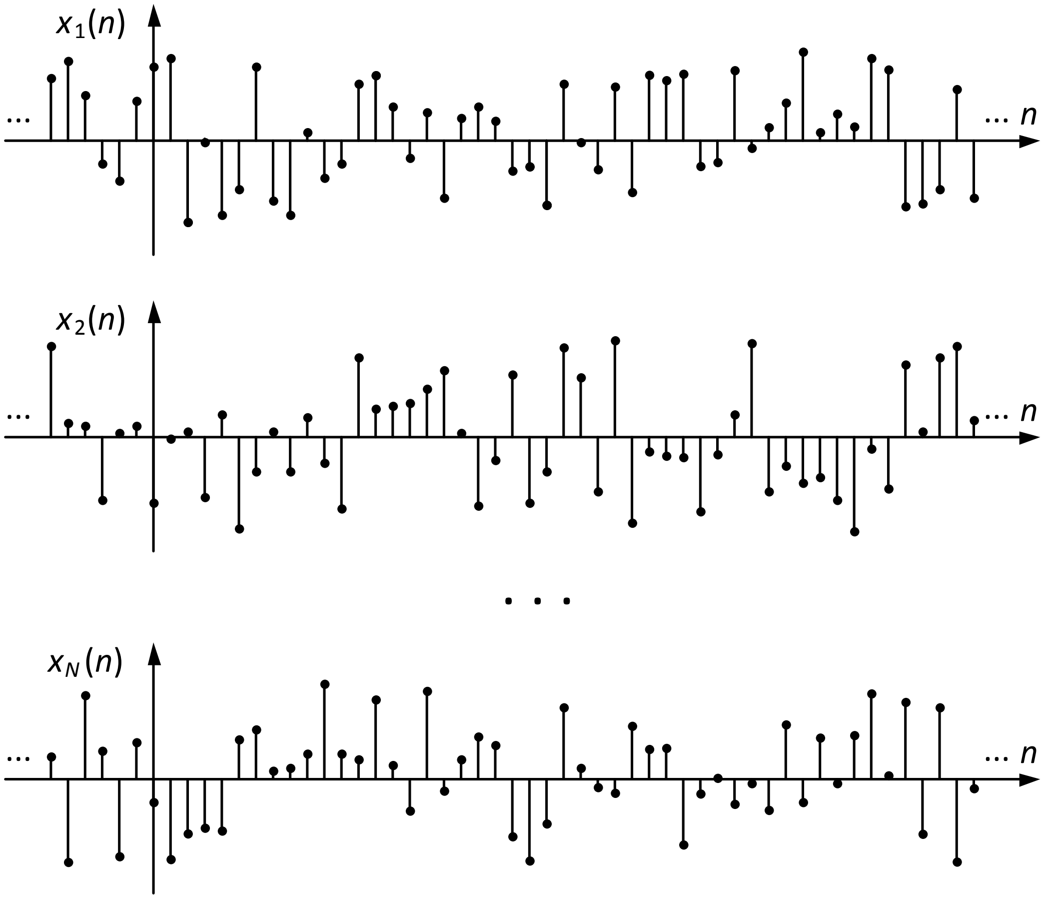
SPECTRAL ANALYSIS
OF DISCRETE-TIME SIGNALS

Random discrete signals

- Exact values are unknown outside of the range in which the signal is observed
- Described through statistical parameters
  - This approach is sensible for many real-world signals, which occur as a consequence of mutual interaction of a number a complex causes
- A random signal is one specific realization of a *random process*
  - A *discrete-time random process* X_n represents a time series of random variables

Statistical ensemble

- A set of all possible realizations of a random process
  - different signals but same statistical properties



Random variables and random processes

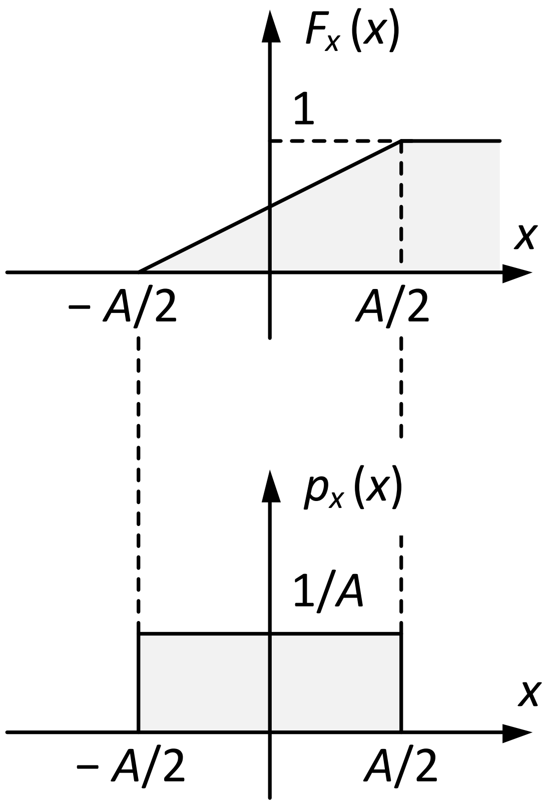
- A random variable can be described through its *cumulative probability distribution function* (CDF):

$$F_X(x) = \mathrm{P}(X \le x)$$

  and if it is continuous, it can also be described through its *probability density function* (PDF):

$$p_X(x) = \frac{dF_X(x)}{dx}$$

- In this way we can also describe any random process at a particular moment



Random processes

- A random process at a moment *n* can be described using a CDF dependent on *n*:

$$F_{X_n}(x,n) = \mathrm{P}(X_n \leq x)$$

  and if it is continuous, it can also be described using a PDF dependent on *n*:

$$p_{X_n}(x,n) = \frac{dF_{X_n}(x,n)}{dx}$$

- The values of X_n are generally mutually dependent, and to fully describe a random process we should know a joint CDF for any subset of moments:

$$F_{X_{n_1}X_{n_2}\dots X_{n_N}}(x_1,x_2,\dots x_N,n_1,n_2,\dots n_N) = \mathrm{P}(X_{n_1} \leq x_1, X_{n_2} \leq x_2,\dots X_{n_N} \leq x_N)$$

  - In practice this function is most often unknown, and we describe random processes using statistical parameters such as mean and autocorrelation function

Stationarity of a random process

- A random process is (*strict-sense*) *stationary* if its joint CDF does not change in time, i.e. if, for any integer *k*, the following holds:

$$F_{X_{n_1}X_{n_2}\dots X_{n_N}}(x_1, x_2, \dots x_N, n_1, n_2, \dots n_N)$$
$$= F_{X_{n_1+k}X_{n_2+k}\dots X_{n_N+k}}(x_1, x_2, \dots x_N, n_1+k, n_2+k, \dots n_N+k)$$

  - There are many random processes for which this does not hold, and which go through various *states*, in which their statistical properties may be quite different

Mean and autocorrelation function

- The *mean* of a random process is its mathematical expectation:

$$m_{X_n}(n) = E\{X_n\} = \int_{-\infty}^{\infty} x p_{X_n}(x,n)dx$$

  and if it does not depend on n, the process is *stationary with respect to its mean*

- The *autocorrelation function* of a random process is the measure of the similarity of its values at the time instants n and $n+k$:

$$r_{X_n X_{n+k}}(n,n+k) = E\{X_n X_{n+k}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_{X_n X_{n+k}}(x_1, x_2, n, n+k) dx_1 dx_2$$

  and if it depends solely on k, the process is *stationary with respect to its autocorrelation function*

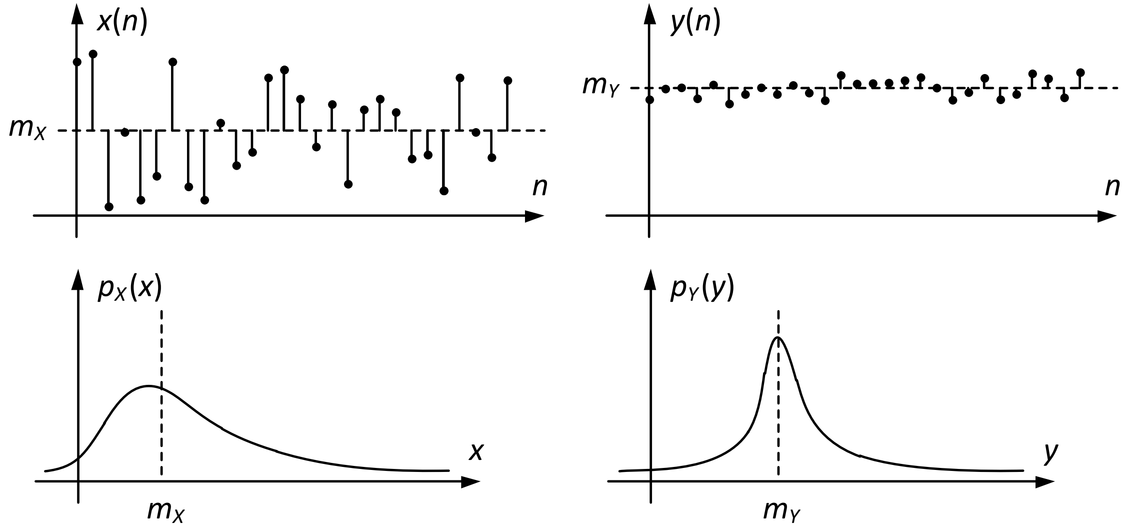
- A process stationary both in terms of its mean and its autocorrelation function is considered *wide-sense stationary*

Variance

- The variance of a wide-sense stationary (WSS) random process is the measure of its expected deviation from its mean:

$$\sigma_X^2 = E\left\{(X_n - m_X)^2\right\} = \int_{-\infty}^{\infty} (x - m_X)^2 p_{X_n}(x,n)dx$$

$\sigma_X^2 > \sigma_Y^2$



- If $m_X = 0$, the variance is equal to the *average power* of the signal

Estimation of statistical parameters

- We usually have only one member of the statistical ensemble $x(n)$ at our disposal, from which we should infer the values of the statistical parameters of a WSS random process
- The mean of the discrete signal $x(n)$ is:

$$\overline{x(n)} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)$$

and its autocorrelation is:

$$r_{xx}(k) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)x(n+k)$$

and if the process is *ergodic*, they are equal to the ensemble mean and autocorrelation function

- In practice we cannot calculate either of these parameters because we would need to know infinitely many values of signal samples and calculate infinite sums

Examples of parameter estimation

Mean estimation

- The standard estimate of the mean is:

$$\hat{m}_x = \frac{1}{N}\sum_{n=0}^{N-1} x(n)$$

- This estimate is *unbiased* (its expectation is equal to the actual mean m_X):

$$E\{\hat{m}_x\} = \frac{1}{N} E\left\{\sum_{n=0}^{N-1} X_n\right\} = \frac{1}{N}\sum_{n=0}^{N-1} E\{X_n\} = m_X$$

- This estimate is also *consistent* (variance goes to 0 when *N* goes to infinity)
  - Consistency is a highly desirable property of an estimate because with a consistent estimate we can expect the same result from any member of the ensemble

Examples of parameter estimation

Estimation of the autocorrelation function

- The standard estimate of the autocorrelation function is:

$$\hat{r}_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x(n)x(n+k)$$

- This estimate is *biased* (its expectation is not equal to the actual autocorrelation function $r_{XX}(k)$):

$$E\{\hat{r}_{xx}(k)\} = \frac{1}{N} \sum_{n=0}^{N-|k|-1} E\{X_n X_{n+k}\} = \frac{1}{N} \sum_{n=0}^{N-|k|-1} r_{XX}(k) = \left(1 - \frac{|k|}{N}\right) r_{XX}(k)$$

  but it is *asymptotically unbiased* (its expectation converges to $r_{XX}(k)$ as N approaches infinity)

- This estimate is *consistent*

Spectrum of a discrete-time random process

- Discrete-time random signals have infinite energy and thus have no DTFT, and we cannot calculate the energy spectral density (which is equal to the square of the magnitude spectrum)
- Discrete-time random signals have finite *average power*:

$$P_x = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x(n)|^2$$

and it is possible to define their *power spectral density* (*power spectrum*), by analogy with a corresponding definition for deterministic signals:

$$S_X(\omega) = \lim_{N\to\infty} E\left\{ \frac{1}{2N+1} \left| \sum_{n=-N}^{N} X_n e^{-j\omega n} \right|^2 \right\}$$

Spectrum of a discrete-time random process

- Power spectral density (PSD) is equal to the Fourier transform of the autocorrelation function:

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} r_{XX}(k) e^{-j\omega k}$$

  and it is a real, non-negative function, and in the case of real signals also even
- The estimation of power spectral density is one of the most frequent and most important tasks in digital signal processing
  - The problem is that we usually have only one member of the statistical ensemble at our disposal, and we usually know the values of only a limited number of its samples

Methods for power spectrum estimation

Non-parametric (classical) methods

- The spectrum is directly calculated by applying DTFT to either a finite-length segment of the input singal or its autocorrelation
- We do not introduce any assumptions as to the shape of the power spectrum

Parametric methods

- We introduce a *model* of power spectral density function and the problem amounts to the estimation of the parameters of this model
- If an adequate model is used, these methods can give much better results than non-parametric ones, particularly if only a short segment of the input signal is available

Non-parametric methods for estimating $S_X(\omega)$

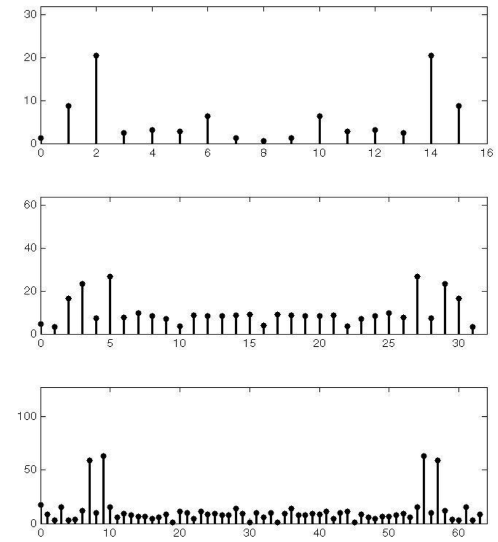
- The estimation of $S_X(\omega)$ is more accurate if a longer segment of the input signal is available (as was the case with deterministic signals)

Example: estimation of the power spectrum of a signal composed from two sinusoids contaminated with noise

$N = 16$

$N = 32$

$N = 64$



- If N is too small, we can increase the density of DFT samples by zero padding, although this does not affect frequency resolution
  - Zero padding, as a means for increasing the density of DFT samples is of particular importance for non-stationary signals

Periodogram

- Direct estimate based on the definition of $S_X(\omega)$, with the time range limited to the interval from 0 to $N\!-\!1$:

$$S_x(\omega) = \lim_{N\to\infty} E\left\{ \frac{1}{2N+1} \left| \sum_{n=-N}^{N} X_n e^{-j\omega n} \right|^2 \right\} \quad \rightarrow \quad \hat{S}_x(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j\omega n} \right|^2$$

which is a real, non-negative function, and in the case of real signals also even, just as was the case for $S_X(\omega)$

Periodogram

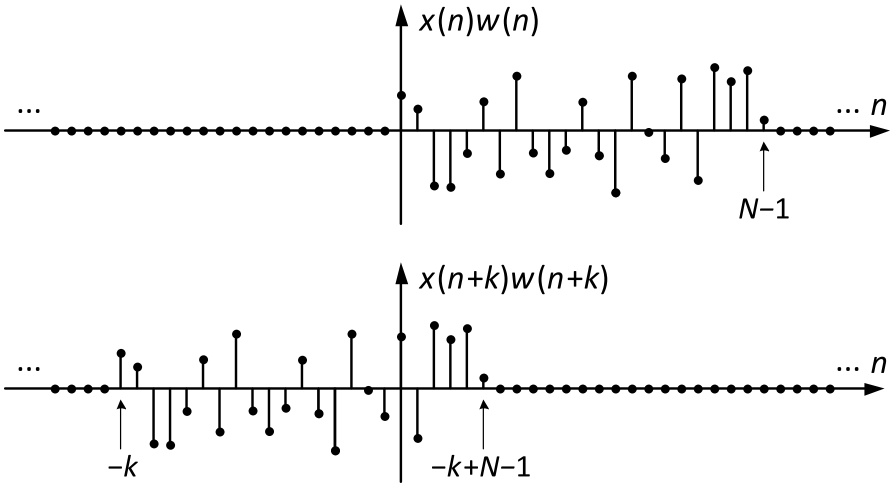
$$\hat{S}_x(\omega) = \frac{1}{N}\left|\overline{X}(\omega)\right|^2 = \frac{1}{N}\overline{X}(\omega)\overline{X}^*(\omega)$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-j\omega n}\sum_{m=0}^{N-1} x(m)e^{j\omega m}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} x(n)x(m)e^{-j\omega(m-n)}$$

$$= \sum_{k=-N+1}^{N-1}\left(\frac{1}{N}\sum_{n=0}^{N-|k|-1} x(n)x(n+k)\right)e^{-j\omega k}$$

$$= \sum_{k=-N+1}^{N-1}\hat{r}_{xx}(k)e^{-j\omega k}$$



- Periodogram represents the Fourier transform of the standard estimate of the autocorrelation function of a random process

Properties of the periodogram

$$E\left\{\hat{S}_x(\omega)\right\} = E\left\{\sum_{k=-N+1}^{N-1} \hat{r}_{xx}(k)e^{-j\omega k}\right\} = \sum_{k=-N+1}^{N-1} E\left\{\hat{r}_{xx}(k)\right\}e^{-j\omega k}$$

$$= \sum_{k=-N+1}^{N-1}\left(1-\frac{|k|}{N}\right)r_{XX}(k)e^{-j\omega k}$$

- The expectation of the periodogram $\hat{S}_X(\omega)$ is not equal to the actual power spectrum $S_X(\omega)$ = DTFT$\{r_{XX}(k)\}$, but to DTFT$\{r_{XX}(k)w_B(k)\}$, where $w_B(k)$ is the triangular window function
  - Periodogram is a *biased* estimate of $S_X(\omega)$, but still *asymptotically unbiased*
- When estimating $S_X(\omega)$ on the basis of a periodogram, the problem of *spectral leakage* occurs

$$E\left\{\hat{S}_x(\omega)\right\} = S_X(\omega) \circledast W_B(\omega) \qquad W_B(\omega) = \frac{1}{N}\left(\frac{\sin N\omega/2}{\sin\omega/2}\right)^2$$

Properties of the periodogram

- Variance of the periodogram $\hat{S}_X(\omega)$ for large values of N in most cases of practical interest is proportional to the square of $S_X(\omega)$

$$\lim_{N\to\infty} \mathrm{Var}\left\{\hat{S}_x(\omega)\right\} \sim S_X^2(\omega)$$

  - Periodogram is an *inconsistent* estimate of $S_X(\omega)$, which means that it can greatly depend on the realization of the random process
  - Due to its inconsistency, periodogram is not considered to be a high quality estimate of $S_X(\omega)$
  - An idea: to split the available segment of the input signal into K equal parts, to calculate the periodogram on each of them and to calculate the average of the results

Periodogram averaging

- Let the available segment of the input signal $x(n)$ of length N be divided into K equal parts of length $M = N/K$, with no overlap (basic version, proposed by Bartlett)
- Let $\hat{S}_x^{(i)}(\omega)$ be the periodogram calculated on the i-th segment:

$$\hat{S}_x^{(i)}(\omega) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_i(n)e^{-j\omega n}\right|^2$$

- The averaged periodogram is obtained as the average of thus obtained K periodograms:

$$\hat{S}_x^{B}(\omega) = \frac{1}{K}\sum_{i=0}^{K-1} \hat{S}_x^{(i)}(\omega)$$

Properties of the averaged periodogram

- Mathematical expectation of $\hat{S}_x^B(\omega)$ is equal to the mathematical expectation of a single periodogram $\hat{S}_x^{(i)}(\omega)$:

$$E\left\{\hat{S}_x^B(\omega)\right\} = \sum_{k=-M+1}^{M-1}\left(1-\frac{|k|}{M}\right)r_{XX}(k)e^{-j\omega k} = S_X(\omega) \circledast W_B(\omega)$$

  and since the triangular window function is now *K* times shorter, the frequency resolution is also *K* times worse
- The variance is reduced by a factor of *K* with respect to the standard periodogram (the variance of the mean of *K* identically distributed random variables is *K* times smaller than the variance of a single one of them):

$$\lim_{N\to\infty} \mathrm{Var}\left\{\hat{S}_x(\omega)\right\} \sim \frac{1}{K} S_X^2(\omega)$$

- By averaging in this way, we have improved the consistency of the estimate at the expense of the frequency resolution

Modifications of the averaged periodogram

- The basic idea of Bartlett can be modified in two ways (both proposed by Welch):
  - Consecutive segments can overlap
    - This increases the number of segments and decreases variance further
    - Overlap should not be too significant because that would undermine the assumption of independence between consecutive segments (if the overlap is above 50% the variance will abruptly stop decreasing)
  - Every segment $x_i(n)$ can be shaped by applying a non-rectangular window function before calculating the periodogram
    - This reduces the correlation between consecutive segments, and also mitigates the problem of lower reliability of the estimate of autocorrelation function for large k
- Without affecting the frequency resolution, the variance is now lower: for a large N it still behaves as $S_X^2(\omega)/K$, but K is now greater owing to the overlap

Periodogram smoothing

- The idea is based on windowing the estimate of the autocorrelation function (as proposed by Blackman and Tukey):
  - This reduces the influence of less reliable values of $\hat{r}_{xx}(k)$, obtained for large $|k|$

$$\hat{S}_x^{BT}(\omega) = \text{FTD}\left\{\hat{r}_{xx}(k)w(k)\right\} = \sum_{k=-K+1}^{K-1} \hat{r}_{xx}(k)w(k)e^{-j\omega k}$$

- Mathematical expectation of a thus smoothed periodogram is:

$$E\left\{\hat{S}_x^{BT}(\omega)\right\} = \text{FTD}\left\{E\left\{\hat{r}_{xx}(k)\right\}w(k)\right\} = \text{FTD}\left\{\hat{r}_{xx}(k)w_B(k)w(k)\right\}$$
$$= S_X(\omega) \circledast W_B(\omega) \circledast W(\omega) = E\left\{\hat{S}_x(\omega)\right\} \circledast W(\omega)$$

where $w(k)$ is usually much shorter than $w_B(k)$, and thus the influence of convolution with $W_B(\omega)$ is negligible

- The resulting spectral leakage depends on $W(\omega)$
  - The choice of $W(\omega)$ affects the trade-off between the variance and the frequency resolution

Parametric methods for estimating $S_X(\omega)$

- We introduce a *model* of the random process, and the problem of estimating $S_X(\omega)$ amounts to the problem of estimating model parameters
- We assume that the signal $x(n)$ is the output of an LTI system with transfer function:

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{q} b_i z^{-i}}{1 + \sum_{i=1}^{p} a_i z^{-i}}$$

where the input is white noise $s(n)$ of zero mean and variance (average power) equal to σ^2

  - The term *white* implies that the samples of noise are mutually uncorrelated, i.e.:

$$r_{SS}(k) = \sigma^2 \delta(k)$$

- This reduces the problem to the modeling of a *system*
- Power spectral density of the output signal is equal to: $S_X(\omega) = \sigma^2 \frac{|B(\omega)|^2}{|A(\omega)|^2}$

Parametric methods for estimating $S_X(\omega)$

- First we need to choose the values for p and q, and then to estimate a_i and b_i
  - q = 0: AR (autoregressive) models
  - p = 0: MA (*moving average*) models
  - $q \neq 0, p \neq 0$: ARMA models

$$\frac{X(z)}{S(z)} = \frac{\sum_{i=0}^{q} b_i z^{-i}}{1 + \sum_{i=1}^{p} a_i z^{-i}}$$

- The choice is usually based on some previous knowledge of the random process
  - Owing to their simplicity, AR models are most commonly used (the others yield non-linear systems of equations)
  - An example of a problem where AR is the obvious solution is the signal which is composed of a sinusoid of an unknown frequency and noise
    - Poles of the model should match the unknown frequency of the sinusoid

Estimation of $S_X(\omega)$ for AR random processes

$$x(n) = -\sum_{i=1}^{p} a_i x(n-i) + s(n)$$

$$E\{x(n)x(n-k)\} = -\sum_{i=1}^{p} a_i E\{x(n-i)x(n-k)\} + E\{s(n)x(n-k)\}$$

$$r_{XX}(k) = -\sum_{i=1}^{p} a_i r_{XX}(k-i) + E\left\{ s(n)\sum_{m=0}^{\infty} h(m)s(n-k-m) \right\}$$

$$= -\sum_{i=1}^{p} a_i r_{XX}(k-i) + \sum_{m=0}^{\infty} h(m)E\{s(n)s(n-k-m)\}$$

we assume $k \geq 0$, and the values for negative k can be easily found because $r_{XX}(k)$ is even

$$= -\sum_{i=1}^{p} a_i r_{XX}(k-i) + \sum_{m=0}^{\infty} h(m)\sigma^2\delta(m+k)$$

$$= \begin{cases} -\sum_{i=1}^{p} a_i r_{XX}(k-i), & k > 0 \\ -\sum_{i=1}^{p} a_i r_{XX}(k-i) + \sigma^2, & k = 0 \end{cases}$$

$h(m)$ exists only for $m \geq 0$, and thus for $k \geq 0$ the only case when the right sum is non-zero is when $k = m = 0$

we have used $h(0) = 1$, which follows from $B(z) = 1$

Estimation of $S_X(\omega)$ for AR random processes

- Coefficients a_i and σ^2 are the solutions of the Yule-Walker equations:

$$\begin{bmatrix} r_{XX}(0) & r_{XX}(-1) & \dots & r_{XX}(-p) \\ r_{XX}(1) & r_{XX}(0) & \dots & r_{XX}(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{XX}(p) & r_{XX}(p-1) & \dots & r_{XX}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- The system matrix is a symmetrical Toeplitz matrix (each descending diagonal from left to right is constant), and thus the system can be solved efficiently, in N^2 instead of N^3 time, using the Levinson-Durbin algorithm
  - The same matrix is obtained if the problem is posed as a *linear prediction* problem, i.e. if the following is assumed:

$$\hat{x}(n) = -\sum_{i=1}^{p} a_i x(n-i)$$

- In practice the autocorrelation function $r_{XX}(k)$ is not available and we only have its estimate, but the values of this estimate for $k \le p$ are still much more reliable than the values for relatively large k