3F1 Signals and Systems: Handout 15

The Power Spectral Density for Discrete Time Signals

Jossy Sayir

Michaelmas Term 2025

Lessons learned in the last lecture...

- random processes are **easy**: all the material that we need to learn in this course was summarised on one single slide
- random processes are **hard** to define and explain precisely because they involve "sloppy" limits and require one to illegally invert the order of limits and expectations

Reminder: what we will learn...

For stationary random processes,

		Discrete time	Continuous time
Time	Auto/cross correlation	$r_{XX}[k] = E[X_\ell X_{\ell+k}]$ $r_{XY}[k] = E[X_\ell Y_{\ell+k}]$	$r_{XX}(\tau) = E[X(t)X(t+\tau)]$ $r_{XY}(\tau) = E[X(t)Y(t+\tau)]$
Frequency	PSD, CSD	$\mathcal{S}_{XX}(\theta), \mathcal{S}_{XY}(\theta)$	$\mathcal{S}_{XX}(\omega), \mathcal{S}_{XY}(\omega)$
Linear filtering		$\begin{cases} \mathcal{S}_{YY}(\theta) = \lvert G(\theta)\rvert^2 \mathcal{S}_{XX}(\theta) \\ \mathcal{S}_{XY}(\theta) = G(\theta)\mathcal{S}_{XX}(\theta) \end{cases}$	$\begin{cases} \mathcal{S}_{YY}(\omega) = \lvert G(\omega)\rvert^2 \mathcal{S}_{XX}(\omega) \\ \mathcal{S}_{XY}(\omega) = G(\omega)\mathcal{S}_{XX}(\omega) \end{cases}$

- PSD: power spectral density, CSD: cross spectral density

Stationarity

- signal statistics are **shift-invariant**
- for any n and any collection $(t_1, t_2, \ldots, t_n)$ of n signal locations, and for any shift τ and for any values $(x_1, x_2, \ldots, x_n)$,

$$f_{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}}(x_1, x_2, \ldots, x_n) = f_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_1, x_2, \ldots, x_n)$$

- in other words there is no "absolute" time in the signal and its statistics only depend on the relative positions of the signal locations, not on its absolute locations
- this definition is the same for discrete and continuous random processes. For discrete random processes, the locations $(t_1, t_2, \ldots, t_n)$ and the shift τ are integers
- we described the statistics of the signal using densities, implicitly assuming that the random variabes involved are continuous, but in fact the same definition applies if dealing with discrete random variables, with densities replaced by probability mass functions

Stationarity: examples

- independent and identically distributed (i.i.d.) discrete-time random process $X_1, X_2, X_3, \ldots \rightarrow$ **stationary**
- "Markov" source: $P(x_k|x_1, x_2, \ldots, x_{k-1}) = P(x_k|x_{k-1})$ for all k and values $x_1, \ldots, x_k \rightarrow$ **stationary**
- "Random walk": $X_1, X_2, X_3, \ldots$ i.i.d. Gaussian random process, and let $Y_1 = X_1$ and $Y_k = Y_{k-1} + X_k$ for $k > 1$ $\rightarrow$ **not stationary: the variance grows with k**

Wide-sense stationarity

- A random process is wide-sense stationary if
  1. the mean $E[X_t]$ is independent of time
  2. the **autocorrelation** depends only on the relative interval

$$r_{XX}(t_1, t_2) = E[X_{t_1} X_{t_2}] = E[X_1 X_{t_2 - t_1 + 1}] = r_{XX}(t_2 - t_1)$$

- wide-sense stationarity is a "poor" stationarity: "I'm not sure the process is stationary because I don't know all its densities, but I am able to verify that the mean and the autocorrelation behave as they would if the process were stationary"
- a stationary process is also weak-sense stationary: shift invariant densities implies that the mean is time-invariant and the autocorrelation depends only on relative intervals
- it would be quite a coincidence if a real world process were weak-sense stationary but not stationary. It easy to construct a pathological example, however, e.g., a signal independently Gaussian $\mathcal{N}(0, 1)$ at even times and independently uniform over $\{-1, 1\}$ at odd times: $E[X_k] = 0$ for all k and $E[X_k X_m] = 0$ for $k \neq m$ and 1 for $k = m$.

Ergodicity

- a stationary random process is ergodic if its statistics observed over time converge to its true probabilities ("ensemble statistics")
- in particular, a stationary process is "mean ergodic" if

$$E[X] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} X_k$$

and "autocorrelation ergodic" if

$$r_{XX}(k) = E[X_1 X_{k+1}] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} X_0 X_k$$

Ergodicity (continued)

- "mean ergodic" and "autocorrelation ergodic" are to "ergodic" as "wide sense stationary" is to "stationary": "I could only verify ergodicity for the mean and autocorrelation and they behave as they would for an ergodic process."
- the definition is the same for continuous random processes, where sums are replaced by integrals
- a good way to understand ergodicity is to consider a pathologically non-ergodic stationary random process: $X_k = 0$ for all k with probability $1/2$ and $X_k = 1$ for all k with probability $1/2$.
- For this process, $E[X_k] = 1/2$ but time averaging will yield either 0 or 1 depending on which realisation ("sample path") of the random process is observed
- for the rest of this course, we will only consider stationary ergodic random processes

Discrete-time white "noise"

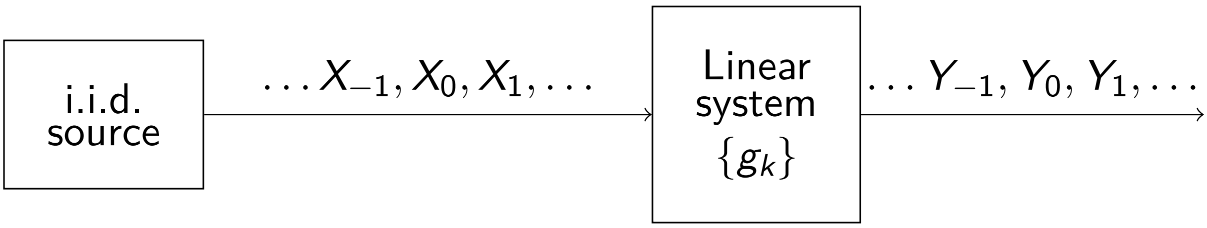
- the term **"white noise"** has been universally adopted to describe an **independent and identically distributed (i.i.d.)** discrete-time random process $\{\ldots, X_{-1}, X_0, X_1, X_2, \ldots\}$ with

$$f_{X_{k_1}, X_{k_2}, \ldots, X_{k_n}}(x_1, \ldots, x_n) = f_X(x_1) \cdot f_X(x_2) \cdots f_X(x_n)$$

  for any n, any $(k_1, \ldots, k_n)$ and any $(x_1, \ldots, x_n)$. White noise is **stationary and ergodic**
- the term is misleading. Take it as a synonym for "i.i.d."
- **white noise isn't necessarily "noise"** in the sense of "something disturbing that you want to get rid of". Those of you doing 3F7 know that data, when properly compressed, is uniform and identically distributed: **data is white noise!**
- **Noise isn't necessarily white!** The "Additive White Gaussian Noise" (AWGN) channel taught in 2P6 Comms is a simplification. Most real-world channels exhibit "echo" which results in non-independent ("coloured") noise.

Setup

- we consider the following scenario



- the system convolves its delta response with the input signal $\{Y_k\} = \{g_k\} \star \{X_k\}$
- the input is i.i.d. (white noise) with $E[X_k] = 0$ and $E[X_k^2] = 1$ (in fact, we only need independence for this derivation, not necessarily identical distributions)
- we observed last lecture that the expected DTFT of X is 0 everywhere
- we would like to analyse the **expected squared DTFT magnitude** of the signals X and Y

Approach

- to avoid taking "sloppy" limits, we will begin by considering a signal X_k consisting of i.i.d. random variables from time $\lfloor -N/2+1 \rfloor$ to time $\lfloor N/2 \rfloor$, and zero elsewhere
- its DTFT is

$$X(\theta) = \sum_{k=\lfloor -N/2+1 \rfloor}^{\lfloor N/2 \rfloor} X_k e^{-jk\theta}$$

- consider first the extreme case $N = 1$ of a signal that is a random variable with mean 0 and variance 1 at time 0 and 0 everywhere else. For this input,

$$X(\theta) = X_0 e^{-j0\theta} = X_0$$

and hence $E[|X(\theta)|^2] = E[X_0^2] = 1$ is constant for all θ

Derivation

- for other N, we compute

$$\begin{aligned}
\frac{1}{N} E\left[|X(\theta)|^2\right] &= \frac{1}{N} E\left[X(\theta) X^*(\theta)\right] \\
&= \frac{1}{N} E\left[\left(\sum_{k=\lfloor -N/2+1 \rfloor}^{\lfloor N/2 \rfloor} X_k e^{-jk\theta}\right)\left(\sum_{m=\lfloor -N/2+1 \rfloor}^{\lfloor N/2 \rfloor} X_m e^{jm\theta}\right)\right] \\
&= \frac{1}{N} \sum_{k=\lfloor -N/2+1 \rfloor}^{\lfloor N/2 \rfloor} \left(E[X_k^2] + \sum_{m \neq k} E[X_k X_m] e^{j(m-k)\theta}\right) \\
&= 1 \text{ for all } \theta
\end{aligned}$$

Power spectral density of white noise

- we've established that for any N, $\frac{1}{N}E[|X(\theta)|^2] = 1$ is a constant for all θ
- taking the limit as N goes to infinity is delicate (the definition of expectation doesn't extend to infinite dimensional vectors), but given that the result is the same for all N it's fairly unproblematic
- the **power spectral density** of white noise

$$\mathcal{S}_{XX}(\theta) = \lim_{N\to\infty} \frac{1}{N}|X(\theta)|^2 = 1$$

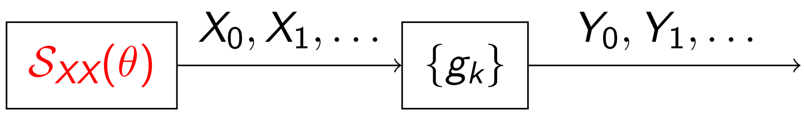
  is constant for all θ. In other words, all frequencies are present in equal measure in an i.i.d. signal.
- this is why it's called "white" noise, in analogy with white light that can be obtained by mixing all colours

Power spectral density of discrete random processes

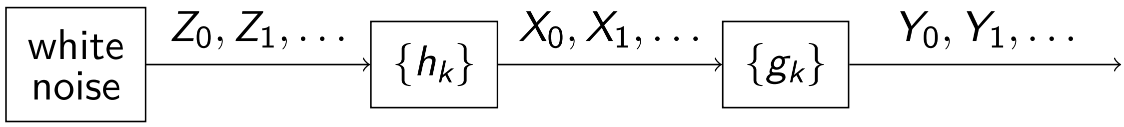
- the convolution property of the DTFT still applies for random signals
- hence, $Y(\theta) = G(\theta)X(\theta)$
- in consequence, the **power spectral density** of the output of the linear system is

$$\begin{aligned}\mathcal{S}_{YY}(\theta) &= \lim_{N\to\infty} \frac{1}{N}E[|Y(\theta)|^2] \\ &= \lim_{N\to\infty} \frac{1}{N}E[|G(\theta)X(\theta)|^2] = |G(\theta)|^2 \lim_{N\to\infty} \frac{1}{N}E[|X(\theta)|^2] \\ &= |G(\theta)|^2\mathcal{S}_{XX}(\theta) = |G(\theta)|^2\end{aligned}$$

Stationary random process through a linear system



- the "special case" of white noise through a linear filter gives the spectral analysis of any random process going through a linear system: if the input X is *not* white, consider it as if it were the output of a linear system with transfer function $H(\theta) = \sqrt{\mathcal{S}_{XX}(\theta)}$ and white noise Z at its input.



- hence this is the same as the white noise case, with filter $H(\theta)G(\theta)$ and therefore

$$\mathcal{S}_{YY}(\theta) = |(H(\theta)G(\theta))|^2 \mathcal{S}_{ZZ}(\theta) = |G(\theta)|^2 \mathcal{S}_{XX}(\theta)$$

Power spectral density vs. autocorrelation

- what is the time-domain equivalent of $\mathcal{S}_{XX}(\theta)$?
- remember that

$$\begin{aligned}
\frac{1}{N} E\left[|X(\theta)|^2\right] &= \frac{1}{N} E[X(\theta)X^*(\theta)] \\
&= \frac{1}{N} E\left[\left(\sum_k X_k e^{-jk\theta}\right)\left(\sum_m X_m e^{jm\theta}\right)\right] \\
&= \frac{1}{N} \sum_k \sum_m E[X_k X_m] e^{-jk\theta} e^{jm\theta} \quad (n = k - m) \\
&= \frac{1}{N} \sum_k \sum_n E[X_k X_{k-n}] e^{-jk\theta} e^{j(k-n)\theta} \\
&= \frac{1}{N} \sum_k \sum_n r_{XX}(n) e^{-jn\theta} = \sum_n r_{XX}(n) e^{-jn\theta}
\end{aligned}$$

Time domain relation

- the **power spectral density (PSD)** is the DTFT of the **autocorrelation** function (ACF)
- the PSD is real by definition and the ACF is symmetric $r_{XX}(n) = r_{XX}(-n)$
- hence, the relation that we derived in the frequency domain when filtering white noise through a linear system

$$\mathcal{S}_{YY}(\theta) = |G(\theta)|^2$$

becomes in the time domain

$$r_{YY}(n) = \{g_k\} * \{g_{-k}\}$$

Cross-correlation and cross-spectral-density

- another quantity of interest is the cross-correlation between input and output

$$r_{XY}(n) = E[X_k Y_{k+n}]$$

- Its DTFT is the cross-spectral density $\mathcal{S}_{XY}(\theta)$ for which

$$\mathcal{S}_{XY}(\theta) = G(\theta)\mathcal{S}_{XX}(\theta)$$

- power spectral density (PSD) and cross spectral density (CSD) can be used to estimate the response of an unknown system.
- A measurement using PSD alone yields only the magnitude square of the system response and cannot be used to estimate the phase and hence the delta response.
- A measurement using the CSD gives magnitude and phase but require the ability to synchronise the output signal Y against the input signal X, which is not always doable in practice