3 The Wiener Filter

The Wiener filter solves the signal estimation problem for stationary signals.

The filter was introduced by Norbert Wiener in the 1940's. A major contribution was the use of a statistical model for the estimated signal (the Bayesian approach!). The filter is optimal in the sense of the MMSE.

As we shall see, the Kalman filter solves the corresponding filtering problem in greater generality, for non-stationary signals.

We shall focus here on the discrete-time version of the Wiener filter.

Contents

3.1 Preliminaries: Random Signals and Spectral Densities

Recall that a (two-sided) discrete-time signal is a sequence

$$x := (\dots, x_{-1}, x_0, x_1, \dots) = (x_k,\ k \in \mathbb{Z})\,.$$

Assume $x_k \in \mathbb{R}^{d_x}$ (vector-valued signal).
Its Fourier and (two-sided) $\mathcal{Z}$ transforms:

$$\begin{aligned} X(e^{j\omega}) &= \sum_{k=-\infty}^{\infty} x_k e^{-j\omega k}\,, \quad \omega \in [-\pi, \pi] \\ X(z) &= \sum_{k=-\infty}^{\infty} x_k z^{-k}\,, \quad z \in D_x \end{aligned}$$

A random signal x is a sequence of random variables. Its probability law may be specified through the joint distributions of all finite vectors $(x_{k_1}, \dots, x_{k_N})$.

The correlation function is defined as:

$$R_x(k, m) = E(x_k x_m^T)\,.$$

The signal is wide-sense stationary (WSS) if, $\forall k, m$

(*i*) $E[x_k] = E[x_0]$

(*ii*) $R_x(k, m) = R_x(m - k)$.

We then consider $R_x(k)$ as the auto-correlation function.
Note that R_x is even: $R_x(-k) = R_x(k)^T$.

The power spectral density of a WSS signal:

$$S_x(e^{j\omega}) \doteq \mathcal{F}\{R_x\} = \sum_{k=-\infty}^{\infty} R_x(k) e^{-j\omega k}\,, \quad \omega \in [-\pi, \pi]$$

It is well defined when $\sum_{k=-\infty}^{\infty} |R_x(k)| < \infty$ (component-wise).

Basic properties:

- $S_x(e^{-j\omega})^T = S_x(e^{j\omega})$
- $S_x(e^{j\omega}) \geq 0$ (non-negative definite).

In the scalar case, $S_x(e^{j\omega})$ is real and non-negative.

Two WSS stationary processes are jointly-WSS if

$$R_{xy}(k,m) \doteq E(x_k y_m^T) = R_{xy}(m-k)\,.$$

Their joint power spectral density is defined as above:

$$S_{xy}(e^{j\omega}) \doteq \mathcal{F}\{R_{xy}\} = S_{yx}(e^{-j\omega})^T$$

.

A linear time-invarant system may be defined through its kernel (or impulse response) h:

$$y_n = \sum_{k=-\infty}^{\infty} h_{n-k} x_k\,.$$

In the frequency domain this gives (for a stable system):

$$Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})\,.$$

If $x_k \in \mathbb{R}^{d_x}$ and $y_k \in \mathbb{R}^{d_y}$, then H is a $d_y \times d_x$ matrix.

The system is (BIBO) stable if $\sum_{k=-\infty}^{\infty} |h_k| < \infty$. Equivalently, all poles of (each component of) $H(z)$ are with $|z| < 1$.

When x is WSS stationary and the system is stable, then y is WSS and

$$S_y(e^{j\omega}) = H(e^{j\omega})S_x(e^{j\omega})H(e^{-j\omega})^T\,,$$
$$S_{yx}(e^{j\omega}) = H(e^{j\omega})S_x(e^{j\omega}) = S_{xy}(e^{-j\omega})^T\,.$$

The power spectral density can be extended to the z domain, and similar relations hold (wherever the transforms are well defined):

$$S_x(z) \doteq \mathcal{Z}(R_x) = \sum_{k=-\infty}^{\infty} R_x(k) z^{-k}$$
$$S_{xy}(z^{-1})^T = S_{yx}(z)$$
$$S_y(z) = H(z)S_x(z)H(z^{-1})^T\,.$$

White noise: A 0-mean process (w_k) with $R_w(k) = R_0\delta_k$ is called a (stationary, wide sense) *white noise*.
Obviously, $S_w(\omega) = R_0$: the spectral density is constant.

A common way to create a 'colored' noise sequence is to filter a white noise sequence, namely $y = h * w$. Then

$$S_y(z) = H(z) R_0 H(z^{-1})^T \,.$$

3.2 Wiener Filtering - Problem Formulation

We are given two processes:

- s_k, the signal to be estimated
- y_k, the observed process

which are jointly wide-sense stationary, with known covariance functions: $R_s(k)$, $R_y(k)$, $R_{sy}(k)$.

A particular case is that of a signal corrupted by additive noise:

$$y_k = s_k + n_k$$

with $(s_k,\, n_k)$ jointly stationary, and $R_s(k)$, $R_n(k)$, $R_{sn}(k)$ given.

Goal: Estimate s_k as a function of y. Specifically:

- Find the <u>linear MMSE</u> estimate of s_k based on (all or part of) (y_k).

There are three versions of this problem:

a. The causal filter: $\hat{s}_k = \sum_{m=-\infty}^{k} h_{k-m} y_m .$

b. The non-causal filter: $\hat{s}_k = \sum_{m=-\infty}^{\infty} h_{k-m} y_m .$

c. The FIR filter: $\hat{s}_k = \sum_{m=k-N}^{k} h_{k-m} y_m$

The first problem is the hardest.

Note: We consider in this chapter the *scalar* case for simplicity.

3.3 The "FIR Wiener Filter"

Consider the FIR filter of length $N+1$:

$$\hat{s}_k = \sum_{m=k-N}^{k} h_{k-m} y_m = \sum_{i=0}^{N} h_i y_{k-i} \, .$$

We need to find the coefficients (h_i) that minimize the MSE:

$$E(s_k - \hat{s}_k)^2 \quad \longrightarrow \quad \min$$

To find (h_i), we can differentiate the error. More conveniently, start with the orthogonality principle:

$$E[(s_k - \hat{s}_k) y_{k-j}] = 0 \, , \quad j = 0, 1, \ldots, N$$

This gives

$$\sum_{i=0}^{N} h_i E[y_{k-i} y_{k-j}] = E(s_k y_{k-j})$$
$$\sum_{i=0}^{N} h_i R_y(i-j) = R_{sy}(j) \, .$$

In matrix form:

$$\begin{bmatrix} R_y(0) & R_y(1) & \ldots & R_y(N) \\ R_y(1) & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & R_y(1) \\ R_y(N) & \ddots & R_y(1) & R_y(0) \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ \vdots \\ h_N \end{bmatrix} = \begin{bmatrix} R_{sy}(0) \\ \vdots \\ \vdots \\ R_{sy}(N) \end{bmatrix}$$

or

$$R_y h = r_{sy} \quad \Rightarrow h = R_y^{-1} r_{sy} \, .$$

These are the <u>Yule-Walker equations</u>.

- We note that $R_y \geq 0$ (positive semi-definite), and non-singular except for degenerate cases. Further, it is a Toeplitz matrix (constant along the diagonals). There exist efficient algorithms (Levinson-Durbin and others) that utilize this structure to efficiently compute h.

The MMSE may now be easily computed:

$$\begin{aligned} E[(\hat{s}_k - s_k)^2] &= E[(\hat{s}_k - s_k)(-s_k)] \qquad \text{(orthogonality)} \\ &= R_s(0) - E(\hat{s}_k s_k) \\ &= R_s(0) - h^T r_{sy} \end{aligned}$$

3.4 The Non-causal Wiener Filter

- Assume that $\hat{s}_k$ may depend on the entire observation signal, $(y_k,\ k \in Z)$.
- The "FIR" approach now fails since the transversal filter has infinitely many coefficients. We therefore use spectral methods.

The required filter has the form:

$$\hat{s}_k = \sum_{i=-\infty}^{\infty} h_i \, y_{k-i} \, .$$

Orthogonality implies: $(\hat{s}_k - s_k) \perp y_{k-j}$, $j \in Z$. Therefore

$$E(s_k \, y_{k-j}) = E(\hat{s}_k \, y_{k-j}) = \sum_{i=-\infty}^{\infty} h_i \, E(y_{k-i} \, y_{k-j}) \, ,$$

or

$$R_{sy}(j) = \sum_{i=-\infty}^{\infty} h_i \, R_y(j-i) \, , \qquad -\infty < j < \infty \, .$$

Using (two-sided) $\mathcal{Z}$-transform:

$$H(z) = \frac{S_{sy}(z)}{S_y(z)} \, .$$

This gives the optimal filter as a transfer function (which is rational if the signal spectra are rational in z).

3.5 The Causal Wiener Filter

We now look for a causal estimator of the form:

$$\hat{s}_k = \sum_{i=0}^{\infty} h_i y_{k-i} \ .$$

Proceeding as above we obtain

$$R_{sy}(j) = \sum_{i=0}^{\infty} h_i \, R_y(j-i) \, , \qquad j \geq 0.$$

This is the Wiener-Hopf equation.

Because of its "one-sidedness", a direct solution via $\mathcal{Z}$ transform does not work. We next outline two approaches for its solution, starting with some background on spectral factorization.

3.5.1 Some Spectral Factorizations

A. Canonical (Wiener-Hopf) factorization:

Given: a spectral density function $S_x(z)$.
Goal: Factor it as

$$S_x(z) = S_x^+(z) \, S_x^-(z) \, ,$$

where S_x^+ is stable and with stable inverse, and $S_x^-(z) = S_x^+(z^{-1})^T$.

Note: this factorization gives a "model" of the signal x as filtered white noise (with unit covariance).

Example: $X(z) = H(z)W(z)$, with $H(z) = \frac{z-3}{z-0.5}$, and $S_w(z) = 1$ (white noise). Then

$$S_x(z) = H(z) \cdot 1 \cdot H(z^{-1}) = \frac{(z-3)(z^{-1}-3)}{(z-0.5)(z^{-1}-0.5)}$$

and

$$S_x^+(z) = \frac{z^{-1}-3}{z-0.5} \, , \quad S_x^-(z) = \frac{z-3}{z^{-1}-0.5} \, .$$

Theorem. Assume:

(i) $S_x(z)$ is rational.

(ii) $S_x(z) = S_x(z^{-1})^T$ (hence, $S_x(e^{j\omega}) = S_x^T(e^{-j\omega})$).

(iii) S_x has no poles on $|z| = 1$, and $S_x(e^{j\omega}) > 0$ for all ω.

Then S_x may be factored as above, with $S_x^+(z)$:

(1) stable: all poles in $|z| < 1$.

(2) stable-inverse: all zeros of det $S^+(z)$ in $|z| < 1$.

Assume henceforth that $S_y(z)$ admits such a decomposition.

Remark: In the non-scalar case, obtaining this decomposition by transfer-function methods is hard. For rational transfer functions, the most efficient methods use state-space models (and Kalman filter equations).

B. Additive Decomposition:

Suppose $H(z)$ has no poles on $|z| = 1$.
We can write it as

$$H(z) = \{H(z)\}_+ + \{H(z)\}_- \,,$$

where: $\{(H(z)\}_+$ has only poles with $|z| < 1$,
$\{H(z)\}_-$ has only poles with $|z| > 1$.

By convention, a constant factor is included in the $\{\ \ \}_+$ term.

Suppose $H(z)$ is a (two-sided) Z transform of a finite-energy signal: $h(k) \in \ell^2(\mathbb{R})$. Then the above decomposition corresponds to the "future" and "past" projections of $h(k)$:

$$\begin{aligned} \{H(z)\}_+ &= \mathcal{Z}\{h_+(k)\}\,, \quad h_+(k) = h(k)u(k) \\ \{H(z)\}_- &= \mathcal{Z}\{h_-(k)\}\,, \quad h_-(k) = h(k)u(-k+1) \end{aligned}$$

3.5.2 The Wiener-Hopf Approach

Recall the Winer-Hopf equation for h:

$$R_{sy}(j) = \sum_{i=0}^{\infty} h_i R_y(j-i) \ , \qquad \underline{j \geq 0} \ .$$

Define

$$g_j = R_{sy}(j) - \sum_{i=0}^{\infty} h_i R_y(j-i) \ , \qquad \underline{-\infty < j < \infty} \ .$$

Obviously, $g_j = 0$ for $j \geq 0$.

Assume that $g_j \to 0$ as $j \to -\infty$ (which holds under "reasonable" assumptions on the covariances).
It follows that $G(z) = \mathcal{Z}\{g_j\}$ has no poles in $|z| < 1$.
Defining $h_j \doteq 0$ for $j < 0$, we obtain:

$$G(z) = S_{sy}(z) - H(z)\, S_y(z) \ .$$

Using the canonical decomposition $S_y = S_y^+ S_y^-$ gives:

$$G(z)\, [S_y^-(z)]^{-1} = S_{sy}(z)\, [S_y^-(z)]^{-1} - H(z)\, S_y^+(z) \ .$$

Note that the first term has no poles with $|z| < 1$, and the last term has no poles with $|z| > 1$. Applying $\{ \quad \}_+$ to both sides gives:

$$0 = \left\{ S_{sy}(z)\, [S_y^-(z)]^{-1} \right\}_+ - H(z)\, S_y^+(z)$$

and

$$\boxed{H(z) = \left\{ S_{sy}(z)\, [S_y^-(z)]^{-1} \right\}_+ [S_y^+(z)]^{-1}} \ .$$

Remark: A similar solution holds for the prediction problem:

- Estimate $s(k+N)$ based on $\{y(m),\ m \leq k\}$.

The same formula is obtained for $H(z)$, with an additional z^N term inside the $\{ \quad \}_+$.
The required additive decomposition can usually be handled using the time-domain interpretation.

3.5.3 The Innovations Approach

[Bode & Shannon, Zadeh & Ragazzini]

The idea: "whitening" the observation process before estimation.

Let

$$G(z) := [S_y^+(z)]^{-1}\,, \qquad (g_k) = \mathcal{Z}^{-1}\{G(z)\}\,.$$

Then $\tilde{y}_k := g_k * y_k$ has $S_{\tilde{y}}(z) \equiv I$. That is, $\tilde{y}$ is a normalized white noise process.

Note that $G(z)$ is causal with causal inverse, so that $\tilde{y}$ holds the same information as y.

Consider the filtering problem with white noise measurement $\{\tilde{y}\}$.
The Wiener-Hopf equation for the corresponding filter $\tilde{h}$ gives:

$$R_{s\tilde{y}}(j) = \sum_{i=0}^{\infty} \tilde{h}_i R_{\tilde{y}}(j-i) \;= \tilde{h}_j\,, \qquad\qquad j \geq 0\,.$$

Thus,

$$\tilde{H}(z) = \{S_{s\tilde{y}}(z)\}_+$$

Now

$$S_{s\tilde{y}}(z) = S_{sy}(s)G(z^{-1})^T = S_{sy}(z)[S_y^-(z)]^{-1}\,,$$

and $H(z) = \tilde{H}(z)\,G(z)$ gives the Wiener filter.

We note that $\tilde{y}_k$ is the *innovations process* corresponding to y_k.

3.6 The Error Equation

For the optimal filter, the orthogonality principle implies the following expression for the minimal MSE:

$$\text{MMSE} = E(s_t - \hat{s}_t)^2 = R_s(0) - R_{\hat{s}s}(0)$$

Therefore

$$\text{MMSE} = \frac{1}{2\pi}\int_0^{2\pi} [S_s(e^{j\omega}) - S_{\hat{s}s}(e^{j\omega})]d\omega$$

where $S_{\hat{s}s}(z) = H(z)S_{ys}(z) = H(z)S_{sy}(z^{-1})$ since $\hat{s} = Hy$.

This can also be expressed in terms of the Z-transform: Let

$$z = e^{j\omega}, \text{ hence } d\omega = \frac{dz}{jz}\,.$$

Then

$$\text{MMSE} = \frac{1}{j2\pi}\oint_C [S_s(z) - S_{\hat{s}s}(z)]z^{-1}dz$$

where C is the unit circle. This expression can be evaluated using the residue theorem.

3.7 Continuous time

The derivations for continuous-time signals are almost identical.
The causal filter has the form:

$$\hat{s}_t = \int_{-\infty}^{t} h_{t-t'} y_{t'} \, dt' \ .$$

The above derivations and results hold by replacing the Z transform with Laplace, z with s, $|z| < 1$ with $Re(s) < 0$, etc.

In particular, the solution for the *prediction* problem:

$$\hat{s}_{t+T} = \int_{-\infty}^{t} h_{t-t'} y_{t'} \, dt' \ .$$

is given by

$$H(s) = \left\{ S_{sy}(s) \, [S_y^-(s)]^{-1} e^{sT} \right\}_+ [S_y^+(s)]^{-1} \ .$$

Remark: For the simple (scalar) problem of filtering in white noise:

$$y_t = s_t + n_t \ ; \quad n \perp s \ , \quad S_n(s) = \rho^2$$

with $S_s(s)$ strictly proper, the Wiener Filter simplifies to

$$H(s) = 1 - \frac{\rho}{S_y^+(s)} \ .$$