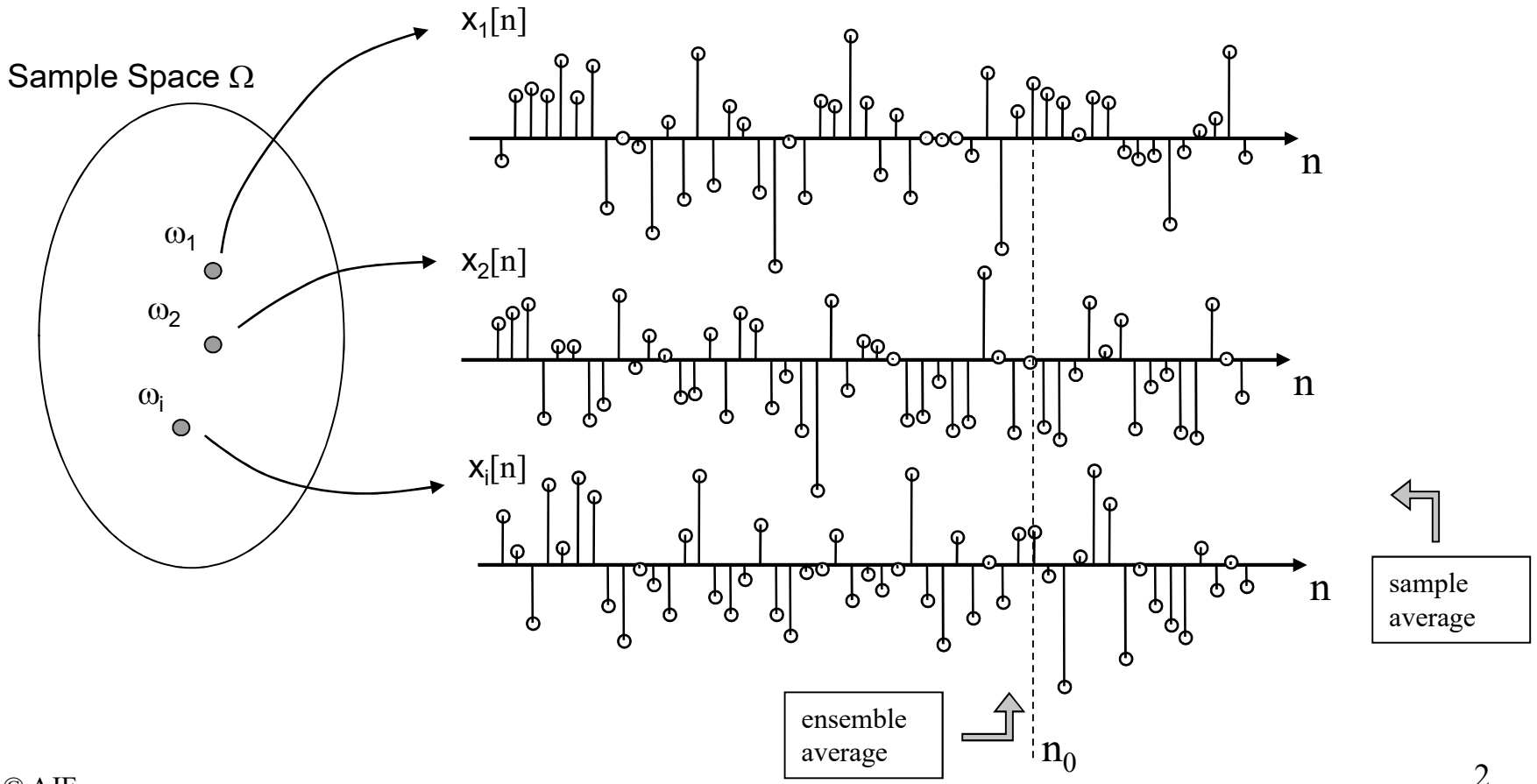
Overview

- *discrete-time random processes*
  - ensemble averages
  - autocorrelation of the sum of uncorrelated random processes
    - the harmonic process, Gaussian processes
  - stationary processes
    - strict sense and wide sense stationarity
  - autocorrelation and autocovariance matrices
    - eigenvalues and eigenvectors of autocorrelation matrices
  - linear transformation of random vectors
  - innovations representations of random vectors
    - eigenanalysis and the Karhunen-Loève transform
    - the KLT of periodic random sequences
    - LDU decomposition and UDL decomposition
    - generation of real-valued random vectors with given second order moments
  - ergodicity
    - ergodicity in the mean and ergodicity in the autocorrelation
  - white noise
  - the power spectrum
    - Properties
  - *sum of independent random variables*

review of discrete-time random processes

- *discrete-time random processes*
  - is a mapping from the sample space Ω (of experimental outcomes) into a set of discrete-time signals x[n]; thus a discrete-time random process is a collection (i.e. ensemble) of discrete-time signals



review of discrete-time random processes

– each sample of the discrete-time signal x[n] is a random variable that is characterized by a probability distribution function

$$F_{x[n]}(\alpha) = \Pr\{x[n] \leq \alpha\}$$

and a probability density function

$$f_{x[n]}(\alpha) = \frac{d}{d\alpha} F_{x[n]}(\alpha)$$

In general, a complete statistical characterization of a random process, in addition to the first-order density functions, requires that joint probability distribution or density functions be specified for collections of random variables (describing how they relate to each other) -which may be difficult or impossible to get...

- *ensemble averages*
  – first-order statistics that depend on *n* :

mean of the process $m_x[n] = E\{x[n]\}$

variance of the process $\sigma_x^2[n] = E\left\{\left(x[n] - m_x[n]\right)^2\right\}$

Note 1: the mean represents the average value of the process as a function of n

Note 2: the variance represents the average squared deviation of the process away from the mean

review of discrete-time random processes

- *ensemble averages*
  - relating the random variables *x[k]* and *x[ℓ]* (of the same process):

autocorrelation : $r_x[k,\ell] = E\{x[k]x^*[\ell]\}$

autocovariance : $c_x[k,\ell] = E\{(x[k]-m_x[k])(x[\ell]-m_x[\ell])^*\}$

both are related as : $c_x[k,\ell] = r_x[k,\ell] - m_x[k]m_x^*[\ell]$

Note: for zero-mean random processes, the autocovariance and the autocorrelation are equal

$\therefore$ the autocovariance and the autocorrelation functions provide information about the degree of linear dependence between two random variables of the same process (i.e. they characterize the 'memory' of the process)

review of discrete-time random processes

- *ensemble averages*
  - if the random variables *x[k]* and *y[ℓ]* belong to different processes:

cross-correlation : $r_{xy}[k,\ell] = E\{x[k]\,y^*[\ell]\}$

cross-covariance : $c_{xy}[k,\ell] = E\{(x[k]-m_x[k])(y[\ell]-m_y[\ell])^*\}$

and also : $c_{xy}[k,\ell] = r_{xy}[k,\ell] - m_x[k]m_y^*[\ell]$

  - two random processes *x[n]* and *y[n]* are uncorrelated if, for all *k* and ℓ :

$$c_{xy}[k,\ell]=0 \quad \Leftrightarrow \quad r_{xy}[k,\ell] = E\{x[k]\}E\{y[\ell]^*\} = m_x[k]m_y^*[\ell]$$

  - two random processes *x[n]* and *y[n]* are orthogonal if, for all *k* and ℓ :

$$r_{xy}[k,\ell]=0$$

Note: zero-mean uncorrelated processes are always orthogonal, however, orthogonal processes are not necessarily uncorrelated

review of discrete-time random processes

- *autocorrelation of the sum of uncorrelated random processes*
  - data acquisition is usually contaminated by noise which typically is considered as additive, if *x[n]* is the signal and *w[n]* is the noise, then the signal is recorded as $y[n] = x[n] + w[n]$

    and thus : $r_y[k,\ell] = E\{(x[k] + w[k])(x[\ell] + w[\ell])^*\}$

    $$r_y[k,\ell] = E\{x[k]x^*[\ell] + w[k]w^*[\ell] + x[k]w^*[\ell] + w[k]x^*[\ell]\}$$

    $$r_y[k,\ell] = E\{x[k]x^*[\ell]\} + E\{w[k]w^*[\ell]\} + E\{x[k]w^*[\ell]\} + E\{w[k]x^*[\ell]\}$$

    usually, noise is assumed to be zero-mean and uncorrelated with the signal :

    $$r_y[k,\ell] = r_x[k,\ell] + r_w[k,\ell] + E\{x[k]\} \times 0 + 0 \times E\{x^*[\ell]\} = r_x[k,\ell] + r_w[k,\ell]$$

    which reveals that the autocorrelation of the sum of uncorrelated (and zero-mean) random processes is equal to sum of the autocorrelation of the individual processes

review of discrete-time random processes

- *example: the harmonic process*
  - the harmonic random process is important radar and sonar signal processing, a simple example of a real-valued harmonic is the random phase sinusoid (*A* and ω_0 are constants and ϕ is a random variable uniformly distributed in the range [-π, π[ )

$$x[n] = A\sin(n\omega_0 + \phi)$$

the PDF for ϕ is $f_\phi(\alpha) = \begin{cases} 1/(2\pi) & ; \quad -\pi \le \alpha < \pi \\ 0 & ; \quad \text{otherwise} \end{cases}$

and thus : $m_x[n] = \int_{-\infty}^{\infty} A\sin(n\omega_0 + \alpha) f_\phi(\alpha) d\alpha = \frac{A}{2\pi}\int_{-\pi}^{\pi} \sin(n\omega_0 + \alpha) d\alpha = 0$

concerning the autocorrelation:

$$r_x[k,\ell] = E\{x[k]x^*[\ell]\} = \int_{-\infty}^{\infty} A^2 \sin(k\omega_0 + \alpha)\sin(\ell\omega_0 + \alpha) f_\phi(\alpha) d\alpha$$

$$r_x[k,\ell] = \frac{A^2}{4\pi}\int_{-\pi}^{\pi} [\cos((k-\ell)\omega_0) - \cos((k+\ell)\omega_0 + 2\alpha)] d\alpha = \frac{A^2}{2}\cos((k-\ell)\omega_0)$$

i.e. the autocorrelation is only a function of the difference between *k* and ℓ

review of discrete-time random processes

- *example: the harmonic process*
  - if a more general case is considered (A_m and ω_m are constants and ϕ_m are uncorrelated random variables uniformly distributed in the range [-π, π[, *v[n]* represents zero-mean additive noise that is uncorrelated with the sinusoids)

$$x[n] = \sum_{m=1}^{M} A_m \sin\left(n\omega_m + \phi_m\right) + v[n]$$

since the random variables ϕ_m are uncorrelated, a given sinusoidal process is uncorrelated with the others, and since the noise is also uncorrelated with the sinusoids, it can be shown that

$$r_x[k,\ell] = \sum_{m=1}^{M} \frac{A_m^2}{2} \cos\left((k-\ell)\omega_m\right) + r_v[k,\ell]$$

where $r_v[k,\ell]$ is the autocorrelation of the additive noise

review of discrete-time random processes

- *Gaussian processes*
  - we extend here the definition (in slide 24) of two jointly Gaussian random variables to a collection of *n* jointly Gaussian random variables

given a vector of n real-valued random variables $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$

whose mean values are $\mathbf{m}_x = [m_1, m_2, \ldots, m_n]^T$

the x_i random variables are jointly Gaussian if the joint probability density function of the *n* random variables is

$$f_x(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\mathbf{C}_x|}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}-\mathbf{m}_x\right)^T \mathbf{C}^{-1}\left(\mathbf{x}-\mathbf{m}_x\right)\right\}$$

where $\mathbf{C}_x$ represents the covariance matrix (which is symmetric positive definite) and its element c_{ij} represents the covariance between x_i and x_j :

$$c_{ij} = E\left\{\left(x_i - m_i\right)\left(x_j - m_j\right)\right\}$$

$|\mathbf{C}_x|$ represents the determinant of the covariance matrix.

A discrete-time random process *x[n]* is Gaussian if every finite collection of samples of *x[n]* are jointly Gaussian

review of discrete-time random processes

- *Gaussian processes*
  - it should be noted that if the random variables are uncorrelated, then

$$\mathbf{C}_x = \mathrm{diag}\left(c_{11}, c_{22}, \ldots, c_{nn}\right)$$

and the Gaussian joint probability density function simplifies to

$$f_x(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\prod_{k=1}^{n} c_{kk}}} \exp\left\{-\frac{1}{2}\sum_{\ell=1}^{n} \frac{\left(x_\ell - m_\ell\right)^2}{c_{\ell\ell}}\right\}$$

or

$$f_x(\mathbf{x}) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{c_{kk}}} \exp\left\{-\frac{1}{2}\frac{\left(x_k - m_k\right)^2}{c_{kk}}\right\}$$

as one would expect, which also means the random variables are statistically independent.

Many of the processes that are found in applications are Gaussian, or approximately Gaussian as a result of the Central Limit theorem.

review of discrete-time random processes

- *stationary processes*
  - if the first-order density function of a random process *x[n]* is independent of time, i.e. for all *k*

$$f_{x[n]}(\alpha) = f_{x[n+k]}(\alpha)$$

the process is said to be first-order stationary, as a result, the first order statistics are independent of time:

$$\sigma_x^2[n] = \sigma_x^2$$

$$m_x[n] = m_x$$

a process is second-order stationary if the second-order joint density function depends only on the time difference n_2-n_1 and not on the individual times n_1 and n_2, i.e.

$$f_{x[n1],x[n2]}(\alpha 1, \alpha 2) = f_{x[n1+k],x[n2+k]}(\alpha 1, \alpha 2)$$

Note: a second-order stationary process is also first-order stationary

second-order stationary processes have second-order statistics that are invariant to a time shift of the process, e.g.

$$r_x[k,\ell] = \int_{-\infty}^{+\infty} \alpha\beta\, f_{x[k],x[\ell]}(\alpha,\beta)\, d\alpha\, d\beta = \int_{-\infty}^{+\infty} \alpha\beta\, f_{x[k+n],x[\ell+n]}(\alpha,\beta)\, d\alpha\, d\beta = r_x[k+n,\ell+n]$$

review of discrete-time random processes

this means that the correlation between the random variables depends only on the time difference (i.e. the lag) separating them:

$$r_x[k,\ell] = r_x[k-\ell,0] \equiv r_x[k-\ell]$$

- in general, a process is said to be stationary of order *L* if the processes *x[n]* and *x[n+k]* have the same *L*th-order joint density function
- a process that is stationary for all orders *L*>0 is said to be stationary in the *strict sense*
- a weaker definition of stationarity involves only the mean and autocorrelation (i.e. it presumes stationarity only up to order two) of a process and is called *wide-sense stationarity*

- *wide sense stationary processes*
  - A random process *x[n]* is said to be *wide-sense* stationary (WSS) if:
    - the mean of the process is constant and finite: $m_x[n] = m_x$
    - the autocorrelation $r_x[k,\ell]$ depends only on the difference $k-\ell$
    - the variance of the process is constant and finite, i.e. $|c_x[\ell]| \le c_x[0] = \sigma_x^2 < \infty$

Note: in the case of a Gaussian process, wide-sense stationarity is equivalent to strict sense stationary since a Gaussian random process is completely defined in terms of the mean and covariance

review of discrete-time random processes

- two processes *x[n]* and *y[n]* are jointly wide-sense stationary if they are wide-sense stationary and if the cross-correlation $r_{xy}[k,\ell]$ depends only on the time difference $k-\ell$

$$r_{xy}[k,\ell] = r_{xy}[k-\ell,0] \equiv r_{xy}[k-\ell] = E\left\{x[k]y^*[\ell]\right\}$$

- properties of the autocorrelation of a *wide-sense* stationary process:
  - Symmetry: the autocorrelation sequence of a WSS process is a conjugate symmetric function of *k* (or a even sequence if the process is real), i.e.

$$r_x[k] = r_x^*[-k]$$

  - Mean square value: the autocorrelation sequence of a WSS process at lag *k=0* is equal to the mean-square value of the process, i.e.

$$r_x[0] = E\left\{x[n]x^*[n]\right\}$$

  - Maximum value: the magnitude of the autocorrelation sequence of a WSS process at lag *k* is upper bounded by its value at lag *k=0*, i.e.

$$\left|r_x[k]\right| \le r_x[0]$$

  - Periodicity: the autocorrelation sequence of a WSS process is periodic with period k_0 if

$$r_x[k_0] = r_x[0]$$

  - in this case *x[n]* is said to be mean-square-periodic:

$$E\left\{\left|x[n]-x[n-k_0]\right|^2\right\} = 0$$

review of discrete-time random processes

– a brief summary extending from the previous slides

considering WSS random processes:

$$r_x[\ell] = E\{x[n]x^*[n-\ell]\} \qquad r_{xy}[\ell] = E\{x[n]y^*[n-\ell]\}$$

$$c_x[\ell] = r_x[\ell] - m_x^2 \qquad c_{xy}[\ell] = r_{xy}[\ell] - m_x m_y^*$$

it can be shown that:

$$r_x[\ell] = r_x^*[-\ell] \qquad c_x[\ell] = c_x^*[-\ell]$$

$$r_{xy}[\ell] = r_{yx}^*[-\ell] \qquad c_{xy}[\ell] = c_{yx}^*[-\ell]$$

$$|r_x[\ell]| \le r_x[0] \qquad |c_x[\ell]| \le c_x[0] = \sigma_x^2$$

$$|r_{xy}[\ell]| \le \sqrt{r_x[0] r_y[0]} \qquad |c_{xy}[\ell]| \le \sqrt{c_x[0] c_y[0]}$$

Note: any strict-sense stationary signal is wide-sense stationary, the inverse is not always true, except if the signal is Gaussian

if the processes are uncorrelated:

$$r_x[\ell] = \sigma_x^2 \delta[\ell] + m_x^2 \qquad r_{xy}[\ell] = m_x m_y^* \qquad c_{xy}[\ell] = 0$$

if the processes are orthogonal:

$$r_x[\ell] = \sigma_x^2 \delta[\ell] + m_x^2 \qquad r_{xy}[\ell] = 0$$

review of discrete-time random processes

- *autocorrelation and autocovariance matrices*
  - are important second-order statistical characterizations of discrete-time random processes and are often represented in a matrix form

the outer product of $\mathbf{x} = [x[0], x[1], \ldots, x[p]]^T$ is a (p+1)×(p+1) matrix:

$$\mathbf{x}\mathbf{x}^H = \begin{bmatrix} x[0]x^*[0] & x[0]x^*[1] & \cdots & x[0]x^*[p] \\ x[1]x^*[0] & x[1]x^*[1] & \cdots & x[1]x^*[p] \\ \vdots & \vdots & \ddots & \vdots \\ x[p]x^*[0] & x[p]x^*[1] & \cdots & x[p]x^*[p] \end{bmatrix}$$

if *x[n]* is WSS, taking the expected value and considering the conjugate symmetry of $r_x^*[\ell]=r_x[-\ell]$, the autocorrelation matrix is obtained:

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\} = \begin{bmatrix} r_x[0] & r_x^*[1] & \cdots & r_x^*[p] \\ r_x[1] & r_x[0] & \cdots & r_x^*[p-1] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[p] & r_x[p-1] & \cdots & r_x[0] \end{bmatrix}$$

Note: since $r_x^*[\ell]=r_x[-\ell]$ then $\mathbf{R}_x$ is Hermitian, in addition it is also Toeplitz (i.e. the elements along each diagonal, parallel to the main diagonal, are equal)

review of discrete-time random processes

- *autocorrelation and autocovariance matrices*

similarly, if $\mathbf{m}_x = [m_x, m_x, \ldots, m_x]^T$ is the (length (p+1)) column vector of the mean of a WSS random process, forming the outer product of $(\mathbf{x} - \mathbf{m}_x)$ leads to a (p+1)×(p+1) covariance matrix:

$$\mathbf{C}_x = E\{(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^H\}$$

or:

$$\mathbf{C}_x = \mathbf{R}_x - \mathbf{m}_x \mathbf{m}_x^H$$

in most practical situations, processes are assumed to be zero-mean and thus

$$\mathbf{C}_x = \mathbf{R}_x$$

the structure of the $\mathbf{R}_x$ matrix reveals it is Hermitian Toeplitz

Note: in the case of real-valued random processes, the autocorrelation matrix is symmetric Toeplitz

– properties of the autocorrelation matrix $\mathbf{R}_x$ of a WSS random process
  1. $\mathbf{R}_x$ is Hermitian (i.e. $\mathbf{R}_x = \mathbf{R}_x^H$) and Toeplitz, $\mathbf{R}_x = \text{Toep}\{r_x[0], r_x[1], \ldots, r_x[p]\}$
  2. $\mathbf{R}_x$ is nonnegative definite, i.e. $\mathbf{R}_x \geq 0$
  3. the eigenvalues λ_k of $\mathbf{R}_x$ are real-valued and nonnegative

review of discrete-time random processes

– property 2 results from the fact that a Hermitian matrix is nonnegative if for any nonzero column vector **a** it can be verified that $\mathbf{a}^H \mathbf{R}_x \mathbf{a} \geq 0$

– in fact, since $\mathbf{R}_x = E\{\mathbf{x}\,\mathbf{x}^H\}$

then $\mathbf{a}^H E\{\mathbf{x}\,\mathbf{x}^H\}\mathbf{a} = E\{\mathbf{a}^H \mathbf{x}\,\mathbf{x}^H \mathbf{a}\} = E\{(\mathbf{a}^H\mathbf{x})(\mathbf{a}^H\mathbf{x})^H\} = E\{(\mathbf{a}^H\mathbf{x})(\mathbf{a}^H\mathbf{x})^*\} = E\{|\mathbf{a}^H\mathbf{x}|^2\} \geq 0$

– property 3 results from the fact that $\mathbf{R}_x$ is Hermitian

Eigenvalues and eigenvectors of $\mathbf{R}_x$

given an Hermitian matrix $\mathbf{R}_x$ (i.e. $\mathbf{R}_x = \mathbf{R}_x^H$), it is important to find the (N×1) vector **q** whose direction is not modified by the linear transformation:

$$\mathbf{R}_x \mathbf{q} = \lambda \mathbf{q}$$

where λ is a constant; vector **q** can be found using $(\mathbf{R}_x - \lambda \mathbf{I})\mathbf{q} = \mathbf{0}$

where **I** is the (N×N) identity matrix and **0** is a (N×1) vector of zeros;

since **q** is arbitrary, the only way this equation is satisfied is if

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

review of discrete-time random processes

This equation is an N-th order polynomial in λ, known as the characteristic equation of $\mathbf{R}_x$. The N roots are called the eigenvalues and, in general, are distinct

$$\lambda_i , \qquad i = 0,1,\ldots,N-1$$

for each eigenvalue λ_i there is an (N×1) eigenvector satisfying

$$\mathbf{R}_x \mathbf{q}_i = \lambda_i \mathbf{q}_i , \qquad i = 0,1,\ldots,N-1$$

Note: the eigenvalues and eigenvectors of a matrix R are conveniently computed in Matlab using `[Lambda, Q]=eig(R)`

Further properties of the autocorrelation matrix $\mathbf{R}_x$:

- the eigenvalues are real and nonnegative: as shown in slide 30, since **R** is positive semidefinite, the quadratic form is positive:

$$\mathbf{q}_i^H \mathbf{R}_x \mathbf{q}_i = \lambda_i \mathbf{q}_i^H \mathbf{q}_i \geq 0$$

- if the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ are distinct, then the corresponding eigenvectors are linearly independent, i.e. given N scalars $\alpha_0, \alpha_1, \ldots, \alpha_{N-1}$

$$\sum_{k=0}^{N-1} \alpha_k \mathbf{q}_k = \mathbf{0}$$

only if all α_k are zero.

Note: linear independence does not imply orthogonality but is a requirement so that the matrix consisting of all eigenvectors is invertible

review of discrete-time random processes

Further properties of the autocorrelation matrix $\mathbf{R}_x$:

- if the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ are distinct, then the corresponding eigenvectors are orthogonal to one another, i.e.

$$\lambda_i \neq \lambda_j \quad \Rightarrow \quad \mathbf{q}_i^H \mathbf{q}_j = 0 \qquad \text{for } i \neq j$$

Note: proof appears e.g. in Manolakis, page 122

- if $\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_{N-1}$ is an orthonormal set of eigenvectors corresponding to $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ distinct eigenvalues, of an N×N correlation matrix **R**, then **R** can be diagonalized as

$$\boldsymbol{\Lambda} = \mathbf{Q}^H \mathbf{R} \mathbf{Q}$$

$$\mathbf{Q} \overset{\Delta}{=} (\mathbf{q}_0 \quad \mathbf{q}_1 \quad \ldots \quad \mathbf{q}_{N-1}),$$

$$\boldsymbol{\Lambda} \overset{\Delta}{=} \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{N-1})$$

Note: **Q** is an eigenmatrix

this results from $\mathbf{RQ} = \mathbf{Q}\boldsymbol{\Lambda}$ and since $\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_{N-1}$ is an orthonormal set of eigenvectors, then Q is unitary, that is $\mathbf{Q}^{-1} = \mathbf{Q}^H$

The correlation matrix R can also be written as

$$\mathbf{R} = \mathbf{Q}\,\boldsymbol{\Lambda}\,\mathbf{Q}^H = \sum_{k=0}^{N-1} \lambda_k \, \mathbf{q}_k \mathbf{q}_k^H$$

which is known as the *spectral theorem*

review of discrete-time random processes

Further properties of the autocorrelation matrix $\mathbf{R}_x$:

- it also follows from linear algebra that if **R** is positive definite (and thus invertible), then, since Λ is diagonal, the inverse of **R** can be obtained as

$$\mathbf{R}^{-1} = \left(\mathbf{Q}\,\Lambda\,\mathbf{Q}^H\right)^{-1} = \mathbf{Q}\,\Lambda^{-1}\,\mathbf{Q}^H = \sum_{k=0}^{N-1} \frac{1}{\lambda_k}\mathbf{q}_k\mathbf{q}_k^H$$

- the trace of **R** :

$$\mathrm{tr}(\mathbf{R}) = \sum_{k=0}^{N-1} \lambda_k$$

- the determinant of **R** :

$$\det(\mathbf{R}) = |\mathbf{R}| = \prod_{k=0}^{N-1} \lambda_k$$

- Eigenvalue spread and dynamic range: the ill conditioning of a matrix $\mathbf{R}_x$ increases with its condition number given by the ratio $\lambda_{max}/\lambda_{min}$, the larger the spread in eigenvalues, the wider (or less flat) the variation of the PSD function, according to the results in slide 31, the condition number is obtained as

$$\chi(\mathbf{R}) \stackrel{\Delta}{=} \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{\max\limits_{\omega} P_x\left(e^{j\omega}\right)}{\min\limits_{\omega} P_x\left(e^{j\omega}\right)}$$

review of discrete-time random processes

- Example (autocorrelation matrix)
  - It has been shown that the autocorrelation function of a random-phase sinusoid is given by

$$r_x[\ell] = \frac{A^2}{2}\cos(\ell\omega_0)$$

  - the corresponding 2×2 correlation matrix is easily obtained as

$$\mathbf{R}_x = \frac{A^2}{2}\begin{bmatrix} 1 & \cos(\omega_0) \\ \cos(\omega_0) & 1 \end{bmatrix}$$

Note: if $\omega_0 \neq 0,\pi$ then $\mathbf{R}_x$ is positive definite

  - and the eigenvalues result from $\det(\mathbf{R}_x - \lambda\mathbf{I}) = 0$

which leads to $\lambda = 1 \pm \cos(\omega_0) \geq 0$

review of discrete-time random processes

- *linear transformation of random vectors*
  - many signal processing applications involve linear operations on random vectors; linear transformations are simple mappings given by the matrix operation

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

where **x** and **y** are the input and output random vector, respectively, and **A** is the transformation matrix (which we assume is rectangular and nonsingular)

thus: $\mathbf{m}_y = E\{\mathbf{y}\} = E\{\mathbf{A}\mathbf{x}\} = \mathbf{A}E\{\mathbf{x}\} = \mathbf{A}\mathbf{m}_x$

and $\mathbf{R}_y = E\{\mathbf{y}\mathbf{y}^H\} = E\{\mathbf{A}\mathbf{x}(\mathbf{A}\mathbf{x})^H\} = E\{\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H\} = \mathbf{A}E\{\mathbf{x}\mathbf{x}^H\}\mathbf{A}^H = \mathbf{A}\mathbf{R}_x\mathbf{A}^H$

similarly: $\mathbf{C}_y = \mathbf{R}_y - \mathbf{m}_y\mathbf{m}_y^H = \mathbf{A}\mathbf{R}_x\mathbf{A}^H - \mathbf{A}\mathbf{m}_x(\mathbf{A}\mathbf{m}_x)^H$

$$\mathbf{C}_y = \mathbf{A}\mathbf{R}_x\mathbf{A}^H - \mathbf{A}\mathbf{m}_x\mathbf{m}_x^H\mathbf{A}^H = \mathbf{A}(\mathbf{R}_x - \mathbf{m}_x\mathbf{m}_x^H)\mathbf{A}^H = \mathbf{A}\mathbf{C}_x\mathbf{A}^H$$

concerning the cross-correlation and cross-variance:

$$\mathbf{R}_{xy} = E\{\mathbf{x}\mathbf{y}^H\} = E\{\mathbf{x}(\mathbf{A}\mathbf{x})^H\} = E\{\mathbf{x}\mathbf{x}^H\mathbf{A}^H\} = E\{\mathbf{x}\mathbf{x}^H\}\mathbf{A}^H = \mathbf{R}_x\mathbf{A}^H$$

$$\mathbf{R}_{yx} = E\{\mathbf{y}\mathbf{x}^H\} = E\{\mathbf{A}\mathbf{x}\mathbf{x}^H\} = E\{\mathbf{A}\mathbf{x}\mathbf{x}^H\} = \mathbf{A}E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{A}\mathbf{R}_x$$

$$\mathbf{C}_{xy} = \mathbf{C}_x\mathbf{A}^H \qquad \mathbf{C}_{yx} = \mathbf{A}\mathbf{C}_x$$

review of discrete-time random processes

- *innovations representation of random vectors*
  - in many practical and theoretical signal processing applications, it is desirable to represent a random vector with a *linearly equivalent* vector consisting of *uncorrelated* components

$$\mathbf{w} = \mathbf{A}\mathbf{x}$$

w is an uncorrelated random vector and expresses an innovations representation, i.e. the random vector **w** contains the same "information" of **x** but because (contrarily to vector **x**) it is uncorrelated, each component w_i of can be thought of as adding "new information" to **w** that is not present in the remaining components

since $\mathbf{C}_w$ must be a diagonal matrix, and since $\mathbf{C}_w = \mathbf{A}\mathbf{C}_x\mathbf{A}^H$, matrix **A** must be such that it diagonalizes matrix $\mathbf{C}_x$ through the transformation **A**. This may be achieved using two approaches:

- eigenanalysis which leads to the (well-known) Karhunen-Loève transform (KLT)
- triangularization methods which lead to LDU decomposition and UDL decomposition

in order to simplify the analysis, we assume that if **x** is a random vector with mean vector μ_x and covariance matrix $\mathbf{C}_x$, then we will just deal with the zero-mean random vector $\mathbf{x}_0=\mathbf{x}-\mu_x$, the resulting covariance matrix is preserved (and $\mathbf{R}_x=\mathbf{C}_x$)

review of discrete-time random processes

- *eigenanalysis and the Karhunen-Loève transform (KLT)*
  - since $\mathbf{w} = \mathbf{A}\mathbf{x}$ and presuming a zero-mean random vector **x**, according to the results in slide 19 and slide 22

$$\boldsymbol{\Lambda}_x = \mathbf{Q}_x^H \mathbf{C}_x \mathbf{Q}_x \qquad \mathbf{C}_w = \mathbf{A}\mathbf{C}_x\mathbf{A}^H$$

let's assume that the diagonalization of $\mathbf{C}_x$ is achieved by making $\mathbf{A}=\mathbf{Q}_x^H$, i.e. the linear transformation matrix **A** is the eigenmatrix $\mathbf{Q}_x^H$:

$$\mathbf{w} = \mathbf{Q}_x^H \mathbf{x}$$

thus $$\mathbf{C}_w = \mathbf{R}_w = E\left\{\mathbf{Q}_x^H\mathbf{x}\left(\mathbf{Q}_x^H\mathbf{x}\right)^H\right\} = E\left\{\mathbf{Q}_x^H\mathbf{x}\mathbf{x}^H\mathbf{Q}_x\right\} = \mathbf{Q}_x^H E\left\{\mathbf{x}\mathbf{x}^H\right\}\mathbf{Q}_x = \mathbf{Q}_x^H\mathbf{C}_x\mathbf{Q}_x = \boldsymbol{\Lambda}_x$$

which means that the covariance matrix $\mathbf{C}_w$ is diagonal, hence several properties result

- the random vector **w** has zero mean and its components are mutually uncorrelated (and therefore orthogonal), also if **x** is $\mathcal{N}(0, \mathbf{C}_x)$,then **w** is $\mathcal{N}(0, \Lambda_x)$ with independent components
- the variances of random variables w_i are equal to the eigenvalues of $\mathbf{C}_x$
- since the transformation matrix $\mathbf{A}=\mathbf{Q}_x^H$ is orthonormal, the transformation is called an orthonormal transformation and the distance measure known as *Mahalanobis distance*:

$$d^2(\mathbf{x}) \overset{\Delta}{=} \mathbf{x}^H\mathbf{C}_x^{-1}\mathbf{x} = \left(\mathbf{Q}_x\mathbf{w}\right)^H\mathbf{C}_x^{-1}\mathbf{Q}_x\mathbf{w} = \mathbf{w}^H\mathbf{Q}_x^H\mathbf{C}_x^{-1}\mathbf{Q}_x\mathbf{w} = \mathbf{w}^H\boldsymbol{\Lambda}_x^{-1}\mathbf{w}$$

is preserved under the transformation

Note: in the case of normal random vectors, the Mahalanobis distance is related to the log-likelihood function

review of discrete-time random processes

- *eigenanalysis and the Karhunen-Loève transform (KLT)*

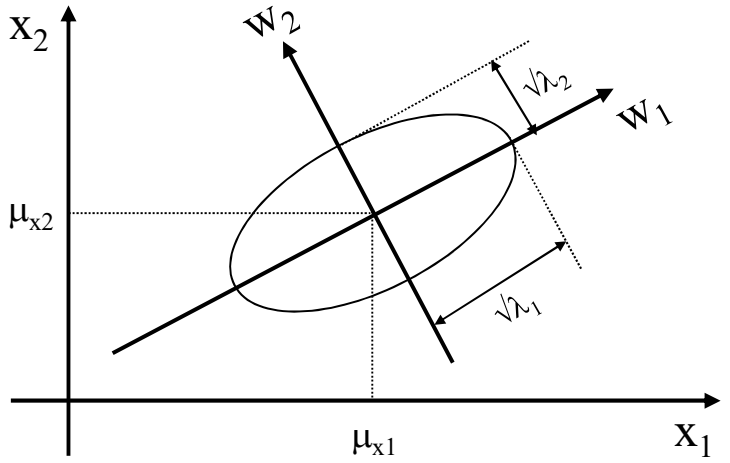
  - since in general $\mathbf{w}=\mathbf{Q}_x^H(\mathbf{x}-\mu_x)$ we have

$$w_i = \mathbf{q}_i^H(\mathbf{x}-\mu_x) = \|\mathbf{x}-\mu_x\| \cos\angle(\mathbf{x}-\mu_x, \mathbf{q}_i)$$

which is the projection of the vector $\mathbf{x}-\mu_x$ onto the unit vector $\mathbf{q}_i$, this means $\mathbf{w}$ represents $\mathbf{x}$ in a new coordinate system that is shifted to μ_x and that is spanned by the set of orthogonal vectors $\mathbf{q}_i$,

geometric interpretation of the transformation for a two-dimensional case, where the contour $d^2(\mathbf{x})$ is displayed

$$d^2(\mathbf{x}) = \mathbf{x}^H\mathbf{C}_x^{-1}\mathbf{x} = \mathbf{w}^H\boldsymbol{\Lambda}_x^{-1}\mathbf{w}$$



Note: an isotropic transformation may be obtained that not only makes $\mathbf{R}_w$ diagonal but also makes it an identity matrix, this is achieved by an additional mapping using the inverse of the square root of the eigenvalue matrix of x, see Manolakis, page 126, a consequence however is that the transformation is orthogonal but not orthonormal, which means the distance measure is not preserved

review of discrete-time random processes

- *eigenanalysis and the Karhunen-Loève transform (KLT)*

KLT: in many signal processing applications, it is convenient to represent the samples of a random signal in another set of number (or coefficients) so that

- as a result of a signal transformation, the energy of the signal is concentrated only in a few coefficients (typical situation in coding/compression)
- the samples are uncorrelated, which is useful e.g. in optimal filtering in order to reduce complexity and improve the signal to noise ratio

main idea 1: to expand a signal as a linear combination of *orthogonal basis functions* so that components of the signal with respect to basis functions do not interfere with one another

- examples of basis functions: the DFT, the DCT, the DST, the Haar transform

main idea 2: a set of orthogonal basis functions for which the signal components are statistically uncorrelated to one another, is based on the second-order properties of the random process and, in particular, on the *diagonalization of the covariance matrix*; it is also an *optimal representation* of the signal in the sense that it provides a representation with the *smallest mean square error* among all other orthogonal transforms

review of discrete-time random processes

- *eigenanalysis and the Karhunen-Loève transform (KLT)*

  – it is shown in Manolakis, section 3.5.3 (pages 130-131) that when a zero-mean random vector **x** with autocorrelation matrix $\mathbf{R}_x$ is transformed using an M×M unitary matrix **A** (i.e. $\mathbf{A}^{-1}=\mathbf{A}^H$):

$$\mathbf{w}=\mathbf{A}^H\mathbf{x} \quad \Leftrightarrow \quad \mathbf{x}=\mathbf{A}\mathbf{w}=\sum_{k=1}^{M} w_k\mathbf{a}_k, \qquad \mathbf{a}_j^H\mathbf{a}_k=0, \quad j\neq k$$

and when only the first m coefficients (m<M) are used to synthesize vector **x**, an error is obtained

$$\mathbf{e}_m=\mathbf{x}-\hat{\mathbf{x}}=\sum_{k=1}^{M} w_k\mathbf{a}_k-\sum_{j=1}^{m} w_j\mathbf{a}_j$$

whose mean square error:

$$E\left\{\mathbf{e}_m^H\mathbf{e}_m\right\}=\sum_{k=1+m}^{M} E\left\{|w_k|^2\right\}\mathbf{a}_k^H\mathbf{a}_k=\sum_{k=1+m}^{M} E\left\{|w_k|^2\right\}=\sum_{k=1+m}^{M}\mathbf{a}_k^H\mathbf{R}_x\mathbf{a}_k=\sum_{k=1+m}^{M}\mathbf{a}_k^H\lambda_k\mathbf{a}_k=\sum_{k=1+m}^{M}\lambda_k$$

is minimized (subject to $\mathbf{a}_k^H\mathbf{a}_k=1$) when **A** is chosen as the eigenmatrix $\mathbf{Q}_x$, i.e. when

$$\mathbf{w}=\mathbf{Q}_x^H\mathbf{x}$$

Note: the optimal solution is found using Lagrange multipliers and is obtained as $\mathbf{R}_x\mathbf{a}_k=\lambda_k\mathbf{a}_k$

  – this orthogonal transformation is the *Discrete Karhunen-Loève transform* (DKLT) and produces a zero-mean uncorrelated random vector with autocorrelation **Λ**

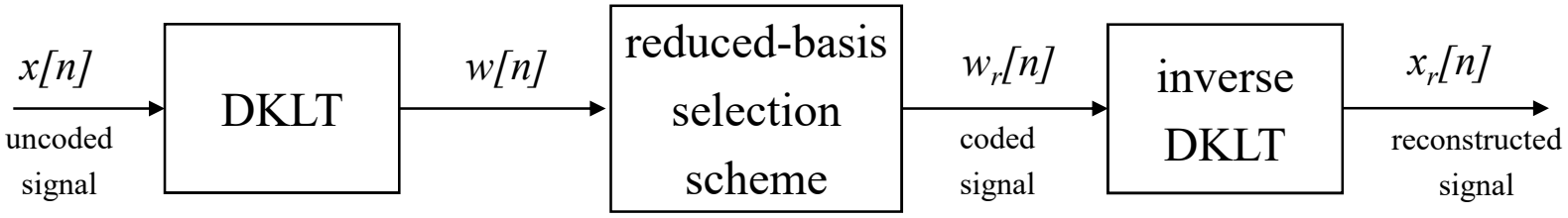
review of discrete-time random processes

- *eigenanalysis and the Karhunen-Loève transform (KLT)*

  – it should be noted that when $\mathbf{A}=\mathbf{Q}_x$, then

$$E\left\{\mathbf{e}_m^H\mathbf{e}_m\right\}=\sum_{k=1+m}^{M}\mathbf{a}_k^H\mathbf{R}_x\mathbf{a}_k=\sum_{k=1+m}^{M}\mathbf{a}_k^H\lambda_k\mathbf{a}_k=\sum_{k=1+m}^{M}\lambda_k$$

  i.e. the MSE in the reduced-basis representation, when only the first *m* basis vectors are used, is the sum of the remaining eigenvalues (which are never negative). Therefore, to obtain a minimum MSE (i.e. an optimum) representation, *the procedure is to choose m eigenvectors corresponding to the m largest eigenvalues*



  – popular application: data compression in communications
    - speech and image (both the transmitter and the receiver must have information about the eigenvectors)

review of discrete-time random processes

the KLT of periodic random sequences (with circulant correlation matrix)

– taking as a reference the definition in slide 15 of the autocorrelation matrix of a random process, if the autocorrelation is periodic with period M, i.e. if in addition to $r_x^*[\ell]=r_x[-\ell]$, it is also true that $r_x[\ell]=r_x[M+\ell]$, then $r_x[\ell]=r_x^*[-\ell]=r_x^*[M-\ell]$, and hence the autocorrelation matrix

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\} = \begin{bmatrix} r_x[0] & r_x^*[1] & \cdots & r_x^*[M-1] \\ r_x[1] & r_x[0] & \cdots & r_x^*[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[M-1] & r_x[M-2] & \cdots & r_x[0] \end{bmatrix}$$

becomes a circular matrix since any row (column) is obtained as circular rotation of the first row (column)

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\} = \begin{bmatrix} r_x[0] & r_x^*[1] & \cdots & r_x^*[M-2] & r_x^*[M-1] \\ r_x^*[M-1] & r_x[0] & \cdots & r_x^*[M-3] & r_x^*[M-2] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_x^*[2] & r_x^*[3] & \cdots & r_x[0] & r_x^*[1] \\ r_x^*[1] & r_x^*[2] & \cdots & r_x^*[M-1] & r_x[0] \end{bmatrix}$$

Note: a circulant matrix is Toeplitz but a Toeplitz matrix is not necessarily circulant

review of discrete-time random processes

the KLT of periodic random sequences (with circulant correlation matrix)

– taking the M-point DFT random sequence $r_x^*[\ell]$:

$$R_x[k] = \sum_{\ell=0}^{M-1} r_x^*[\ell] W_M^{k\ell}, \qquad W_M = \exp(-j\frac{2\pi}{M})$$

and if we define

$$\mathbf{w}_k = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & W_M^k & W_M^{2k} & \dots & W_M^{(M-1)k} \end{bmatrix}^T$$

then, recognizing that $\mathbf{W}_k^{-1} = \mathbf{W}_k^{M-1}$:

- the product of the first row of $\mathbf{R}_x$ by $\mathbf{W}_k$ leads to products $r_x^*[\ell] W_M^{k\ell}$ and to $R_x[k]/\sqrt{M}$
- the product of the second row $\mathbf{R}_x$ by $\mathbf{W}_k$ leads to products

of the type $r_x^*[\ell-1] W_M^{k\ell} = r_x^*[\ell-1] W_M^{k(\ell-1)} W_M^k$ and hence to $R_x[k] W_M^k / \sqrt{M}$

in general the product of the *i*-th row $\mathbf{R}_x$ by $\mathbf{W}_k$ leads to $R_x[k] W_M^{(i-1)k} / \sqrt{M}$

which means:

$$\mathbf{R}_x \mathbf{w}_k = R_x[k] \mathbf{w}_k \quad , \quad k = 0, 1, \dots, M-1$$

that is, the normalized DFT vectors $\mathbf{W}_k$ are the eigenvectors of the circulant matrix $\mathbf{R}_x$, and the corresponding eigenvalues are the DFT coefficients $R_x[k]$, therefore, *the DFT provides the KLT of periodic random sequences*

review of discrete-time random processes

the KLT of periodic random sequences (with circulant correlation matrix)

– If we define the M×M matrix $\mathbf{W} = \begin{bmatrix} \mathbf{w}_0 & \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_{M-1} \end{bmatrix}$

it can be shown that it is unitary:

$$\mathbf{W}^H \mathbf{W} = \mathbf{W}\mathbf{W}^H = \mathbf{I}$$

and the results of the previous slides may be expressed as

$$\mathbf{W}^H \mathbf{R}_x \mathbf{W} = \text{diag}\{R_x[0], R_x[1], \dots, R_x[M-1]\}$$

which shows that the DFT performs the diagonalization of circulant matrices

- in many cases, we can use the DFT to approximate the KLT of stationary random sequences, particularly when the correlation is negligible for $|\ell| > M$

review of discrete-time random processes

- *transformation using triangular decomposition*
  - triangular decomposition leads to transformations that result in *causal* or *anticausal* linear filtering of associated sequences, we will consider
    - lower-diagonal-upper (LDU) leading to causal filtering
    - upper-diagonal-lower (UDL) leading to anticausal filtering

LDU decomposition

- any Hermitian, positive definite matrix **R** can be factored as

$$\mathbf{R} = \mathbf{L}\mathbf{D}_L\mathbf{L}^H \quad \Leftrightarrow \quad \mathbf{L}^{-1}\mathbf{R}\mathbf{L}^{-H} = \mathbf{D}_L$$

Note: $\det(\mathbf{R})=\det(\mathbf{D}_L)$

where **L** is a unit (meaning $\det(\mathbf{L})=1$) lower triangular matrix, $\mathbf{D}_L$ is a diagonal matrix with positive elements, and $\mathbf{L}^H$ is a unit upper triangular matrix

If a linear transformation is defined as $\mathbf{w} = \mathbf{L}^{-1}\mathbf{x} \overset{\Delta}{=} \mathbf{B}\mathbf{x}$ then

$$\mathbf{R}_\mathbf{w} = E\{\mathbf{w}\mathbf{w}^H\} = \mathbf{L}^{-1}E\{\mathbf{x}\mathbf{x}^H\}\mathbf{L}^{-H} = \mathbf{L}^{-1}\mathbf{R}\mathbf{L}^{-H} = \mathbf{D}_L$$

Note: the provided Matlab function `[L,D]=ldlt(R)` computes the LDU decomposition

which means the components of **w** are orthogonal and the elements of $\mathbf{D}_L$ (along the diagonal) are their second moments; also, as w_i is a linear combination of samples of **x** (up to index i), it can be interpreted as a result of *causal* linear filtering (Manolakis, page 128):

$$w_i = \sum_{k=0}^{i} b_{ik}x_k = \sum_{k=0}^{i} b_i[k]x[i-k] \quad , \quad i = 0,1,\ldots,M-1$$

review of discrete-time random processes

- *transformation using triangular decomposition*

UDL decomposition

– similar to LDU, involves factorization of a Hermitian, positive definite matrix **R** as

$$\mathbf{R} = \mathbf{U}\mathbf{D}_U\mathbf{U}^H \quad \Leftrightarrow \quad \mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-H} = \mathbf{D}_U$$

Note: $\det(\mathbf{R})=\det(\mathbf{D}_U)$

where **U** is a unit upper triangular matrix, $\mathbf{D}_U$ is a diagonal matrix with positive elements, and $\mathbf{U}^H$ is a unit lower triangular matrix

Note: $\mathbf{L}\neq\mathbf{U}^H$ and $\mathbf{D}_U\neq\mathbf{D}_L$

If a linear transformation is defined as $\mathbf{w} = \mathbf{U}^{-1}\mathbf{x} \overset{\Delta}{=} \mathbf{A}\mathbf{x}$ then

$$\mathbf{R}_\mathbf{w} = E\{\mathbf{w}\mathbf{w}^H\} = \mathbf{U}^{-1}E\{\mathbf{x}\mathbf{x}^H\}\mathbf{U}^{-H} = \mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-H} = \mathbf{D}_U$$

Note: if **U** is upper triangular, $\mathbf{U}^{-1}$ is also upper triangular

which means the components of **w** are orthogonal and the elements of $\mathbf{D}_U$ (along the diagonal) are their second moments; also, as w_i is a linear combination of samples of **x** (from index i up to M-1), it can be interpreted as a result of *anticausal* linear filtering (Manolakis, page 128):

$$w_i = \sum_{k=i}^{M-1} a_{ik}x_k = \sum_{k=i}^{M-1} a_i[k]x[i-k] \quad , \quad i = 0,1,\ldots,M-1$$

as, if in the LDU decomposition $\mathbf{x} = \begin{bmatrix} x[n] & x[n-1] & x[n-2] & \ldots & x[n-M+1] \end{bmatrix}^T$

review of discrete-time random processes

- *comparison of eigenanalysis and (lower) triangular decompositions for zero-mean random vectors*

eigenanalysis	(lower) triangular decomposition
$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$	$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$
$\mathbf{R}\mathbf{q}_k = \lambda_k \mathbf{q}_k$	
$\mathbf{Q} = [\mathbf{q}_0 \quad \mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_{M-1}]$	$\mathbf{L} = \text{unit lower triangular}$
$\mathbf{\Lambda} = \text{diag}\{\lambda_0, \lambda_1, \ldots, \lambda_{M-1}\}$	$\mathbf{D}_L = \text{diag}\{\xi_0, \xi_1, \ldots, \xi_{M-1}\}$
$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \sum_{k=0}^{M-1} \lambda_k \mathbf{q}_k \mathbf{q}_k^H$	$\mathbf{R} = \mathbf{L}\mathbf{D}_L\mathbf{L}^H$
$\mathbf{R}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^H = \sum_{k=0}^{M-1} \frac{1}{\lambda_k} \mathbf{q}_k \mathbf{q}_k^H$	$\mathbf{R}^{-1} = \mathbf{L}^{-H}\mathbf{D}_L^{-1}\mathbf{L}^{-1}$
$\mathbf{\Lambda} = \mathbf{Q}^H\mathbf{R}\mathbf{Q}$ $\quad$ $\mathbf{\Lambda}^{-1} = \mathbf{Q}^H\mathbf{R}^{-1}\mathbf{Q}$	$\mathbf{D}_L = \mathbf{L}^{-1}\mathbf{R}\mathbf{L}^{-H}$ $\quad$ $\mathbf{D}_L^{-1} = \mathbf{L}^{-H}\mathbf{R}^{-1}\mathbf{L}^{-1}$
$\det(\mathbf{R}) = \det(\mathbf{\Lambda}) = \prod_{k=0}^{M-1} \lambda_k$ $\quad$ $\text{tr}(\mathbf{R}) = \text{tr}(\mathbf{\Lambda}) = \sum_{k=0}^{M-1} \lambda_k$	$\det(\mathbf{R}) = \det(\mathbf{D}_L) = \prod_{k=0}^{M-1} \xi_k$
$\mathbf{w} = \mathbf{Q}_x^H\mathbf{x}$ $\quad$ $E\{\mathbf{w}\mathbf{w}^H\} = \mathbf{\Lambda}$	$\mathbf{w} = \mathbf{L}^{-1}\mathbf{x}$ $\quad$ $E\{\mathbf{w}\mathbf{w}^H\} = \mathbf{D}_L$

review of discrete-time random processes

- *generation of real-valued random vectors with given second order moments (Manolakis, pages 128-129)*
  - problem: to generate M samples of a zero-mean real-valued random vector **x** with a given symmetric positive definite autocorrelation matrix $\mathbf{R}_x$

the innovations representation of the previous slides suggest three approches, the common idea is to factor $\mathbf{R}_x$ using either the eigenanalysis (orthonormal transformation) or the triangularization transformation, to obtain the diagonal matrix (Λ_x or $\mathbf{D}_L$ or $\mathbf{D}_U$), generate M samples of an IID sequence (vector **w**) with the obtained diagonal variances (a normal pseudo number generator is presumed in order to ensure preservation the original distribution of the IID samples), and then transform the samples by using the inverse transformation matrix ($\mathbf{Q}_x$ or $\mathbf{L}_x$ or $\mathbf{U}_x$)

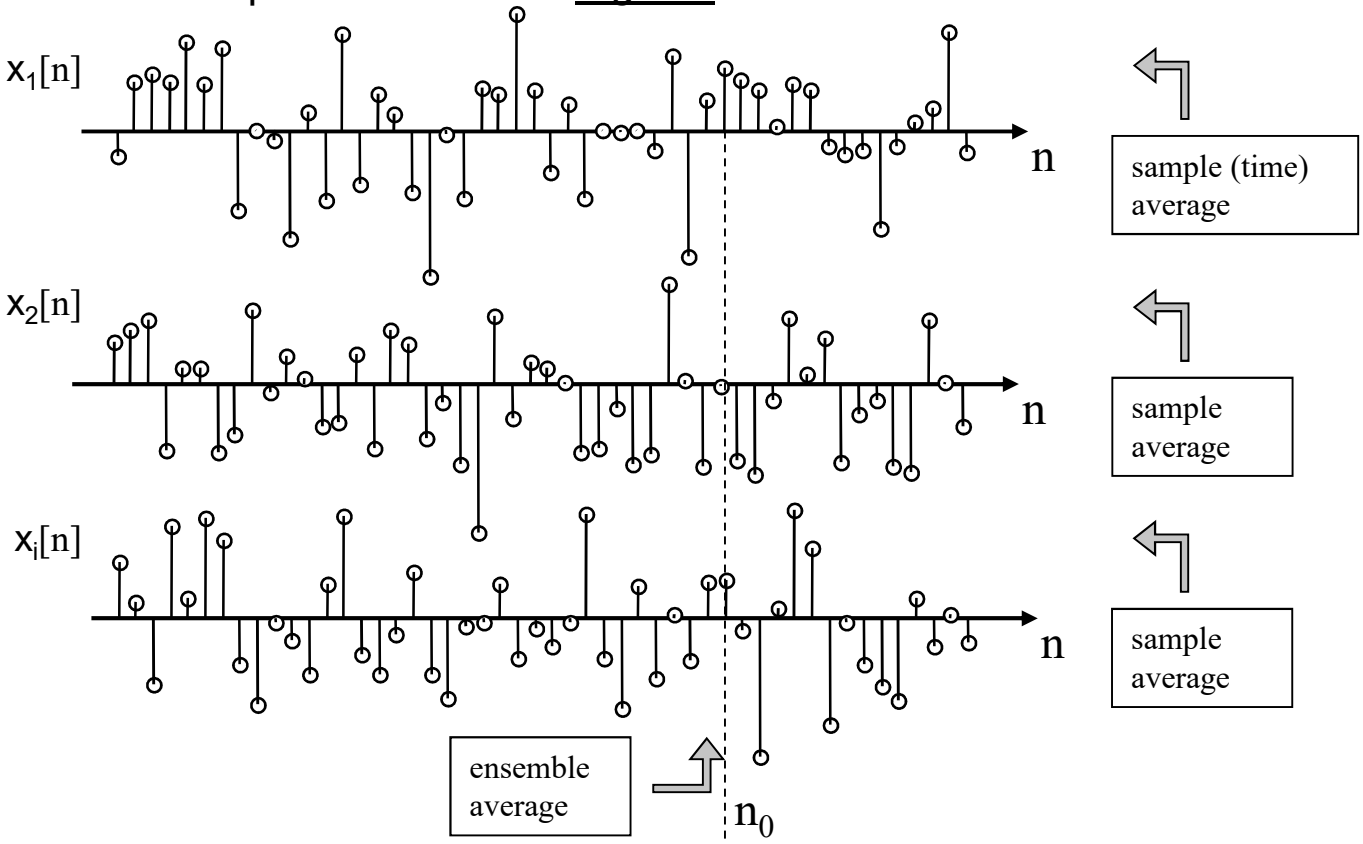
1. Eigendecomposition approach: first factor $\mathbf{R}_x$ as $\mathbf{R}_x=\mathbf{Q}_x\Lambda_x\mathbf{Q}_x^H$, then generate **w** using $\mathcal{N}(0, \Lambda_x)$, finally compute **x** using $\mathbf{x}=\mathbf{Q}_x\mathbf{w}$,

2. LDU triangularization approach: first factor $\mathbf{R}_x$ as $\mathbf{R}_x=\mathbf{L}_x\mathbf{D}_L\mathbf{L}_x^H$, then generate **w** using $\mathcal{N}(0, \mathbf{D}_L)$, finally compute **x** using $\mathbf{x}=\mathbf{L}_x\mathbf{w}$,

3. UDL triangularization approach: first factor $\mathbf{R}_x$ as $\mathbf{R}_x=\mathbf{U}_x\mathbf{D}_U\mathbf{U}_x^H$, then generate **w** using $\mathcal{N}(0, \mathbf{D}_U)$, finally compute **x** using $\mathbf{x}=\mathbf{U}_x\mathbf{w}$.

review of discrete-time random processes

- *ergodicity*
  - all the statistical averages considered so far are ensemble averages (i.e. evaluated for a particular time instant n_0) but these are usually not available or accessible
  - usually, only a single (sample) realization of the random process is available
  - if the statistical averages obtained from a single realization of the random process (i.e. time averages) are representative (and thus, valid estimates) of ensemble averages, then the random process is called ***ergodic***



review of discrete-time random processes

- *ergodicity in the mean*
  - if the sample mean of a WSS random process converges to m_x in the mean square sense, then the process is *ergodic in the mean*

sample mean: $\langle x[n] \rangle_N = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$

Note: only stationary signals can be ergodic, WSS does not imply ergodicity

ergodic in the mean if $\lim_{N\to\infty} \langle x[n] \rangle_N = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} x[n] = E\{x[n]\} = m_x$

convergence in the mean square sense requires that

1. the sample mean be asymptotically unbiased (which is true for any WSS process):

$$E\{\langle x[n] \rangle_N\} = \frac{1}{N} \sum_{n=0}^{N-1} E\{x[n]\} = m_x$$

2. the variance of the estimate goes to zero (i.e. vanishes) as N$\to\infty$

$$Var\{\langle x[n] \rangle_N\} = E\{(\langle x[n] \rangle_N - m_x)(\langle x[n] \rangle_N - m_x)^*\} = E\left\{\left(\frac{1}{N} \sum_{n=0}^{N-1} (x[n] - m_x)\right)\left(\frac{1}{N} \sum_{k=0}^{N-1} (x[k] - m_x)\right)^*\right\}$$

$$Var\{\langle x[n] \rangle_N\} = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} E\{(x[n] - m_x)(x[k] - m_x)^*\} = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} c_x[n-k]$$

review of discrete-time random processes

$$Var\left\{\langle x[n]\rangle_N\right\} = \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{k=0}^{N-1} c_x[n-k] = \frac{1}{N^2}\sum_{k=-(N-1)}^{N-1}\left(N-|k|\right)c_x[k] = \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)c_x[k]$$

where $c_x[k]$ denotes the autocovariance of $x[n]$

thus, $x[n]$ is ergodic in the mean if and only if $\lim_{N\to\infty}\frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)c_x[k]=0$

a (simpler) necessary and sufficient condition is: $\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}c_x[k]=0$

other sufficient conditions are: $c_x[0]<\infty$ and $\lim_{k\to\infty}c_x[k]=0$ i.e. a WSS process is ergodic in the mean if it is asymptotically uncorrelated

Example:

the random phase sinusoid: $x[n]=A\sin\left(n\omega_0+\phi\right)$

its autocovariance function is $c_x[\ell]=\frac{A^2}{2}\cos\left(\ell\omega_0\right)$ and, if $\omega_0\neq 0$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\ell=0}^{N-1}c_x[\ell]=\lim_{N\to\infty}\frac{A^2}{2N}\mathrm{Re}\left\{\sum_{\ell=0}^{N-1}\exp\left(j\ell\omega_0\right)\right\}=\lim_{N\to\infty}\frac{A^2}{2N}\frac{\sin\left(N\omega_0/2\right)}{\sin\left(\omega_0/2\right)}\cos\left[\left(N-1\right)\omega_0/2\right]=0$$

review of discrete-time random processes

- *ergodicity in the autocorrelation*

  estimating the autocorrelation sequence $r_x[\ell] = E\{x[n]x^*[n-\ell]\}$ using a single realization of the process *x[n]* may be achieved using

$$\langle r_x[\ell] \rangle_N = \frac{1}{N}\sum_{n=0}^{N-1} x[n]x^*[n-\ell] = \frac{1}{N}\sum_{n=0}^{N-1} y_\ell[n]$$

  if $\langle r_x[\ell] \rangle_N$ converges in the mean-square sense to $r_x[\ell]$ as N$\to\infty$, the process is autocorrelation ergodic and:

$$\lim_{N\to\infty} \langle r_x[\ell] \rangle_N = r_x[\ell]$$

  since $\langle r_x[\ell] \rangle_N$ is the sample mean of $y_k[n]$, *x[n]* is autocorrelation ergodic if $y_k[n]$ is ergodic in the mean

  Note: if x[n] is ergodic in both the mean and autocorrelation, it is also WSS, however, WSS does not imply ergodicity

  particular case for a Gaussian process: a necessary and sufficient condition for a WSS process with covariance $c_x[\ell]$ to be autocorrelation ergodic, in the mean square sense, is $\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} c_x^2[k] = 0$

review of discrete-time random processes

- *white noise*
  - a wide-sense stationary process *v[n]* is white if the autocovariance function is zero for all k≠0, i.e. it consists of a sequence of uncorrelated random variables (each having a variance σ_v^2) :

$$c_v[k] = \sigma_v^2 \, \delta[\ell]$$

    - white Gaussian noise (WGN) consists of a sequence of uncorrelated real-valued Gaussian random variables

    - if the noise is complex $v[n] = v_R[n] + jv_I[n]$

    then $E\left\{|v[n]|^2\right\} = E\left\{|v_R[n]|^2\right\} + E\left\{|v_I[n]|^2\right\}$

- *the power spectrum*
  - the Fourier transform of a random process can not be computed because the process is an ensemble of discrete signals, however, it is possible to obtain a frequency-domain representation of the process by taking the Fourier transform of ensemble averages

review of discrete-time random processes

- *the power spectrum*
  - the power spectral density (or power spectrum) of a random process is the Fourier transform of the (deterministic sequence) $r_x[\ell]$

$$P_x\left(e^{j\omega}\right) = \sum_{\ell=-\infty}^{\infty} r_x[\ell] e^{-j\omega\ell}$$

    - if *x[n]* is not zero-mean, then the power spectrum has a spike at $\omega=0$, for nonzero-mean random processes the power spectrum is defined to be the discrete-time Fourier transform of the autocovariance

  - the Fourier pair is completed with the inverse Fourier transform of $P_x(e^{j\omega})$

$$r_x[\ell] = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right) e^{j\omega\ell} d\omega$$

    - the power spectrum provides a frequency-domain description of the second-order moment of the process

  - as a more general alternative, the power spectrum of *x[n]* can also be obtained as the Z-transform of $r_x[\ell]$
  -

$$P_x(z) = \sum_{\ell=-\infty}^{\infty} r_x[\ell] z^{-\ell}$$

review of discrete-time random processes

- *properties of the power spectrum*
  - Positivity: the power spectrum of a WSS process is positive

$$P_x\left(e^{j\omega}\right) \geq 0$$

  - Total power: the power in a zero mean WSS random process is proportional to the area under the power spectral density curve

$$E\left\{\left|x[n]\right|^2\right\} = r_x[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right)d\omega$$

Note: the power spectrum can be seen as a density function that describes how the power of x[n] is distributed in ω

  - Symmetry: since the autocorrelation of a WSS random process is conjugate symmetric, the power spectrum is a real function of ω

$$r_x[\ell] = r_x^*[-\ell] \quad \Rightarrow \quad P_x\left(e^{j\omega}\right) = P_x^*\left(e^{j\omega}\right) \qquad P_x(z) = P_x^*\left(1/z^*\right)$$

In addition, is *x[n]* is real, then $P_x(e^{j\omega})$ is even and therefore

$$r_x[\ell] = r_x[-\ell] \quad \Rightarrow \quad P_x\left(e^{j\omega}\right) = P_x\left(e^{-j\omega}\right) \qquad P_x(z) = P_x\left(1/z\right)$$

review of discrete-time random processes

- *properties of the power spectrum*
  - Eigenvalue extremal property: the eigenvalues of the n×n autocorrelation matrix of a zero mean WSS random process are upper and lower bounded by the maximum and minimum values, respectively of the power spectrum

$$\min_{\omega} P_x\left(e^{j\omega}\right) \leq \lambda_i \leq \max_{\omega} P_x\left(e^{j\omega}\right)$$

in fact, if λ_i and $\mathbf{q}_i$ are the eigenvalues and eigenvectors of the autocorrelation matrix $\mathbf{R}_x$, we have:

$$\mathbf{R}_x\mathbf{q}_i = \lambda_i\mathbf{q}_i\ , \qquad i = 0,1,\ldots,N-1$$

$$\Leftrightarrow \quad \mathbf{q}_i^H\mathbf{R}_x\mathbf{q}_i = \lambda_i\mathbf{q}_i^H\mathbf{q}_i$$

and thus

$$\lambda_i = \frac{\mathbf{q}_i^H\mathbf{R}_x\mathbf{q}_i}{\mathbf{q}_i^H\mathbf{q}_i}$$

also, since $\mathbf{q}_i = \left[q_i[0], q_i[1], \ldots, q_i[N-1]\right]$ and $r_x[k-\ell] = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right)e^{j\omega(k-\ell)}d\omega$

$$\mathbf{q}_i^H\mathbf{R}_x\mathbf{q}_i = \sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} q_i^*[k]\,r_x[k-\ell]\,q_i[\ell]$$

review of discrete-time random processes

we have $$\mathbf{q}_i^H \mathbf{R}_x \mathbf{q}_i = \frac{1}{2\pi} \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} q_i^*[k] q_i[\ell] \int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right) e^{j\omega(k-\ell)} d\omega$$

$$\mathbf{q}_i^H \mathbf{R}_x \mathbf{q}_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right) \sum_{k=0}^{N-1} q_i^*[k] e^{j\omega k} \sum_{\ell=0}^{N-1} q_i[\ell] e^{-j\omega\ell} d\omega$$

and if $$\sum_{\ell=0}^{N-1} q_i[\ell] e^{-j\omega\ell} = Q\left(e^{j\omega}\right)$$

then $$\mathbf{q}_i^H \mathbf{R}_x \mathbf{q}_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right) Q_i^*\left(e^{j\omega}\right) Q_i\left(e^{j\omega}\right) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right) \left|Q_i\left(e^{j\omega}\right)\right|^2 d\omega$$

also, by the Parseval theorem $$\mathbf{q}_i^H \mathbf{q}_i = \sum_{k=0}^{N-1} \left|q_i[k]\right|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|Q_i\left(e^{j\omega}\right)\right|^2 d\omega$$

we obtain $$\lambda_i = \frac{\mathbf{q}_i^H \mathbf{R}_x \mathbf{q}_i}{\mathbf{q}_i^H \mathbf{q}_i} = \frac{\int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right) \left|Q_i\left(e^{j\omega}\right)\right|^2 d\omega}{\int_{-\pi}^{\pi} \left|Q_i\left(e^{j\omega}\right)\right|^2 d\omega}$$

and since $$\min_{\omega} P_x\left(e^{j\omega}\right) \le P_x\left(e^{j\omega}\right) \le \max_{\omega} P_x\left(e^{j\omega}\right)$$ then $$\min_{\omega} P_x\left(e^{j\omega}\right) \le \lambda_i \le \max_{\omega} P_x\left(e^{j\omega}\right)$$

review of discrete-time random processes

it can also be shown [Monson, page 98] that, in addition to providing a frequency-domain representation of the second-order moment, the power spectrum may also be related to the ensemble average of the squared (discrete-time) Fourier magnitude $|X(e^{j\omega})|^2$, in fact, it may be viewed as the expected value of $P_N(e^{j\omega})$ in the limit as $N \to \infty$

$$P_x\left(e^{j\omega}\right) = \lim_{N\to\infty} E\left\{P_N\left(e^{j\omega}\right)\right\} = \frac{1}{2N+1} E\left\{\left|\sum_{k=-N}^{N} x[n]e^{-j\omega n}\right|^2\right\}$$

Summary note concerning stationarity and ergodicity: in practice, stationarity means that the most important statistical properties do not change over the time when we observe the signal; whereas stationarity ensures the time invariance of the statistics of a random signal, ergodicity implies that any statistics can be calculated either by averaging over all members of the ensemble at a fixed time, or by time averaging over any single representative member of the ensemble

Summary note concerning white noise: the term white noise is used to emphasize that all frequencies contribute the same amount of power, as in the case of white light, which is obtained by mixing all possible colors by the same amount

Summary note concerning uncorrelatedness and PDF: the conditions of uncorrelatedness or independence do not put any restriction on the form of the probability density function (PDF), thus an IID discrete-time random process can have any type of probability distribution; what about deterministic signals ?

review of discrete-time random processes

- *sum of independent random variables*
  - a random variable *y* may be expressed as a combination of M statistically independent random variables x_k (c_k are constants):

$$y = \sum_{k=1}^{M} c_k x_k$$

it can be shown [Manolakis, page 90] that if the individual means are m_{xk}, the individual variances are σ_{xk}^2, and the individual PDFs are $f_{xk}(\alpha)$, then

$$m_y = \sum_{k=1}^{M} c_k m_{xk}$$

$$\sigma_y^2 = \sum_{k=1}^{M} |c_k|^2 \sigma_{xk}^2$$

$$f_y(\alpha) = \frac{1}{|c_1|} f_{x1}\left(\frac{\alpha}{c_1}\right) * \frac{1}{|c_2|} f_{x2}\left(\frac{\alpha}{c_2}\right) * \quad \cdots \quad * \frac{1}{|c_M|} f_{xM}\left(\frac{\alpha}{c_M}\right)$$

the *Central Limit Theorem* states that the sum of an infinite number of statistically independent random variables (with finite mean and variance) leads to a combined distribution that converges to the Gaussian distribution