Taxonomy

Figure: Family tree

The Discrete Kalman Filter

These notes are taken from *Statistical Digital Signal Processing and Modeling* by Monson Hayes, John Wiley, New York, NY, 1996. ISBN 0-47159431-8

We have already considered the problem of designing a causal Wiener filter, **w** to estimate an AR process $x(n)$ given statistical information

Figure: Modeling an AR process

The solution involves formulating the Wiener-Hopf equations

$$\underbrace{\begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix}}_{\mathbf{R}} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{M-1} \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} p^*(0) \\ p^*(-1) \\ \vdots \\ p^*(1-M) \end{bmatrix}}_{p}$$

where (using this notation)

$$r(l) = E[y(n)y^*(n-l)]$$
$$p(-l) = E[y(n)x^*(n-l)]$$

and solving for the Wiener filter

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{p}.$$

Processing our noisy measurements with the Wiener filter gives us our optimal (in the MSE sense) estimate, $\hat{x}(n)$.

The primary limitation of the Wiener filter is that it requires $x(n)$ and $y(n)$ to be jointly wide sense stationary processes. Since most processes encountered in practice are nonstationary, this constraint limits the usefulness of the Wiener filter. We thus reexamine estimation within the context of nonstationary processes and derive what is known as the Discrete Kalman Filter.

Estimation of Order 1 AR Processes

Consider the estimation of a 1st order AR process of the form

$$x(n) = a(1)x(n-1) + w(n)$$

based on the *observations*, $y(n) = x(n) + v(n)$ where the process noise, $w(n)$ and the measurement noise, $v(n)$ are uncorrelated white noise processes. It can be shown from the Wiener-Hopf equations, the optimum linear estimate of $x(n)$ using all observations $y(k)$, for $k \le n$ can be computed as

$$\hat{x}(n) = a(1)\hat{x}(n-1) + K[y(n) - a(1)\hat{x}(n-1)]$$

where K is a constant referred to as the Kalman gain. This recursion minimizes the MSE, $E[|x(n) - \hat{x}(n)|^2]$.

There are two problems with this solution:

1) $x(n)$ and $y(n)$ are required to be jointly wide-sense stationary processes, i.e. if the AR coefficient is time-varying, the recursion does not yield the optimum linear estimate.
2) the filter is not allowed to "turn on" at $n = 0$, i.e. the recursion assumes observations $y(k)$, for $k \le n$ are available.

Estimation of Order p AR Processes

Although we have only been considering order 1 AR processes, using state variables we may extend these results to the estimation of more general processes. Let $x(n)$ be a pth order AR process generated with

$$x(n) = \sum_{k=1}^{p} a(k)x(n-k) + w(n)$$

and suppose $x(n)$ is measured (observed) in the presence of additive noise

$$y(n) = x(n) + v(n).$$

Let

$$\mathbf{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-p+1) \end{bmatrix}$$

then we have

$$\mathbf{x}(n) = \begin{bmatrix} a(1) & a(2) & \cdots & a(p-1) & a(p) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \mathbf{x}(n-1) + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w(n)$$

and

$$y(n) = [1 \quad 0 \quad \cdots \quad 0]\mathbf{x}(n) + v(n).$$

These equations can be rewritten in state-space form

$$\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1) + \mathbf{w}(n)$$
$$y(n) = \mathbf{C}\mathbf{x}(n) + v(n)$$

where **A** is a $p \times p$ state transition matrix, $\mathbf{w}(n) = [\,\omega(\nu) \quad 0 \quad \Lambda \quad 0]^T$ is a vector noise process, and $\mathbf{C} = [1 \quad 0 \quad \mathrm{L} \quad 0]$.

For the order p AR process, the optimum linear estimate of the "state", $\mathbf{x}(n)$ using all measurements up to time n may then be expressed in the form

$$\hat{\mathbf{x}}(n) = \mathbf{A}\hat{\mathbf{x}}(n-1) + \mathbf{K}[y(n) - \mathbf{C}\mathbf{A}\hat{\mathbf{x}}(n-1)]$$

where **K** is a constant referred to as the Kalman gain vector.

Estimation of General Processes

We may generalize the above optimum linear estimate to *nonstationary* processes as follows. Let $x(n)$ be a $p \times 1$ state vector that evolves according to (state equation)

$$\mathbf{x}(n) = \mathbf{A}(\nu - 1)\boldsymbol{\xi}(\nu - 1) + \boldsymbol{\omega}(\nu)$$

where $\mathbf{A}(n-1)$ is a time-varying $p \times p$ state transition matrix and $\mathbf{w}(n)$ is a vector of zero-mean, white noise processes with

$$E[\mathbf{w}(n)\boldsymbol{\omega}^H(\kappa)] = \begin{cases} \boldsymbol{\Theta}_{\omega}(\nu), & \kappa = \nu \\ \mathbf{0}, & \kappa \neq \nu \end{cases}$$

In addition, let $y(n)$ be a vector of observations that are formed according to (observation equation)

$$\mathbf{y}(n) = \mathbf{X}(\nu)\boldsymbol{\xi}(\nu) + \boldsymbol{\varpi}(\nu)$$

where $\mathbf{y}(n)$ is a vector of length q, $\mathbf{C}(n)$ is a time-varying $q \times q$ matrix, and $\mathbf{v}(n)$ is a vector of zero mean white processes that are statistically independent of $\mathbf{w}(n)$ with

$$E[\mathbf{v}(n)\boldsymbol{\varpi}^H(\kappa)] = \begin{cases} \boldsymbol{\Theta}_{\varpi}(\nu), & \kappa = \nu \\ \mathbf{0}, & \kappa \neq \nu \end{cases}$$

Thus we expect the optimum linear estimate of $\mathbf{x}(n)$ to be expressible in the form

$$\hat{\mathbf{x}}(n) = \mathbf{A}(\nu - 1)\hat{\boldsymbol{\xi}}(\nu - 1) + \mathbf{K}(\nu)[\boldsymbol{\psi}(\nu) - \mathbf{X}(\nu)\mathbf{A}(\nu - 1)\hat{\boldsymbol{\xi}}(\nu - 1)]\,.$$

With the appropriate Kalman gain matrix, $\mathbf{K}(n)$, this recursion corresponds to the discrete Kalman filter. We will now show the optimum linear estimate of $\mathbf{x}(n)$ has the above form and derive the Kalman gain $\mathbf{K}(n)$ that minimizes the mean-square estimation error.

Discrete Kalman Filter Development

In the derivation of the Kalman filter we make the following assumptions and use the following notation:

- The state-transition matrix, $\mathbf{A}(n)$; measurement matrix, $\mathbf{C}(n)$ are known
- covariance matrix of the process noise, $\mathbf{Q}_w(n)$; and covariance matrix of the measurement noise, $\mathbf{Q}_v(n)$ are known.

- $\hat{\mathbf{x}}(n|n)$ denotes the best linear estimate of **x**(n) at time n given the observations **y**(i) for i = 1, 2, …, n
- $\hat{\mathbf{x}}(n|n-1)$ denotes the best linear estimate of **x**(n) at time n given the observations up to $n - 1$
- $\mathbf{e}(n|n) = \boldsymbol{\xi}(\nu) - \hat{\boldsymbol{\xi}}(\nu|\nu)$ and $\mathbf{e}(n|n-1) = \boldsymbol{\xi}(\nu) - \hat{\boldsymbol{\xi}}(\nu|\nu-1)$ are the state estimation errors
- $\mathbf{P}(n|n) = E[\boldsymbol{\varepsilon}(\nu|\nu)\boldsymbol{\varepsilon}^H(\nu|\nu)]$ and $\mathbf{P}(n|n-1) = E[\boldsymbol{\varepsilon}(\nu|\nu-1)\boldsymbol{\varepsilon}^H(\nu|\nu-1)]$ are the error covariance matrices

---

Problem Statement

Suppose we are given an estimate $\hat{\mathbf{x}}(0|0)$ of the state **x**(0) and the error covariance matrix $\mathbf{P}(0|0)$. When the measurement **y**(1) becomes available the goal is to update $\hat{\mathbf{x}}(0|0)$ and find the estimate $\hat{\mathbf{x}}(1|1)$ of the state at time n = 1 that minimizes the MSE

$$\begin{aligned}\xi(1) &= E\left[\left\|\mathbf{e}(1|1)\right\|^2\right] \\ &= \sum_{i=0}^{p-1} E\left[\left|e_i(1|1)\right|^2\right] \\ &= \mathrm{tr}[\mathbf{P}(1|1)]\end{aligned}$$

After $\hat{\mathbf{x}}(1|1)$ has been determined and the error covariance $\mathbf{P}(1|1)$ is found, the estimation is repeated for the next observation **y**(2). Thus for each $n > 0$, given $\hat{\mathbf{x}}(n-1|\nu-1)$ and $\mathbf{P}(n-1|\nu-1)$, when a new observation, $y(n)$ becomes available, the problem is to find the minimum mean square error (MMSE) estimate $\hat{\mathbf{x}}(n|n)$ of the state vector **x**(n).

We will solve this problem in two steps:
1) Given $\hat{\mathbf{x}}(n-1|\nu-1)$ we will find $\hat{\mathbf{x}}(n|n-1)$ which is the best estimate of **x**(n) without the observation **y**(n).
2) Given **y**(n) and $\hat{\mathbf{x}}(n|n-1)$ we will estimate **x**(n) with $\hat{\mathbf{x}}(n|n)$.

We will also need to figure out how to initialize algorithm.