Statistical Digital Signal Processing

Chapter 1: Introduction and Review of Stochastic Process

Objective of the Course

- To enable students analyze, represent and manipulate random signals with LTI systems.
  - Understand challenges posed by random signals,
  - Understand how to model random signals,
  - Understand how to represent random signals,
  - Understand how to design LTI system to estimate random signals,
  - Understand efficient algorithms to estimate random signals.

Content of the course

1. Introduction and Review of Stochastic Process
2. Linear Signal Modelling
3. Nonparametric Power Spectrum Estimation
4. Optimum Linear Estimation of Signals
5. Algorithms for Optimum Linear Filter
6. Adaptive Filters

References

- Statistical Digital Signal Processing and Modeling, M. Hayes, Wiley, 1996.
- Statistical and Adaptive signal Processing, Dimitris G. Manolakis, Vinay K. Ingle, Stephen M. Kogon, Artech House, 2005
- Optimum Signal Processing, Sophocles J. Orfanidis, McGraw-Hill, 2007

Evaluation

- Assignment (20%)
- Mid Exam (30%)
- Final Exam (50%)

Random Variables

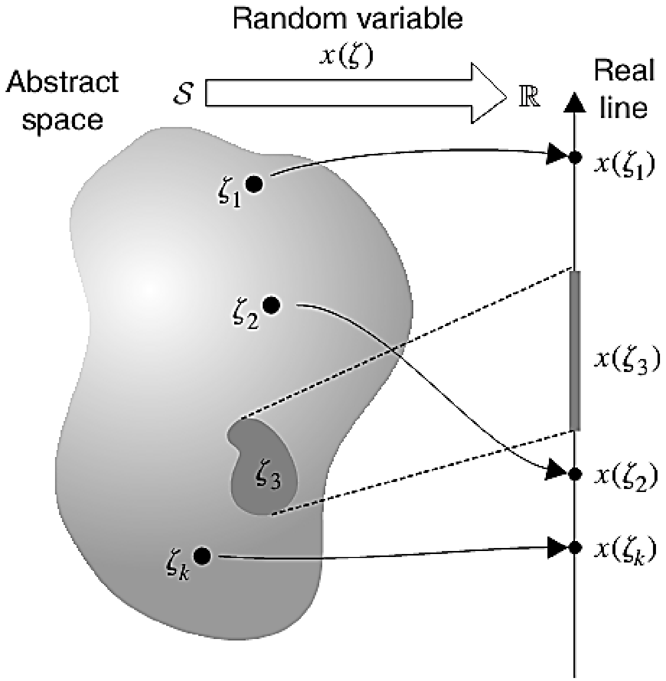
- Any random situation can be studied by the axiomatic definitions of probability by defining $(S, \mathcal{F}, Pr)$.
  - $S = \{\zeta_1, \zeta_2, \ldots\}$ – Universal set of unpredictable outcomes
  - $\mathcal{F}$ – collection of subset of S whose elements are called events.
  - $\Pr\{\zeta_k,\ \ k = 1,2, \ldots\}$ – probability representing the unpredictability of these events.

- Difficult to work with this probability space for two reasons.
  - The basic space contains abstract events and outcomes that are difficult to manipulate.
    - We want random outcomes that can be measured and manipulated in a meaningful way by using numerical operations.
  - The probability function Pr $\{\cdot\}$ is a set function that again is difficult, to manipulate by using calculus.

- A random variable x(ζ) is a mapping that assigns a real number x to every outcome ζ from an abstract probability space.

- A complex valued random variable represented as

$$x(\zeta) = x_R(\zeta) + jx_I(\zeta$$



- This mapping should satisfy the following two conditions:
  - the interval $\{X(\zeta) \leq x\}$ is an event in the abstract probability space for every x ;
  - $\Pr\{X(\zeta) = \infty\} = 0$ and $\Pr\{X(\zeta) = -\infty\} = 0$.
- A random variable is called *discrete-valued* if x takes a discrete set of values $\{x_k\}$;
- Otherwise, it is termed a *continuous-valued* random variable.

Representation of Random Variables

Cumulative distribution function (CDF)	$F_X(x) = Pr\{X(\zeta) \le x\}$
Probability density function (pdf)	$f_X(x) = \dfrac{dF_X(x)}{dx}$
Expectation of a random variable	$E\{x(\zeta)\} = \mu_x = \begin{cases} \sum_x x_k p_x \\ \int_{-\infty}^{\infty} x f_X(x) dx \end{cases}$
Expectation of a function of random variable	$E\{g[x(\zeta)]\} = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
Moments	$r_x^{(m)} = E\{x^m(\zeta)\} = \int_{-\infty}^{\infty} x^m f_X(x) dx$
Central Moments	$\gamma_x^{(m)} = E\{(x(\zeta) - \mu_x)^m\} = \int_{-\infty}^{\infty} (x - \mu_x)^m f_X(x) dx$

Second central moment or variance	$\sigma_x^2 = \gamma_x^{(2)} = E\{(x(\zeta) - \mu_x)^2\}$
Skewness	$k_x^{(3)} = \frac{1}{\sigma_x^3}\gamma_x^{(3)}$
Kurtosis	$k_x^{\sim(4)} = \frac{1}{\sigma_x^4}\gamma_x^{(4)} - 3$
Characteristic functions	$\mathbf{\Phi}_x(\xi) = E\left\{e^{j\xi x(\zeta)}\right\} = \int_{-\infty}^{\infty} f_X(x) e^{j\xi x} dx$
Moment generating functions	$\bar{\mathbf{\Phi}}_x(s) = E\left\{e^{sx(\zeta)}\right\} = \int_{-\infty}^{\infty} f_X(x) e^{sx} dx$
Cumulants generating functions	$\bar{\mathbf{\Psi}}_x(s) = \ln \bar{\mathbf{\Phi}}_x(s) = \ln E\left\{e^{sx(\zeta)}\right\}$
Cumulants	$k_x^{(m)} = \left.\frac{\mathrm{d}^m[\bar{\mathbf{\Psi}}_x(s)]}{\mathrm{d}s^m}\right\rvert_{s=0}$

Useful Random Variables

Uniformly distributed RV

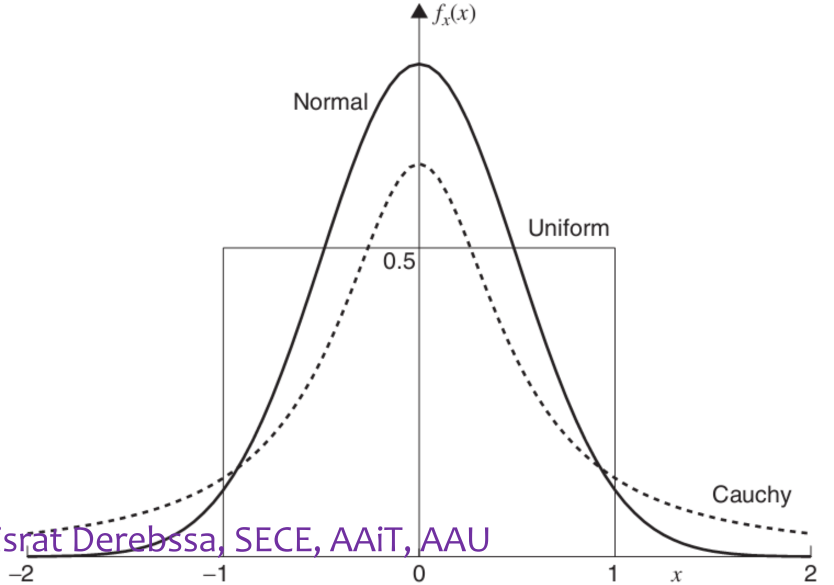
$$f_x(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$$

Normal RV

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma_x}\right)^2\right]}$$

Cauchy RV

$$f_x(x) = \frac{\beta}{\pi} \frac{1}{(x-\mu)^2 + \beta^2}$$



Random Vectors

- A real-valued **random vector** containing M RV is represented as:

$$\mathbf{X}(\zeta) = [x_1(\zeta), x_2(\zeta), \dots, x_M(\zeta)]^T$$

- A random vector is completely characterized by its joint CDF

$$F_{\mathbf{x}}(x_1, x_2, \dots, x_M) = \Pr\{X_1(\zeta) \le x_1, X_2(\zeta) \le x_2, \dots, X_M(\zeta) \le x_M\}$$

- Often written as

$$F_{\mathbf{X}}(\boldsymbol{x}) = \Pr\{\mathbf{X}(\zeta) \le \boldsymbol{x}\}$$

- Two random variables $X_1(\zeta)$ and $X_2(\zeta)$ are independent if the events $\{X_1(\zeta) \le x_1\}$ and $\{X(\zeta) \le x_2\}$ are jointly independent. That is,

$$\Pr\{X_1(\zeta) \le x_1, X_2(\zeta), \le x_2\} = \Pr\{X_1(\zeta), \le x_1\} \Pr\{X_2(\zeta), \le x_2\}$$

- The probability functions require an enormous amount of information that is not easy to obtain or is too complex mathematically for practical use.
- In practical applications, random vectors are described by less complete but more manageable statistical averages.

Statistical Description of Random Vector

Mean vector	$E\{\mathbf{x}(\zeta)\} = \boldsymbol{\mu}_\mathbf{x} = \begin{bmatrix} E\{x_1(\zeta)\} \\ \vdots \\ E\{x_M(\zeta)\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_M \end{bmatrix}$
Autocorrelation matrix	$\mathbf{R}_\mathbf{x} = E\{\mathbf{x}(\zeta)\mathbf{x}^H(\zeta)\} = \begin{bmatrix} r_{11} & \cdots & r_{1M} \\ \vdots & \ddots & \vdots \\ r_{M1} & \cdots & r_{MM} \end{bmatrix}$ $r_{ij} \triangleq E\{x_i(\zeta)x_j^*(\zeta)\} = r_{ji}^*$
Autocovariace matrix	$\boldsymbol{\Gamma}_\mathbf{x} = E\{[\mathbf{x}(\zeta) - \boldsymbol{\mu}_\mathbf{x}][\mathbf{x}(\zeta) - \boldsymbol{\mu}_\mathbf{x}]^H\} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1M} \\ \vdots & \ddots & \vdots \\ \gamma_{M1} & \cdots & \gamma_{MM} \end{bmatrix}$ $\boldsymbol{\Gamma}_\mathbf{x} = \mathbf{R}_\mathbf{x} - \boldsymbol{\mu}_\mathbf{x}\boldsymbol{\mu}_\mathbf{x}{}^H$
Correlation coefficient	$\rho_{ij} = \frac{\gamma_{ij}}{\sigma_i \sigma_j} = \rho_{ji}$
Uncorrelatedness	$\gamma_{ij} = 0, \text{for } i \neq j$
Orthogonal	$r_{ij} = 0, \text{for } i \neq j$

Statistical Description of Two Random Vectors

Cross-correlation matrix	$\mathbf{R_{xy}} = E\{\mathbf{x}(\zeta)\mathbf{y}^{\boldsymbol{H}}(\zeta)\} = \begin{bmatrix} E\{x_1(\zeta)y_1^*(\zeta)\} & \dots & E\{x_1(\zeta)y_L^*(\zeta)\} \\ \vdots & \ddots & \vdots \\ E\{x_M(\zeta)y_1^*(\zeta)\} & \dots & E\{x_M(\zeta)y_L^*(\zeta)\} \end{bmatrix}$
Cross-covariance matrix	$\boldsymbol{\Gamma}_{\mathbf{x}y} = E\left\{[\mathbf{x}(\zeta) - \boldsymbol{\mu}_{\mathbf{x}}]\left[\mathbf{y}(\zeta) - \boldsymbol{\mu}_{\mathbf{y}}\right]^{\boldsymbol{H}}\right\} = \mathbf{R_{xy}} - \boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{y}}{}^{H}$
Uncorrelated	$\boldsymbol{\Gamma}_{\mathbf{x}y} = \mathbf{0} \rightarrow \mathbf{R_{xy}} = \boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{y}}{}^{H}$
Orthogonal	$\mathbf{R_{xy}} = \mathbf{0}$

Linear Transformations of Random Vector

- Linear transformations are relatively simple mappings and are given by the matrix operation

  $\mathbf{y}(\zeta) = \mathbf{A}\mathbf{x}(\zeta)$, A is an LxM matrix and x is M dimensional vector
- Assuming L=M, and both are real valued.

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{f_x(\mathbf{A}^{-1}\mathbf{y})}{|\det \boldsymbol{A}|}$$

- If L>M, only M random variables $\mathbf{y_i}(\zeta)$ can be independently determined from $\mathbf{x}(\zeta)$,
  - The remaining L-M can be obtained from the first $\mathbf{y_i}(\zeta)$
- If L<M, we can augment $\mathbf{y}(\zeta)$ into an M-vector by introducing auxiliary random variables.

- For complex valued RV,

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{f_x(\mathbf{A}^{-1}\mathbf{y})}{|\det \boldsymbol{A}|^2}$$

- The determination of $f_{\mathbf{y}}(\mathbf{y})$ is tedious and in practice not necessary.

Statistical Description of Linear Transformation of Random Vector

Mean vector	$\boldsymbol{\mu}_{\mathbf{y}} = E\{\mathbf{y}(\zeta)\} = E\{\mathbf{A}\mathbf{x}(\zeta)\} = \mathbf{A}E\{\mathbf{x}(\zeta)\} = \mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}$
Autocorrelation matrix	$\mathbf{R}_{\mathbf{y}} = E\{\mathbf{y}\boldsymbol{y}^{\boldsymbol{H}}\} = E\{\mathbf{A}\mathbf{x}\mathbf{x}^{H}\boldsymbol{A}^{H}\}=\mathbf{A}E\{\mathbf{x}\mathbf{x}^{H}\}\boldsymbol{A}^{H}=\mathbf{A}\mathbf{R}_{\mathbf{x}}\boldsymbol{A}^{H}$
Autocovariace matrix	$\boldsymbol{\Gamma}_{\mathbf{y}} = \mathbf{A}\boldsymbol{\Gamma}_{\mathbf{x}}\boldsymbol{A}^{H}$
Cross-correlation	$\mathbf{R}_{\mathbf{xy}} = E\{\mathbf{x}\boldsymbol{y}^{\boldsymbol{H}}\} = E\{\mathbf{x}\mathbf{x}^{H}\boldsymbol{A}^{H}\}=E\{\mathbf{x}\mathbf{x}^{H}\}\boldsymbol{A}^{H}=\mathbf{R}_{\mathbf{x}}\boldsymbol{A}^{H}$ $\mathbf{R}_{\mathbf{yx}} = \mathbf{A}\mathbf{R}_{\mathbf{x}}$
Cross-covariance	$\boldsymbol{\Gamma}_{\mathbf{xy}} = \boldsymbol{\Gamma}_{\mathbf{x}}\boldsymbol{A}^{H}$ $\boldsymbol{\Gamma}_{\mathbf{yx}} = \mathbf{A}\boldsymbol{\Gamma}_{\mathbf{x}}$

Normal Random Vectors

- If the components of the random vector x (ζ) are jointly normal, then x (ζ) is a normal random *M* -vector.
- For real valued normal random vector

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{M/2}|\boldsymbol{\Gamma}_{\mathbf{x}}|^{1/2}} e^{\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Gamma}_{\mathbf{x}}{}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})\right]}$$

- Its characteristic equation is

$$\Phi_{\mathbf{x}}(\boldsymbol{\xi}) = e^{\left(j\boldsymbol{\xi}^T\boldsymbol{\mu}_{\mathbf{x}} - \frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\Gamma}_{\mathbf{x}}\boldsymbol{\xi}\right)}$$

Properties of normal random vector

- Pdf and all higher order moments completely specified from mean vector and covariance matrix.
- If the components of $\mathbf{x}(\zeta)$ are mutually uncorrelated, they are also independent.
- A linear transformation of a normal random vector is also normal.

Sum of Independent Random Variables

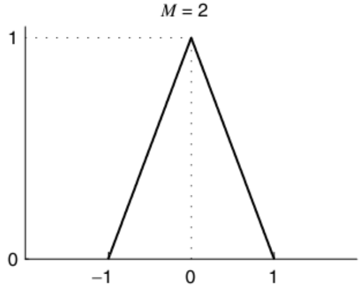
- If a random variable is a linear combination of M statistically independent random variables, the pdf and statistical descriptors are easy.

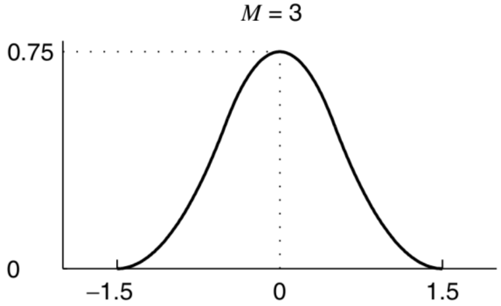
$$y = c_1 x_1 + c_2 x_2 + \cdots + c_M x_M$$

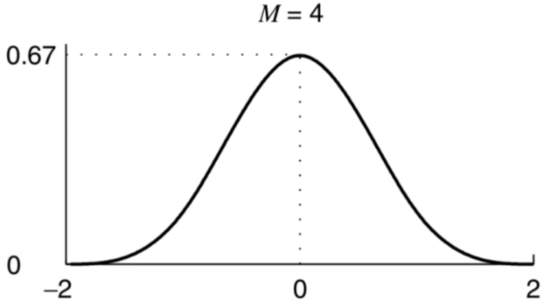
Mean	$\mu_y = \sum_{k=1}^{M} c_k \mu_{x\,k}$
Variance	$\sigma_y^2 = E\left\{\left\lvert\sum_{k=1}^{M} c_k\left[x_k - \mu_{x\,k}\right]\right\rvert^2\right\} = \sum_{k=1}^{M} \lvert c_k\rvert^2 \sigma_{x_k}^2$
Probability density function	$f_y(y) = f_{x_1}(y) * f_{x_2}(y) * \cdots * f_{x_M}(y)$

- Example: What is the pdf of y if its is the sum of four identical independent random variables uniformly distributed over [-0.5, 0.5].
- Solution:

U[-0.5, 0.5]*U[-0.5, 0.5]= f_{x12} f_{x12}*U[-0.5, 0.5]=f_{x123} f_{x123}*U[-0.5, 0.5]=f_{x1234}







Conditional Density

- Provides a measure of the degree of dependence of the variables on each other.
- From Bayes' rule, the joint pdf is given as

$$P(x_1, x_2) = P(x_1|x_2)P(x_2) = P(x_2|x_1)P(x_1)$$

$$P(x_1|x_2) = \frac{P(x_2|x_1)P(x_1)}{P(x_2)}$$

- If they are independent

$$P(x_1|x_2) = P(x_1)$$

Ensemble Averages

- A discrete-time random process is a sequence of random variables, x(n)
- The mean of the process, mean of each of these random variables may be calculated as

$$m_x(n) = E\{x(n)\}$$

- The variance is

$$\sigma_x^2(n) = E\{|x(n) - m_x(n)|^2\}$$

- These are ensemble averages.

- The autocorrelation of the process is

$$r_x(k,l) = E\{x(k)x^*(l)\}$$

- This provides the statistical relationship between the random variables x(k) and x(l).
- Wide-sense stationary
  - Mean of process is constant,
  - Autocorrelation dependent only on (k-l)
  - Variance is finite

Ergodicity

- The mean and autocorrelation of a random process are obtained from actual observed data instead of from probability density function.
- If a large number of observations is available

$$\hat{m}_x(n) = \frac{1}{L} \sum_{i=1}^{L} x_i(n)$$

- Since the sample mean is average of random variables, it is itself a random variable.

- If the ensemble statistic approaches the actual statistic, it is called unbiased estimator.

$$\lim_{N\to\infty} E\left\{\hat{m}_x(N)\right\} = m_x$$

- If the variance of the estimator is very small it is called consistent estimator.

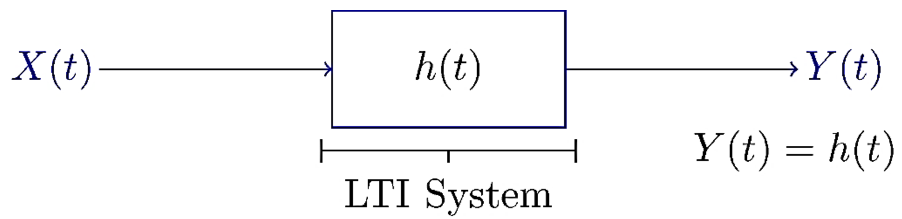
$$\lim_{N\to\infty} \mathrm{Var}\left\{\hat{m}_x(N)\right\} = 0$$

- If both are satisfied, it is ergodic to the mean.

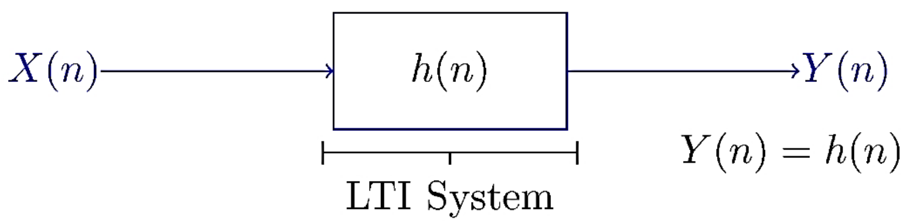
- This ergodicity principle may be generalized to other ensemble averages.

Random Processes through Linear Time-invariant Systems

- Consider a linear time-invariant system with impulse response h(t) driven by a random process input X(t)



$$Y(t) = h(t) * X(t)$$



$$Y(n) = h(n) * X(n)$$

- It is difficult to obtain a complete specification of the output process in general,
  - The input is known only probabilistically.
- The mean and autocorrelation of the output can be determined in terms of the mean and autocorrelation of the input.

- The mean of the output is

$$\mu_Y(t) = E[Y(t)] = E\left[\int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)\,d\alpha\right]$$
$$= \int_{-\infty}^{\infty} h(\alpha)E[X(t-\alpha)]\,d\alpha$$

- If the input is WSS

$$\mu_Y(t) = E[Y(t)] = \int_{-\infty}^{\infty} h(\alpha)\mu_X\,d\alpha$$
$$= \mu_X \int_{-\infty}^{\infty} h(\alpha)\,d\alpha.$$

- Note that mean of output is not function of time.

- The cross correlation between X and Y

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h(\alpha) R_X(\tau + \alpha)\, d\alpha$$
$$= h(\tau) * R_X(-\tau) = h(-\tau) * R_X(\tau)$$

- The autocorrelation of the output is

$$R_Y(\tau) = h(\tau) * h(-\tau) * R_X(\tau)$$

Power Spectrum

- The Fourier transform is important in the representation of random processes.
- Since random signals are only known probabilistically, it is not possible to compute the Fourier transform directly.
- For a wide-sense stationary random process, the autocorrelation is a deterministic function of time.

- The periodogram is an estimation of the power spectrum

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jk\omega}$$

- The autocorrelation sequence can be obtained from the periodogram

$$r_x(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})e^{jk\omega}d\omega$$

- Properties of the periodogram
  - It is real valued and symmetric, $P_x(e^{j\omega}) = P_x(e^{-j\omega})$
  - It is non-negative, $P_x(e^{j\omega}) \geq 0$
  - The power in a zero-mean WSS process is proportional to the area under the curve of the PSD

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

Spectral Factorization

- The power spectrum evaluated by the z-transform

$$P_x(z) = \sum_{k=-\infty}^{\infty} r_x(k) z^{-k}$$

- The power spectrum of a WSS process maybe factorized as

$$P_x(z) = \sigma_0^2 Q(z) Q^*(1/z^*)$$

- Q(z) is a minimum phase
  - All poles and zeros of Q(z) are inside the unit circule.

- From this representation

$$P_x(z) = \sigma_0^2 Q(z) Q^*(1/z^*)$$

  - Any regular random process may be realized as the output of a causal stable filter driven by white noise.
  - The inverse filter 1/Q(z) can be seen as a whitening filter.
  - The inverse filter retains all the information of x(n).

- For a rational P(z), the spectral factorization is

$$P_x(z) = \frac{N(z)}{D(z)}$$

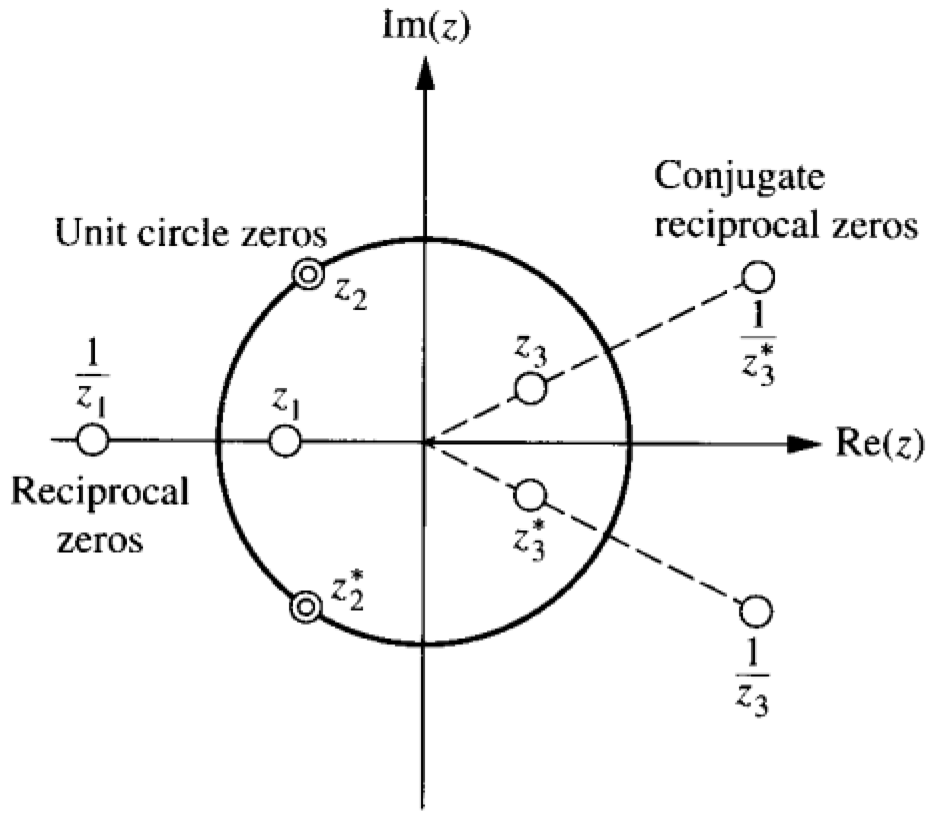
$$P_x(z) = \sigma_0^2 Q(z) Q^*(1/z^*) = \sigma_0^2 \left[\frac{B(z)}{A(z)}\right]\left[\frac{B^*(1/z^*)}{A^*(1/z^*)}\right]$$

- Where both A(z) and B(z) are polynomials with roots inside the unit circle

$$A(z) = 1 + a(1)z^{-1} + \cdots + a(p)z^{-p}$$

$$B(z) = 1 + b(1)z^{-1} + \cdots + b(q)z^{-q}$$

- This is due to the symmetric property of PSD.



Assignment 1

- 1.1 Show that if the m^{th} derivative of the moment generating function with respect to s evaluated at $s = 0$ results in the m^{th} moment.
- 1.2 Find the mean, variance, moments and moment generating functions of Uniform, Normal and Cauchy RV.
- 1.3 Show that a linear transformation of a normal random vector is also normal.
- 1.4 Find the spectral factorization of the following function.

$$R_x(z) = \frac{8}{5}\frac{(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)}{(1-\frac{4}{5}z^{-1})(1-\frac{4}{5}z)}$$

- 1.5 The input to a linear shift-invariant filter with unit sample response h(n) is a zero-mean wide-sense stationary processes with autocorrelation $r_x(k)$. Find the autocorrelation of the output processes for all k and its variance.

$$h(n) = \delta(n) - \frac{1}{3}\delta(n-1) + \frac{1}{4}\delta(n-2)$$

$$r_x(k) = \left(\frac{1}{2}\right)^{|k|}$$