3. ESTIMATION OF SIGNALS USING A LEAST SQUARES TECHNIQUE

3.0 INTRODUCTION

The purpose of this chapter is to introduce estimators shortly. More elaborated courses on System Identification, which are given by Prof. Johan Schoukens, exist. When one speaks about "*measurement*" in practice, "*estimation*" is meant. Indeed, besides the deterministic component to be measured, noise will be present as well, which is a stochastic component. One will limit the study here to *Least Square Estimators* only, although more powerful versions exist (e.g. *Maximum Likelihood Estimators*). The orthogonality principle will be repeated in order to derive some filters.

3.1 LEAST SQUARES ESTIMATION OF THE VALUE OF A STOCHASTIC VALUE BY A CONSTANT

Let x be a stochastic variable and a a constant. The estimation of x in the sense of the least squares approach means that a constant a as to be found so that the following error becomes minimal:

$$\overline{(x - a)^2} \triangleq \int_{-\infty}^{+\infty} (x - a)^2 \,.\, w_{(1)}(x)\, dx$$

The solution to this problem is given by:

$$\boxed{a = \overline{x} = \mu}$$

Indeed:

$$\overline{(x - a)^2} = a^2 - 2a\,\overline{x} + \overline{x^2}$$

Minimizing yields:

$$\frac{d\left[\overline{(x - a)^2}\right]}{da} = 0 \qquad \left(\frac{\partial}{\partial a} = \frac{d}{da} \text{ here}\right)$$

and hence:

$$2a - 2\overline{x} = 0 \supset a = \overline{x} = \mu$$

3.2 LINEAR LEAST SQUARES ESTIMATION

Suppose that a stochastic variable y has to be estimated by a linear function ax + b, where x also is a stochastic variable and where a and b are constants. What will be the values of the coefficients a and b when a Least Squares approach is followed? To solve this problem, the following cost function (mean square of the error between observation y and linear model ax + b) must be minimized:

$$\overline{\left\{ y - (ax+b) \right\}^2}$$

Remark: The resulting least squares error between the observation and the linear model can be larger than when a nonlinear model of higher order would be applied; e.g. a polynomial in x.

The constants a and b that will minimise the Least Squares Error:

$$e \triangleq \overline{\left\{ y - (ax+b) \right\}^2} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y - ax - b)^2 \cdot w_{(2)}(x,y)\ dx\, dy$$

are given by:

$$a = \frac{r\,\sigma_y}{\sigma_x}$$

$$b = \overline{y} - a\,\overline{x} = \mu_y - a\,\mu_x$$

The minimised error will be given by:

$$e_m = \sigma_y^2\,(1 - r^2)$$

where r represents the *correlation function of Pearson*:

$$r = \frac{\overline{(x-\mu_x)\cdot(y-\mu_y)}}{\sqrt{\overline{(x-\mu_x)^2}\cdot\overline{(y-\mu_y)^2}}}$$

This result has been shown already often earlier[1]. To proof this, suppose

[1] Note that the correlation function r is identical here to the normalized covariance ρ in the study of the white Gaussian noise in Chapter 2, but applied here to two stochastic variables x and y.

that the value of a is known. Then b is found as the Least Squares Estimation of the variable (y-ax) by a constant. And hence:

$$b = \overline{y - ax} = \bar{y} - a\bar{x} = \mu_y - a\mu_x$$

With b determined as such, the least squares error e_m becomes:

$$e = \overline{(y - ax - b)^2} = \overline{(y - ax - \mu_y + a\mu_x)^2} = \overline{\left[(y - \mu_y) - a(x - \mu_x)\right]^2}$$

This can be rewritten as:

$$\overline{(y - \mu_y)^2 + a^2(x - \mu_x)^2 - 2a(y - \mu_y).(x - \mu_x)}$$

$$= \overline{(y - \mu_y)^2} + a^2\overline{(x - \mu_x)^2} - 2a\overline{(y - \mu_y).(x - \mu_x)}$$

Minimising yields:

$$\frac{\partial e}{\partial a} = 0 = 2a\overline{(x - \mu_x)^2} - 2\overline{(y - \mu_y).(x - \mu_x)}$$

$$0 = a\left(\overline{x^2} - 2\bar{x}\mu_x + \mu_x^2\right) - \overline{(y - \mu_y).(x - \mu_x)}$$

$$0 = a\left(\overline{x^2} - \mu_x^2\right) - r\sqrt{\overline{(x - \mu_x)^2}\,.\,\overline{(y - \mu_y)^2}}$$

$$\frac{\partial e}{\partial a} = 0 \Rightarrow -r\sigma_x\sigma_y + \sigma_x^2 a = 0$$

$$\Rightarrow a = \frac{r\sigma_y}{\sigma_x}$$

and hence:

$$e_m = \overline{\left[(y - \mu_y) - \frac{r\sigma_y}{\sigma_x}\,.\,(x - \mu_x)\right]^2}$$

$$= \overline{(y - \mu_y)^2} + \frac{r^2\sigma_y^2}{\sigma_x^2}\,.\,\overline{(x - \mu_x)^2} - 2\frac{r\sigma_y}{\sigma_x}\overline{(x - \mu_x).(y - \mu_y)}$$

$$= \sigma_y^2 + r^2\frac{\sigma_y^2}{\sigma_x^2}\,.\,\sigma_x^2 - 2\frac{r\sigma_y}{\sigma_x}\,.\,r\sigma_x\sigma_y$$

$$e_m = \sigma_y^2 - 2\, r\, \sigma_x \sigma_y \left[\frac{r\,\sigma_y}{\sigma_x} \right] + \sigma_x^2 \left[\frac{r^2 \sigma_y^2}{\sigma_x^2} \right]$$

$$\Rightarrow \quad \boxed{e_m = \sigma_y^2 \left[1 - r^2 \right]}$$

3.3 PRINCIPLE OF ORTHOGONALITY

Suppose that x and y are zero mean processes. The constant a that is found via the Least Squares Estimator minimising the error between y and the model ax:

$$e = \overline{\left(y - ax \right)^2}$$

is the so that (y-ax) is orthogonal, i.e. uncorrelated with x [2].

$$\overline{\left(y - ax \right) . x} = 0$$

and e_m is given by:

$$e_m = \overline{\left(y - ax \right) . y}$$

Proof:

Suppose (y-ax) and x are uncorrelated; i.e. that:

$$\boxed{\overline{\left(y - ax \right) . x} = 0}$$

One has to demonstrate that then the error e_m is minimal. For any arbitrary A different from a ($A \neq a$)one finds:

$$\overline{\left(y - Ax \right)^2} = \overline{\left(y - ax + ax - Ax \right)^2} = \overline{\left[\left(y - ax \right) + \left(a - A \right) x \right]^2}$$

$$= \overline{\left(y - ax \right)^2} + 2 \left(a - A \right) . \left[\overline{\left(y - ax \right) . x} \right] + \left(a - A \right)^2 \overline{x^2}$$

$$\Downarrow \qquad\qquad \Downarrow \qquad \Downarrow$$

$$0 \qquad\qquad > 0 \qquad > 0$$

[2] Only in the case that x and y are zero mean processes, uncorrelatedness and orthogonality are equal properties. Two variables x and y are orthogonal if their inner product is zero. Two variables x and y are uncorrelated if the inner product of ($x - \mu_x$) and ($y - \mu_y$) is zero. Hence for zero mean processes uncorrelated and orthogonal are the same.

and hence:

$$\overline{(y - Ax)^2} > \overline{(y - ax)^2}$$

So that only with a the error is minimal.

The minimal error e_m becomes further:

$$e_m = \overline{(y^2 - axy - axy + a^2x^2)}$$

$$= \overline{(y - ax).y} - a . \overline{(y - ax).x}$$

$$\Downarrow$$

$$0$$

Geometrical interpretation.

The quantity:

$$\overline{(y - ax)^2}$$

is the square of the length (y-ax). That length is minimal if is orthogonal to x: (y-ax) ⊥ x. This is shown in Figure 3.1.

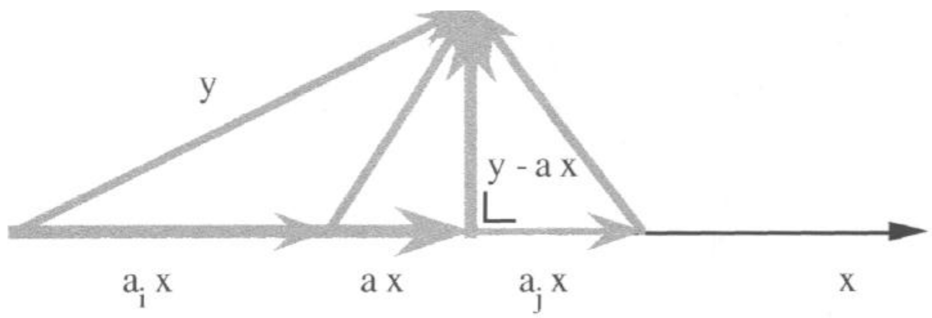

Figure 3.1: Geometric interpretation of the orthogonality principle.

3.4 THE WIENER FILTER

Suppose a stochastic signal s(t) that is corrupted with noise n(t). The stochastic signal x(t) = s(t) + n(t) has 2 components, the signal s(t) and the noise n(t); or the noise is considered to be purely additive. Suppose further that x(t) can be observed at any time instant; i.e. $\forall t \in [-\infty,+\infty]$. The goal is to estimate s(t) applying a linear operator (filter) on the data x(t).

The problem posed can be rephrased now in the framework of telecommunications. The signal s(t) is generated by the source at the side of the transmitter. It is the information bearing signal that has to be transmitted to a receiver over a communication channel. The latter will introduce noise n(t). The received signal at the end of the channel is represented by x(t). The communication channel, therefore, is regarded as being linear and distortionless. Radio communication channels, where the transmission is achieved via electromagnetic wave propagation of modulated signals in the medium air, can be modelled – in first order attempts – as such.

The signals x(t) and s(t) are supposed to be wide sense stationary and ergodic. The basic principle for the application of the Wiener filter is sketched in Figure 3.2.

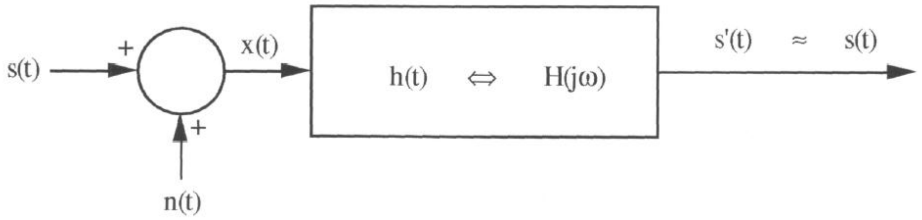

Figure 3.2: The application of the Wiener filter.

The response s'(t) of the linear time invariant system is given by the convolution of x(t) with the impulse response h(t) of the Wiener filter. The fact that the filter can be modeled by a linear time invariant system motivates the assumption that both s(t) and x(t) should be stationary.

$$s'(t) \quad = \quad \int_{-\infty}^{+\infty} x(t-\alpha)\ .\ h(\alpha)\ d\alpha$$

The impulse response that minimises the error between s(t) and the output of the filter s'(t):

$$\overline{|\ s(t)\ -\ s'(t)\ |^2}$$

is the one that corresponds to the Wiener filter.

One will now try to find the impulse response h(t). Hereto, the orthogonality principle will be applied. The condition to estimate in an optimal linear way a signal s(t) in the sense of the minimisation of the least squares error between estimate s'(t) and the signal s(t) is given by the orthogonality principle: the *optimal estimation error* (s(t)- s'(t)) and the *observation* x(t) should be *orthogonal*; i.e. *uncorrelated*. The latter means that their cross correlation should be zero:

$$\overline{\left[\, s(t) - s'(t) \,\right] . x(t)} = 0$$

Or:

$$\overline{\left\{ \left[s(t) - \int_{-\infty}^{+\infty} x(t-\alpha) . h(\alpha)\, d\alpha \right] . x(\xi) \right\}} = 0 \qquad \forall\, \xi$$

With the ergodicity requirement for the signals this can be rewritten as:

$$R_{sx}(t-\xi) = \int_{-\infty}^{+\infty} R_{xx}(t-\alpha-\xi)\, h(\alpha)\, d\alpha \qquad \forall\, \xi$$

Let $\tau = t - \xi$, then:

$$R_{sx}(\tau) = \int_{-\infty}^{+\infty} R_{xx}(\tau-\alpha)\; h(\alpha)\, d\alpha \qquad \forall\, \tau$$

This integral equation notes the condition to which the impulse response of the *Wiener filter* should satisfy. It is known as the *Wiener-Hopf* integral equation, which can be solved easily when operating in the frequency domain.

Let:

$$S_{xx}(\omega) = \Im\left(R_{xx}(\tau)\right) \qquad (1)$$
$$S_{sx}(\omega) = \Im\left(R_{sx}(\tau)\right) \qquad (2)$$
$$H(j\omega) = \Im\left(h(t)\right) \qquad (3)$$

Relation (1) is simply the relation of Wiener-Khintchine for ergodic signals. Relation (2) is the extension of the theorem of Wiener-Khintchine, which is also valid for the cross correlation (instead of the auto correlation). It can be shown, without proof here, that the cross power spectrum density is found from the application of the Fourier integral on the cross correlation function. Relation (3) is the result of the application of the Fourier integral on the impulse response h(t) of a linear time invariant system, which yields the transfer function $H(j\omega)$.

Applying the Fourier integral on both sides of the Wiener-Hopf integral equations, taking into account the convolution theorem, results in:

$$S_{sx}(\omega) = S_{xx}(\omega) . H(j\omega)$$

Hence, the transfer function and the impulse response of the filter can be found:

$$H(j\omega) = \frac{S_{sx}(\omega)}{S_{xx}(\omega)} \quad \Leftrightarrow \quad h(t) = \Im^{-1}\left\{ \frac{S_{sx}(\omega)}{S_{xx}(\omega)} \right\}$$

The least squares error (LSE) is given by:

$$e = \overline{\left[s(t) - \int_{-\infty}^{+\infty} x(t-\alpha)\, h(\alpha)\, d\alpha \right] . s(t)}$$

$$= \overline{s^2(t)} - \int_{-\infty}^{+\infty} \overline{x(t-\alpha) . s(t)}\ . h(\alpha)\, d\alpha$$

Or, expressed in terms of correlation functions, the LSE can be written as:

$$e = R_{ss}(0) - \int_{-\infty}^{+\infty} R_{sx}(\alpha)\, h(\alpha)\, d\alpha$$

In case one would prefer to express the error in the frequency domain, this can be computed easily also. Let:

$$\chi(\tau) = R_{ss}(\tau) - R_{sx}(-\tau) * h(\tau)$$

Then, it is clear that the LSE can be written as: $e = \chi(0)$. The application of the Fourier integral on $\chi(\tau)$ yields:

$$\Im\big(\chi(\tau)\big) = S_{ss}(\omega) - S_{sx}(-\omega) . H(j\omega) = S_{ss}(\omega) - \frac{S_{sx}(-\omega) . S_{sx}(\omega)}{S_{xx}(\omega)}$$

Remember, further, that:

$e^{j\omega\tau} = 1$ since $\tau = 0$

Hence:

$$e = \chi(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[S_{ss}(\omega) - \frac{S_{sx}(\omega) . S_{sx}(-\omega)}{S_{xx}(\omega)} \right] d\omega$$

3.5 THE WIENER FILTER IN CASE THAT THE SIGNAL AND THE NOISE ARE UNCORRELATED

In case that the signal s and the noise n are uncorrelated, their cross correlation, and hence also their cross power spectrum is zero:

$$S_{sn}(\omega) = 0$$

Because of the linearity, one can write also that:

$$S_{xx}(\omega) = S_{ss}(\omega) + S_{nn}(\omega)$$

and also:

$$S_{sx}(\omega) = S_{ss}(\omega) + S_{sn}(\omega)$$

So that $S_{sx}(\omega) = S_{ss}(\omega)$.

Hence:

$$H(j\omega) = \frac{S_{ss}(\omega)}{S_{ss}(\omega) + S_{nn}(\omega)}$$

The LSE can be formulated further as:

$$e = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{S_{ss}(\omega) \,.\, S_{nn}(\omega)}{S_{ss}(\omega) + S_{nn}(\omega)} d\omega$$

In case that the power spectra of the signal and the noise do not overlap, i.e. in case that $S_{ss}(\omega).S_{nn}(\omega) = 0$, one finds that:

$H(j\omega) = 1$ for $\omega \supset\!- S_{ss}(\omega) \neq 0$;

$H(j\omega) = 0$ for $\omega \supset\!- S_{nn}(\omega) \neq 0$;

$H(j\omega)$: arbitrary for all other ω

Examples:

In the Figure 3.3 some examples are sketched.

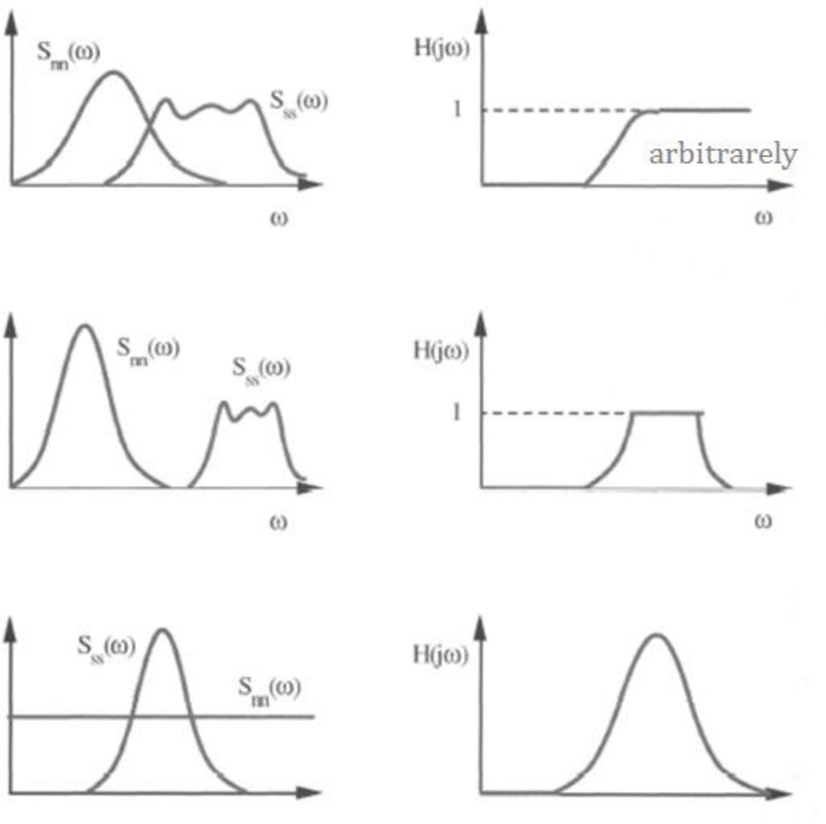

Figure 3.3: The application of the Wiener filter on some examples.

3.6 THE MATCHED FILTER

The problem that will be treated here is part of the *detection theory* that will be discussed in Chapter 5. The linear detection problem will be partly covered in this paragraph in order to compare the Matched filter with the already developed Wiener filter.

The problem to be solved is to detect a signal s(t) in the presence of additive noise n(t).

If n(t) is *white noise* then the optimal linear filter that optimizes the Signal-to-Noise Ratio (SNR) in a certain predefined time instance t_0, will be given by the Matched filter. Suppose that x(t) is observed:

$$x(t) = s(t) + n(t)$$

where the noise n(t) has a power spectral density (PSD) $W_n(\omega)$ and t_0 is the time instance, where the SNR of the output of the filter has be maximised. h(t) will be noted as the impulse response of a *linear time invariant system*, and hence, the filter will have a fixed structure, with constant, i.e. non time variant parameters. This implies that the applied signals must be stationary as well. Filters that adapt their parameters in function of a time varying statistic of the stochastic signals are used in practice as well; e.g. (extended)

Kalman filters, particle filters etc. However, in this analysis the stochastic signals are supposed to be *wide sense stationary* and *ergodic*.

The signal at the output of the filter $y(t_0)$ will have two components:

- $y_s(t_0)$ that describes the signal $s(t)$;
- $y_n(t_0)$ that is provided by the noise $n(t)$.

$$y_s(t_0) = \int_{-\infty}^{+\infty} s(\tau) \, . \, h(t_0 - \tau) \, d\tau$$

$$y_n(t_0) = \int_{-\infty}^{+\infty} n(\tau) \, . \, h(t_0 - \tau) \, d\tau$$

Problem: Maximise the SNR at time instance t_0 :

$$\left(\frac{\mathfrak{S}}{\mathfrak{N}}\right)_0 \triangleq \frac{\overline{y_s^2(t_0)}}{\overline{y_n^2(t_0)}}$$

Since the component $y_s(t_0)$ will have a fixed value ($s(t)$ is known $\forall t$, and hence is regarded as being deterministic) and will depend only on the choice of the time instance t_0, the SNR becomes:

$$\left(\frac{\mathfrak{S}}{\mathfrak{N}}\right)_0 \triangleq \frac{y_s^2(t_0)}{\overline{y_n^2(t_0)}}$$

This is the ratio of the power of the signal at time instance $t = t_0$ to the average power of the noise at the same time instance. This ratio must be maximised when considering all possible impulse responses $\{h(t)\}$ of the linear time invariant filter, and retaining the one that results into maximal SNR.

The solution of this problem can be found easily when the *Schwarz inequality* is written:

$$\left| \int_{-\infty}^{+\infty} A_1(\omega) \, . \, A_2(\omega) \, d\omega \right|^2 \leq \int_{-\infty}^{+\infty} \left| A_1(\omega) \right|^2 d\omega \, . \int_{-\infty}^{+\infty} \left| A_2(\omega) \right|^2 d\omega$$

with equality only and only if:

$$A_1(\omega) = k \, . \, A_2^*(\omega)$$

where k = a constant and $A_i(\omega)$ must be a quadratic integrable (square-integrable) function, else the existence of the integrals will not be guaranteed.

The SNR can be rewritten because of the ergodicity as:

$$\left(\frac{\mathfrak{S}}{\mathfrak{N}}\right)_0 \triangleq \frac{y_s^2(t_0)}{\overline{y_n^2(t_0)}} = \frac{\left| \int_{-\infty}^{+\infty} H(\omega) \, . \, S(\omega) \, . \, e^{j\omega t_0} \, d\omega \right|^2}{2\pi \int_{-\infty}^{+\infty} \left| H(\omega) \right|^2 . \, W_n(\omega) \, d\omega}$$

Since:

$$y_s^2(t_0) = \left| h(t) * s(t) \right|^2_{t=t_0} = \left| \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(\omega) \, . \, S(\omega) \, . \, e^{j\omega t_0} \, d\omega \right|^2$$

and:

$$\overline{y_n^2(t_0)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| H(\omega) \right|^2 . \, W_n(\omega) \, d\omega$$

where $S(\omega) = \Im(s(t))$ and $H(\omega) = \Im(h(t))$.

This can be found also when considering the convolution theory valid for time invariant systems, as is expressed in the Figure 3.4.

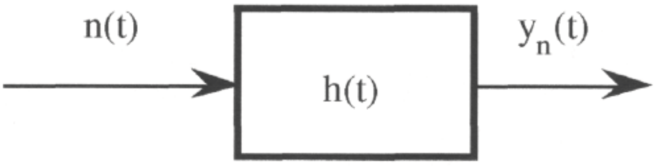

Figure 3.4: The noise signal is applied to the Matched filter.

$$y_n(t) = h(t) * n(t)$$

Hence, the power spectrum density of the noise can be written using the convolution theorem as:

$$Y_n(\omega) = \left| H(\omega) \right|^2 . \, W_n(\omega)$$

Using the relations of Wiener-Khintchine the autocorrelation function can be retrieved, when ergodicity is taken into account, as:

$$R_n(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| H(\omega) \right|^2 . \, W_n(\omega) \, . \, e^{j\omega\tau} \, d\omega$$

The power of the noise signal, however, corresponds to the value of the autocorrelation function at the time origin. And hence:

$$P = \overline{y_n^2(t)} = R_n(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(\omega)|^2 \, W_n(\omega)\, d\omega$$

Let in the inequality of Schwarz the functions $A_i(\omega)$ be:

$$A_1(\omega) \equiv H(\omega) \,.\, \sqrt{W_n(\omega)}$$

$$A_2(\omega) \equiv S(\omega) \,.\, e^{j\omega t_0} \,/\, \sqrt{W_n(\omega)}$$

Then, one finds that:

$$\left| \int_{-\infty}^{+\infty} H(\omega) \,.\, S(\omega) \,.\, e^{j\omega t_0}\, d\omega \right|^2 \leq \int_{-\infty}^{+\infty} |H(\omega)|^2 \,.\, W_n(\omega)\, d\omega \,.\, \int_{-\infty}^{+\infty} \frac{|S(\omega)|^2}{W_n(\omega)}\, d\omega$$

since:

$$\left| e^{j\omega t_0} \right| = 1$$

Since both factors in the right hand side are positive, dividing by 2π yields:

$$\frac{\left| \int_{-\infty}^{+\infty} H(\omega) \,.\, S(\omega) \,.\, e^{j\omega t_0}\, d\omega \right|^2}{2\pi \int_{-\infty}^{+\infty} |H(\omega)|^2 \,.\, W_n(\omega)\, d\omega} \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{|S(\omega)|^2}{W_n(\omega)}\, d\omega$$

The left hand side, obviously, corresponds to the SNR that has to be maximised! Hence:

$$\left(\frac{\mathfrak{S}}{\mathfrak{N}}\right)_0 \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{|S(\omega)|^2}{W_n(\omega)}\, d\omega$$

The maximum of the SNR, therefore, is known. The equality is valid only and only if:

$$H(\omega) \,.\, \sqrt{W_n(\omega)} = \frac{k \,.\, S^*(\omega) \,.\, e^{-j\omega t_0}}{\sqrt{W_n(\omega)}}$$

The optimal transfer function can be written now then as:

$$H(\omega) \equiv \frac{k \cdot S^*(\omega) \cdot e^{-j\omega t_0}}{W_n(\omega)}$$

In case that the additive noise is white, then the PSD of n(t) becomes : $W_n(\omega) = \sigma^2$. The transfer function of the Matched filter then becomes:

$H(\omega) \equiv K \cdot S^*(\omega) \cdot e^{-j\omega t_0}$; with $K \equiv k/\sigma^2$, a gain.

The impulse response then becomes:

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(\omega) \cdot e^{j\omega t} \, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} K \cdot S^*(\omega) \cdot e^{j\omega(t - t_0)} \, d\omega$$

$$= s^*(t - t_0)$$

and hence, if the signal is real:

$$h(t) = s(t - t_0)$$

What are the differences in a priori knowledge between the Wiener filter and the Matched filter? In case the noise is white, than the impulse response of the matched filter is equal to the signal shifted over the delay t_0. And hence, the filter is ‘matched’ to the signal.