3F3 - Random Processes, Optimal Filtering and Model-based Signal Processing

3F3 - Random Processes

February 3, 2015

Overview of course

This course extends the theory of 3F1 Random processes to Discrete-time random processes. It will make use of the discrete-time theory from 3F1. The main components of the course are:

- Section 1 (2 lectures) Discrete-time random processes
- Section 2 (3 lectures): Optimal filtering
- Section 3 (1 lectures): Signal Modelling and parameter estimation

Discrete-time random processes form the basis of most modern communications systems, digital signal processing systems and many other application areas, including speech and audio modelling for coding/noise-reduction/recognition, radar and sonar, stochastic control systems, ... As such, the topics we will study form a very vital core of knowledge for proceeding into any of these areas.

Section 1: Discrete-time Random Processes

- We define a discrete-time random process in a similar way to a continuous time random process, i.e. an ensemble of functions

$$\{X_n(\omega)\}, \quad n = -\infty \ldots, -1, 0, 1, \ldots, \infty$$

- ω is a random variable having a probability density function $f(\omega)$.
- Think of a *generative* model for the waveforms you might observe (or measure) in practice:
  1. First draw a random value $\tilde{\omega}$ from the density $f()$.
  2. The observed waveform for this value $\omega = \tilde{\omega}$ is given by

$$X_n(\tilde{\omega}), \quad n = -\infty \ldots, -1, 0, 1, \ldots, \infty$$

  3. The 'ensemble' is built up by considering all possible values $\tilde{\omega}$ (the 'sample space') and their corresponding time waveforms $X_n(\tilde{\omega})$.
  4. $f(\omega)$ determines the relative frequency (or probability) with which each waveform $X_n(\omega)$ can occur.
  5. Where no ambiguity can occur, ω is left out for notational simplicity, i.e. we refer to 'random process $\{X_n\}$'

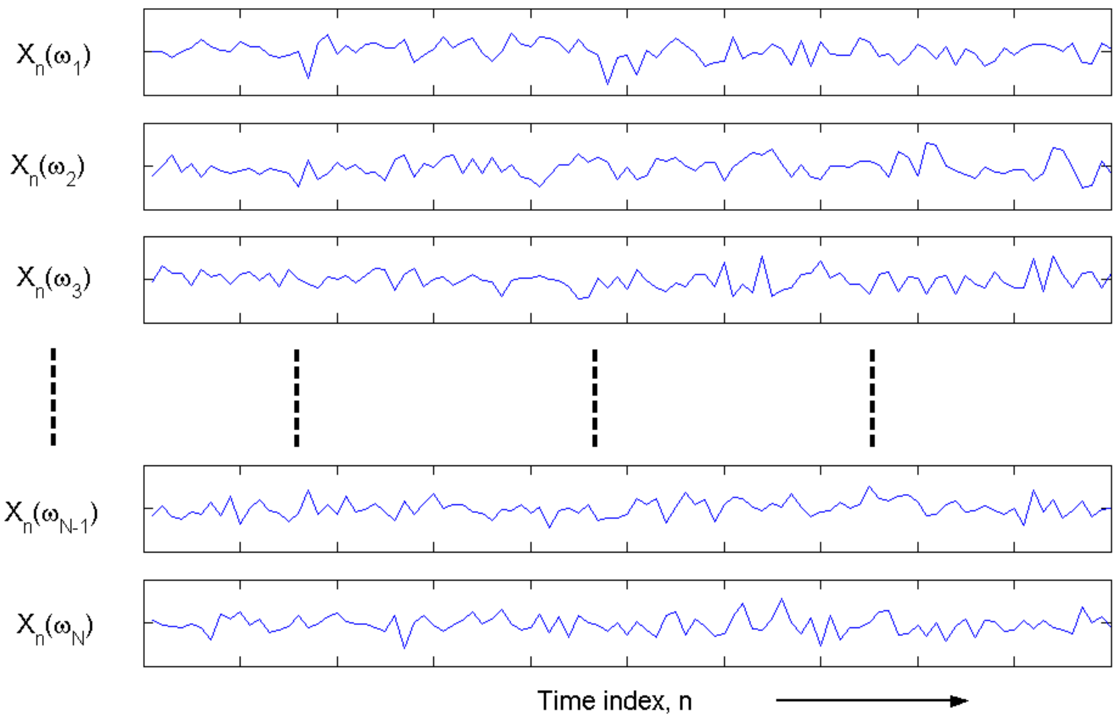


Figure 1: Ensemble representation of a discrete-time random process

Most of the results for continuous time random processes follow through almost directly to the discrete-time domain. A discrete-time random process (or 'time series') $\{X_n\}$ can often conveniently be thought of as a continuous-time random process $\{X(t)\}$ evaluated at times $t = nT$, where T is the sampling interval.

Example: the harmonic process

- The harmonic process is important in a number of applications, including radar, sonar, speech and audio modelling. An example of a real-valued harmonic process is the random phase sinusoid.
- Here the signals we wish to describe are in the form of sine-waves with known amplitude A and frequency Ω_0. The phase, however, is unknown and random, which could correspond to an unknown delay in a system, for example.
- We can express this as a random process in the following form:

$$x_n = A\sin(n\Omega_0 + \phi)$$

Here A and Ω_0 are fixed constants and ϕ is a random variable having a uniform probability distribution over the range $-\pi$ to $+\pi$:

$$f(\phi) = \begin{cases} 1/(2\pi) & -\pi < \phi \leq +\pi \\ 0, & \text{otherwise} \end{cases}$$

A selection of members of the ensemble is shown in the figure.

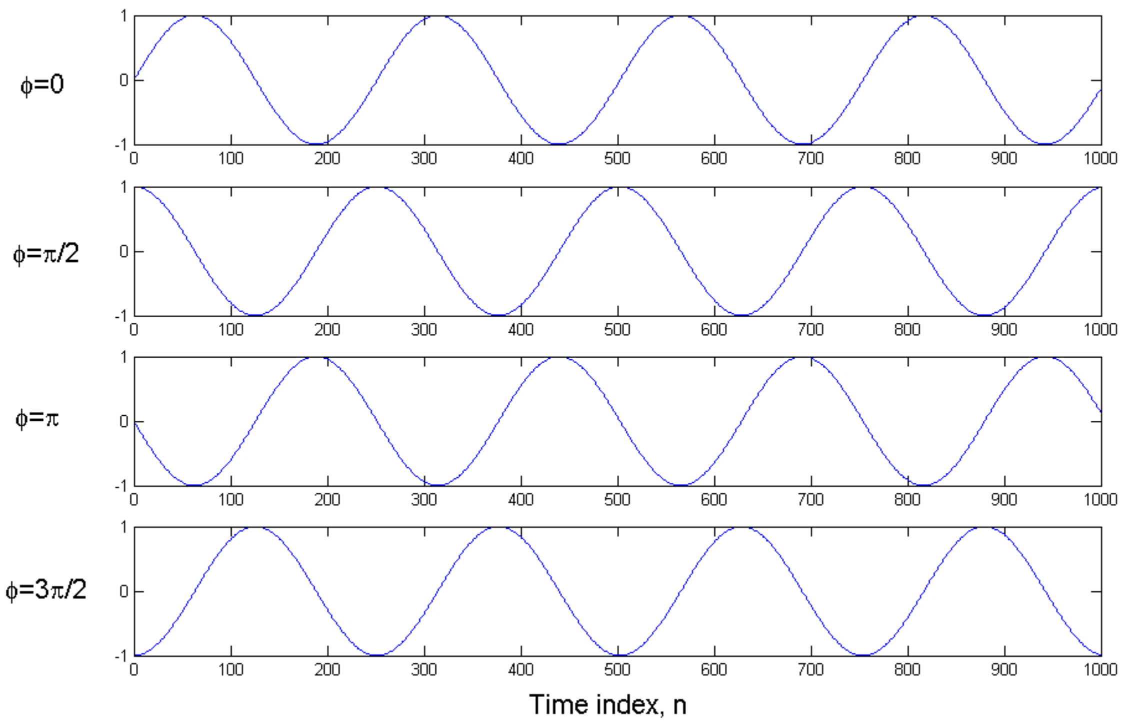


Figure 2: A few members of the random phase sine ensemble. $A = 1$, $\Omega_0 = 0.025$.

Correlation functions

- The mean of a random process $\{X_n\}$ is defined as $E[X_n]$ and the autocorrelation function as

$$\boxed{r_{XX}[n,m] = E[X_n X_m]}$$

Autocorrelation function of random process

- The cross-correlation function between two processes $\{X_n\}$ and $\{Y_n\}$ is:

$$\boxed{r_{XY}[n,m] = E[X_n Y_m]}$$

Cross-correlation function

Stationarity

A stationary process has the same statistical characteristics *irrespective of shifts along the time axis*. To put it another way, an observer looking at the process from sampling time n_1 would not be able to tell the difference in the *statistical* characteristics of the process if he moved to a different time n_2. This idea is formalised by considering the Nth order density for the process:

$$f_{X_{n_1}, X_{n_2}, \ldots, X_{n_N}}(x_{n_1}, x_{n_2}, \ldots, x_{n_N})$$

Nth order density function for a discrete-time random process

which is the joint probability density function for N arbitrarily chosen time indices $\{n_1, n_2, \ldots n_N\}$. Since the probability distribution of a random vector contains all the statistical information about that random vector, we should expect the probability distribution to be unchanged if we shifted the time axis any amount to the left or the right, for a stationary signal. This is the idea behind *strict-sense stationarity* for a discrete random process.

A random process is strict-sense stationary if, *for any finite* c, N *and* $\{n_1,\ n_2,\ \ldots\ n_N\}$:

$$f_{X_{n_1}, X_{n_2}, \ldots, X_{n_N}}(\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_N) = f_{X_{n_1+c}, X_{n_2+c}, \ldots, X_{n_N+c}}(\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_N)$$

Strict-sense stationarity for a random process

Strict-sense stationarity is hard to prove for most systems. In this course we will typically use a less stringent condition which is nevertheless very useful for practical analysis. This is known as *wide-sense stationarity*, which only requires first and second order moments (i.e. mean and autocorrelation function) to be invariant to time shifts.

A random process is *wide-sense stationary (WSS)* if:

1. $\mu_n = E[X_n] = \mu$, (mean is constant)
2. $r_{XX}[n, m] = r_{XX}[m - n]$, (autocorrelation function depends only upon the difference between n and m).
3. The variance of the process is finite:

$$E[(X_n - \mu)^2] < \infty$$

Wide-sense stationarity for a random process

Note that strict-sense stationarity *plus finite variance (condition 3)* implies wide-sense stationarity, but not *vice versa*.

Example: random phase sine-wave

Continuing with the same example, we can calculate the mean and autocorrelation functions and hence check for stationarity. The process was defined as:

$$x_n = A\sin(n\Omega_0 + \phi)$$

A and Ω_0 are fixed constants and ϕ is a random variable having a uniform probability distribution over the range $-\pi$ to $+\pi$:

$$f(\phi) = \begin{cases} 1/(2\pi) & -\pi < \phi \leq +\pi \\ 0, & \text{otherwise} \end{cases}$$

1. **Mean:**

$$
\begin{aligned}
E[X_n] &= E[A\sin(n\Omega_0+\phi)] \\
&= A\,E[\sin(n\Omega_0+\phi)] \\
&= A\,\{E[\sin(n\Omega_0)\cos(\phi)+\cos(n\Omega_0)\sin(\phi)]\} \\
&= A\,\{\sin(n\Omega_0)E[\cos(\phi)]+\cos(n\Omega_0)E[\sin(\phi)]\} \\
&= 0
\end{aligned}
$$

since $E[\cos(\phi)] = E[\sin(\phi)] = 0$ under the assumed uniform distribution $f(\phi)$.

2. **Autocorrelation:**

$$
\begin{aligned}
& r_{XX}[n,m] \\
& = E[X_n, X_m] \\
& = E[A\sin(n\Omega_0+\phi).A\sin(m\Omega_0+\phi)] \\
& = 0.5A^2\left\{E\left[\cos[(n-m)\Omega_0]-\cos[(n+m)\Omega_0+2\phi]\right]\right\} \\
& = 0.5A^2\{\cos[(n-m)\Omega_0]-E\left[\cos[(n+m)\Omega_0+2\phi]\right]\} \\
& = 0.5A^2\cos[(n-m)\Omega_0]
\end{aligned}
$$

Hence the process satisfies the three criteria for wide sense stationarity.

Power spectra

For a wide-sense stationary random process $\{X_n\}$, the power spectrum is defined as the discrete-time Fourier transform (DTFT) of the discrete autocorrelation function:

$$\mathcal{S}_X(e^{j\Omega}) = \sum_{m=-\infty}^{\infty} r_{XX}[m]\, e^{-jm\Omega} \tag{1}$$

Power spectrum for a random process

where $\Omega = \omega T$ is used for convenience.
The autocorrelation function can thus be found from the power spectrum by inverting the transform using the inverse DTFT:

$$r_{XX}[m] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{S}_X(e^{j\Omega})\, e^{jm\Omega}\, d\Omega \tag{2}$$

Autocorrelation function from power spectrum

- The power spectrum is a real, positive, even and periodic function of frequency.
- The power spectrum can be interpreted as a density spectrum in the sense that the mean-squared signal value at the output of an ideal band-pass filter with lower and upper cut-off frequencies of ω_l and ω_u is given by

$$\frac{1}{\pi}\int_{\omega_l T}^{\omega_u T} \mathcal{S}_X(e^{j\Omega})\, d\Omega$$

Here we have assumed that the signal and the filter are real and hence we add together the powers at negative and positive frequencies.

EXAMPLE: POWER SPECTRUM

The autocorrelation function for the random phase sine-wave was previously obtained as:

$$r_{XX}[m] = 0.5A^2 \cos[m\Omega_0]$$

Hence the power spectrum is obtained as:

$$\begin{aligned}
\mathcal{S}_X(e^{j\Omega}) &= \sum_{m=-\infty}^{\infty} r_{XX}[m]\, e^{-jm\Omega} \\
&= \sum_{m=-\infty}^{\infty} 0.5A^2 \cos[m\Omega_0]\, e^{-jm\Omega} \\
&= 0.25A^2 \\
&\quad \times \sum_{m=-\infty}^{\infty} (\exp(jm\Omega_0) + \exp(-jm\Omega_0))\, e^{-jm\Omega} \\
&= 0.5\pi A^2 \\
&\quad \times \sum_{m=-\infty}^{\infty} \delta(\Omega - \Omega_0 - 2m\pi) \\
&\qquad\qquad + \delta(\Omega + \Omega_0 - 2m\pi)
\end{aligned}$$

where $\Omega = \omega T$ is once again used for shorthand.

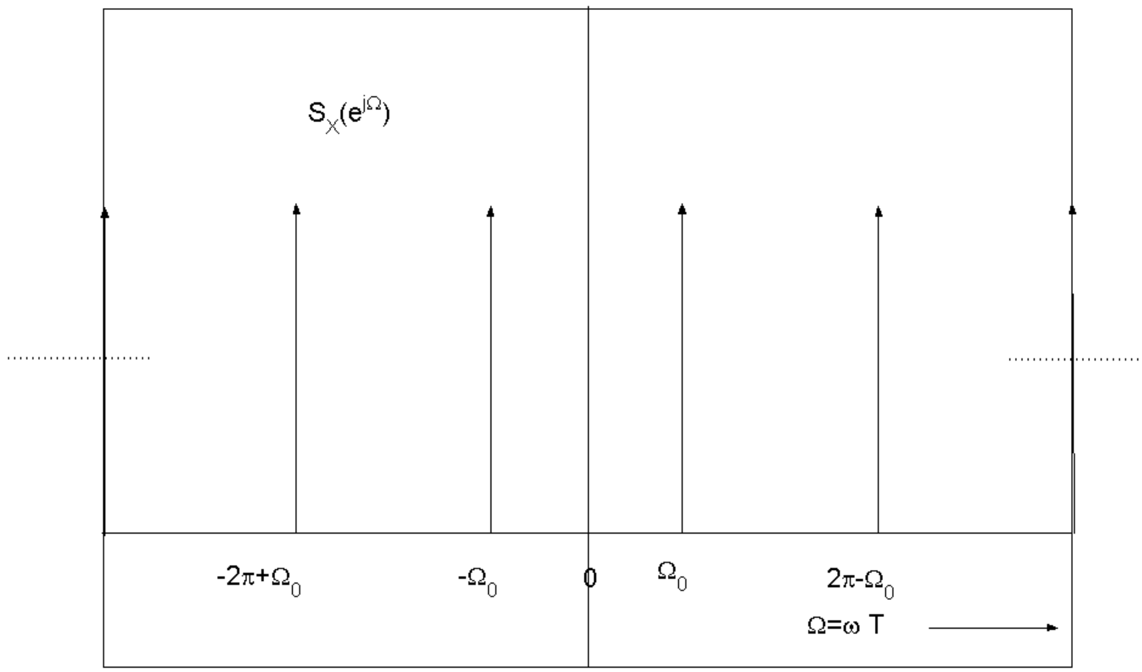


Figure 3: Power spectrum of harmonic process

White noise

White noise is defined in terms of its auto-covariance function.
A wide sense stationary process is termed white noise if:

$$c_{XX}[m] = E[(X_n - \mu)(X_{n+m} - \mu)] = \sigma_X^2 \delta[m]$$

where $\delta[m]$ is the discrete impulse function:

$$\delta[m] = \begin{cases} 1, & m = 0 \\ 0, & \text{otherwise} \end{cases}$$

$\sigma_X^2 = E[(X_n - \mu)^2]$ is the variance of the process. If $\mu = 0$ then σ_X^2 is the *mean-squared* value of the process, which we will sometimes refer to as the 'power'.

The power spectrum of zero mean white noise is:

$$\begin{aligned} \mathcal{S}_X(e^{j\omega T}) &= \sum_{m=-\infty}^{\infty} r_{XX}[m]\, e^{-jm\Omega} \\ &= \sigma_X^2 \end{aligned}$$

i.e. flat across all frequencies.

Example: white Gaussian noise (WGN)

There are many ways to generate white noise processes, all having the property

$$c_{XX}[m] = E[(X_n - \mu)(X_{n+m} - \mu)] = \sigma_X^2 \delta[m]$$

The ensemble illustrated earlier in Fig. was the *zero-mean Gaussian white noise* process. In this process, the values X_n are drawn *independently* from a Gaussian distribution with mean 0 and variance σ_X^2.
The N^{th} order pdf for the Gaussian white noise process is:

$$f_{X_{n_1}, X_{n_2}, \ldots, X_{n_N}}(\alpha_1, \, \alpha_2, \, \ldots, \, \alpha_N) = \prod_{i=1}^{N} \mathcal{N}(\alpha_i | 0, \sigma_X^2)$$

where

$$\mathcal{N}(\alpha | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2 \right)$$

is the univariate normal pdf.

We can immediately see that the Gaussian white noise process is *Strict sense stationary*, since:

$$
\begin{aligned}
&f_{X_{n_1}, X_{n_2}, \ldots, X_{n_N}}(\alpha_1, \alpha_2, \ldots, \alpha_N) \\
&\quad = \prod_{i=1}^{N} \mathcal{N}(\alpha_i | 0, \sigma_X^2) \\
&\quad\quad = f_{X_{n_1+c}, X_{n_2+c}, \ldots, X_{n_N+c}}(\alpha_1, \alpha_2, \ldots, \alpha_N)
\end{aligned}
$$

Linear systems and random processes

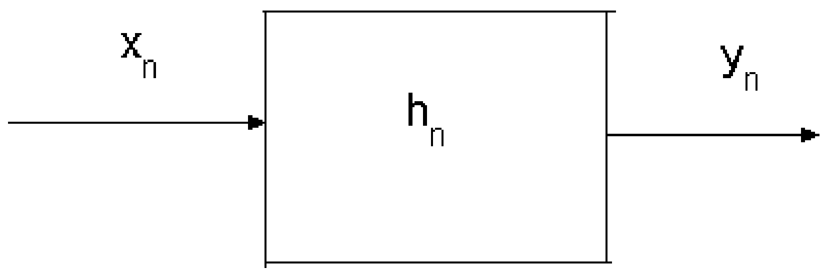


Figure 4: Linear system

When a wide-sense stationary discrete random process $\{X_n\}$ is passed through a *stable*, linear time invariant (LTI) system with digital impulse response $\{h_n\}$, the output process $\{Y_n\}$, i.e.

$$y_n = \sum_{k=-\infty}^{+\infty} h_k \, x_{n-k} = x_n * h_n$$

is also wide-sense stationary.

We can express the output correlation functions and power spectra in terms of the input statistics and the LTI system:

$$r_{XY}[k] = E[X_n\, Y_{n+k}] = \sum_{l=-\infty}^{\infty} h_l\, r_{XX}[k-l] = h_k * r_{XX}[k] \qquad (3)$$

Cross-correlation function at the output of a LTI system

$$r_{YY}[l] = E[Y_n\, Y_{n+l}]$$

$$\sum_{k=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} h_k\, h_i\, r_{XX}[l+i-k] = h_l * h_{-l} * r_{XX}[l] \qquad (4)$$

Autocorrelation function at the output of a LTI system

Note: these are convolutions, as in the continuous-time case. This is easily remembered through the following figure:

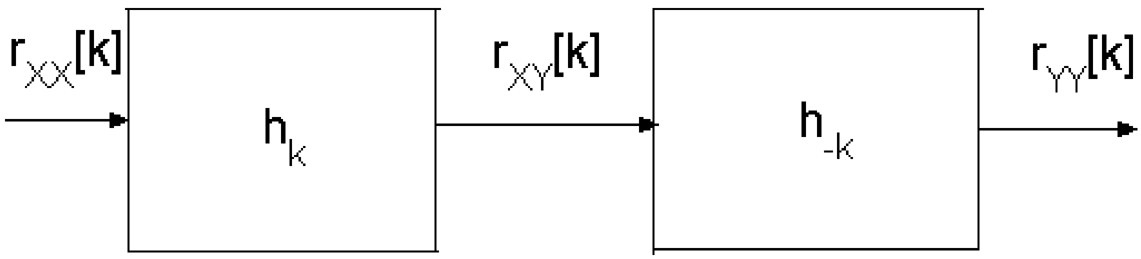


Figure 5: Linear system - correlation functions

Taking DTFT of both sides of (4):

$$\boxed{\mathcal{S}_Y(e^{j\omega T}) = |H(e^{j\omega T})|^2 \mathcal{S}_X(e^{j\omega T})} \qquad (5)$$

Power spectrum at the output of a LTI system

EXAMPLE: FILTERING WHITE NOISE

Suppose we filter a zero mean white noise process $\{X_n\}$ with a first order *finite impulse response (FIR) filter*:

$$y_n = \sum_{m=0}^{1} b_m \, x_{n-m}, \qquad \text{or} \qquad Y(z) = (b_0 + b_1 z^{-1}) X(z)$$

with $b_0 = 1$, $b_1 = 0.9$. This an example of a *moving average (MA)* process.

The impulse response of this causal filter is:

$$\{h_n\} = \{b_0, \, b_1, \, 0, \, 0, \, ...\}$$

The autocorrelation function of $\{Y_n\}$ is obtained as:

$$r_{YY}[l] = E[Y_n \, Y_{n+l}] = h_l * h_{-l} * r_{XX}[l] \qquad (6)$$

This convolution can be performed directly. However, it is more straightforward in the frequency domain.

The frequency response of the filter is:

$$H(e^{j\Omega}) = b_0 + b_1 e^{-j\Omega}$$

The power spectrum of $\{X_n\}$ (white noise) is:

$$\mathcal{S}_X(e^{j\Omega}) = \sigma_X^2$$

Hence the power spectrum of $\{Y_n\}$ is:

$$\begin{aligned}\mathcal{S}_Y(e^{j\Omega}) &= |H(e^{j\Omega})|^2 \mathcal{S}_X(e^{j\Omega}) \\ &= |b_0 + b_1 e^{-j\Omega}|^2 \sigma_X^2 \\ &= (b_0 b_1 e^{+j\Omega} + (b_0^2 + b_1^2) + b_0 b_1 e^{-j\Omega})\sigma_X^2\end{aligned}$$

as shown in the figure overleaf. Comparing this expression with the DTFT of $r_{YY}[m]$:

$$\mathcal{S}_Y(e^{j\Omega}) = \sum_{m=-\infty}^{\infty} r_{YY}[m]\, e^{-jm\Omega}$$

we can identify non-zero terms in the summation only when $m = -1, 0, +1$, as follows:

$$r_{YY}[-1] = \sigma_X^2 b_0 b_1, \quad r_{YY}[0] = \sigma_X^2 (b_0^2 + b_1^2)$$
$$r_{YY}[1] = \sigma_X^2 b_0 b_1$$

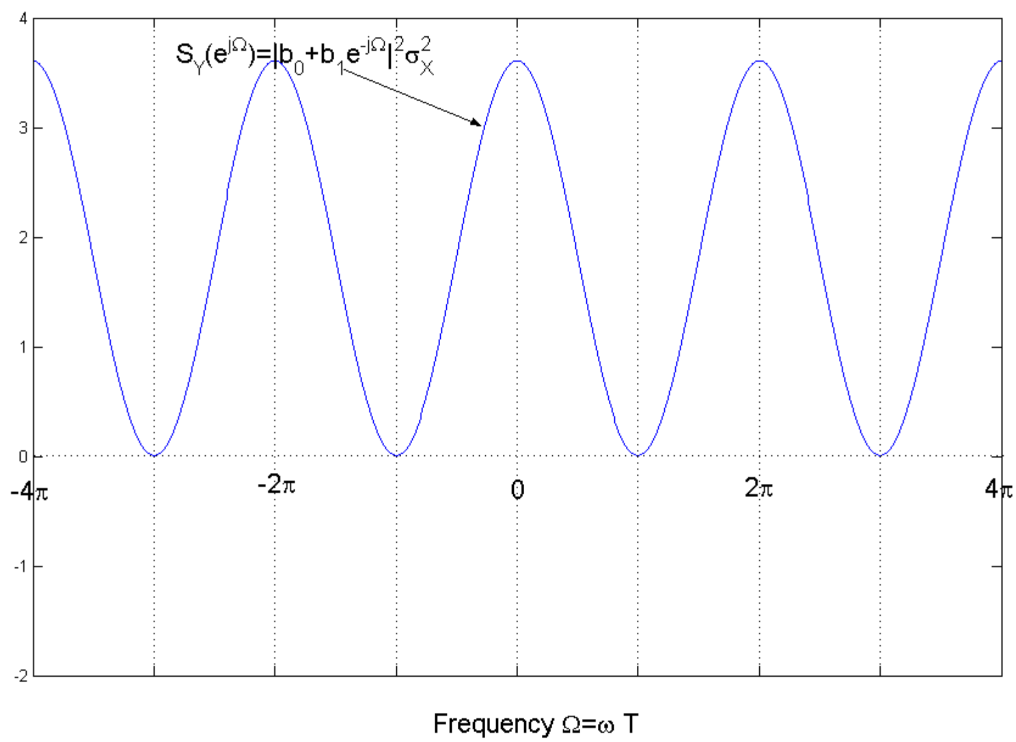


Figure 6: Power spectrum of filtered white noise

Ergodic Random processes

- For an **Ergodic** random process we can estimate expectations by performing time-averaging on a single sample function, e.g.

$$\mu = E[X_n] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} x_n \quad \text{(Mean ergodic)}$$

$$r_{XX}[k] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} x_n x_{n+k} \quad \text{(Correlation ergodic)} \tag{7}$$

- As in the continuous-time case, these formulae allow us to make the following estimates, for 'sufficiently' large N:

$$\mu = E[X_n] \approx \frac{1}{N} \sum_{n=0}^{N-1} x_n \quad \text{(Mean ergodic)}$$

$$r_{XX}[k] \approx \frac{1}{N} \sum_{n=0}^{N-1} x_n x_{n+k} \quad \text{(Correlation ergodic)} \tag{8}$$

  Note, however, that this is implemented with a simple computer code loop in discrete-time, unlike the continuous-time case which requires an approximate integrator circuit.

Under what conditions is a random process ergodic?

- It is hard in general to determine whether a given process is ergodic.
- A necessary and sufficient condition for *mean* ergodicity is given by:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} c_{XX}[k] = 0$$

  where c_{XX} is the *autocovariance* function:

$$c_{XX}[k] = E[(X_n - \mu)(X_{n+k} - \mu)]$$

  and $\mu = E[X_n]$.
- A simpler sufficient condition for mean ergodiciy is that $c_{XX}[0] < \infty$ and

$$\lim_{N\to\infty} c_{XX}[N] = 0$$

- Correlation ergodicity can be studied by extensions of the above theorems. We will not require the details here.
- Unless otherwise stated, we will always assume that the signals we encounter are both wide-sense stationary and ergodic. Although neither of these will always be true, it will usually be acceptable to assume so in practice.

EXAMPLE

Consider the very simple 'd.c. level' random process

$$X_n = A$$

where A is a random variable having the *standard* (i.e. mean zero, variance=1) Gaussian distribution

$$f(A) = \mathcal{N}(A|0, 1)$$

The mean of the random process is:

$$E[X_n] = \int_{-\infty}^{\infty} x_n(a) f(a) da = \int_{-\infty}^{\infty} a f(a) da = 0$$

Now, consider a random sample function measured from the random process, say

$$x_t = a_0$$

The mean value of this particular sample function is $E[a_0] = a_0$. Since in general $a_0 \neq 0$, the process is clearly *not* mean ergodic.

Check this using the *mean ergodic* theorem. The autocovariance function is:

$$\begin{aligned} c_{XX}[k] &= E[(X_n - \mu)(X_{n+k} - \mu)] \\ &= E[X_n X_{n+k}] = E[A^2] = 1 \end{aligned}$$

Now

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} c_{XX}[k] = (1 \times N)/N = 1 \neq 0$$

Hence the theorem confirms our finding that the process is not ergodic in the mean.

While this example highlights a possible pitfall in assuming ergodicity, most of the processes we deal with will, however, be ergodic, see examples paper for the random phase sine-wave and the white noise processes.

Comment: Complex-valued processes.

The above theory is easily extended to complex valued processes $\{X_n = X_n^{Re} + jX_n^{Im}\}$, in which case the autocorrelation function is defined as:

$$r_{XX}[k] = E[X_n^* X_{n+k}]$$

Revision: Continuous time random processes

Figure 7: Ensemble representation of a random process

- A random process is an *ensemble* of functions $\{X(t), \omega\}$, representing the set of all possible waveforms that we might observe for that process. [N.B. ω is identical to the random variable α in the 3F1 notes].

  ω is a random variable with its own probability distribution P_ω which determines randomly which waveform $X(t, \omega)$ is observed. ω may be continuous- or discrete-valued.

  As before, we will often omit ω and refer simply to random process $\{X(t)\}$.

- The mean of a random process is defined as $\mu(t) = E[X(t)]$ and the autocorrelation function as $r_{XX}[t_1, t_2] = E[X(t_1)X(t_2)]$. The properties of expectation allow us to calculate these in (at least) two ways:

$$\begin{aligned}
\mu(t) &= E[X(t_1)] \\
&= \int_x x\, f_{X(t)}(x)\, dx \\
&= \int_\omega X(t, \omega)\, f_\omega(\omega) d\omega
\end{aligned}$$

$$
\begin{aligned}
r_{XX}(t_1, t_2) &= E[X(t_1)X(t_2)] \\
&= \int_{x_1}\int_{x_2} x_1 x_2 \, f_{X(t_1),X(t_2)}(x_1, x_2)\, dx_1 dx_2 \\
&= \int_{\omega} X(t_1, \omega) X(t_2, \omega) f_{\omega}(\omega) d\omega
\end{aligned}
$$

i.e. directly in terms of the density functions for $X(t)$ or indirectly in terms of the density function for ω.

- A *wide-sense stationary (WSS)* process $\{X(t)\}$ is defined such that its mean is a constant, and $r_{XX}(t_1, t_2)$ depends only upon the difference $\tau = t_2 - t_1$, i.e.

$$E[X(t)] = \mu, \qquad r_{XX}(\tau) = E[X(t)X(t+\tau)] \tag{9}$$

  [Wide-sense stationarity is also referred to as 'weak stationarity']

- The *Power Spectrum* or *Spectral Density* of a WSS random process is defined as the Fourier Transform of $r_{XX}(\tau)$,

$$S_X(\omega) = \int_{-\infty}^{\infty} r_{XX}(\tau)\, e^{-j\omega\tau}\, d\tau \tag{10}$$

- For an **Ergodic** random process we can estimate expectations by performing time-averaging on a single sample function, e.g.

$$\mu = E[X(t)] = \lim_{D\to\infty} \frac{1}{2D} \int_{-D}^{+D} x(t) dt$$

$$r_{XX}(\tau) = \lim_{D\to\infty} \frac{1}{2D} \int_{-D}^{+D} x(t) x(t+\tau) dt$$

Therefore, for ergodic random processes we can make estimates for these quantities by integrating over some suitably large (but finite) time interval $2D$, e.g.:

$$E[X(t)] \approx \frac{1}{2D} \int_{-D}^{+D} x(t)dt$$

$$r_{XX}(\tau) \approx \frac{1}{2D} \int_{-D}^{+D} x(t)x(t+\tau)dt$$

where $x(t)$ is a waveform measured at random from the process.

Section 2: Optimal Filtering

Parts of this section and Section 3. are adapted from material kindly supplied by Prof. Peter Rayner.

- Optimal filtering is an area in which we design filters that are optimally adapted to the statistical characteristics of a random process. As such the area can be seen as a combination of standard filter design for deterministic signals with the random process theory of the previous section.
- This remarkable area was pioneered in the 1940's by Norbert Wiener, who designed methods for optimal estimation of a signal measured in noise. Specifically, consider the system in the figure below.

Figure 8: The general Wiener filtering problem

- A desired signal d_n is observed in noise v_n:

$$x_n = d_n + v_n$$

- Wiener showed how to design a linear filter which would optimally estimate d_n given just the noisy observations x_n and some assumptions about the statistics of the random signal and noise processes. This class of filters, the *Wiener filter*, forms the basis of many fundamental signal processing applications.
- Typical applications include:
  - Noise reduction e.g. for speech and music signals
  - Prediction of future values of a signal, e.g. in finance
  - Noise cancellation, e.g. for aircraft cockpit noise
  - Deconvolution, e.g. removal of room acoustics (dereverberation) or echo cancellation in telephony.
- The Wiener filter is a very powerful tool. However, it is only the optimal *linear* estimator for stationary signals. The *Kalman filter* offers an extension for non-stationary signals via *state space models*. In cases where a linear filter is still not good enough, non-linear filtering techniques can be adopted. See 4th year Signal Processing and Control modules for more advanced topics in these areas.

The Discrete-time Wiener Filter

In a minor abuse of notation, and following standard conventions, we will refer to both random variables and their possible values in lower-case symbols, as this should cause no ambiguity for this section of work.

Figure 9: Wiener Filter

- In the most general case, we can filter the observed signal x_n with an *Infinite impulse response (IIR)* filter, having a non-causal impulse response h_p:

$$\{h_p;\ p = -\infty, \ldots, -1, 0, 1, 2, \ldots, \infty\} \tag{11}$$

- We filter the observed noisy signal using the filter $\{h_p\}$ to obtain an estimate $\hat{d}_n$ of the desired signal:

$$\hat{d}_n = \sum_{p=-\infty}^{\infty} h_p \, x_{n-p} \tag{12}$$

- Since both d_n and x_n are drawn from random processes $\{d_n\}$ and $\{x_n\}$, we can only measure performance of the filter in terms of *expectations*. The criterion adopted for Wiener filtering is the *mean-squared error (MSE)* criterion. First, form the error signal ϵ_n:

$$\epsilon_n = d_n - \hat{d}_n = d_n - \sum_{p=-\infty}^{\infty} h_p \, x_{n-p}$$

The *mean-squared error (MSE)* is then defined as:

$$\boxed{J = E[\epsilon_n^2]} \tag{13}$$

- The Wiener filter minimises J with respect to the filter coefficients $\{h_p\}$.

Derivation of Wiener filter

The Wiener filter assumes that $\{x_n\}$ and $\{d_n\}$ are *jointly wide-sense stationary*. This means that the means of both processes are constant, and all autocorrelation functions/cross-correlation functions (e.g. $r_{xd}[n,m]$) depend only on the time difference $m-n$ between data points.

The expected error (13) may be minimised with respect to the impulse response values h_q. A sufficient condition for a minimum is:

$$\frac{\partial J}{\partial h_q} = \frac{\partial E[\epsilon_n^2]}{\partial h_q} = E\left[\frac{\partial \epsilon_n^2}{\partial h_q}\right] = E\left[2\epsilon_n \frac{\partial \epsilon_n}{\partial h_q}\right] = 0$$

for each $q \in \{-\infty, \ldots, -1, 0, 1, 2, \ldots\infty\}$.

The term $\frac{\partial \epsilon_n}{\partial h_q}$ is then calculated as:

$$\frac{\partial \epsilon_n}{\partial h_q} = \frac{\partial}{\partial h_q}\left\{d_n - \sum_{p=-\infty}^{\infty} h_p \, x_{n-p}\right\} = -x_{n-q}$$

and hence the coefficients must satisfy

$$\boxed{E\left[\epsilon_n\, x_{n-q}\right] = 0; \qquad -\infty < q < +\infty} \tag{14}$$

This is known as the *orthogonality principle*, since two random variables X and Y are termed *orthogonal* if

$$E[XY] = 0$$

Now, substituting for ϵ_n in (14) gives:

$$\begin{aligned}
E\left[\epsilon_n\, x_{n-q}\right] &= E\left[\left(d_n - \sum_{p=-\infty}^{\infty} h_p\, x_{n-p}\right)\, x_{n-q}\right] \\
&= E\left[d_n\, x_{n-q}\right] - \sum_{p=-\infty}^{\infty} h_p\; E\left[x_{n-q}\, x_{n-p}\right] \\
&= r_{xd}[q] - \sum_{p=-\infty}^{\infty} h_p\, r_{xx}[q-p] \\
&= 0
\end{aligned}$$

Hence, rearranging, the solution must satisfy

$$\boxed{\sum_{p=-\infty}^{\infty} h_p\, r_{xx}[q-p] = r_{xd}[q], \qquad -\infty < q < +\infty} \tag{15}$$

This is known as the *Wiener-Hopf* equations.

The *Wiener-Hopf* equations involve an infinite number of unknowns h_q. The simplest way to solve this is in the frequency domain. First note that the *Wiener-Hopf* equations can be rewritten as a discrete-time convolution:

$$h_q * r_{xx}[q] = r_{xd}[q], \qquad -\infty < q < +\infty \tag{16}$$

Taking *discrete-time Fourier transforms (DTFT)* of both sides:

$$H(e^{j\Omega})\mathcal{S}_x(e^{j\Omega}) = \mathcal{S}_{xd}(e^{j\Omega})$$

where $\mathcal{S}_{xd}(e^{j\Omega})$, the DTFT of $r_{xd}[q]$, is defined as the *cross-power spectrum* of d and x.

Hence, rearranging:

$$\boxed{H(e^{j\Omega}) = \frac{\mathcal{S}_{xd}(e^{j\Omega})}{\mathcal{S}_x(e^{j\Omega})}} \tag{17}$$

Frequency domain Wiener filter

Mean-squared error for the optimal filter

The previous equations show how to calculate the optimal filter for a given problem. They don't, however, tell us how well that optimal filter performs. This can be assessed from the mean-squared error value of the optimal filter:

$$J = E[\epsilon_n^2] = E[\epsilon_n(d_n - \sum_{p=-\infty}^{\infty} h_p\, x_{n-p})]$$

$$= E[\epsilon_n\, d_n] - \sum_{p=-\infty}^{\infty} h_p\, E[\epsilon_n\, x_{n-p}]$$

The expectation on the right is zero, however, for the optimal filter, by the orthogonality condition (14), so the minimum error is:

$$J_{\mathrm{min}} = E[\epsilon_n\, d_n]$$

$$= E[(d_n - \sum_{p=-\infty}^{\infty} h_p\, x_{n-p})\, d_n]$$

$$= r_{dd}[0] - \sum_{p=-\infty}^{\infty} h_p\, r_{xd}[p]$$

IMPORTANT SPECIAL CASE: UNCORRELATED SIGNAL AND NOISE PROCESSES

An important sub-class of the Wiener filter, which also gives considerable insight into filter behaviour, can be gained by considering the case where the desired signal process $\{d_n\}$ is uncorrelated with the noise process $\{v_n\}$, i.e.

$$r_{dv}[k] = E[d_n v_{n+k}] = 0, \qquad -\infty < k < +\infty$$

Consider the implications of this fact on the correlation functions required in the *Wiener-Hopf* equations:

$$\sum_{p=-\infty}^{\infty} h_p \, r_{xx}[q-p] = r_{xd}[q], \qquad -\infty < q < +\infty$$

1. r_{xd}.

$$r_{xd}[q] = E[x_n d_{n+q}] = E[(d_n + v_n) d_{n+q}] \tag{18}$$

$$= E[d_n d_{n+q}] + E[v_n d_{n+q}] = r_{dd}[q] \tag{19}$$

since $\{d_n\}$ and $\{v_n\}$ are uncorrelated.
Hence, taking DTFT of both sides:

$$\mathcal{S}_{xd}(e^{j\Omega}) = \mathcal{S}_d(e^{j\Omega})$$

2. r_{xx}.

$$
\begin{aligned}
r_{xx}[q] = E[x_n x_{n+q}] &= E[(d_n + v_n)(d_{n+q} + v_{n+q})] \\
&= E[d_n d_{n+q}] + E[v_n v_{n+q}] + E[d_n v_{n+q}] + E[v_n d_{n+q}] \\
&= E[d_n d_{n+q}] + E[v_n v_{n+q}] = r_{dd}[q] + r_{vv}[q]
\end{aligned}
$$

Hence

$$\mathcal{S}_x(e^{j\Omega}) = \mathcal{S}_d(e^{j\Omega}) + \mathcal{S}_v(e^{j\Omega})$$

Thus the Wiener filter becomes

$$H(e^{j\Omega}) = \frac{\mathcal{S}_d(e^{j\Omega})}{\mathcal{S}_d(e^{j\Omega}) + \mathcal{S}_v(e^{j\Omega})}$$

From this it can be seen that the behaviour of the filter is intuitively reasonable in that at those frequencies where the noise power spectrum $\mathcal{S}_v(e^{j\Omega})$ is small, the gain of the filter tends to unity whereas the gain tends to a small value at those frequencies where the noise spectrum is significantly larger than the desired signal power spectrum $\mathcal{S}_d(e^{j\Omega})$.

Example: AR Process

An autoregressive process $\{D_n\}$ of order 1 (see section 3.) has power spectrum:

$$\mathcal{S}_D(e^{j\Omega}) = \frac{\sigma_e^2}{(1 - a_1 e^{-j\Omega})(1 - a_1 e^{j\Omega})}$$

Suppose the process is observed in zero mean white noise with variance σ_v^2, which is uncorrelated with $\{D_n\}$:

$$x_n = d_n + v_n$$

Design the Wiener filter for estimation of d_n.

Since the noise and desired signal are uncorrelated, we can use the form of Wiener filter from the previous page. Substituting in the various terms and rearranging, its frequency response is:

$$H(e^{j\Omega}) = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_v^2(1 - a_1 e^{-j\Omega})(1 - a_1 e^{j\Omega})}$$

The impulse response of the filter can be found by inverse Fourier transforming the frequency response. This is studied in the examples paper.

The FIR Wiener filter

Note that, in general, the Wiener filter given by equation (17) is non-causal, and hence physically unrealisable, in that the impulse response h_p is defined for values of p less than 0. Here we consider a practical alternative in which a causal FIR Wiener filter is developed.

In the FIR case the signal estimate is formed as

$$\hat{d}_n = \sum_{p=0}^{P-1} h_p \, x_{n-p} \tag{20}$$

and we minimise, as before, the objective function

$$J = E[(d_n - \hat{d}_n)^2]$$

The filter derivation proceeds as before, leading to an *orthogonality* principle:

$$\boxed{E\left[\epsilon_n \, x_{n-q}\right] = 0; \qquad q = 0, \ldots, P-1} \tag{21}$$

and *Wiener-Hopf* equations as follows:

$$\boxed{\sum_{p=0}^{P-1} h_p \, r_{xx}[q-p] = r_{xd}[q], \qquad q = 0, 1, \ldots, P-1} \tag{22}$$

The above equations may be written in matrix form as:

$$\mathbf{R}_x \mathbf{h} = \mathbf{r}_{xd}$$

where:

$$\mathbf{h} = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{P-1} \end{bmatrix} \qquad \mathbf{r}_{xd} = \begin{bmatrix} r_{xd}[0] \\ r_{xd}[1] \\ \vdots \\ r_{xd}[P-1] \end{bmatrix}$$

and

$$\mathbf{R}_x = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \cdots & r_{xx}[P-1] \\ r_{xx}[1] & r_{xx}[0] & \cdots & r_{xx}[P-2] \\ \vdots & \vdots & & \vdots \\ r_{xx}[P-1] & r_{xx}[P-2] & \cdots & r_{xx}[0] \end{bmatrix}$$

$\mathbf{R}_x$ is known as the *correlation matrix*.

Note that $r_{xx}[k] = r_{xx}[-k]$ so that the *correlation matrix* $\mathbf{R}_x$ is symmetric and has constant diagonals (a symmetric *Toeplitz* matrix).
The coefficient vector can be found by matrix inversion:

$$\mathbf{h} = \mathbf{R}_x{}^{-1} \mathbf{r}_{xd} \tag{23}$$

This is the FIR Wiener filter and as for the general Wiener filter, it requires *a-priori* knowledge of the autocorrelation matrix $\mathbf{R}_x$ of the input process $\{x_n\}$ and the cross-correlation $\mathbf{r}_{xd}$ between the input $\{x_n\}$ and the desired signal process $\{d_n\}$.

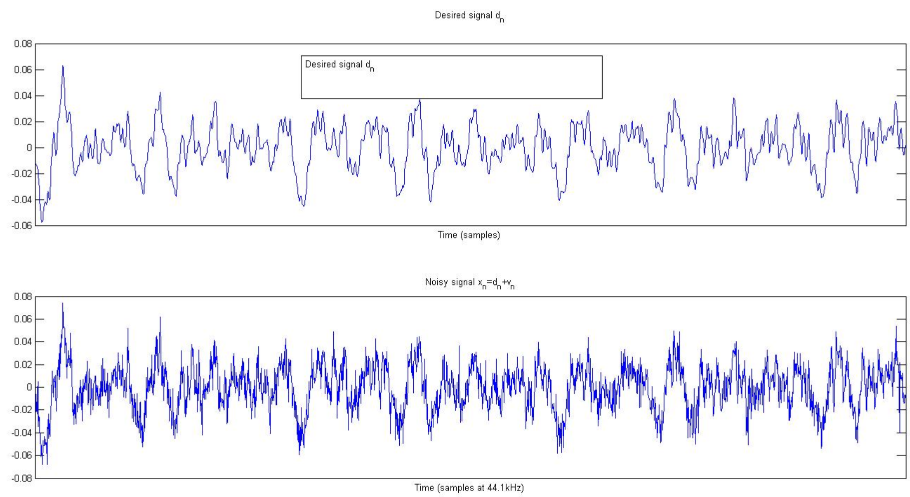
As before, the minimum mean-squared error is given by:

$$\begin{aligned} J_{\min} &= E[\epsilon_n \, d_n] \\ &= E[(d_n - \sum_{p=0}^{P-1} h_p \, x_{n-p}) \, d_n] \\ &= r_{dd}[0] - \sum_{p=0}^{P-1} h_p \, r_{xd}[p] \\ &= r_{dd}[0] - \mathbf{r}_{xd}^T \mathbf{h} = r_{dd}[0] - \mathbf{r}_{xd}^T \mathbf{R}_x{}^{-1} \mathbf{r}_{xd} \end{aligned}$$

Case Study: Audio Noise Reduction

- Consider a section of acoustic waveform (music, voice, ...) d_n that is corrupted by additive noise v_n

$$x_n = d_n + v_n$$



- We could try and noise reduce the signal using the FIR Wiener filter.
- Assume that the section of data is wide-sense stationary and ergodic (approx. true for a short segment around 1/40 s). Assume also that the noise is white and uncorrelated with the audio signal - with variance σ_v^2, i.e.

$$r_{vv}[k] = \sigma_v^2 \delta[k]$$

- The Wiener filter in this case needs (see eq. (23)):

$$r_{xx}[k], \quad \text{Autocorrelation of noisy signal}$$

$$r_{xd}[k] = r_{dd}[k] \quad \text{Autocorrelation of desired signal}$$

  [since noise uncorrelated with signal, as in eq. (19)]

- Since signal is assumed ergodic, we can estimate these quantities:

$$r_{xx}[k] \approx \frac{1}{N} \sum_{n=0}^{N-1} x_n x_{n+k}$$

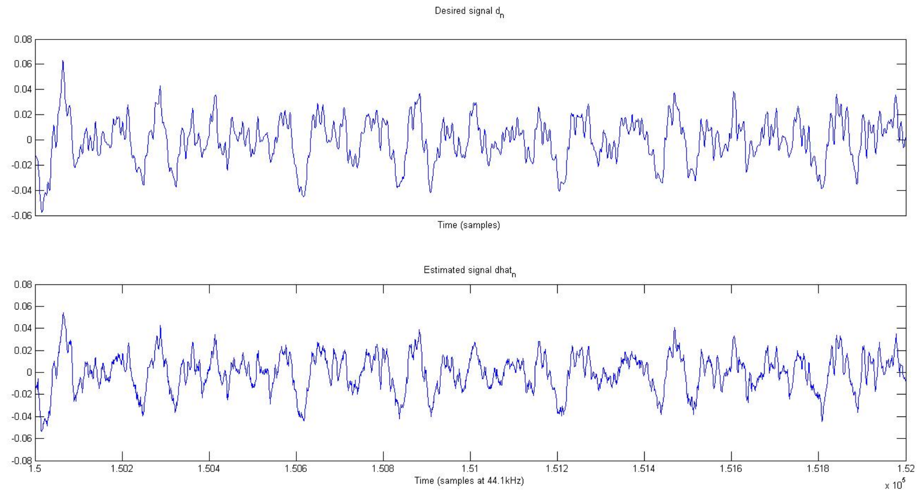
$$r_{dd}[k] = r_{xx}[k] - r_{vv}[k] = \begin{cases} r_{xx}[k], & k \neq 0 \\ r_{xx}[0] - \sigma_v^2, & k = 0 \end{cases}$$

  [Note have to be careful that $r_{dd}[k]$ is still a valid autocorrelation sequence since r_{xx} is just an estimate.]

- Choose the filter length P, form the autocorrelation matrix and cross-correlation vector and solve in e.g. Matlab:

$$\mathbf{h} = \mathbf{R}_x^{-1} \mathbf{r}_{xd}$$

- The output looks like this, with $P = 350$:



- The theoretical mean-squared error is calculated as:

$$J_{\min} = r_{dd}[0] - \mathbf{r}_{xd}^T \mathbf{h}$$

- We can compute this for various filter lengths, and in this artificial scenario we can compare the theoretical error performance with the actual mean-squared error, since we have access to the true d_n itself:

$$J_{\text{true}} = \frac{1}{N} \sum_{n=0}^{N-1} (d_n - \hat{d}_n)^2$$

Not necessarily equal to the theoretical value since we estimated the autocorrelation functions from finite pieces of data and assumed stationarity of the processes.

Example: AR process

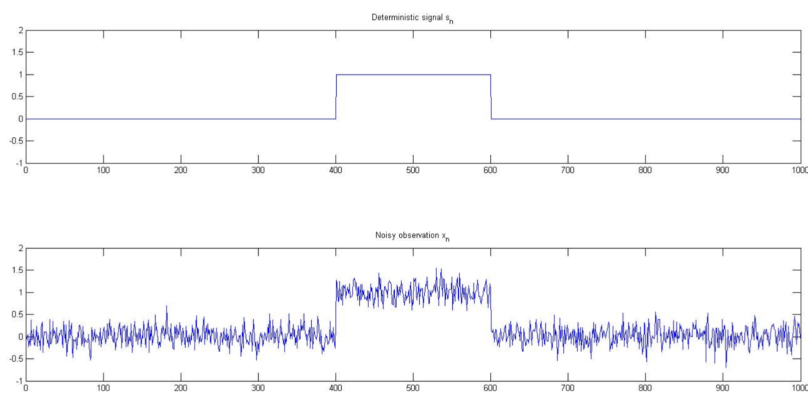
Now take as a numerical example, the exact same setup, but we specify...

Example: Noise cancellation

Matched Filtering

- The Wiener filter shows how to extract a random signal from a random noise environment.
- How about the (apparently) simpler task of detecting a known deterministic signal s_n, $n = 0, \ldots, N-1$, buried in random noise v_n:

$$x_n = s_n + v_n$$



- The technical method for doing this is known as the *matched filter*
- It finds extensive application in detection of pulses in communications data, radar and sonar data.

- To formulate the problem, first 'vectorise' the equation

$$\mathbf{x} = \mathbf{s} + \mathbf{v}$$

$$\mathbf{s} = [s_0,\ s_1, \ldots, s_{N-1}]^T,\ \ \mathbf{x} = [x_0,\ x_1, \ldots, x_{N-1}]^T$$

- Once again, we will design an optimal FIR filter for performing the detection task. Suppose the filter has coefficients h_m, for $m = 0, 1, \ldots, N-1$, then the output of the filter at time $N-1$ is:

$$y_{N-1} = \sum_{m=0}^{N-1} h_m x_{N-1-m} = \mathbf{h}^T \tilde{\mathbf{x}} = \mathbf{h}^T (\tilde{\mathbf{s}} + \tilde{\mathbf{v}}) = \mathbf{h}^T \tilde{\mathbf{s}} + \mathbf{h}^T \tilde{v}$$

where $\tilde{\mathbf{x}} = [x_{N-1},\ x_{N-2}, \ldots, x_0]^T$, etc. ('time-reversed' vector).

- This time, instead of minimising the mean-squared error as in the Wiener Filter, we attempt to maximise the signal-to-noise ratio (SNR) at the output of the filter, hence giving best possible chance of detecting the signal s_n.
- Define output SNR as:

$$\frac{E[|\mathbf{h}^T \tilde{\mathbf{s}}|^2]}{E[|\mathbf{h}^T \tilde{\mathbf{v}}|^2]} = \frac{|\mathbf{h}^T \tilde{\mathbf{s}}|^2}{E[|\mathbf{h}^T \tilde{\mathbf{v}}|^2]}$$

[since numerator is not a random quantity].

Signal output energy

- The signal component at the output is $\mathbf{h}^T\tilde{\mathbf{s}}$, with energy

$$|\mathbf{h}^T\tilde{\mathbf{s}}|^2 = \mathbf{h}^T\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T\mathbf{h}$$

- To analyse this, consider the matrix $\mathbf{M} = \tilde{\mathbf{s}}\tilde{\mathbf{s}}^T$. What are its eigenvectors/eigenvalues ?
- Recall the definition of eigenvectors ($\mathbf{e}$) and eigenvalues (λ):

$$\mathbf{M}\mathbf{e} = \lambda\mathbf{e}$$

- Try $\mathbf{e} = \tilde{\mathbf{s}}$:

$$\mathbf{M}\tilde{\mathbf{s}} = \tilde{\mathbf{s}}\tilde{\mathbf{s}}^T\tilde{\mathbf{s}} = (\tilde{\mathbf{s}}^T\tilde{\mathbf{s}})\tilde{\mathbf{s}}$$

  Hence the unit length vector $\mathbf{e}_0 = \tilde{\mathbf{s}}/|\tilde{\mathbf{s}}|$ is an eigenvector and $\lambda = (\tilde{\mathbf{s}}^T\tilde{\mathbf{s}})$ is the corresponding eigenvalue.
- Now, consider any vector $\mathbf{e}'$ which is orthogonal to $\mathbf{e}_0$ (i.e. $\tilde{\mathbf{s}}^T\mathbf{e}' = 0$):

$$\mathbf{M}\mathbf{e}' = \tilde{\mathbf{s}}\tilde{\mathbf{s}}^T\mathbf{e}' = 0$$

  Hence $\mathbf{e}'$ is also an eigenvector, but with eigenvalue $\lambda' = 0$. Since we can construct a set of $N-1$ orthonormal (unit length *and* orthogonal to each other) vectors which are orthogonal to $\tilde{\mathbf{s}}$, call these $\mathbf{e}_1$, $\mathbf{e}_2$, ..., $\mathbf{e}_{N-1}$, we have now discovered all N eigenvectors/eigenvalues of $\mathbf{M}$.

- Since the N eigenvectors form an orthonormal basis, we may represent *any* filter coefficient vector $\mathbf{h}$ as a linear combination of these:

$$\mathbf{h} = \alpha\mathbf{e}_0 + \beta\mathbf{e}_1 + \gamma\mathbf{e}_2 + \ \ldots \ + \ldots\mathbf{e}_{N-1}$$

- Now, consider the signal output energy again:

$$\begin{aligned}
\mathbf{h}^T\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T\mathbf{h} &= \mathbf{h}^T\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T(\alpha\mathbf{e}_0 + \beta\mathbf{e}_1 + \gamma\mathbf{e}_2 + \ \ldots \ + \ldots\mathbf{e}_{N-1}) \\
&= \mathbf{h}^T(\alpha\tilde{\mathbf{s}}^T\tilde{\mathbf{s}})\mathbf{e}_0, \text{ because } \mathbf{M}\mathbf{e}_0 = (\tilde{\mathbf{s}}^T\tilde{\mathbf{s}})\mathbf{e}_0 \\
&= (\alpha\mathbf{e}_0 + \beta\mathbf{e}_1 + \gamma\mathbf{e}_2 + \ \ldots \ + \ldots\mathbf{e}_{N-1})^T(\alpha\tilde{\mathbf{s}}^T\tilde{\mathbf{s}})\mathbf{e}_0 \\
&= \alpha^2\tilde{\mathbf{s}}^T\tilde{\mathbf{s}}
\end{aligned}$$

since $\mathbf{e}_i^T\mathbf{e}_j = \delta[i-j]$.

Noise output energy

- Now, consider the expected noise output energy, which may be simplified as follows:

$$E[|\mathbf{h}^T\tilde{\mathbf{v}}|^2] = E[\mathbf{h}^T\tilde{\mathbf{v}}\tilde{\mathbf{v}}^T\mathbf{h}] = \mathbf{h}^T E[\tilde{\mathbf{v}}\tilde{\mathbf{v}}^T]\mathbf{h}$$

- We will here consider the case where the noise is white with variance σ_v^2. Then, for any time indexes $i = 0, ..., N-1$ and $j = 0, ..., N-1$:

$$E[v_i v_j] = \begin{cases} \sigma_v^2, & i = j \\ 0, & i \neq j \end{cases}$$

and hence

$$E[\tilde{\mathbf{v}}\tilde{\mathbf{v}}^T] = \sigma_v^2\mathbf{I}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix
[diagonal elements correspond to $i = j$ terms and off-diagonal to $i \neq j$ terms].

- So, we have finally the expression for the noise output energy:

$$E[|\mathbf{h}^T\tilde{\mathbf{v}}|^2] = \mathbf{h}^T\sigma_v^2\mathbf{I}\mathbf{h} = \sigma_v^2\mathbf{h}^T\mathbf{h}$$

and once again we can expand $\mathbf{h}$ in terms of the eigenvectors of $\mathbf{M}$:

$$\sigma_v^2\mathbf{h}^T\mathbf{h} = \sigma_v^2(\alpha^2 + \beta^2 + \gamma^2 + ...)$$

again, since $\mathbf{e}_i^T\mathbf{e}_j = \delta[i-j]$

SNR MAXIMISATION

- The SNR may now be expressed as:

$$\frac{|\mathbf{h}^T\tilde{\mathbf{s}}|^2}{E[|\mathbf{h}^T\tilde{\mathbf{v}}|^2]} = \frac{\alpha^2\tilde{\mathbf{s}}^T\tilde{\mathbf{s}}}{\sigma_v^2(\alpha^2+\beta^2+\gamma^2+\ldots)}$$

- Clearly, scaling $\mathbf{h}$ by some factor ρ will not change the SNR since numerator and denominator will both then scale by ρ^2. So, we can arbitrarily fix $|\mathbf{h}| = 1$ (any other value except zero would do, but 1 is a convenient choice) and then maximise.
- With $|\mathbf{h}| = 1$ we have $(\alpha^2+\beta^2+\gamma^2+\ldots) = 1$ and the SNR becomes just equal to $\alpha^2\tilde{\mathbf{s}}^T\tilde{\mathbf{s}}/(\sigma_v^2 \times 1)$.
- The largest possible value of α given that $|\mathbf{h}| = 1$ corresponds to $\alpha = 1$, which implies then that $\beta = \gamma = \ldots = 0$, and finally we have the solution as:

$$\mathbf{h}^{\text{opt}} = 1 \times \mathbf{e}_0 = \frac{\tilde{\mathbf{s}}}{|\tilde{\mathbf{s}}|}, \text{ since } \mathbf{e}_0 = \frac{\tilde{\mathbf{s}}}{|\tilde{\mathbf{s}}|}$$

  i.e. the optimal filter coefficients are just the (normalised) time-reversed signal!
- The SNR at the optimal filter setting is given by

$$\text{SNR}^{\text{opt}} = \frac{\tilde{\mathbf{s}}^T\tilde{\mathbf{s}}}{\sigma_v^2}$$

and clearly the performance depends (as expected) very much on the energy of the signal $\boldsymbol{s}$ and the noise $\boldsymbol{v}$.

Practical Implementation of the matched filter

- We chose a batch of data of same length as the signal $\mathbf{s}$ and optimised a filter $\mathbf{h}$ of the same length.
- In practice we would now run this filter over a much longer length of data $\mathbf{x}$ which contains $\mathbf{s}$ at some unknown position and find the time at which maximum energy occurs. This is the point at which $\mathbf{s}$ can be detected, and optimal thresholds can be devised to make the decision on whether a detection of $\mathbf{s}$ should be declared at that time.
- Example (like a simple square pulse radar detection problem):

$$s_n = \text{Rectangle pulse} = \begin{cases} 1, & n = 0, 1, \ldots, T-1 \\ 0, & \text{otherwise} \end{cases}$$

- Optimal filter is the (normalised) time reversed version of s_n:

$$h_n^{\text{opt}} = \begin{cases} 1/\sqrt{T}, & n = 0, 1, \ldots, T-1 \\ 0, & \text{otherwise} \end{cases}$$

- SNR achievable at detection point:

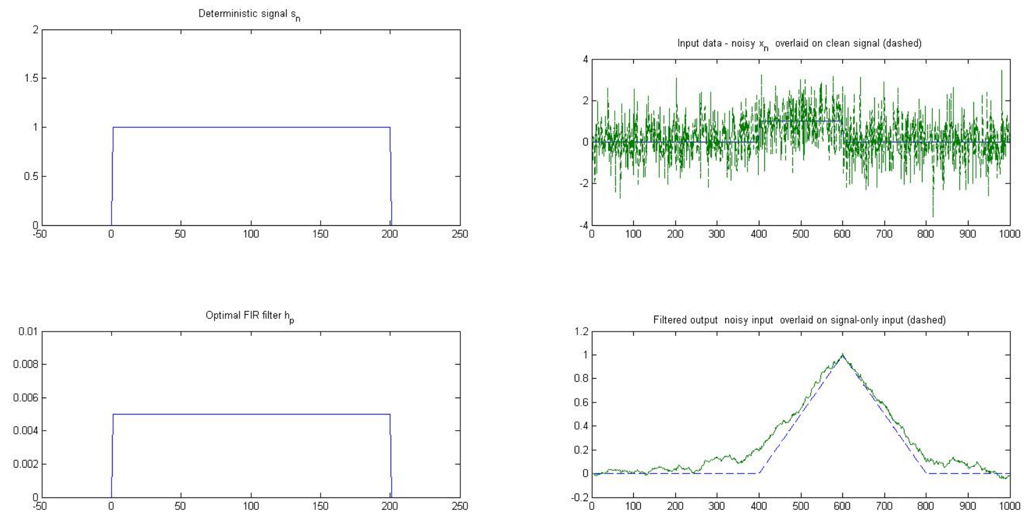
$$\text{SNR}^{\text{opt}} = \frac{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}}{\sigma_v^2} = \frac{T}{\sigma_v^2}$$

Compare with best SNR attainable before matched filtering:

$$\text{SNR} = \frac{\text{Max signal value}^2}{\text{Average noise energy}} = \frac{1}{\sigma_v^2}$$

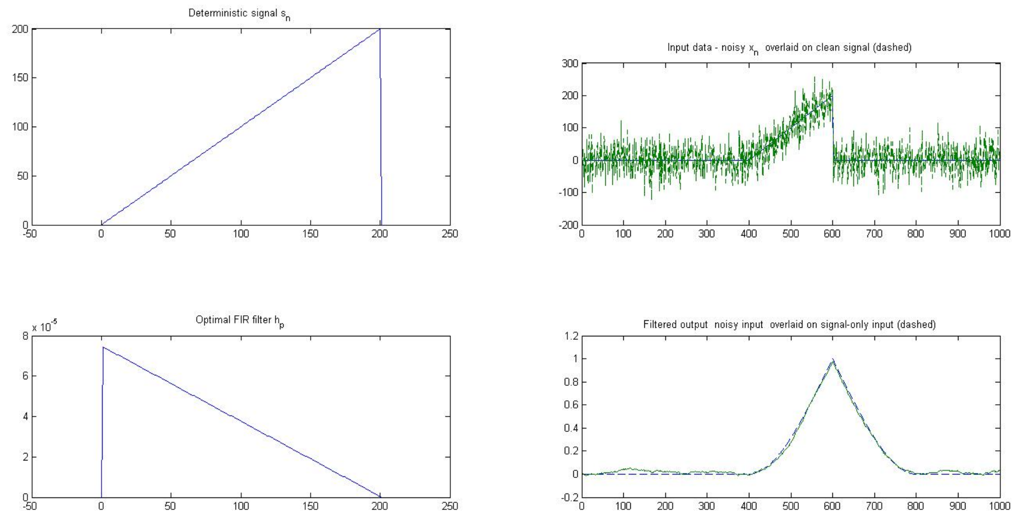
i.e. a factor of T improvemment, which could be substantial for long pulses $T >> 1$.

- See below for example inputs and outputs:



- See below for a different case where the signal is a saw-tooth pulse:

$$s_n = \text{Sawtooth pulse} = \begin{cases} n+1, & n = 0, 1, ..., T-1 \\ 0, & \text{otherwise} \end{cases}$$


Deterministic signal s_n
Input data - noisy x_n overlaid on clean signal (dashed)
Optimal FIR filter h_p
x 10^-5
Filtered output noisy input overlaid on signal-only input (dashed)

Section 3: Model-based signal processing

In this section some commonly used signal models for random processes are described, and methods for their parameter estimation are presented.

- If the physical process which generated a set of data is known or can be well approximated, then a parametric model can be constructed
- Careful estimation of the parameters in the model can lead to very accurate estimates of quantities such as power spectrum.
- We will consider the *autoregressive moving-average (ARMA)* class of models in which a LTI system (digital filter) is driven by a white noise input sequence.
- If a random process $\{X_n\}$ can be modelled as white noise exciting a filter with frequency response $H(e^{j\Omega})$ then the spectral density of the process is:

$$\boxed{S_X(e^{j\Omega}) = \sigma_w^2 \, |H(e^{j\Omega})|^2}$$

  where σ_w^2 is the variance of the white noise process. In its general form, this is known as the *innovations* representation of a random process. It can be shown that any regular (i.e. not predictable with zero error) stationary random process can be expressed in this form.

- We will study models in which the frequency response $H(e^{j\Omega})$ is that of an IIR digital filter, in particular the all-pole case (known as an AR model).
- Parametric models need to be chosen carefully - an inappropriate model for the data can give misleading results
- We will also study parameter estimation techniques for signal models.

ARMA Models

A quite general representation of a stationary random process is the autoregressive moving-average (ARMA) model:

- The ARMA(P,Q) model difference equation representation is:

$$x_n = -\sum_{p=1}^{P} a_p \, x_{n-p} + \sum_{q=0}^{Q} b_q \, w_{n-q} \tag{24}$$

  where:

  a_p are the AR parameters,

  b_q are the MA parameters

  and $\{W_n\}$ is a zero-mean stationary white noise process with variance, σ_w^2. Typically $b_0 = 1$ as otherwise the model is redundantly parameterised (any arbitrary value of b_0 could be equivalently modelled by changing the value of σ_v)

- Clearly the ARMA model is a pole-zero IIR filter-based model with transfer function

$$H(z) = \frac{B(z)}{A(z)}$$

where:

$$A(z) = 1 + \sum_{p=1}^{P} a_p z^{-p}, \quad B(z) = \sum_{q=0}^{Q} b_q z^{-q}$$

- Unless otherwise stated we will always assume that the filter is stable, i.e. the poles (solutions of $A(z) = 0$) all lie *within* the unit circle (we say in this case that $A(z)$ is *minimum phase*). Otherwise the autocorrelation function is undefined and the process is technically *non-stationary*.
- Hence the power spectrum of the ARMA process is:

$$S_X(e^{j\omega T}) = \sigma_w^2 \frac{|B(e^{j\omega T})|^2}{|A(e^{j\omega T})|^2}$$

- Note that, as for standard digital filters, $A(z)$ and $B(z)$ may be factorised into the *poles* d_q and *zeros* c_q of the system. In which case the power spectrum can be given a geometric interpretation (see e.g. 3F3 laboratory):

$$\boxed{|H(e^{j\Omega})| = G \frac{\prod_{q=1}^{Q} \text{Distance from } e^{j\Omega} \text{ to } c_q}{\prod_{p=1}^{P} \text{Distance from } e^{j\Omega} \text{ to } d_p}}$$

- The ARMA model, and in particular its special case the AR model ($Q = 0$), are used extensively in applications such as speech and audio processing, time series analysis, econometrics, computer vision (tracking), ...

The ARMA model is quite a flexible and general way to model a stationary random process:

- The poles model well the *peaks* in the spectrum (sharper peaks implies poles closer to the unit circle)
- The zeros model troughs or nulls in the power spectrum
- Highly complex power spectra can be approximated well by large model orders P and Q

AR models

The AR model, despite being less general than the ARMA model, is used in practice far more often, owing to the simplicity of its mathematics and the efficient ways in which its parameters may be estimated from data. Hence we focus on AR models from here on. ARMA model parameter estimation in contrast requires the solution of troublesome nonlinear equations, and is prone to local maxima in the parameter search space. It should be noted that an AR model of sufficiently high order P can approximate any particular ARMA model.

The AR model is obtained in the all-pole filter case of the ARMA model, for which $Q = 0$. Hence the model equations are:

$$x_n = -\sum_{p=1}^{P} a_p \, x_{n-p} + w_n \tag{25}$$

Autocorrelation function for AR Models

The autocorrelation function $r_{XX}[r]$ for the AR process is:

$$r_{XX}[r] = E[x_n x_{n+r}]$$

Substituting for x_{n+r} from Eq. (25) gives:

$$\begin{aligned} r_{XX}[r] &= E\left[x_n \left\{ -\sum_{p=1}^{P} a_p \, x_{n+r-p} + w_{n+r} \right\} \right] \\ &= -\sum_{p=1}^{P} a_p \, E[x_n x_{n+r-p}] + E[x_n w_{n+r}] \\ &= -\sum_{p=1}^{P} a_p \, r_{XX}[r-p] + r_{XW}[r] \end{aligned}$$

Note that the auto-correlation and cross-correlation satisfy the same AR system difference equation as $\boldsymbol{x}_n$ and $\boldsymbol{w}_n$ (Eq. 25).
Now, let the impulse response of the system $H(z) = \frac{1}{A(z)}$ be h_n, then:

$$x_n = \sum_{m=-\infty}^{\infty} h_m \, w_{n-m}$$

Then, from the linear system results of Eq. (3),

$$r_{XW}[k] = r_{WX}[-k]$$
$$= h_{-k} * r_{WW}[-k] = \sum_{m=-\infty}^{\infty} h_{-m}\, r_{WW}[m-k]$$

[Note that we have used the results that $r_{XY}[k] = r_{YX}[-k]$ and $r_{XX}[k] = r_{XX}[-k]$ for stationary processes.]

$\{W_n\}$ is, however, a zero-mean white noise process, whose autocorrelation function is:

$$r_{WW}[m] = \sigma_W^2 \delta[m]$$

and hence

$$r_{XW}[k] = \sigma_W^2 h_{-k}$$

Substituting this expression for $r_{XW}[k]$ into equation **??** gives the so-called *Yule-Walker Equations* for an AR process,

$$r_{XX}[r] = -\sum_{p=1}^{P} a_p\, r_{XX}[r-p] + \sigma_W^2 h_{-r} \qquad (26)$$

Since the AR system is *causal*, all terms h_k are zero for $k < 0$. Now consider inputing the unit impulse $\delta[k]$ in to the AR difference equation. It is then clear that

$$h_0 = -\sum_{i=1}^{P} a_i(h_{-i} = 0) + 1 = 1$$

Hence the *Yule-Walker* equations simplify to:

$$\boxed{r_{XX}[r] + \sum_{p=1}^{P} a_p\, r_{XX}[r-p] = \begin{cases} \sigma_W^2, & r = 0 \\ 0, & \text{Otherwise} \end{cases}} \qquad (27)$$

or in matrix form:

$$\begin{bmatrix} r_{XX}[0] & r_{XX}[-1] & \dots & r_{XX}[-P] \\ r_{XX}[1] & r_{XX}[0] & \dots & r_{XX}[1-P] \\ \vdots & \vdots & & \vdots \\ r_{XX}[P] & r_{XX}[P-1] & \dots & r_{XX}[0] \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_W^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This may be conveniently partitioned as follows:

$$\left[\begin{array}{c|ccc} r_{XX}[0] & r_{XX}[-1] & \dots & r_{XX}[-P] \\ \hline r_{XX}[1] & r_{XX}[0] & \dots & r_{XX}[1-P] \\ \vdots & \vdots & & \vdots \\ r_{XX}[P] & r_{XX}[P-1] & \dots & r_{XX}[0] \end{array}\right] \left[\begin{array}{c} 1 \\ \hline a_1 \\ \vdots \\ a_P \end{array}\right]$$

$$= \left[\begin{array}{c} \sigma_W^2 \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]$$

Taking the bottom P elements of the right and left hand sides:

$$\begin{bmatrix} r_{XX}[0] & r_{XX}[-1] & \dots & r_{XX}[1-P] \\ r_{XX}[1] & r_{XX}[0] & \dots & r_{XX}[2-P] \\ \vdots & \vdots & & \vdots \\ r_{XX}[P-1] & r_{XX}[P-2] & \dots & r_{XX}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}$$

$$= - \begin{bmatrix} r_{XX}[1] \\ r_{XX}[2] \\ \vdots \\ r_{XX}[P] \end{bmatrix}$$

or

$$\mathbf{R}\mathbf{a} = -\mathbf{r} \tag{28}$$

and

$$\sigma_W^2 = \left[\begin{array}{cccc} r_{XX}[0] & r_{XX}[-1] & \ldots & r_{XX}[-P] \end{array}\right] \left[\begin{array}{c} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_P \end{array}\right]$$

This gives a relationship between the AR coefficients, the autocorrelation function and the noise variance terms. If we can form an *estimate* of the autocorrelation function using the observed data then (28) above may be used to solve for the AR coefficients as:

$$\mathbf{a} = -\mathbf{R}^{-1}\mathbf{r} \tag{29}$$

Example: AR coefficient estimation