Stochastic Signals and Systems

Contents

1 Probability Theory
2 Stochastic Processes
3 Parameter Estimation
4 Signal Detection
5 Spectrum Analysis
6 Optimal Filtering

6	Optimal Filtering	3
6.1	Introduction	3
6.2	Matched Filtering	4
6.2.1	Matched Filtering for White Noise	10
6.2.2	Matched Filtering as Correlation Processing	17
6.3	Wiener Filtering	19
6.3.1	Wiener-Hopf Equation	22
6.3.2	Finite Wiener Filtering	26
6.3.3	Noncausal Wiener Filtering	32
6.3.4	Causal Wiener Filtering	39
6.4	Kalman Filtering	49
6.4.1	State Space Model	50
6.4.2	State Estimation	53
6.4.3	Kalman Filter Approach	55
	References to Chapter 6	65

6 Optimal Filtering

6.1 Introduction

A basic problem in the application of stochastic processes is the estimation of a signal in the presence of additive noise. The signal may be random or deterministic, and the noise may be colored or white.

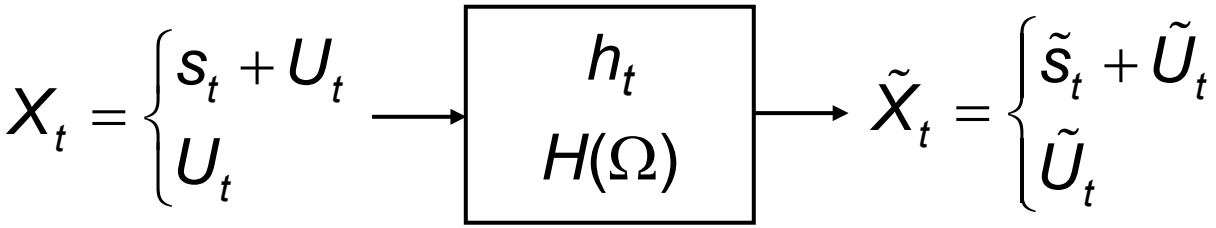
The problem consists of establishing the presence of the signal or of estimating its form. The solution of this problem depends on the state of prior knowledge concerning the signal and the noise, e.g. we may be able to specify signal and noise covariance functions, power spectra or probability densities.

Furthermore, system constraints define the form of the solution. For example we might allow the system to be nonlinear/linear, time-variant/invariant, realizable, etc.

In the following, we shall be exclusively concerned with linear time-variant/invariant systems but will not necessarily require that they be realizable.

6.2 Matched Filtering

In Chapter 2 we considered stochastic processes and described the impact of linear systems on these processes. Now, we develop techniques for designing linear filters to minimize the effect of noise.



The signal X_t could be either a signal in noise or noise only. The signal s_t is assumed to be deterministic. Additionally, we suppose that $\mu_U = \mathrm{E}(U_t) = 0$ and the spectrum $C_{UU}(\Omega)$ of the input noise U_t is known.

Now, we wish to determine the filter characteristics such that the instantaneous ratio of the output signal power to the output noise power is maximized at sampling time t_0, i.e.

$$\left(\frac{S}{N}\right)_{\text{out,max}} = \max_h \left(\frac{S}{N}\right)_{\text{out}} = \max_h \left|\tilde{s}_{t_0}\right|^2 \Big/ \mathrm{E}\left(\tilde{U}_t^2\right)$$

This problem typically arises in sonar and radar applications, where we wish to establish the presence and location of a signal s_t returning from a distant target.

Remark:

The matched filter does not preserve the waveform of the input signal. The objective is to distort the waveform and filter the noise such that at the sampling time t_0 the output signal level will be as large as possible compared to the output noise level.

Theorem:

The matched filter that maximizes

$$\left(\frac{S}{N}\right)_{\text{out}} = \left|\tilde{s}_{t_0}\right|^2 \Big/ \mathrm{E}\left(\tilde{U}_t^2\right)$$

has a transfer function given by

$$H_{opt}(\Omega) = k\,\frac{S(\Omega)^*}{C_{UU}(\Omega)}\,e^{-j\Omega t_0},$$

where

$$S(\Omega) = \mathcal{F}\{s_t\} \quad \text{and} \quad C_{UU}(\Omega) = \mathcal{F}\{c_{UU}(\tau)\}$$

are the Fourier transform of s_t and the spectrum of U_t, respectively, k is an arbitrary constant and t_0 is the sampling time when (S/N) is evaluated.

Exercise 6.2-1:
Proof of the Theorem

Remarks:

k is an arbitrary constant since the signal and the noise at the input are both multiplied by *k*. Thus *k* cancels in the relation for $(S/N)_{out}$.

The filter found may or may not be causal. If it is not causal, it has to be approximated by a causal filter.

The transfer function of the optimum filter is proportional to the complex conjugate of the spectrum of the input signal. Hence, we might say that the linear system is matched to the specified signal.

6.2.1 Matched Filtering for White Noise

Theorem:

Suppose the input noise is white. Then the impulse response of the matched filter is given by

$$h_{t,opt} = c \cdot s^*_{t_0 - t},$$

where c is an arbitrary constant, t_0 is the time of the peak signal output, and s_t is the known input signal waveform.

Consequently, the impulse response of the matched filter is simply a time reversed, complex conjugated and by t_0 translated version of the known signal waveform.

Therefore, the filter is said to be "matched to the signal".

Exercise 6.2-2:

Proof of the Theorem

$$x_t = \begin{cases} s_t + Z_t \\ Z_t \end{cases} \longrightarrow \boxed{\begin{matrix} h_t \\ H(\Omega) \end{matrix}} \longrightarrow \tilde{x}_t = \begin{cases} \tilde{s}_t + \tilde{Z}_t \\ \tilde{Z}_t \end{cases}$$

The signal-to-noise ratio at the output is given by

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|S(\Omega)\right|^2 d\Omega \Big/ \sigma_z^2 = \frac{1}{\sigma_z^2}\sum_{t=0}^{T-1}\left|s_t\right|^2 = \frac{E_s}{\sigma_z^2},$$

where $E_s = \sum_{t=0}^{T-1}\left|s_t\right|^2$ is the energy of the input signal of finite length T.

Remarks:

The signal-to-noise ratio at the output of the filter depends on the signal energy and power level of the noise and not on the particular signal waveform used.

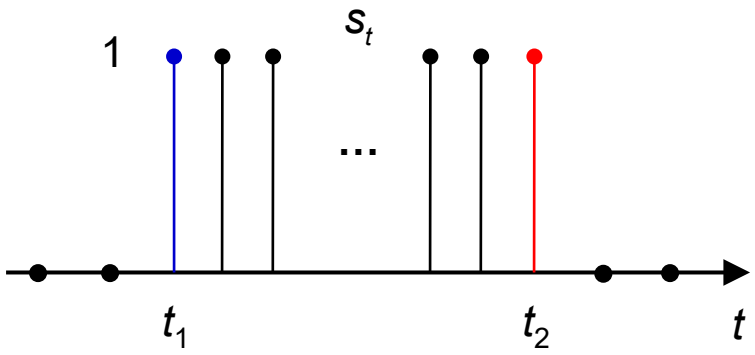
To improve the signal-to-noise ratio, we can increase the signal amplitude or the signal length.

Example:

We want to find the matched filter for the known signal

$$s_t = \begin{cases} 1 & \text{if } t_1 \le t \le t_2 \\ 0 & \text{elsewhere} \end{cases}$$

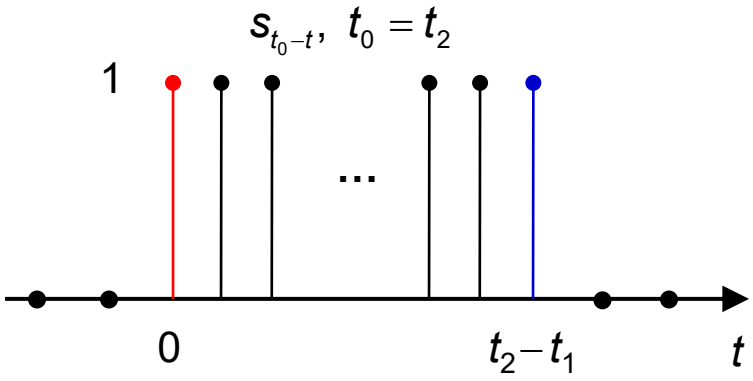
of finite extent $T = t_2 - t_1 + 1$, as visualized below,



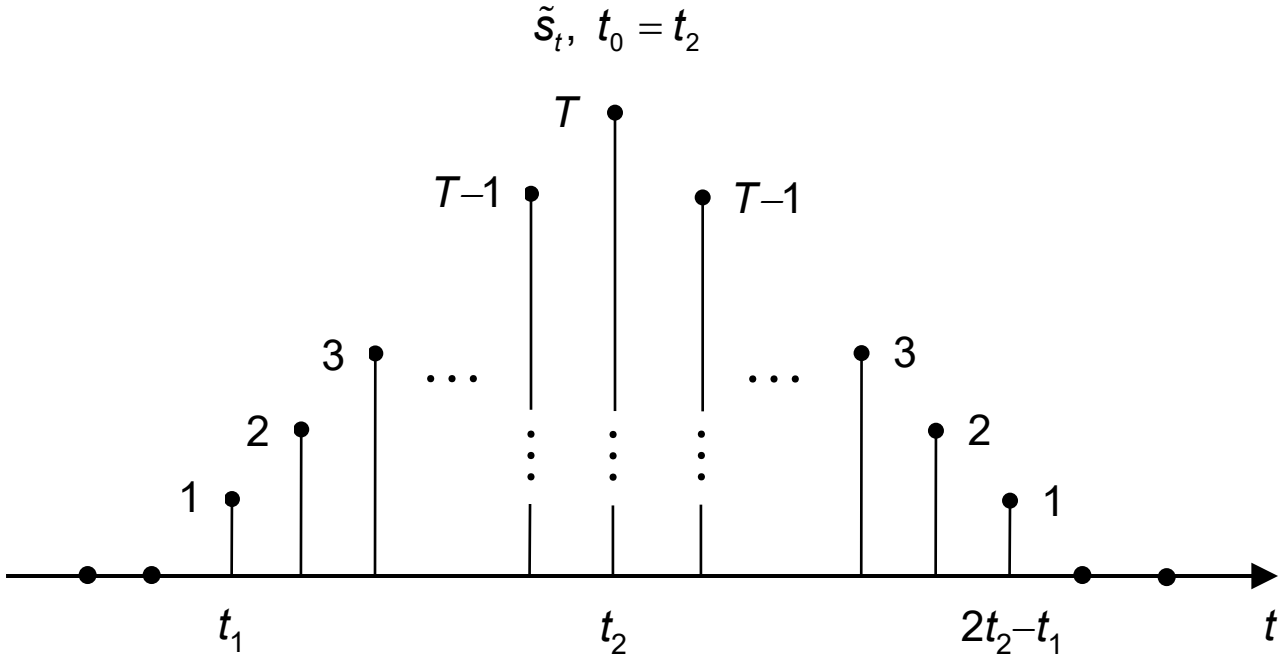
that is imbedded in additive white noise. Hence, the impulse response of the matched filter is given by

$$h_{t,opt} = s_{t_0 - t} = s_{-(t - t_0)}$$

In order to obtain a causal matched filter, we have to require $t_0 \geq t_2$. Choosing $t_0 = t_2$, the impulse response of the matched filter is shown below.



The signal component of the matched filter output, $\tilde{s}_t$, is depicted in the following figure.



The peak output level occurs at $t = t_0$. The input signal waveform has been distorted by the filter in order to peak up the output signal at $t = t_0$.

Exercise 6.2-3:

Matched filter design for exponentially decaying signals

6.2.2 Matched Filtering as Correlation Processing

Consider a known signal waveform s_t of finite support $t \in \{t_1, \ldots, t_2\}$ embedded in white noise Z_t, i.e. the signal extent is $T = t_2 - t_1 + 1$. The output of the matched filter at t_0 is

$$\tilde{X}_{t_0} = \sum_{\tau} X_\tau \, h_{t_0 - \tau, opt}.$$

Using the matched filter for white noise with $t_0 = t_2$, i.e.

$$h_{t,opt} = \begin{cases} s_{t_0 - t} & \text{if } 0 \le t < T \\ 0 & \text{elsewhere} \end{cases}$$

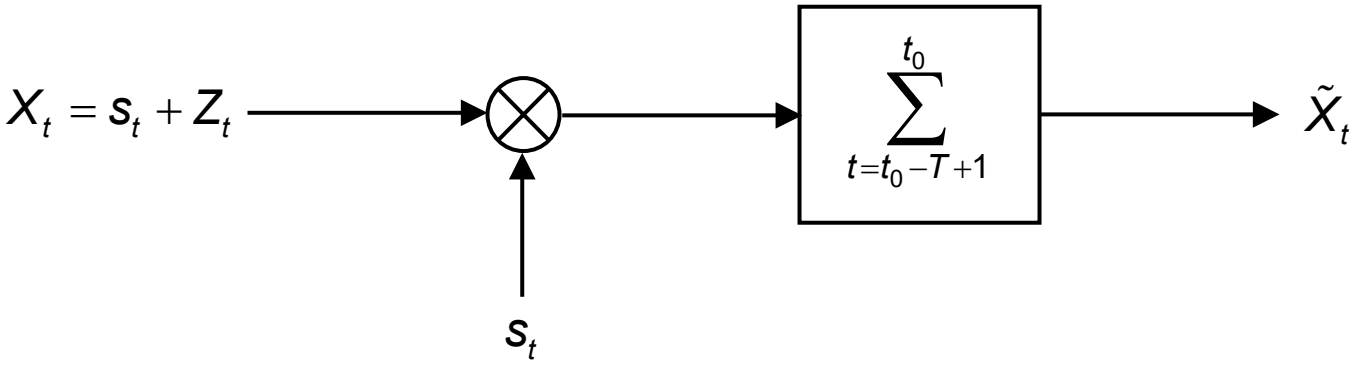
the summation is non-zero for

$$0 \le t_0 - \tau < T \;\Rightarrow\; t_0 - T + 1 \le \tau \le t_0.$$

and the convolution can be reformulated to

$$\tilde{X}_{t_0} = \sum_{\tau=t_0-T+1}^{t_0} X_\tau \, h_{t_0-\tau,opt} = \sum_{\tau=t_0-T+1}^{t_0} X_\tau \, s_{t_0-(t_0-\tau)} = \sum_{\tau=t_0-T+1}^{t_0} X_\tau \, s_\tau .$$

Hence, matched filtering can be interpreted as a correlation operation which is illustrated in the following figure.

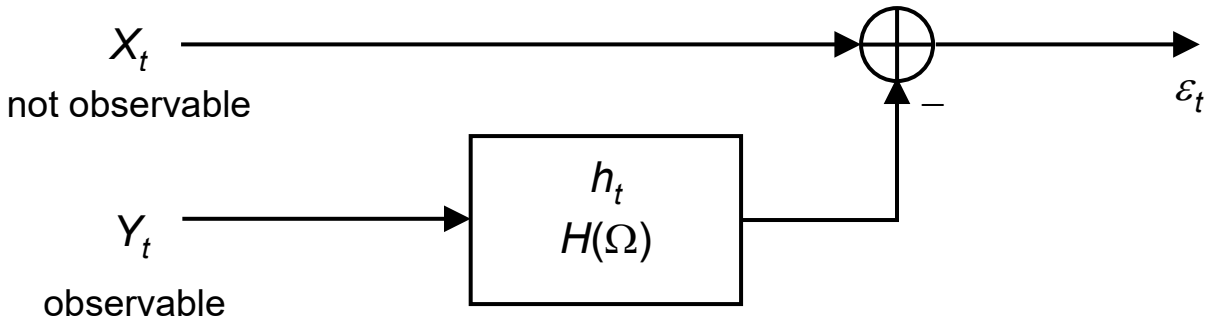


6.3 Wiener Filtering

The matched filter considered in the previous section is an optimal filter in the sense that it provides the highest SNR at the output for detecting the presence of a known signal.

The Wiener filter considered now aims to provide an optimal estimation of the realization of one stochastic process from observations of another stochastic process.

More specifically, we consider a system configuration as shown in the figure below, where X_t, Y_t and ε_t denote the stochastic process to be estimated, the observed stochastic process and the error process, respectively.



The goal is to design a linear time-invariant filter with impulse response h_t such that the expected value of the squared-error process, i.e. the MSE, is minimized. The filter which minimizes the MSE is known as Wiener filter.

For the following considerations we suppose that X_t and Y_t are real valued, zero-mean and jointly wide-sense stationary (WSS) stochastic processes.

Since the processes X_t and Y_t are jointly WSS and the filter with impulse response h_t is assumed to be stable, the error process ε_t is also WSS.

Hence, the MSE, which is the second-order moment of ε_t, does not depend on the index t. The MSE can be expressed in terms of the filter response h_t by

$$q(h_t) = \mathrm{E}\left(\varepsilon_t^2(h_t)\right) = \mathrm{E}\left(X_t - \sum_\tau h_\tau Y_{t-\tau} \right)^2$$

$$= \mathrm{E}(X_t^2) + \sum_{\tau_1}\sum_{\tau_2} h_{\tau_1} h_{\tau_2} \mathrm{E}(Y_{t-\tau_1} Y_{t-\tau_2}) - 2\sum_\tau h_\tau \mathrm{E}(X_t Y_{t-\tau})$$

$$= c_{XX}(0) + \sum_{\tau_1}\sum_{\tau_2} h_{\tau_1} h_{\tau_2} c_{YY}(\tau_2 - \tau_1) - 2\sum_\tau h_\tau c_{XY}(\tau).$$

The impulse response of the optimal (Wiener) filter $h_{t,opt}$ is defined by

$$h_{t,opt} = \underset{h_t \in \mathcal{H}}{\arg\min}\, q(h_t),$$

where $\mathcal{H}$ denotes the set of all absolutely summable impulse responses.

6.3.1 Wiener-Hopf Equation

Now, we would like to solve the MSE problem

$$\min_{h_t \in \mathcal{H}} q(h_t) = \min_{h_t \in \mathcal{H}} \mathrm{E}\left(X_t - \sum\nolimits_{\tau} h_\tau Y_{t-\tau} \right)^2 .$$

The solution of this problem is provided by exploiting the orthogonality principle stated in the following theorem.

Theorem: (Orthogonality Principle)

Suppose X_t and Y_t are jointly WSS. The impulse response $h_{t,opt} \in \mathcal{H}$ minimizes the MSE if and only if

$$\mathrm{E}\left[\left(X_t - \sum\nolimits_\tau h_{\tau,opt} Y_{t-\tau}\right) \sum\nolimits_\tau h_\tau Y_{t-\tau}\right] = 0 \quad \forall h_t \in \mathcal{H}.$$

For finding the solution of the minimization problem a more convenient form of the orthogonality condition is given by the following result.

Corollary:

$h_{t,opt} \in \mathcal{H}$ minimizes the MSE if and only if

$$\mathrm{E}\left[\left(X_t - \sum\nolimits_\tau h_{\tau,opt} Y_{t-\tau}\right) Y_u\right] = 0 \quad \forall u \in \mathbb{Z}.$$

Exercise 6.3-1:

Proof of the Theorem

Using the corollary, we obtain the equation that specifies the impulse response $h_{t,opt}$ of the optimum estimator.

Reformulation of the orthogonality condition

$$\mathrm{E}\left[\left(X_t - \sum\nolimits_{\tau} h_{\tau,opt} Y_{t-\tau}\right) Y_u\right] = 0 \quad \forall u \in \mathcal{U} \subset \mathbb{Z}$$

provides

$$\mathrm{E}\left(X_t Y_u\right) = \sum\nolimits_{\tau} h_{\tau,opt}\, \mathrm{E}\left(Y_{t-\tau} Y_u\right) \quad \forall u \in \mathcal{U} \subset \mathbb{Z}$$

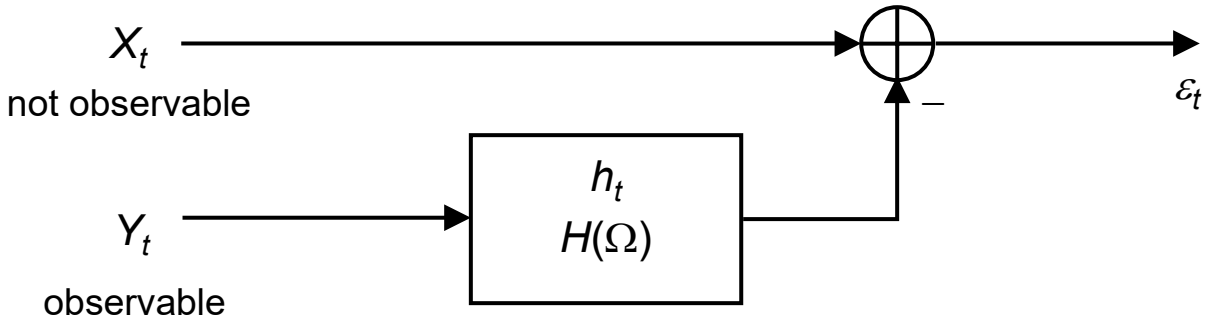
and finally

$$c_{XY}(v) = \sum\nolimits_{\tau} h_{\tau,opt}\, c_{YY}(v-\tau) \quad \forall v = t - u \in \mathcal{V} \subset \mathbb{Z}$$

which is known as Wiener-Hopf equation.

6.3.2 Finite Wiener Filtering

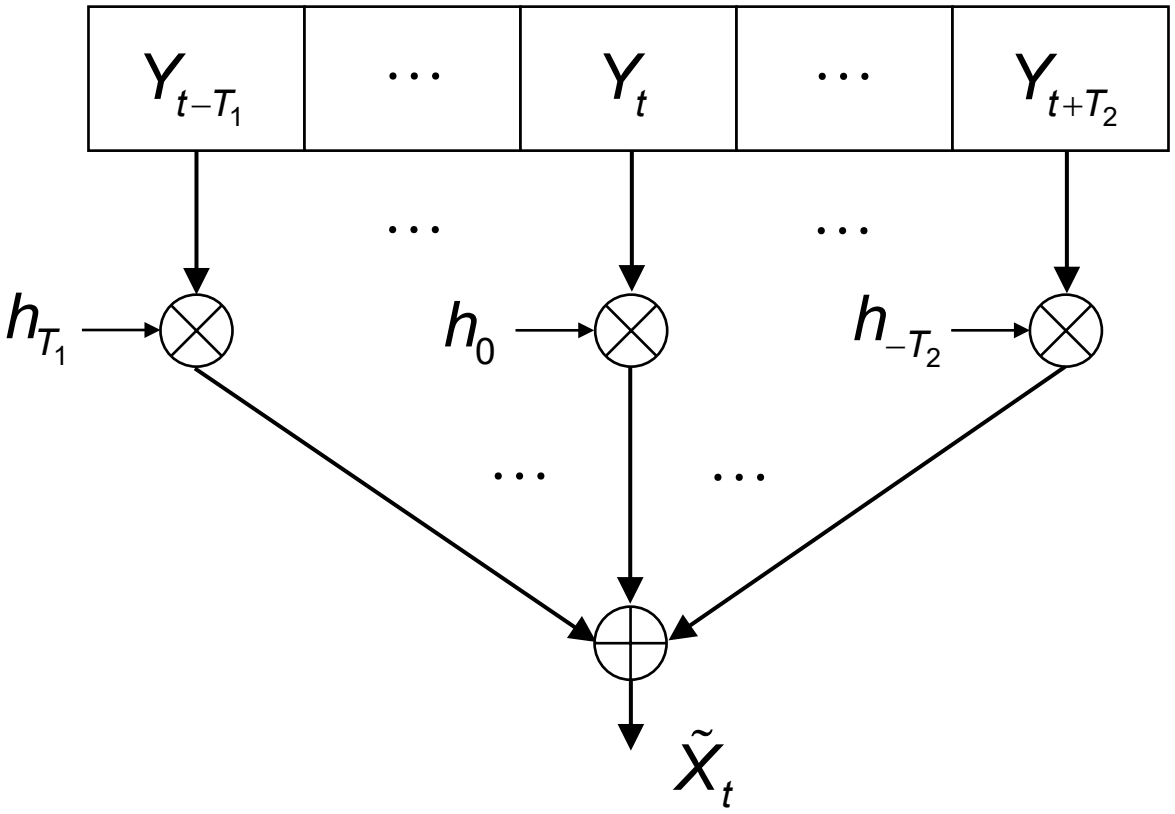
We consider now the problem



where the stochastic process Y_t is observed only over a finite discrete-time interval $\mathcal{U}=\{t-T_1,\ldots,t+T_2\}$ with $-T_1 \le T_2$. It is desired to obtain an estimate $\tilde{X}_t$ for X_t by applying a linear filter with impulse response h_t, i.e.

$$\tilde{X}_t = \sum_{\tau=t-T_1}^{t+T_2} h_{t-\tau} Y_\tau = \sum_{\tau=-T_2}^{T_1} h_\tau Y_{t-\tau}.$$

The filter h_t is therefore of finite length as shown below for $T_1, T_2 > 0$.



As discussed in the previous section, we wish to determine the optimum solution $h_{t,opt}$, which minimizes the MSE of the estimate $\tilde{X}_t$.

The relation between the discrete-time interval $\mathcal{U}$ and the time t at which X_t should be estimated gives rise to the following three types of estimation problems.

Filtering

Suppose that Y_t has been observed over the discrete-time interval $\mathcal{U} = \{t - T_1, \ldots, t\}$ with $T_1 > 0$. Then X_t has to be estimated from the most recent observations. The solution to this problem provides a causal filter which can be implemented in real-time.

Smoothing

If the observations are taken over the discrete-time interval $\mathcal{U} = \{t - T_1, \ldots, t + T_2\}$ with $T_1, T_2 > 0$ then X_t can be estimated from past and future observations. This is applicable in post-processing situations, when a realization of Y_t has been recorded and can be played back.

Prediction

Let Y_t be given over the discrete-time interval $\mathcal{U} = \{t - T_1, \ldots, t - k\}$ with $T_1 > k > 0$. Then X_t has to be predicted from past observations. Since the filtering procedure is defined by a linear operation $\tilde{X}_t$ represents a k-step linear predictor for X_t.

Now, the optimum solution $h_{t,opt}$ has to satisfy the Wiener-Hopf equation

$$c_{XY}(v) = \sum_{\tau=-T_2}^{T_1} h_{\tau,opt}\, c_{YY}(v-\tau) \quad \forall v \in \mathcal{V} = \{-T_2,\ldots,T_1\}$$

which can be expressed in matrix form as

$$\mathbf{C}_{YY}\mathbf{h}_{opt} = \mathbf{c}_{XY},$$

where

$$\mathbf{h}_{opt} = \begin{pmatrix} h_{-T_2,opt} \\ \vdots \\ h_{T_1,opt} \end{pmatrix}, \quad \mathbf{c}_{XY} = \begin{pmatrix} c_{XY}(-T_2) \\ \vdots \\ c_{XY}(T_1) \end{pmatrix}$$

and

$$\mathbf{C}_{YY} = \begin{pmatrix} c_{YY}(0) & c_{YY}(1) & \cdots & c_{YY}(T_1+T_2) \\ c_{YY}(1) & c_{YY}(0) & \cdots & c_{YY}(T_1+T_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ c_{YY}(T_1+T_2) & c_{YY}(T_1+T_2-1) & \cdots & c_{YY}(0) \end{pmatrix}.$$

Assuming that $\mathbf{C}_{YY}$ is positive definite (covariance matrices are at least nonnegative definite) the optimum impulse response is given by

$$\mathbf{h}_{opt} = \mathbf{C}_{YY}^{-1}\,\mathbf{c}_{XY}.$$

The optimum solution $\mathbf{h}_{opt}$ can be efficiently computed by the Levinson-Durbin algorithm which exploits the Toeplitz structure and the symmetry of $\mathbf{C}_{YY}$.

6.3.3 Noncausal Wiener Filtering

The following Wiener Filtering approach is termed non-causal because one wants to estimate X_t based on observations Y_u for all $u \in \mathcal{U} = \mathbb{Z}$. Thus

$$\tilde{X}_t = \sum_{\tau=-\infty}^{\infty} h_{t-\tau} Y_\tau = \sum_{\tau=-\infty}^{\infty} h_\tau Y_{t-\tau},$$

where the filtering operation is not necessarily causal, i.e. the impulse response may not satisfy $h_t = 0$ for $t < 0$.

Hence, the Wiener-Hopf equation can be written as

$$c_{XY}(v) = \sum_{\tau=-\infty}^{\infty} h_{\tau,opt}\, c_{YY}(v-\tau) \quad \forall v \in \mathcal{V} = \mathbb{Z},$$

where the right side of the equation represents a discrete

convolution of $h_{\tau,opt}$ and $c_{YY}(\nu)$ with $\nu = ...–1,0,1,...$

Assuming that the following Fourier transforms exist

$$H_{opt}(\Omega) = \sum_{\tau=-\infty}^{\infty} h_{\tau,opt}\, e^{-j\Omega\tau}, \quad C_{XY}(\Omega) = \sum_{\nu=-\infty}^{\infty} c_{XY}(\nu) e^{-j\Omega\nu} \quad \text{and}$$

$$C_{YY}(\Omega) = \sum_{\nu=-\infty}^{\infty} c_{YY}(\nu) e^{-j\Omega\nu} \quad \text{with} \quad -\pi \le \Omega \le \pi$$

the Wiener-Hopf equation becomes

$$C_{XY}(\Omega) = H_{opt}(\Omega) C_{YY}(\Omega), \quad -\pi \le \Omega \le \pi,$$

where $H_{opt}(\Omega)$, $C_{XY}(\Omega)$ and $C_{YY}(\Omega)$ denote the transfer function of the optimal filter, the power spectral density of Y_t and the cross power spectral density of X_t and Y_t.

Hence, the Wiener-Hopf equation can be easily solved for the transfer function of the optimum filter, i.e.

$$H_{opt}(\Omega) = \frac{C_{XY}(\Omega)}{C_{YY}(\Omega)}, \quad -\pi \le \Omega \le \pi,$$

from which the impulse response is obtained by

$$h_{t,opt} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{C_{XY}(\Omega)}{C_{YY}(\Omega)} e^{j\Omega t} d\Omega, \quad t \in \mathbb{Z}.$$

The minimum mean square error (MMSE)

$$q(h_{t,opt}) = \min_{h_t \in \mathcal{H}} q(h_t) = \min_{h_t \in \mathcal{H}} \mathrm{E}\left(X_t - \sum_{\tau=-\infty}^{\infty} h_\tau Y_{t-\tau} \right)^2$$

can now be expressed by

$$q(h_{t,opt}) = \mathrm{E}\left(X_t - \sum_{\tau=-\infty}^{\infty} h_{\tau,opt} Y_{t-\tau} \right)^2 = \mathrm{E}\left(\left(X_t - \sum_{\tau=-\infty}^{\infty} h_{\tau,opt} Y_{t-\tau} \right) X_t \right)$$

$$= \mathrm{E}(X_t^2) - \sum_{\tau=-\infty}^{\infty} h_{\tau,opt} \mathrm{E}(X_t Y_{t-\tau}) = c_{XX}(0) - \sum_{\tau=-\infty}^{\infty} h_{\tau,opt} c_{XY}(\tau).$$

Since $c_{XY}(\tau) = c_{YX}(-\tau)$ and $C_{YX}(\Omega) = C_{XY}^*(\Omega)$ the infinite sum in the equation above can be written as

$$\sum_{\tau=-\infty}^{\infty} h_{\tau,opt} c_{XY}(\tau) = \sum_{\tau=-\infty}^{\infty} h_{\tau,opt} c_{YX}(-\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_{opt}(\Omega) C_{YX}(\Omega) d\Omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{C_{XY}(\Omega)}{C_{YY}(\Omega)} C_{YX}(\Omega) d\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|C_{XY}(\Omega)|^2}{C_{YY}(\Omega)} d\Omega.$$

Moreover, exploiting

$$c_{XX}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} C_{XX}(\Omega) d\Omega$$

the MMSE can be rewritten as

$$q(h_{t,opt}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(C_{XX}(\Omega) - \frac{|C_{XY}(\Omega)|^2}{C_{YY}(\Omega)} \right) d\Omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1 - |R_{XY}(\Omega)|^2 \right) C_{XX}(\Omega) d\Omega,$$

where

$$|R_{XY}(\Omega)|^2 = \frac{|C_{XY}(\Omega)|^2}{C_{XX}(\Omega) C_{YY}(\Omega)}$$

denotes the so-called magnitude squared coherence.

Applying the results stated in Chapter 2.6.3, $R_{XY}(\Omega)$ can be expressed by

$$R_{XY}(\Omega) = \frac{\mathrm{Cov}\left(dZ_X(\Omega), dZ_Y(\Omega)\right)}{\sqrt{\mathrm{Var}\left(dZ_X(\Omega)\right)\mathrm{Var}\left(dZ_Y(\Omega)\right)}}$$

so that $R_{XY}(\Omega)$ can be interpreted as the correlation coefficient between the random components of X_t and Y_t at frequency Ω. Thus,

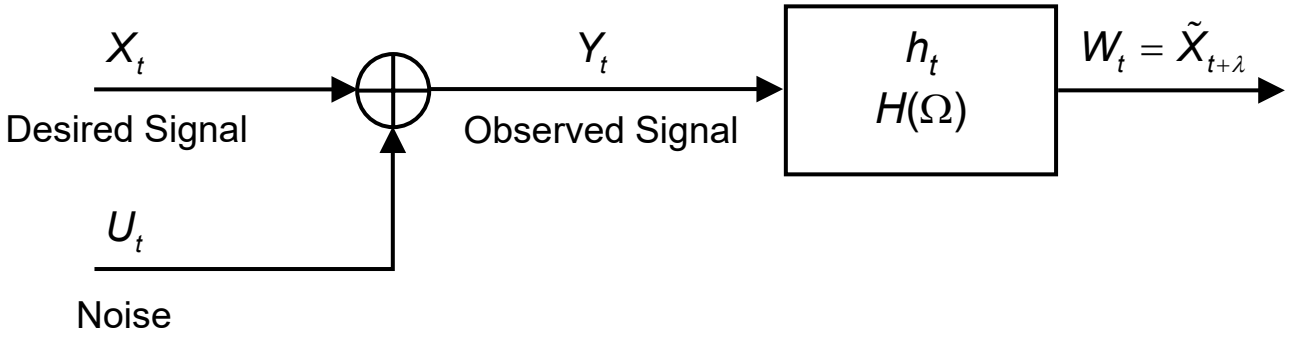
$$|R_{XY}(\Omega)|^2 \le 1, \quad -\pi \le \Omega \le \pi.$$

Moreover, the equality sign holds for all $\Omega \in [-\pi, \pi]$ if and only if X_t and Y_t are related by a linear transformation

$$X_t = \sum_{\tau=-\infty}^{\infty} h_\tau Y_{t-\tau}.$$

Exercise 6.3-2:

Signal estimation in additive noise, noncausal filtering



6.3.4 Causal Wiener Filtering

Noncausal Wiener filtering is improper for applications in which real-time estimation is required. Therefore, it is of practical relevance to consider the causal Wiener filtering problem in which we wish to estimate X_t based on observations Y_u for all $u \in \mathcal{U}=\{\tau \in \mathbb{Z}: \tau \le t\}$.

Hence, with the causal filter approach

$$\tilde{X}_t = \sum_{\tau=-\infty}^{t} h_{t-\tau} Y_\tau = \sum_{\tau=0}^{\infty} h_\tau Y_{t-\tau}$$

the Wiener-Hopf equation becomes

$$c_{XY}(v) = \sum_{\tau=0}^{\infty} h_{\tau,opt}\, c_{YY}(v-\tau) \quad \forall v \in \mathcal{V} = \mathbb{N}_0.$$

Since the Wiener-Hopf equation is only defined for $v \geq 0$ it can not be solved by Fourier transform. Before we can derive the solution we first have to discuss the so-called spectral factorization.

Spectral Factorization and Linear Representation

Suppose Y_t has an absolutely continuous and integrable spectral density $C_{YY}(\Omega)$. Then Y_t can be represented as noncausal filtered white noise

$$Y_t = \sum_{\tau=-\infty}^{\infty} g_\tau Z_{t-\tau} \quad \text{with} \quad \sum_{t=-\infty}^{\infty} |g_t|^2 < \infty, \; E(Z_t) = 0 \text{ and } c_{zz}(t) = \delta_t$$

and $C_{YY}(\Omega)$ can be written as

$$C_{YY}(\Omega) = G(\Omega) G(\Omega)^* \quad \text{with} \quad G(\Omega) = \sum_{t=-\infty}^{\infty} g_t \, e^{-j\Omega t}.$$

Moreover, if $C_{YY}(\Omega)$ satisfies the Paley-Wiener condition

$$\int_{-\pi}^{\pi} \log\left(C_{YY}(\Omega)\right) d\Omega > -\infty$$

an unique causal impulse response g_t with $g_0 > 0$ and $\sum_{t=0}^{\infty} |g_t|^2 < \infty$ exists, such that

$$\bar{G}(z) = \sum_{t=0}^{\infty} g_t \, z^{-t} \quad \text{with} \quad \bar{G}(e^{j\Omega}) = G(\Omega)$$

has no zeros outside the unit circle (minimum phase filter) and provides a factorization for $C_{YY}(\Omega)$ in the form

$$C_{YY}(\Omega) = G(\Omega)G(\Omega)^* = |G(\Omega)|^2 = |\bar{G}(e^{j\Omega})|^2.$$

Furthermore, $\sum_{t=0}^{\infty} |g_t|^2 < \infty$ implies that no poles of $\bar{G}(z)$ are lying outside the unit circle.

The Paley-Wiener condition is a fairly weak condition that may hold in any situation of practical interest.

However, if we impose the additional constraints

$$\sum_{t=0}^{\infty} |g_t| < \infty \quad \text{and} \quad \bar{G}(z) = \sum_{t=0}^{\infty} g_t \, z^{-t} \neq 0 \quad \forall |z| \geq 1$$

an explicit expression can be obtain for $\bar{G}(z)$.

The constraints imply that all zeros and poles of $\bar{G}(z)$ are lying within the unit circle. Thus, $\bar{G}(z)$ can be interpreted as the z-Transform of a linear, causal, stable, minimum phase and invertible stable filter.

Now, the spectral density of X_t given by

$$C_{YY}(\Omega) = |G(\Omega)|^2 = |\bar{G}(e^{j\Omega})|^2$$

can be analytically continued

$$\bar{C}_{YY}(z) = \bar{G}(z)\bar{G}(z^{-1}) = \sum_{t=-\infty}^{\infty} c_{YY}(t) z^{-t}$$

in an annulus that contains $|z| = 1$. Since

$$\bar{C}_{YY}(z) = \bar{C}_{YY}(z^{-1})$$

we can conclude that if $z_{0,i}$ and $z_{\infty,j}$ are zeros and poles of $\bar{C}_{YY}(z)$ respectively then $z_{0,i}^{-1}, z_{0,i}^{*}, z_{0,i}^{*\,-1}$ and $z_{\infty,j}^{-1}, z_{\infty,j}^{*}, z_{\infty,j}^{*\,-1}$ are also zeros and poles of $\bar{C}_{YY}(z)$.

Hence, we obtain the so-called canonical factorization

$$\bar{C}_{YY}(z) = \bar{C}_{YY}^{+}(z)\bar{C}_{YY}^{-}(z),$$

where $\bar{C}_{YY}^{+}(z)$ and $\bar{C}_{YY}^{-}(z)$ contain all zeros and poles of

$\bar{C}_{YY}(z)$ that are lying within or outside the unit circle respectively. Since the zeros and poles of $\bar{C}^{+}_{YY}(z)$ are related to the zeros and poles of $\bar{C}^{-}_{YY}(z)$ by mirroring at the unit circle we can write $\bar{C}^{-}_{YY}(z) = \bar{C}^{+}_{YY}(z^{-1})$.

Solution of the Wiener-Hopf Equation

Due to the preceding assumptions and explanations we are now able to solve the Wiener-Hopf equation. For this purpose we define the sequence

$$q_v = c_{XY}(v) - \sum_{\tau=0}^{\infty} h_{\tau,opt}\, c_{YY}(v-\tau)$$

which has obviously to satisfy $q_v = 0$ for all $v \geq 0$.

After applying the two-sided z-Transform we obtain

$$\begin{aligned}\bar{Q}(z) &= \bar{C}_{XY}(z) - \bar{H}_{opt}(z)\bar{C}_{YY}(z) \\ &= \bar{C}_{XY}(z) - \bar{H}_{opt}\,\bar{C}^{+}_{YY}(z)\bar{C}^{-}_{YY}(z),\end{aligned}$$

where the convolution theorem for z-Transforms and the canonical factorization has been exploited.

Since q_v is anticausal its z-Transform $\bar{Q}(z)$ does not contain a constant component and can only possess poles outside the unit circle.

Dividing the former equation by $\bar{C}^{-}_{YY}(z)$ we can write

$$\frac{\bar{Q}(z)}{\bar{C}^{-}_{YY}(z)} = \frac{\bar{C}_{XY}(z)}{\bar{C}^{-}_{YY}(z)} - \bar{H}_{opt}(z)\bar{C}^{+}_{YY}(z).$$

Due to the aforementioned properties of $\bar{Q}(z)$ and since $\bar{C}_{YY}^{-}(z)$ has only zeros outside the unit circle we can infer that $\bar{Q}(z)/\bar{C}_{YY}^{-}(z)$ does not contain a constant component and possesses only poles outside the unit circle.

Moreover, as the poles of $\bar{H}_{opt}(z)\bar{C}_{YY}^{+}(z)$ are lying within the unit circle its inverse z-Transform represents a causal sequence.

Hence, after defining the operation

$$\left[\bar{F}(z)\right]_{+} = \left[\sum_{t=-\infty}^{\infty} f_t\, z^{-t}\right]_{+} = \sum_{t=0}^{\infty} f_t\, z^{-t}$$

we can state that

$$\left[\bar{Q}(z)/\bar{C}_{YY}^{-}(z)\right]_{+}=0 \text{ and } \left[\bar{H}_{opt}(z)\bar{C}_{YY}^{+}(z)\right]_{+}=\bar{H}_{opt}(z)\bar{C}_{YY}^{+}(z).$$

Finally, the latter provides together with

$$\left[\bar{Q}(z)/\bar{C}_{YY}^{-}(z)\right]_{+}=\left[\bar{C}_{XY}(z)/\bar{C}_{YY}^{-}(z)\right]_{+}-\left[\bar{H}_{opt}(z)\bar{C}_{YY}^{+}(z)\right]_{+}$$

the desired solution for the Wiener filter in the z-domain

$$\bar{H}_{opt}(z)=\frac{1}{\bar{C}_{YY}^{+}(z)}\left[\frac{\bar{C}_{XY}(z)}{\bar{C}_{YY}^{-}(z)}\right]_{+}$$

which in the frequency domain can be expressed by

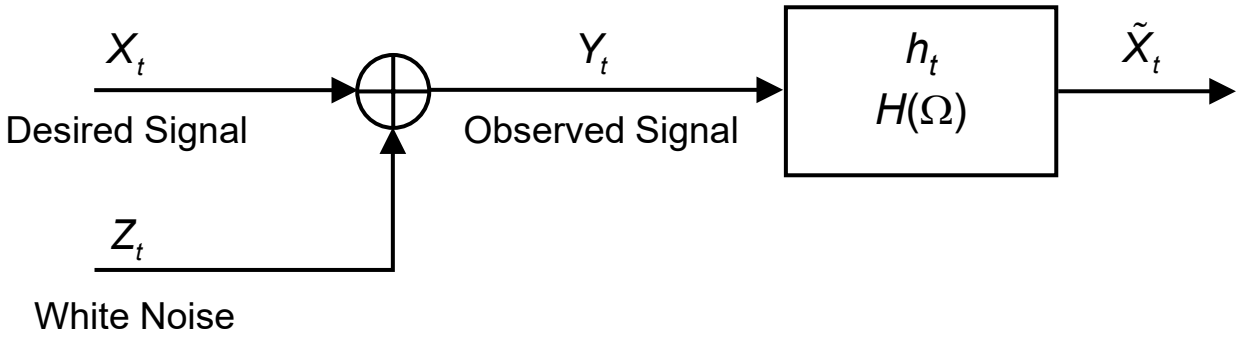
$$\bar{H}_{opt}(e^{j\Omega})=\frac{1}{\bar{C}_{YY}^{+}(e^{j\Omega})}\left[\frac{\bar{C}_{XY}(e^{j\Omega})}{\bar{C}_{YY}^{-}(e^{j\Omega})}\right]_{+}=\frac{1}{C_{YY}^{+}(\Omega)}\left[\frac{C_{XY}(\Omega)}{C_{YY}^{-}(\Omega)}\right]_{+}=H_{opt}(\Omega).$$

Exercise 6.3-3:

Solution of the Wiener-Hopf equation for white noise and its application after prewhitening

Exercise 6.3-4:

Signal estimation in additive white noise, causal filtering



6.4 Kalman Filtering

The Wiener approach considered in the previous section solved the MMSE problem for filtering, prediction and smoothing of scalar wide sense stationary processes, where the derivation of the optimum filter could be primarily considered as an frequency domain approach.

The Kalman approach addresses the filtering, prediction and smoothing problem of not necessarily stationary and vector valued processes. It provides solutions in the time domain by virtue of formulating the problem in the state space.

6.4.1 State Space Model

Discrete-time dynamic systems can be represented by a state equation

$$\mathbf{X}_{t+1} = \mathbf{f}_t(\mathbf{X}_t, \mathbf{U}_t), \quad t = 0,1,2,\ldots$$

and a measurement equation

$$\mathbf{Y}_{t+1} = \mathbf{h}_t(\mathbf{X}_t, \mathbf{V}_t), \quad t = 0,1,2,\ldots,$$

where $(\mathbf{f}_t)_{t\geq 0}$ and $(\mathbf{h}_t)_{t\geq 0}$ are sequences of functions, $(\mathbf{X}_t)_{t\geq 0}$ is a sequence in $\mathbb{R}^p$ describing the states of interest, $(\mathbf{U}_t)_{t\geq 0}$ is a sequence in $\mathbb{R}^q$ acting on $(\mathbf{X}_t)_{t\geq 1}$ and where $(\mathbf{Y}_t)_{t\geq 0}$ and $(\mathbf{V}_t)_{t\geq 0}$ are sequences in $\mathbb{R}^r$ representing the measurements and the measurement noise, respectively.

If the discrete-time system is linear the state and measurement equations can be expressed by

$$\mathbf{X}_{t+1} = \mathbf{F}_t \mathbf{X}_t + \mathbf{G}_t \mathbf{U}_t, \quad t = 0,1,2,\ldots,$$

and

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{V}_t, \quad t = 0,1,2,\ldots,$$

where $\mathbf{F}_t$, $\mathbf{G}_t$ and $\mathbf{H}_t$ are $p \times p$, $p \times q$ and $r \times p$ matrices, respectively, for each t.

Furthermore, if the system is time-invariant the matrices $\mathbf{F}_t$, $\mathbf{G}_t$ and $\mathbf{H}_t$ become constant coefficient matrices which are accordingly denoted by $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$. Hence, the state and measurement equations simplify to

$$\mathbf{X}_{t+1} = \mathbf{F}\mathbf{X}_t + \mathbf{G}\mathbf{U}_t \quad \text{and} \quad \mathbf{Y}_t = \mathbf{H}\mathbf{X}_t + \mathbf{V}_t, \quad t = 0,1,2,\ldots.$$

Exercise 6.4-1:

State space representation of the one-dimensional motion of a particle

6.4.2 State Estimation

Now, we suppose that the sequence $\mathbf{Y}_0,\ldots,\mathbf{Y}_t$ has been observed and that the state $\mathbf{X}_\tau$ should be estimated.

This estimation problem is known as

a) filtering problem if $t = \tau$,
b) smoothing problem if $t > \tau$,
c) prediction problem if $t < \tau$.

To estimate the state $\mathbf{X}_\tau$ by means of realizations of the measurement sequence $\mathbf{Y}_0,\ldots,\mathbf{Y}_t$ in the minimum mean square error sense we have to find an estimating function $\hat{\mathbf{X}}_\tau(\mathbf{y}_0,\ldots,\mathbf{y}_t)$ that minimizes

$$R_{MSE}(\hat{\mathbf{X}}_\tau) = \mathrm{E}\left(\left\| \mathbf{X}_\tau - \hat{\mathbf{X}}_\tau(\mathbf{Y}_0,\ldots,\mathbf{Y}_t) \right\|_2^2 \right).$$

We know from Chapter 1.10.4 or 3.7.1 that the optimum estimating function is given by the conditional mean, i.e.

$$\hat{\mathbf{X}}_\tau(\mathbf{y}_0,\ldots,\mathbf{y}_t) = \mathrm{E}\left(\mathbf{X}_\tau \mid \mathbf{Y}_0 = \mathbf{y}_0,\ldots,\mathbf{Y}_t = \mathbf{y}_t\right).$$

However, we are usually interested in generating estimates in real time as *t* increases. Since the data increase linear with *t,* an efficient calculation of the conditional mean will be impossible unless suitable restrictions on the system model structure are imposed, e.g. restriction to discrete-time linear systems with independent input and measurement noise sequences of independent zero-mean Gaussian random vectors.

6.4.3 Kalman Filter Approach

Within the assumptions mentioned above a computational efficient and simultaneous solution of the filtering and one step prediction problem can be stated.

Theorem:

For the linear, finite-dimensional, discrete-time system

$$\mathbf{X}_{t+1} = \mathbf{F}_t \mathbf{X}_t + \mathbf{G}_t \mathbf{U}_t \quad \text{and} \quad \mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{V}_t, \quad t = 0,1,2,\ldots,$$

where $(\mathbf{U}_t)$ and $(\mathbf{V}_t)$ are independent sequences of independent zero-mean Gaussian vectors which are independent of the Gaussian initial condition $\mathbf{X}_0$ with

$$\mathrm{E}(\mathbf{U}_t \mathbf{U}_t^T) = \boldsymbol{\Sigma}_{\mathbf{UU}}(t),\ \mathrm{E}(\mathbf{V}_t \mathbf{V}_t^T) = \boldsymbol{\Sigma}_{\mathbf{VV}}(t),\ \mathrm{E}(\mathbf{X}_0) = \boldsymbol{\mu}_0,\ \mathrm{Cov}(\mathbf{X}_0) = \boldsymbol{\Sigma}_0$$

the estimators

$$\mathbf{X}_{t|t} = \mathrm{E}(\mathbf{X}_t \,|\, \mathbf{Y}_0, \ldots, \mathbf{Y}_t) \quad \text{and} \quad \mathbf{X}_{t+1|t} = \mathrm{E}(\mathbf{X}_{t+1} \,|\, \mathbf{Y}_0, \ldots, \mathbf{Y}_t)$$

can be recursively determined by the following equations.

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \mathbf{K}_t \left(\mathbf{Y}_t - \mathbf{H}_t \mathbf{X}_{t|t-1} \right) \quad t = 0, 1, 2, \ldots$$

and

$$\mathbf{X}_{t+1|t} = \mathbf{F}_t \mathbf{X}_{t|t} \quad t = 0, 1, 2, \ldots$$

with initialization $\mathbf{X}_{0|-1} = \boldsymbol{\mu}_0$ and Kalman gain matrix

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \, \mathbf{H}_t^T \left(\mathbf{H}_t \, \boldsymbol{\Sigma}_{t|t-1} \, \mathbf{H}_t^T + \boldsymbol{\Sigma}_{\mathbf{VV}}(t) \right)^{-1},$$

where

$$\boldsymbol{\Sigma}_{t|t-1} = \mathrm{Cov}(\mathbf{X}_t \,|\, \mathbf{Y}_0,\ldots,\mathbf{Y}_{t-1})$$
$$= \mathrm{E}\left((\mathbf{X}_t - \mathbf{X}_{t|t-1})(\mathbf{X}_t - \mathbf{X}_{t|t-1})^T \,|\, \mathbf{Y}_0,\ldots,\mathbf{Y}_{t-1}\right)$$

is the covariance of the prediction error which can be computed jointly with the filtering error covariance

$$\boldsymbol{\Sigma}_{t|t} = \mathrm{Cov}(\mathbf{X}_t \,|\, \mathbf{Y}_0,\ldots,\mathbf{Y}_t) = \mathrm{E}\left((\mathbf{X}_t - \mathbf{X}_{t|t})(\mathbf{X}_t - \mathbf{X}_{t|t})^T \,|\, \mathbf{Y}_0,\ldots,\mathbf{Y}_t\right)$$

by the recursion

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1}, \quad t = 0,1,2,\ldots$$

and

$$\boldsymbol{\Sigma}_{t+1|t} = \mathbf{F}_t \boldsymbol{\Sigma}_{t|t} \mathbf{F}_t^T + \mathbf{G}_t \boldsymbol{\Sigma}_{\mathbf{UU}}(t) \mathbf{G}_t^T, \quad t = 0,1,2,\ldots$$

with the initialization $\boldsymbol{\Sigma}_{0|-1} = \boldsymbol{\Sigma}_0$.

Exercise 6.4-2:

Proof of the Theorem

The recursions consist of the following two basic steps.

Measurement update

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \mathbf{K}_t \left(\mathbf{Y}_t - \mathbf{H}_t \mathbf{X}_{t|t-1} \right) \text{ and } \mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{\Sigma}_{t|t-1}$$

which updates the

- state estimate of $\mathbf{X}_t$ by incorporating the new measurement $\mathbf{Y}_t$,
- filter error covariance matrix.

Time update

$$\mathbf{X}_{t+1|t} = \mathbf{F}_t \mathbf{X}_{t|t} \text{ and } \mathbf{\Sigma}_{t+1|t} = \mathbf{F}_t \mathbf{\Sigma}_{t|t} \mathbf{F}_t^T + \mathbf{G}_t \mathbf{\Sigma}_{\mathbf{UU}}(t) \mathbf{G}_t^T$$

which provides the

- one-step prediction of the state estimate,
- prediction error covariance matrix.

The measurement update equation

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \mathbf{K}_t \left(\mathbf{Y}_t - \mathbf{H}_t \mathbf{X}_{t|t-1} \right)$$

can be viewed as a combination of the predicted state vector and a correction term. Since

$$\begin{aligned} \mathbf{Y}_{t|t-1} &= \mathrm{E}(\mathbf{Y}_t \,|\, \mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1}) \\ &= \mathbf{H}_t \, \mathrm{E}(\mathbf{X}_t \,|\, \mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1}) + \mathrm{E}(\mathbf{V}_t \,|\, \mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1}) = \mathbf{H}_t \, \mathbf{X}_{t|t-1} \end{aligned}$$

the difference in the correction term can be interpreted as an error signal

$$\mathbf{I}_t = \mathbf{Y}_t - \mathbf{Y}_{t|t-1} = \mathbf{Y}_t - \mathbf{H}_t \mathbf{X}_{t|t-1},$$

which is known as innovation. The term innovation comes from the fact that $\mathbf{I}_t$ is the part of

$$\mathbf{Y}_t = \mathbf{Y}_{t|t-1} + \mathbf{I}_t$$

that cannot be predicted and therefore contains the new information that is gained by the current observation.

Remarks:

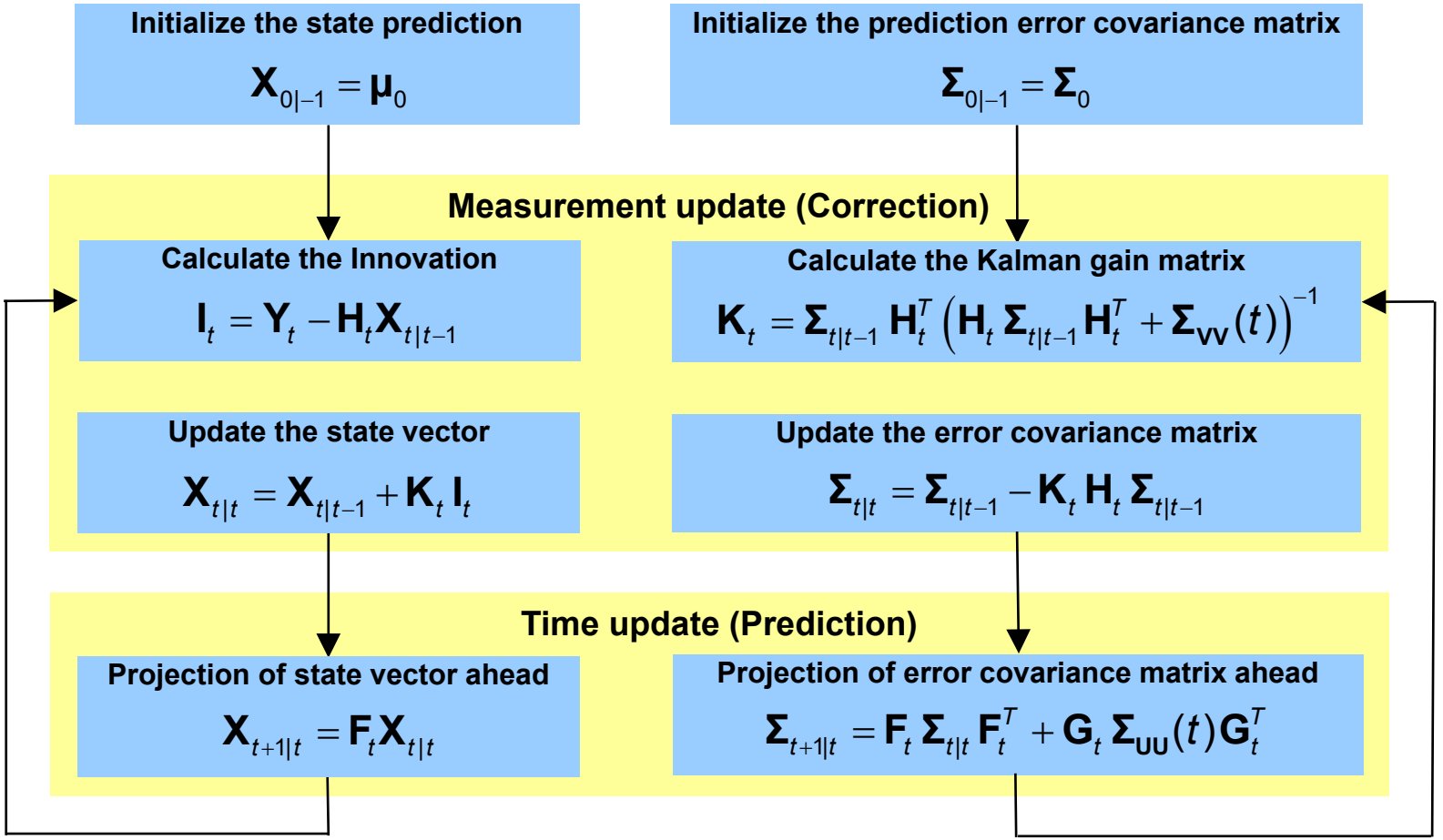
The innovation sequence $\mathbf{I}_t$ is a sequence of independent not identically distributed zero-mean Gaussian random vectors.

$\mathbf{Y}_t$ consists of a part, $\mathbf{Y}_{t|t-1}$, that is completely dependent and a part, $\mathbf{I}_t$, that is completely independent of the past. Thus $\mathbf{I}_t$ provides a set of independent observations that forms suitably scaled the output of a prewhitening operation.

Exercise 6.4-3:

Proof of the remarks

Kalman Filtering Algorithm



Exercise 6.4-4:
Scalar Kalman filter

Exercise 6.4-5:
Track-While-Scan Radar with independent accelerations from scan to scan

Exercise 6.4-6:
Track-While-Scan Radar with dependent accelerations from scan to scan

Exercise 6.4-7:
State space representation of AR(p)-, MA(q)-, and ARMA(p,q)-Processes

References to Chapter 6

[1] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Dover, 2005

[2] A. Gelb, *Applied Optimal Estimation*, MIT Press, 1974

[3] T. Kailath, A.H. Sayed and B. Hassibi, *Linear Estimation*, Prentice Hall, 2000

[4] S. Kay, *Fundamentals of Statistical Signal Processing, Vol. 1: Estimation Theory*, Prentice Hall, 1993

[5] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, 1998

[6] M.B. Priestley, *Spectral Analysis and Time Series*, Academic Press, 1984

[7] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, 1994