Program in Applied Mathematics

Numerical Wiener Filtering using CKMS

Jared McBride

Applied Mathematics
University of Arizona

October 27, 2021

THE UNIVERSITY OF ARIZONA

Motivation

- Noise reduction
- Data-driven model reduction

Outline

Program in Applied Mathematics

- **Stochastic process:** A family of random variables indexed by an index set (discrete or continuous). E.g. $X : \Omega \times \mathbb{R} \to \mathbb{C}^d$ or $X : \Omega \times \mathbb{Z} \to \mathbb{C}^d$
- **Timeseries:** A realization of a stochastic process (usually indexed by time). In this talk this is indexed over a finite set. E.g. $x : \{1, 2, \ldots, N\} \to \mathbb{C}^d$
- **Signal:** A stochastic process.
- **Wide sense stationary (WSS) stochastic process:** A stochastic process satisfying the following conditions:

$$\mathbb{E}X_n = \mu \qquad \text{(no dependence on } t)$$

$$\mathbb{E}[(X_n - \mu)(X_n - \mu)^*] = f(n - m) \qquad \text{(depends only on difference } t - s)$$

Where the asterisks * denote the conjugate transpose.

- **Covariance sequence:** Given a WSS stochastic process $X = \{X_n\}$ it is the (doubly infinite) sequence $R_X(n)$ given by

$$R_X(n,m) = \mathbb{E}[(X_n - \mu)(X_m - \mu)^*] = R_X(n-m)$$

- **Power spectrum:** The Fourier series of the covariance sequence

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} R_X(k)e^{-ik\omega}$$

- ***z*-spectrum:** The z-series of the covariance sequence

$$\bar{S}_X(z) = \sum_{k=-\infty}^{\infty} R_X(k)z^{-k}$$

- **Jointly wide-sense stationary stochastic processes:** Two stochastic processes satisfying the following conditions:

$$\mathbb{E}X_n = \mu_X, \quad \mathbb{E}Y_n = \mu_Y \qquad \text{and} \qquad \mathbb{E}[(X_n - \mu_X)(Y_n - \mu_Y)^*] = f(n-m)$$

- **Cross covariance sequence:** Given two (jointly) WSS stochastic process $X = \{X_n\}, Y = \{Y_n\}$ it is the (doubly infinite) sequence $R_{XY}(n)$ given by

$$R_{XY}(n,m) = \mathbb{E}[(X_n - \mu_X)(Y_m - \mu_Y)^*] = R_{XY}(n-m)$$

- **Cross spectrum:** The Fourier series of the cross covariance sequence

$$S_{XY}(\omega) = \sum_{k=-\infty}^{\infty} R_{XY}(k) e^{-ik\omega}$$

- **z-cross spectrum:** The z-series of the cross covariance sequence

$$\bar{S}_{XY}(z) = \sum_{k=-\infty}^{\infty} R_{XY}(k) z^{-k}$$

Let $X_n \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d

- This is WSS since

$$\mathbb{E}X_n = \mu$$
$$\mathbb{E}[(X_n - \mu)(X_m - \mu)^*] = \sigma^2 \delta_{n,m} = \sigma^2 \delta(n - m)$$

- The autocovariance sequence is

$$R_X(n) = \sigma^2 \delta(n) \qquad \text{for } n \in \mathbb{Z}$$

- The power spectrum and z-spectrum is therefore given by

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} \sigma^2 \delta(k) z^{-k} = \sigma^2 = \bar{S}_X(z)$$

Program in Applied Mathematics

Standard state-space model of a system: A model of the following form:

$$\begin{cases} X_{i+1} &= F_i X_i + G_i u_i \\ Y_i &= H_i X_i + v_i \end{cases}$$

where $F_i \in C^{n\times n}$, $G_i \in C^{n\times m}$, and $H_i \in C^{p\times n}$ are known matrices, and $u = \{u_i\}$, $v = \{v_i\}$, and X_0 are variables with the following property

$$\mathbb{E}\begin{pmatrix} X_0 \\ u_i \\ v_i \end{pmatrix}\begin{pmatrix} X_0 \\ u_j \\ v_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi_0 & 0 & 0 & 0 \\ 0 & Q_i\delta_{ij} & S_i\delta_{ij} & 0 \\ 0 & S_i^*\delta_{ij} & R_i\delta_{ij} & 0 \end{pmatrix}$$

- Y is the output (or observations).
- X is the state variable.
- u is the process (or plant) noise
- v is the measurement noise.

Time-invariant state-space model: A model of the following form:

$$\begin{cases} X_{i+1} & = FX_i + Gu_i \\ Y_i & = HX_i + v_i \end{cases}$$

where $F \in C^{n\times n}$, $G \in C^{n\times m}$, and $H \in C^{p\times n}$ are known matrices, and $u = \{u_i\}$, $v = \{v_i\}$, and X_0 are variables with the following property

$$\mathbb{E}\begin{pmatrix} X_0 \\ u_i \\ v_i \end{pmatrix}\begin{pmatrix} X_0 \\ u_j \\ v_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi_0 & 0 & 0 & 0 \\ 0 & Q\delta_{ij} & S\delta_{ij} & 0 \\ 0 & S^*\delta_{ij} & R\delta_{ij} & 0 \end{pmatrix}$$

Observe $F_i = F$, $H_i = H$, $G_i = G$, $Q_i = Q$, $R_i = R$, *and* $S_i = S$

- **Linear time-invariant system:**
  - (Linear)
    $$L(\alpha u + \beta v) = \alpha L u + \beta L v \qquad \text{for all } \alpha, \beta \in \mathbb{C}$$
  - (Time invariant) Let S be the shift operator $(Su)_n = u_{n+1}$, then L is time invariant if
    $$LSu = SLu$$

These can be represented as a convolution

$$(Lu)_n = \sum_{k=-\infty}^{\infty} l_k u_{n-k}$$

Example: Is a time-invariant state-space model a linear time-invariant system?

The inputs are u, v and the outputs are X, Y, let $\mathcal{L}(u, v) = (X, Y)$

- (Time-invariant) Does $\mathcal{L}(Su, Sv) = (SX, SY)$?

$$\begin{cases} X'_{i+1} & = FX'_i + Gu_{i+1} \\ Y'_i & = HX'_i + v_{i+1} \end{cases} \Rightarrow \begin{cases} X'_j & = FX'_{j-1} + Gu_j \\ Y'_{j-1} & = HX'_{j-1} + v_j \end{cases}$$

So, $X_j = X'_{j-1} = (S^{-1}X')_j$ and $SX = X'$. Same for Y'.

Example: Is a time-invariant state-space model a linear time-invariant system?

The inputs are u, v and the outputs are X, Y, let $\mathcal{L}(u, v) = (X, Y)$

- (Linear) Let $\mathcal{L}(u', v') = (X', Y')$

$$\begin{cases} \alpha X_{i+1} + X'_{i+1} & = \alpha F X_i + F X'_i + \alpha G u_i + G u'_i \\ \alpha Y_i + Y'_i & = \alpha H X_i + H X'_i + \alpha v_{i+1} + v'_i \end{cases} \Rightarrow$$

$$\begin{cases} (\alpha X + X')_{i+1} & = F(\alpha X + X')_i + G(\alpha u + u')_i \\ (\alpha Y + Y')_i & = H(\alpha X + X')_i + (\alpha v + v')_i \end{cases}$$

So, observe that $\mathcal{L}(\alpha u + u', \alpha v + v') = (\alpha X + X', \alpha Y + Y')$

- **Impulse response of a (LTI) system:** The output of the system when the impulse signal $\delta = (\ldots, 0, 1, 0, \ldots)$ is the input.
  - Example: If the system can be written as a convolution with with some element $l = \{l_k\}$, the impulse response recovers that element.

$$L\delta_n = \sum_{k=-\infty}^{\infty} \delta_k l_{n-k} = l_n$$

- **z-series:** Given a sequence a (bilaterally infinite) the z-series is the complex function

$$A(z) = \mathcal{Z}\{a\} = \sum_{k=-\infty}^{\infty} a_k z^{-k}.$$

- **Transfer function of an LTI system:** The z-series of the impulse response of a system

$$L(z) = \sum_{k=-\infty}^{\infty} l_k z^{-k}.$$

- **Causal:** A LTI system is causal if it's impulse response is causal which means $h_k = 0$ for $k < 0$.

- **BIBO Stability:** A LTI system is stable if given a bounded input the output of the system is bounded.

- **Inverse:** The inverse of an LTI system maps the output to the input.

- **Minimum-phase:** A linear time-invariant system is minimum-phase if it and it's inverse are both causal and stable.

A few facts: Suppose an LTI system $\mathcal{L}$ has a rational transfer function $H(z)$, then

- $\mathcal{L}$ is BIBO stable if its impulse response h_k is absolutely summable (think of Hölder $1, \infty$). This means $H(z)$ converges on the unit circle.
- $\mathcal{L}$ is causal if the poles of $H(z)$ lie within the unit circle.
- The transfer function of $\mathcal{L}^{-1}$ is $[H(z)]^{-1}$.
- $\mathcal{L}$ is minimum phase if the poles and zeros of $H(z)$ lie (strictly) within the unit circle.
- If $\mathcal{L}$ is an LTI system and X is WSS, then $\mathcal{L}X$ is stationary and jointly stationary with X.

Example: MA(q)

- Consider system $\mathcal{R}$ given by $\mathcal{R}(u) = v$ where

$$v_n = u_n + r_1 u_{n-1} + \cdots + r_q u_{n-q} = \sum_{i=0}^{q} r_i u_{n-i} = (r \star u)_n$$

What do we know about $\mathcal{R}$

- LTI
- FIR (finite impulse response), only $q + 1$ taps.

$$r = (\ldots, 0, \boxed{1}, r_1, r_2, \ldots, r_q, 0, \ldots)$$

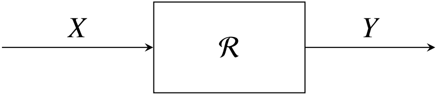
- Transfer function:

$$R(z) = \sum_{k=-\infty}^{\infty} r_k z^{-k} = \sum_{k=0}^{q} r_k z^{-k}$$

- Causal and stable
- Minimum phase? (depends)

Example: MA(q)

- Recall $X_n \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d



$$Y_n = X_n + r_1 X_{n-1} + \cdots + r_q X_{n-q}$$

What do we know about Y,

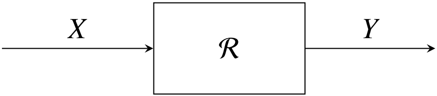
- Bounded
- WSS?
- Covariance sequence:

$$R_X(n) = \sigma^2 \sum_{i=0}^{q} r_i r_{i-n}^* \qquad \text{for } n \in \mathbb{Z}$$

(Observe $R_X(n) = 0$ for $|n| > q$)

Example: MA(q)

- Recall $X_n \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d



$$Y_n = X_n + r_1 X_{n-1} + \cdots + r_q X_{n-q}$$

What do we know about Y,

- z-spectrum

$$\bar{S}_Y(z) = \bar{S}_{r \star X}(z) = R(z)\bar{S}_X(z)R^*(z^{-*}) = \left(\sum_{k=0}^{q} r_k z^{-k}\right) \sigma^2 \left(\sum_{k=0}^{q} r_k^* z^k\right)$$

- Power spectrum,

$$S_Y(\omega) = \bar{S}_Y(e^{i\omega}) = \sigma^2 \left|\sum_{k=0}^{q} r_k e^{-i\omega k}\right|^2$$

- **Canonical Spectral factorization:** Given a z-spectrum $S_x(z)$, for which all S_X and $\log|S_x|$ are integrate on the unit circle, it can be factored as:

$$S_X(z) = L(z)R_e L^*(z^{-*})$$

such that

  - $R_e > 0$,
  - $\lim_{z\to\infty} L(z) = I$, and
  - $L(z)$ and $L(z)^{-1}$ are analytic in $|z| > 1$.

- **Canonical Spectral factorization of Laurent Polynomials:** Furthermore, if S_X is a Laurent Polynomial, that is,

$$S_X(z) = \sum_{k=-m}^{m} r_k z^{-k}, \quad (r_k = r^*_{-k}), \qquad \text{then} \qquad L(z) = I + \sum_{k=1}^{m} l_k z^{-k}$$

Wiener Filter

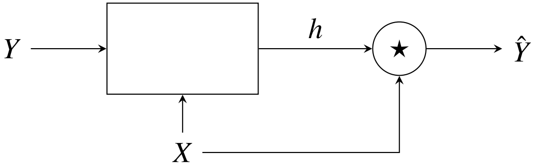
Given two (jointly) **WSS** processes Y,X, the Wiener filter provides the optimal linear estimator of Y given X, that is,

$$\mathbb{E}\|Y_n - (X \star h)_n\|^2 = \text{minimum in } h$$

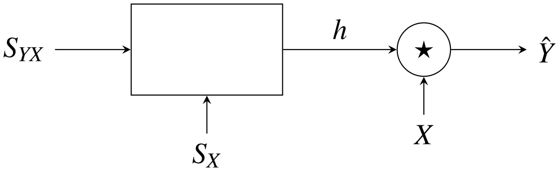
Once constructed the Wiener filter is an LTI system.

- h is the Wiener filter
- Y is the signal,
- X is the predictors.

So,



Or rather,



Assume we have it (h not necessarily causal).
Since, it is a linear least squares estimate,

$$0 = \mathbb{E}\Big[(\mathbf{y}_n - \hat{\mathbf{y}}_n)(\mathbf{x}_m)\Big] = \mathbb{E}\Big[(\mathbf{y}_n - (h \star \mathbf{x})_n)(x_m)^*\Big]$$

This implies

$$\mathbb{E}\mathbf{y}_n\mathbf{x}_m^* = \mathbb{E}\sum_{k=-\infty}^{\infty} h_k\mathbf{x}_{n-k}\mathbf{x}_m^* = \sum_{k=-\infty}^{\infty} h_k\mathbb{E}\mathbf{x}_{n-k}\mathbf{x}_m^*$$

or rather (with relabeling $n - m \mapsto n$)

$$R_{\mathbf{yx}}(n) = \sum_{k=-\infty}^{\infty} h_k R_{\mathbf{x}}(n-k)$$

The form of RHS suggest use of the z-transform.

Applying the z-transform to both sides gives

$$S_{\mathbf{yx}}(z) = H(z)S_{\mathbf{x}}(z)$$

where

$$H(z) = \mathcal{Z}\{h_n\} = \sum_{n=-\infty}^{\infty} h_n z^{-n}$$

So,

$$H(z) = S_{\mathbf{yx}}(z)S_{\mathbf{x}}^{-1}(z)$$

we then apply the inverse z-transform to recover h

$$h_n = \frac{1}{2\pi i}\int_C S_{\mathbf{yx}}(z)S_{\mathbf{x}}^{-1}(z)z^{n-1}\,dz$$

If we require that h is causal this is more difficult. Then

$$0 = \mathbb{E}\Big[(\mathbf{y}_n - \hat{\mathbf{y}}_n)(\mathbf{x}_m)\Big] = \mathbb{E}\Big[(\mathbf{y}_n - (h \star \mathbf{x})_n)(x_m)^*\Big] \qquad \text{only for } m \leq n$$

This implies

$$\mathbb{E}\mathbf{y}_n\mathbf{x}_m^* = \mathbb{E}\sum_{k=-\infty}^{\infty} h_k\mathbf{x}_{n-k}\mathbf{x}_m^* = \sum_{k=-\infty}^{\infty} h_k\mathbb{E}\mathbf{x}_{n-k}\mathbf{x}_m^* \qquad \text{only for } m \leq n$$

or rather (with relabeling $n - m \mapsto n$)

$$R_{\mathbf{yx}}(n) = \sum_{k=-\infty}^{\infty} h_k R_{\mathbf{x}}(n-k) \qquad \text{only for } n \geq 0$$

The form of RHS suggest use of the z-transform. But we can't!

However observe that for

$$g_n = R_{\mathbf{yx}}(n) - \sum_{k=-\infty}^{\infty} h_k R_{\mathbf{x}}(n-k)$$

g is strictly anti-casual since $g_n = 0$ when $n \geq 0$. Now apply the z-transform to both sides. We get

$$G(z) = S_{\mathbf{yx}}(z) - H(z)S_{\mathbf{x}}(z)$$

Now apply the spectral factorization to $S_{\mathbf{x}}(z)$ And proceed as follows

$$G(z) = S_{\mathbf{yx}}(z) - H(z)S_{\mathbf{x}}^{-}(z)S_{\mathbf{x}}^{+}(z)$$

and observe when we apply the inverse

$$\underbrace{G(z){S^{+}_{\mathbf{x}}}^{-1}(z)}_{\textit{strictly anti-causal}} = \underbrace{S_{\mathbf{yx}}(z){S^{+}_{\mathbf{x}}}^{-1}(z)}_{\textit{mixed}} - \underbrace{H(z)S_{\mathbf{x}}^{-}(z)}_{\textit{causal}}$$

And so, as a necessary condition,

$$H(z) = \left\{ S_{\mathbf{yx}}(z) S^{+\,-1}_{\mathbf{x}}(z) \right\}_{+} S^{-\,-1}_{\mathbf{x}}(z)$$

How to implement it and test it?

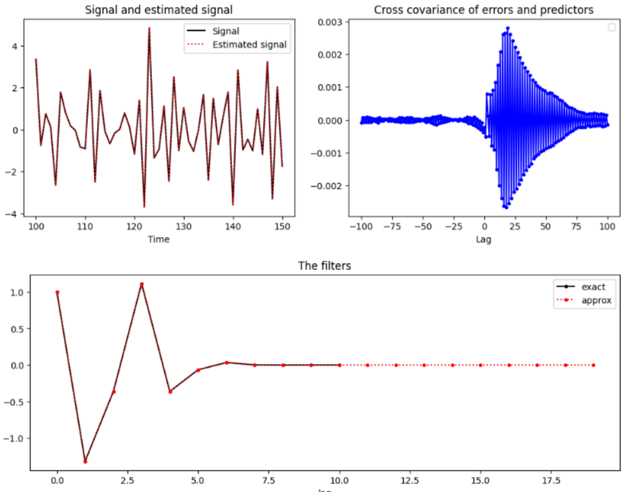
$$h_k = \frac{1}{2\pi i} \int_C H(z) z^{n-1}\, dz$$

Put $\hat{Y} = h \star X$,

- Plot Y and $\hat{Y}$
- Plot $R_{Y-\hat{Y},X}$ over some interval containing zero.
- Global statistic: mean square error (MSE)

Program in Applied Mathematics

- X is an $AR(10)$ process, time series length 10,000.
- $Y = f \star X$ is X filtered by a causal, stable 10-tap, filter f.



- Commonly there is a statespace structure,

$$\begin{cases} X_{i+1} &= FX_i + Gu_i \\ Y_i &= HX_i + v_i \end{cases}$$

- Kalman extended developed statespace estimation for the time **variant** case

$$\begin{cases} X_{i+1} &= F_iX_i + G_iu_i \\ Y_i &= H_iX_i + v_i \end{cases}$$

  The so-called innovations process,
  - $\boxed{e_i = Y_i - \hat{Y}_{i|i-1} = H_i\hat{X}_{i|i-1}}$ where $\hat{Y}_{i|i-1}$ and $\hat{X}_{i|i-1}$ are the l.l.s.e. of Y_i and X_i (respectively) given $Y_{i-1}, \ldots, Y_1$
  - this process is white.
  - It can be used to recursively produce state prediction.

The Innovations Recursions [1, p. 317]

Consider the standard statespace model

$$\begin{cases} X_{i+1} & = F_i X_i + G_i u_i \\ Y_i & = H_i X_i + v_i \end{cases} \qquad i \geq 0$$

The innovations process of Y can be recursively computed using the equations

$$e_i = Y_i - H_i \hat{X}_i, \qquad \hat{X}_0 = 0, \qquad e_0 = Y_0,$$
$$\hat{X}_{i+1} = F_i \hat{X}_i + K_{p,i} e_i, \qquad i \geq 0,$$

where $K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}$, $R_{e,i} = R_i + H_i P_i H_i^*$, and

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}, \qquad P_0 = \Pi_0$$

Here, $P_i = \mathbb{E} \tilde{X}_i \tilde{X}_i^*$ where $\tilde{X}_i = X_i - \hat{X}_i$. When $m \ll n$, $p \ll n$ to go from e_i to e_{i+1} requires $O(n^3)$ operations.

The Innovations Recursions [1, p. 317]

Consider the standard statespace model

$$\begin{cases} X_{i+1} & = F_i X_i + G_i u_i \\ Y_i & = H_i X_i + v_i \end{cases} \qquad i \geq 0$$

The innovations process of Y can be recursively computed using the equations

$$\hat{X}_{i+1} = F_i \hat{X}_i + K_{p,i} e_i, \qquad \hat{X}_0 = 0, \qquad e_0 = Y_0,$$
$$Y_i = H_i \hat{X}_i + e_i, \qquad i \geq 0,$$

where $K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}$, $R_{e,i} = R_i + H_i P_i H_i^*$, and

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \qquad P_0 = \Pi_0$$

Here, $P_i = \mathbb{E}\tilde{X}_i \tilde{X}_i^*$ where $\tilde{X}_i = X_i - \hat{X}_i$. When $m \ll n$, $p \ll n$ to go from e_i to Ee_{i+1} requires $O(n^3)$ operations.

- Knowing $K_{p,i}$ and $R_{e,i}$ gives a **causal and causally invertable** system $\mathcal{L} : e \mapsto Y$ given by

$$\begin{aligned} \hat{X}_{i+1} &= F_i \hat{X}_i + K_{p,i} e_i, \qquad \hat{X}_0 = 0 \\ Y_i &= H_i \hat{X}_i + e_i \end{aligned}$$

  (This is called an innovations model)

- The inverse $\mathcal{L}^{-1} : Y \mapsto e$ is given by

$$\begin{aligned} \hat{X}_{i+1} &= (F_i - K_{p,i} H_i)\hat{X}_i + K_{p,i} Y_i, \qquad \hat{X}_0 = 0 \\ e_i &= -H_i \hat{X}_i + Y_i \end{aligned}$$

- $\mathcal{L}$ is an example of a modeling filter

- $\mathcal{L}^{-1}$ is an example of a whitening filter

Kalman Filtering by Chadrasekhar-Kailath-Morf-Sidhu (CKMS)

- Kailath, Morf and Sidhu (1973) observed that

  though P_i is full rank $\delta P_i := P_{i+1} - P_i$ can have low rank

- So write,

$$\delta P_i = L_i M_i L_i^*$$

- It was shown that if $\delta P_0 = L_0 M_0 L_0^*$ with M_0
  - hermitian
  - nonsingular
  - size $\alpha \times \alpha$

  , then for $i > 0$, $\delta P_i = L_i M_i L_i^*$ with M_i
  - hermitian
  - nonsingular
  - size $\alpha \times \alpha$

Program in Applied Mathematics

The Fast (CKMS) Kalman Recursions [1, p. 409]

The $K_{p,i}$ and $R_{e,i}$ from the Kalman recursion above can be recursively computed by the following set of coupled recursions, for $i \geq 0$

$$
\begin{aligned}
K_{p,i+1} &= K_{p,i} - FL_i R_{r,i}^{-1} L_i^* H^* \\
L_{i+1} &= FL_i - K_{p,i} R_{e,i}^{-1} HL_i \\
R_{e,i+1} &= R_{e,i} - HL_i R_{r,i}^{-1} L_i^* H^* \\
R_{r,i+1} &= R_{r,i} - L_i^* H^* R_{e,i}^{-1} HL_i
\end{aligned}
$$

The recursion is initialized as follows: $K_{p,0} = F\Pi_0 H^* + GS$ and $R_{e,0} = R + H\Pi_0 H^*$. Then factor get L_0 and $R_{r,0}$

$$
\delta P_0 := F\Pi_0 F^* + GQG^* - K_{p,0} R_{e,0}^{-1} K_{p,0}^* - \Pi_0 =: -L_0 R_{r,0}^{-1} L_0^*
$$

where L_0 is $n \times \alpha$ and $R_{r,0}$ is $\alpha \times \alpha$, nonsingular and Hermitian.

We can get away with $\alpha = 1$!

This is a result of stationary,

- There exists a $\bar{\Pi}$ such that

$$\bar{\Pi} = F\bar{\Pi}F * + GQG^*$$

- If we pick $\Pi_0 = \bar{\Pi}$ then

$$\delta P_0 = F\Pi_0 F^* + GQG^* - K_{p,0}R_{e,0}^{-1}K_{p,0}^* - \Pi_0 = -K_{p,0}R_{e,0}^{-1}K_{p,0}^* = -L_0 R_{r,0}^{-1}L_0^*$$

- So, we can initialize
  - $L_0 = K_0 = F\Pi_0 H^* + GS$
  - $R_{r,0} = R_{e,0} = R + H\bar{\Pi}H^*$

Most Numerical algorithms assume $S(z)$ is rational and has the form of a Laurent Polynomial (noted above) (this is what we assume here) Algorithms that use Toeplitz matrices.

- Bauer
- Schur
- Levinson-Durbin

Algorithms that use state-space formulations.

- Riccati Equation
- Kalman Filter
- Chadrasekhar-Kailath-Morf-Sidhu (CKMS)

Sayed, Ali H., and Thomas Kailath. "A survey of spectral factorization methods." *Numerical linear algebra with applications* 8.6-7 (2001): 467-496.

Given $S_Y(z)$, $Y_n \in \mathbb{C}^d$ for $n > -\infty$, (stationary discrete-time stochastic process)

$$S_Y(z) = \sum_{n=-\infty}^{\infty} R_Y(n) z^{-n},$$

Now, if the covariance decays fast enough truncate

$$\tilde{S}_Y(z) = \sum_{n=-m}^{m} R_Y(n) z^{-n}.$$

It is possible to construct $\tilde{Y}_n$ (finite state-spaces process) with

$$S_{\tilde{Y}}(z) = \tilde{S}_Y(z),$$

Program in Applied Mathematics

$$\begin{cases} X_{i+1} & = FX_i + Gv_i \\ \tilde{Y}_i & = HX_i + u_i \end{cases}$$

provided that

$$F = \begin{pmatrix} 0 & & & & & \\ I & 0 & & & & \\ & I & 0 & & & \\ & & \ddots & \ddots & & \\ & & & & I & 0 \end{pmatrix} \in \mathbb{C}^{md \times md}$$

$$H = \begin{pmatrix} 0 & \dots & 0 & I \end{pmatrix} \in \mathbb{C}^{d \times md}$$

$$\mathbb{E} \begin{pmatrix} v_i & u_i \end{pmatrix} \begin{pmatrix} v_j^* \\ u_j^* \end{pmatrix} = \begin{pmatrix} R\delta_{ij} & S\delta_{ij} \\ S^*\delta_{ij} & Q\delta_{ij} \end{pmatrix}$$

$$\Pi = F\Pi F^* + GQG^*$$
$$GS = N - F\Pi H^*$$
$$R = R_Y(0) - H\Pi H^*$$

$$\Pi = \operatorname{cov}(X_i, X_i) = \mathbb{E}X_i X_i^* \quad \left(\in \mathbb{C}^{md \times md}\right)$$

$$N = \begin{pmatrix} R_Y(m) \\ R_Y(m-1) \\ \vdots \\ R_Y(1) \end{pmatrix} \quad \left(\in \mathbb{C}^{md \times d}\right)$$

Program in Applied Mathematics

$$S_{\tilde{Y}}(z) = \sum_{n=-m}^{m} R_Y(n) z^{-n} = \tilde{S}_Y(z)$$

- Original model for $\tilde{Y}$

$$\begin{cases} X_{i+1} &= FX_i + Gv_i \\ \tilde{Y}_i &= HX_i + u_i \end{cases}$$

- Innovations model (modeling filter) for $\tilde{Y}$

$$\begin{cases} \hat{X}_{i+1} &= F\hat{X}_i + K_i e_i, \qquad \hat{X}_0 = 0 \\ \tilde{Y}_i &= H\hat{X}_i + e_i \end{cases}$$

where $\mathbb{E} e_i e_j^* = R_{e,i}\delta_{ij}$,

$$K_i = (N - F\Sigma_i H^*)R_{e,i}^{-1}, \qquad R_{e,i} = R_Y[0] - H\Sigma_i H^*, \qquad \text{and}$$

$$\Sigma_{i+1} = F\Sigma_i F^* + K_i R_{e,i} K_i^* \qquad \Sigma_i = \mathbb{E}\hat{X}_i\hat{X}_i^*$$

- Will K_i, $R_{e,i}$ converge?
- Yes, this is a consequence of F being stable. Let $K = \lim_i K_i$, $R_e = \lim_i R_{e,i}$,

$$\begin{cases} \hat{X}_{i+1} &= F\hat{X}_i + Ke_i \\ \tilde{Y}_i &= H\hat{X}_i + e_i, \qquad \mathbb{E}e_ie_j^* = R_e\delta_{ij} \end{cases}$$

- In our context this system may be represented as a convolution since

$$\tilde{Y}_i = \big(\mathcal{L}(e)\big)_i = e_i + \sum_{j=1}^{m} K^{(j)} e_{i-1-m+j} = (\ell * e)_i$$

where

$$\ell = (I, K^{(m)}, K^{(m-1)}, \ldots, K^{(1)})$$

- So the modeling filter is

$$h_{\text{mod}} = (R_e^{1/2}, K^{(m)}R_e^{1/2}, K^{(m-1)}R_e^{1/2}, \ldots, K^{(1)}R_e^{1/2})$$

Program in Applied Mathematics

- $\tilde{Y}$ is an approximating MA(m).
- Observe that

$$S_{\tilde{Y}}(z) = S_{\ell * e}(z) = L(z)S_e(z)L^*(z^{-*}) = L(z)RL^*(z^{-*})$$

where L is the z-transform of ℓ.

$$L(z) = \sum_{k=1}^{\infty} \ell_k z^{-k+1}$$

- And so,

$$S_Y(z) \approx \tilde{S}_Y(z) = S_{\tilde{Y}}(z) = L(z)RL^*(z^{-*})$$

provides a spectral factorization.

Numerical Wiener filtering

Expensive Wiener filter solvers:

- **Backslash (QR)**
- **Direct (numerical) optimization**

Cheaper Wiener filter solvers

- Wiener-Hopf with CKMS
- Kaczmarz

- In the presents of data, N samples. Pick $M \ll N$, we seek h_n, $n = 0, 1, 2, \ldots, M-1$ so that
  - $\hat{Y}_n = \sum_{k=0}^{M-1} X_{n-k} h_k$ for $n > M$
  - $\mathbb{E}\|Y_n - \hat{Y}_n\|^2 \approx \|Y - \hat{Y}\|_2^2 =$ minimum over all such $\{h_n\}$
- This is a regression problem

$$Y_n \sim X_n, X_{n-1}, \ldots, X_{n-M+1}$$

- Consider the design matrix,

$$\mathbf{X} = \begin{bmatrix} x_{M+1:N} & \left(S^{-1}X\right)_{M+1:N} & \cdots & \left(S^{-M}X\right)_{M+1:N} \end{bmatrix}$$

Numerical Wiener filtering

Backslash (QR) or any least squares solver

Let

$$\mathbf{X} = \begin{bmatrix} X_{M+1:N} & (S^{-1}X)_{M+1:N} & \cdots & (S^{-M}X)_{M+1:N} \end{bmatrix}$$
$$= \begin{bmatrix} X_{M+1:N} & X_{M:N-1} & \cdots & X_{1:N-M} \end{bmatrix}$$
$$= \begin{bmatrix} X_{M+1} & X_M & X_{M-1} & \cdots & X_1 \\ X_{M+2} & X_{M+1} & X_M & \cdots & X_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_N & X_{N-1} & X_{N-2} & \cdots & X_{N-M} \end{bmatrix} \in \mathbb{C}^{(M+1)\times(N-M)},$$

and

$$\mathbf{Y} = Y_{M+1:N} = \begin{bmatrix} Y_{M+1} \\ Y_{M+1} \\ \vdots \\ Y_N \end{bmatrix} \qquad \text{Then} \qquad \boxed{\mathbf{Y} \approx \mathbf{X}\mathbf{h} = \hat{Y}_{M+1:N}}$$

Where $\mathbf{h} = h_{1:M}$

So we have the least square problem

$$\boxed{\mathbf{Y} \approx \mathbf{X}\mathbf{h}}$$

- QR
- SVD
- randomized algorithms

(May benefit from regularization)

Algorithm (three steps)

1. Build spectral factors
   1.1 Approximate autocovariance of X (predictors)
   1.2 Feed this into CKMS
2. Approximate cross spectrum
3. Compute H using

$$H(z) = \left\{ S_{\mathbf{yx}}(z) L^{-*}(z^{-*}) \right\}_{+} L^{-1}(z)$$

   3.1 Divide spectral factor
   3.2 Take causal part
   3.3 Divide
   3.4 Extract filter

- **DFT:** I use `fft` from FFTW.jl (*a Julia wrapper for the FFTW library written in C*).
- Here is what it does:

$$v_k = \texttt{fft}(u)_k = \sum_{j=1}^{N} u_j e^{-\frac{2\pi i}{N}(j-1)(k-1)}$$

$$u_j = \texttt{ifft}(v)_j = \frac{1}{N}\sum_{k=1}^{N} v_k e^{\frac{2\pi i}{N}(k-1)(j-1)}$$

- Here is why I use it so much:
  - Given $S(z) = \sum_{j=1}^{N} c_j z^{-(j-1)}$ $(= \mathcal{Z}\{c\})$
  - evaluated at $z_k = e^{\frac{2\pi i}{N}(k-1)}$ for $k = 1, \ldots, N$ (N equally-spaced, unit-circle points)
  - use `fft`

$$S(z_k) = \sum_{j=1}^{N} c_j e^{-\frac{2\pi i}{N}(j-1)(k-1)} = \texttt{fft}(c)_k.$$

- Compute smoothed autocovariance sequence of predictors:

  For predictors $X = \left(X_n^{(i)};\ i = 1, 2, \ldots, \nu,\ n = 1, 2, \ldots, N\right) \in \mathbb{C}^{\nu \times N}$

  - $$C_{i,j,k} = \begin{cases} \sum_{n=1}^{N-k} X_{n+k}^{(i)} \left(X_n^{(j)}\right)^* & k \geq 0 \\ \sum_{n=1-k}^{N} X_{n+k}^{(i)} \left(X_n^{(j)}\right)^* & k < 0 \end{cases} \qquad k = -P, -P+1, \ldots, P$$
  - $$A_{i,j,k} = \Lambda_k \cdot \frac{1}{2}\left(C_{i,j,k} + C_{i,j,-k}^*\right) \qquad k = 0, 1, \ldots, P$$

    where Λ_k is the (Parzen) smoothing function

$$\Lambda_k = \begin{cases} 1 - 6\left(\frac{k}{P}\right)^2 + 6\left(\frac{|k|}{P}\right)^3, & |k| \leq P/2 \\ 2\left(1 - \frac{|k|}{P}\right)^3, & P/2 < |k| \leq P \\ 0, & |k| > P \end{cases}$$

- Feed into CKMS for spectral factorization:
  - $\texttt{CMKS} : \mathbb{C}^{\nu\times\nu\times(P+1)} \to \mathbb{C}^{\nu\times\nu\times(P+1)}$

$$\texttt{CKMS}(A) = l$$

- Form Spectral factors on unit circle grid. $z_k = e^{\frac{2\pi i}{N_{ex}}(k-1)}$ for $k = 1, \ldots, N_{ex}$
  - $l_k = l_{:,:,k}$, and $l_k = 0_{\nu\times\nu}$ for $P < k \leq N_{ex}$
  - $L(z) = \displaystyle\sum_{k=1}^{N_{ex}} l_k z^{-k}$
  - $L(z_k) = L_k, \qquad L = (L_k;\ k = 0, 1, \ldots, N_{ex})$

  then

$$L = \texttt{fft}(l) \qquad \text{taken only in third dimension}$$

- Right spectral factor $L' = (L'_k;\ k = 0, 1, \ldots, N_{ex})$

Numerical Wiener filtering

Step 2: Approximate cross spectrum

Program in Applied Mathematics

- Compute smoothed cross covariance sequence of signal with predictors:

  Now for $Y = \left(Y_n^{(i)};\ i = 1, 2, \ldots, d,\ n = 1, 2, \ldots, N\right) \in \mathbb{C}^{d\times N}$

  - $$C_{i,j,k} = \begin{cases} \displaystyle\sum_{n=1}^{N-k} Y_{n+k}^{(i)} \left(X_n^{(j)}\right)^* & k \geq 0 \\ \displaystyle\sum_{n=1-k}^{N} Y_{n+k}^{(i)} \left(X_n^{(j)}\right)^* & k < 0 \end{cases} \qquad k = -P, -P+1, \ldots, P$$
  - $A_{i,j,k} = \Lambda_{|k|} \cdot C_{i,j,k} \qquad k = -P, -P+1, \ldots, P$

- Form Cross spectrum
  - Pad: $A_{i,j,k} = 0_{d\times\nu}$ for $P < |k| \leq N_{ex}/2$[1]
  - Evaluate on unit circle

$$S = \texttt{fft}(A) \qquad \text{taken only in third dimension}$$

[1] it is a little more complicated than this

- Divide spectral factor:
  - $K \in \mathbb{C}^{d\times\nu\times N_{ex}}$
  - $K_k = S_k(L'_k)^{-1}$ for $k = 0, 1, \ldots, N_{ex}$
- Take causal part of K
  - $\tilde{K} = \texttt{ifft}(K)$
  - zeros negative lags of $\tilde{K}$
  - $K_+ = \texttt{ifft}(\tilde{K})$
- Divide again
  - $H_k = K_{+k}(L_k)^{-1}$ for $k = 0, 1, \ldots, N_{ex}$
- Extract filter
  - $\tilde{h} = \texttt{ifft}(H)$
  - $h = (\tilde{h}_k \in \mathbb{C}^{d\times\nu}; k = 0, 1, \ldots, M-1)$

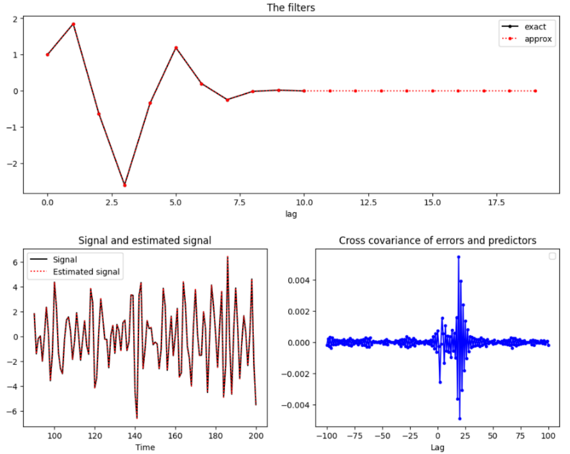
Choices

- P - number of autocovariance terms considered.
- Λ - windowing function.
- N_{ex} - size of grid on unit circle.
- M - number of taps in Wiener filter.

Example 1: AR(2) Predictors

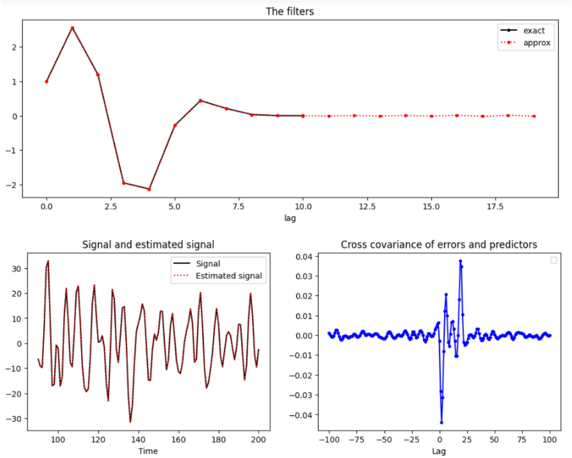
Program in Applied Mathematics

- X is an $AR(2)$ process, time series length 10,000. (zeros: $-0.767, -0.276$)
- $Y = f \star X$ is X filtered by a causal, stable 10-tap, filter f.



Example 2: ARMA(5,5) Predictors

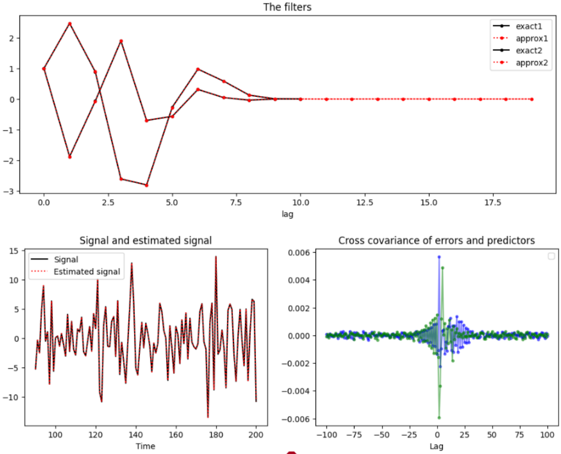
- X is an $ARMA(5,5)$ process, time series length 10,000. (zeros: 0.660198, 0.310362, -0.27755, -0.227891, -0.0424978; poles: -0.444798, -0.854724, -0.779101, -0.0154544, 0.794821)
- $Y = f \star X$ is X filtered by a causal, stable 10-tap, filter f.



Example 3: VAR(2) Predictors

Program in Applied Mathematics

- X is an $VAR(2)$ process, time series length 10,000.
  (Poles: 0.231959, -0.896785; 0.168827, -0.889844
- $Y = f \star X$ is X filtered by a causal, stable 10-tap, filter f.



Thank you!

Ali H Sayed and Thomas Kailath.
A survey of spectral factorization methods.
Numerical linear algebra with applications, 8(6-7):467–496, 2001.