} \\ \cdots & \cdots & \cdots & \cdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ b_N \end{bmatrix}$$

This set of equations can be treated in a number of ways. For the case of N=M i.e. number of equations equals number of unknowns, then one can use Gaussian elimination or matrix inversion.

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Linear equations: Singular case

- Consider the pair of equations $$\mathbf{x}_1 + \mathbf{x}_2 = 1$$ $$\mathbf{x}_1 + \mathbf{x}_2 = 2$$
- In matrix form, we have

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

- Obviously, A is singular (det(A)=0) and no solution exists
- However, for $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- The vector below will satisfy the equations for any constant a

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Linear equations: Underdetermined case

- For a rectangular matrix (n<m) there are fewer equations then unknowns, consequently – many vectors that satisfy the equations.
- A workaround: the <u>minimum norm solution</u>, i.e. the solution to finding the vector x satisfying

$$\min \|\mathbf{x}\| \quad \textbf{such that } \mathbf{Ax} = \mathbf{b}$$

- The solution is

$$\mathbf{x}_0 = \mathbf{A}^{\mathbf{H}} \left(\mathbf{A}\mathbf{A}^{\mathbf{H}} \right)^{-1} \mathbf{b}$$

- $\mathbf{A}^{\mathbf{H}} \left(\mathbf{A}\mathbf{A}^{\mathbf{H}} \right)^{-1}$ is called the pseudo-inverse of A for the underdetermined problem

Quadratic forms

- An important matrix construction in statistical signal/data filtering is the so-called quadratic form.

- The quadratic form of a real square matrix **A** is given by:

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i a_{ij} x_j$$

- and for the complex case:

$$Q(\mathbf{x}) = \mathbf{x}^H \mathbf{A}\mathbf{x} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i^* a_{ij} x_j$$

- The matrix is said to be positive definite if:

$$Q(\mathbf{x}) > 0$$ for all non-zero **x** vectors

- and the matrix is said to be positive semi-definite if:

$$Q(\mathbf{x}) \geq 0$$ for all non-zero **x** vectors

Quadratic forms

example (1), if:

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + 3x_2^2$$

and therefore matrix *A* is positive definite

example (2), if:

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2$$

- and therefore matrix *A* is positive semi-definite (as can be zero e.g. when x = [0 x_2])

Algorithms & Data
Mathematical Preliminaries

Stochastic Processes
and Signal Analysis

- **Stochastic Processes :**

  definition of stochastic, stationarity (SSS and WSS) and ergodicity

- **Statistics for Random Signal Processing**

  mean, correlation and covariance

- **Correlation Matrix**

  correlation matrix for WSS processes, properties of correlation matrix, time averaged values

Stochastic Processes:

- A stochastic process is a ***random*** process.

- A discrete random signal is defined as a sequence of indexed random variables assuming values:

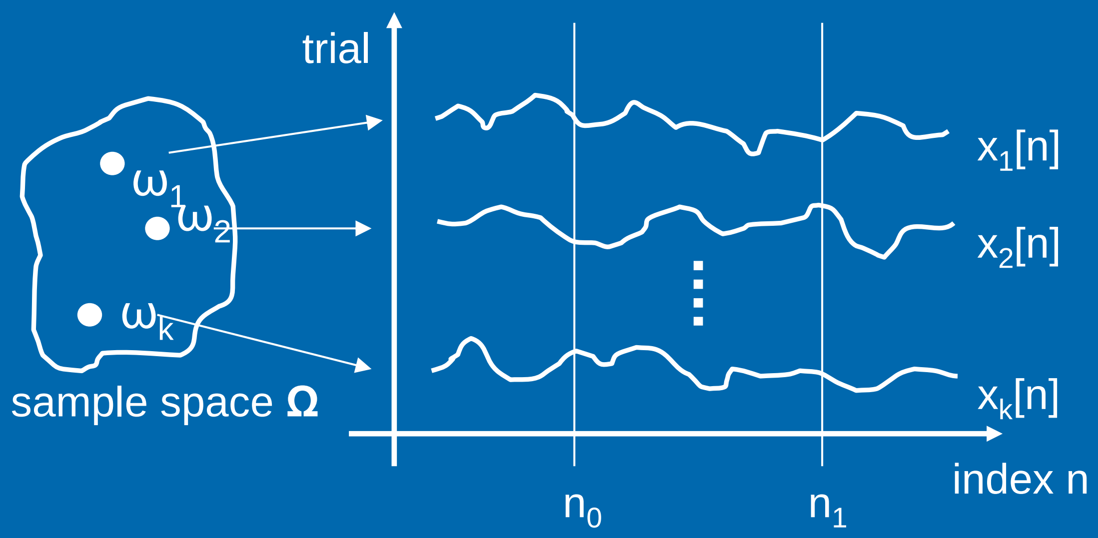
$$\mathbf{x}[0], \mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[i], \ldots$$

where it is assumed that:

- samples are evenly spaced in time
- samples are continuous in amplitude (infinite precision representation)
- samples are taken at a rate greater than twice the highest frequency component present (i.e. Nyquist satisfied)
- sample period is normalised to unity

Stochastic Processes:

Consider an ensemble of sample functions:



- A ***stochastic process*** is an ensemble of time (or spatial) variables together with a probability rule which assigns a probability to any event observed

- The figure shows a set of sample functions, or ***realisations***, $x_k[n]$, corresponding to a sample point ω_k in the sample space $\boldsymbol{\Omega}$.

Observation of sample waveforms at some point n_0. Each sample has a value $x_k[n_0]$ and a probability $P(\omega_k)$. The set of numbers $\{x_k[n_0]\}$ k = 1..K form a ***random variable***. Observation at n_1 results in a second random variable $\{x_k[n_1]\}$ k = 1..K.

Stationarity of a random Process:

•The joint probability density function (JPDF) allows a complete description of a random process.

•For a random process to have ***Strict Sense Stationarity*** (SSS) then the joint probability density function (JPDF) must be invariant to shifts in time (or space). In practice SSS is difficult to apply.

This leads to the weaker definition of ***Wide Sense Stationarity*** (WSS) where a signals first order moment (mean) and second order moment (variance) are invariant to shifts in time (or space).

The expectations of a random process are ensemble averages *across* the process.

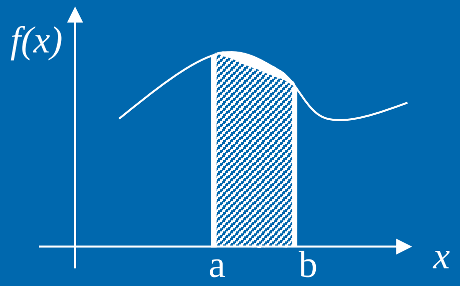
For example, the mean at time n_i is the expectation of the variable $\{x[n_i]\}$ which describes all possible values of the variable at time n_i.

If the time averages taken *along* the process converge as $n \rightarrow \infty$ with probability 1 to the ensemble average then the process is said to be ***ergodic***.

Statistics for random signal processing:

The probability density function (PDF) *f(x)* describes the distribution of probability.
The probability that a random value x[n] will lie in the interval [a, b] is given by:

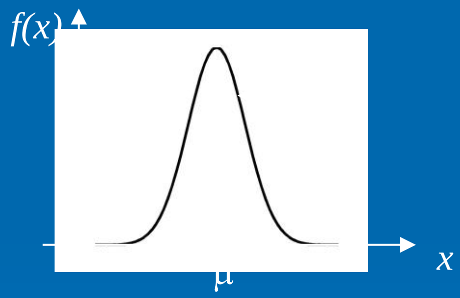
$$\Pr[a \le x \le b] = \int_a^b f(x)\,dx$$



which is the area under the pdf between a and b.

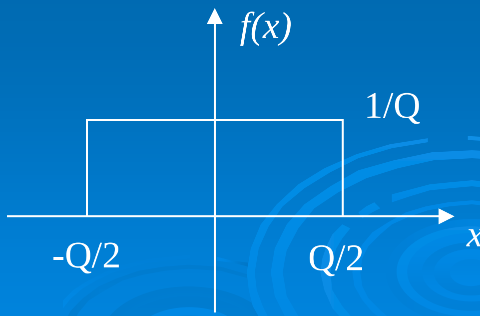
Gaussian PDF:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$



Uniform PDF

$$f(x) = \begin{cases} 1/Q \;\; for\; -Q/2 \le x \le Q/2 \\ \qquad\qquad 0 \end{cases}$$



Statistics for random signal processing:

- Important for Statistical Signal Processing are:

Mean, Correlation & Covariance

Since they are

- well suited to characterising linear operations on random processes
- are amenable to experimental evaluation
- offer tractable mathematical analysis

Mean

The mean of a stochastic process is defined as:

$$\mu_n = E\{x[n]\} = \int_{-\infty}^{\infty} x[n] f(x[n]) dx[n]$$

where *f(x[n])* is the probability density function (PDF) and E{•} is the expectation operator.

In general, distributions may vary with time so that the mean of the process at n_0 is not the same as at n_1, or:

$$\mu_{n_0} \neq \mu_{n_1}$$

(for example, speckle noise in a synthetic aperture radar image)

But, for a ***stationary*** process, the PDF is the same for all n, and therefore:

$$\mu = E\{x[n]\} = \text{constant}$$

The expectation can be interpreted as an average value obtained by repeating the experiment

$$\mu = E\{x[n]\} = \lim_{N \to \infty} \left[\frac{1}{N} \sum_{i=1}^{N} x_i[n] \right]$$

N.B. Average across the process.

Correlation

The correlation of a stochastic process is the second order moment. It gives a measure of the amount of dependence of {x[n]} at n_0 and n_1. It is defined as:

$$R_x(n_0,n_1) = E\{x[n_0]x^*[n_1]\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x[n_0]x^*[n_1]f(x[n_0]x[n_1])dx[n_0]dx[n_1]$$

where $f(x[n_0],x[n_1])$ is the *joint* probability density function (PDF) and E{•} is the expectation operator.

The expectation can again be interpreted as an average value obtained by repeating the experiment

$$R_x(n_0,n_1) = E\{x[n_0]x^*[n_1]\} = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} x_i[n_0]x_i^*[n_1]\right]$$

When n = n_0 = n_1 then $R_x(n, n)$ is the average power of x[n].

The correlation is also referred to as the autocorrelation as it measures the correlation of the signal with itself.

Covariance

The covariance of a stochastic process is the second order moment, but with the mean removed. It is defined as:

$$C_x(n_0, n_1) = E\left\{\left(x[n_0] - \mu_{n_0}\right)\left(x[n_1] - \mu_{n_1}\right)^*\right\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x[n_0] - \mu_{n_0}\right)\left(x[n_1] - \mu_{n_1}\right)^* f\left(x[n_0]x[n_1]\right)dx[n_0]dx[n_1]$$

where $f(x[n_0],x[n_1])$ is the *joint* probability density function (PDF) and E{•} is the expectation operator.

The above expression can be re-written as:

$$C_x(n_0, n_1) = E\left\{\left(x[n_0] - \mu_{n_0}\right)\left(x[n_1] - \mu_{n_1}\right)^*\right\}$$
$$= E\left\{x[n_0]x[n_1]^*\right\} + \mu_{n_0}\mu_{n1} - E\left\{x[n_1]^*\right\}\mu_{n_0} - E\left\{x[n_0]\right\}\mu_{n_1}$$
$$= \mathbf{E}\left\{\mathbf{x}[\mathbf{n}_0]\mathbf{x}[\mathbf{n}_1]^*\right\} - \mu_{\mathbf{n}_0}\mu_{\mathbf{n}_1}$$

When $n = n_0 = n_1$ then $C_x(n, n) = \sigma^2$ is called the variance of x[n].
If process is zero mean then covariance = correlation.

For Wide Sense Stationary (WSS) signals the following simplifications/assumptions are made:

- Constant mean

$$E\{x[n]\} = \mu_x \quad \textit{for all } n$$

- Correlation only dependant upon the lag

$$E\{x[n]x^*[n-k]\} = R_x(n, n-k) = r_x(k)$$

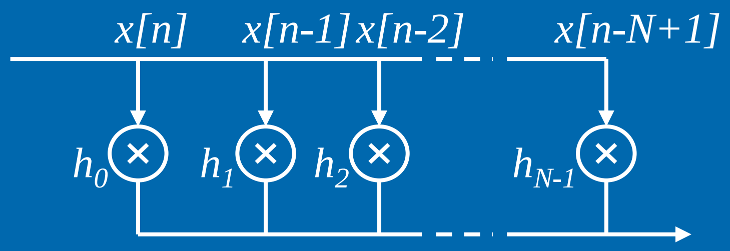
- Signals of finite energy

$$E\{|x[n]|^2\} \leq \infty$$

Correlation matrix for WSS processes:

The correlation matrix is a matrix whose elements are the autocorrelation values at various lags (in time or space).

Vector **x**[n] is the observation of a time series:



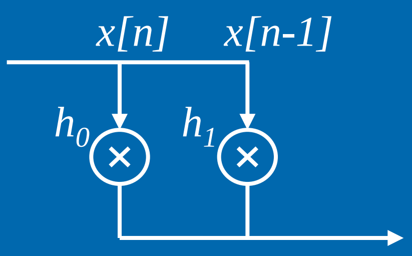
$$\mathbf{x}^T[n] = \begin{bmatrix} x[n] & x[n-1] & \cdots & x[n-N+1] \end{bmatrix}$$

Correlation matrix is given as the expectation of the outer product:

$$\mathbf{R}_x = E\{\mathbf{x}[n]\mathbf{x}^H[n]\} = \begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(N-1) \\ r_x(-1) & r_x(0) & & r_x(N-2) \\ \vdots & & \ddots & \vdots \\ r_x(-N+1) & & & r_x(0) \end{bmatrix}$$

For example, for a two tap FIR filter:

Vector **x**[n] is the observation of a time series:



$$\mathbf{x}^T[n] = \begin{bmatrix} x[n] & x[n-1] \end{bmatrix}$$

Correlation matrix is given as the expectation of the outer product:

$$\mathbf{R}_x = E\{\mathbf{x}[n]\mathbf{x}^H[n]\} = E\left\{\begin{bmatrix} x[n] \\ x[n-1] \end{bmatrix}\begin{bmatrix} x^*[n] & x^*[n-1] \end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} x[n]x^*[n] & x[n]x^*[n-1] \\ x[n-1]x^*[n] & x[n-1]x^*[n-1] \end{bmatrix}\right\}$$

$$= \begin{bmatrix} r_x(0) & r_x(1) \\ r_x(-1) & r_x(0) \end{bmatrix}$$

Correlation matrix for a WSS process has the properties:

- The correlation matrix is *Hermitian*

$$\mathbf{R}_x = \mathbf{R}_x^H$$

- The correlation matrix is *Toeplitz*

- The correlation matrix is always positive semi-definite and generally positive definite.

- Thus, the correlation matrix is almost always non-singular (an inverse exists)