Estimation and Kalman Filtering (Part 2)

ME C231B/EECS C220C

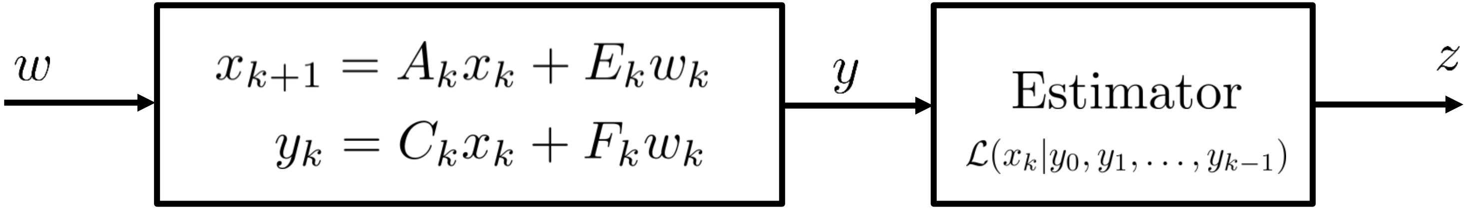
UC Berkeley

Spring 2018

A. Packard and F. Borrelli

Kalman Filter Setup

Linear dynamical system



- Discrete-time model
- (random variable) Internal state, $x_k \in \mathbb{R}^n$, not measured, but to-be-estimated
- (random variable) Driven by noise, $w_k \in \mathbb{R}^p$
- (random variable) Measured output, $y_k \in \mathbb{R}^m$, is noisy, linear combinations of x_k
- <u>Known</u> matrices, $\{A_k, E_k, C_k, F_k\}_{k=0}^{\infty}$ defining relationships between (x, y, w)
- <u>Known</u> statistical properties (means, variances, correlations) of $\{w_k\}_{k=0}^{\infty}$, and x_0
  - All other random variables, $y_0, y_1, \ldots,$ and $x_1, x_2, \ldots,$ are a consequence of the dynamic evolution

Build "optimal estimator" in context described, namely: $z_k = \mathcal{L}(x_k | y_0, y_1, \ldots, y_{k-1})$

Notation problem (just need to be clear)

The variables in play have lots of dependencies:

- Random variables: dependency on $\omega \in \Omega$
- Time-dependent: dependency on discrete-time index
- vector-valued: $\mathbb{R}^n$, so have a 1st element, 2nd element, ...

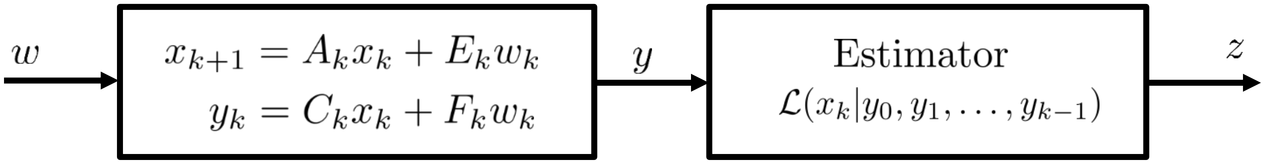
We will suppress the dependency on Ω, and just remember it is present

- several options for time/element (usually k, j for time)

$$x_k = \begin{bmatrix} x_k^{[1]} \\ x_k^{[2]} \end{bmatrix} \qquad x_k = \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} \qquad x_k = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \qquad x_k = \begin{bmatrix} x_k(1) \\ x_k(2) \end{bmatrix}$$

Bad, so no

Summary



Best linear estimator in following sense: at time k,

- the output z_k is an affine combination of $\{y_0, y_1, \ldots, y_{k-1}\}$. The coefficients of the affine combination, $(l_k, \{L_{k,j}\}_{j=1}^{k-1})$ can be functions of the known statistics on x_0 and w, as well as the known matrices A, E, C, F.

$$z_k = l_k + \sum_{j=0}^{k-1} L_{k,j} y_j$$

- For any particular (x_0, w) in the sample space, this estimator results in an estimation error, $e_k := x_k - z_k$. On some (x_0, w), the error might be very small; on other (x_0, w), the error might be large.
- The average of $e_k e_k^T$, taken over all (x_0, w) in the sample space, and weighted by their probabilities represents a matrix-valued cost (ie., the variance of the estimation error).
- Over all possible affine estimators, the estimator derived here yields the minimum **average** error.

Known statistical properties of disturbance and initial condition

$w \rightarrow$

$$\begin{aligned} x_{k+1} &= A_k x_k + E_k w_k \\ y_k &= C_k x_k + F_k w_k \end{aligned}$$

$\xrightarrow{y}$ Estimator $\mathcal{L}(x_k | y_0, y_1, \ldots, y_{k-1})$ $\xrightarrow{z}$

$$\begin{aligned}
\text{mean of initial condition: } & \mathbb{E}x_0 = m_0 \\
\text{variance of initial condition: } & \mathbb{E}\left[(x_0 - m_0)(x_0 - m_0)^T\right] = \Sigma_0 \\
\text{mean of disturbance: } & \mathbb{E}w_k = 0, \quad \forall k \\
\text{variance of disturbance: } & \mathbb{E}w_k w_k^T = W_k, \quad \forall k \\
\text{uncorrelation across time: } & \mathbb{E}w_k w_j^T = 0, \quad \text{for } k \neq j \\
\text{cross correlation: } & \mathbb{E}(x_0 - m_0)w_k^T = 0, \quad \forall k
\end{aligned}$$

The details of the random variables $x_0, w_0, w_1 \ldots$ will not be important in computing the estimator $\mathcal{L}(x_k | y_0, y_1, \ldots, y_{k-1})$

Only these various means and variances and covariances are assumed known, and the optimal unbiased linear estimator can be derived.

Example: probability model of initial condition and disturbance

Sample space with 20 outcomes, different probabilities

6 random variables, $x_0, w_0, w_1, \ldots, w_4$

$x_k \in \mathbb{R}^3, \quad w_k \in \mathbb{R}^2$

P	0.075	0.065	0.0479	0.0759	0.0436	0.0882	0.0707	0.0659	0.0418	0.0185	0.0133	0.0722	0.087	0.00505	0.00718	0.0717	0.0761	0.0269	0.0164	0.0314
x0(1)	-1.6e-01	-5.4e-01	1.3e-02	-2.4e-01	-1.3e+00	1.6e+00	-3.2e-01	1.2e+00	-2.5e-01	5.1e-02	-2.0e+00	6.6e-01	-7.0e-01	6.0e+00	-4.7e+00	8.3e-01	-6.0e-01	-4.0e-01	-5.6e-01	-5.8e-01
x0(2)	2.3e-01	1.3e+00	1.1e+00	-2.3e-01	8.1e-01	1.7e-01	-1.6e+00	3.0e-01	-1.1e+00	5.3e-01	-1.8e+00	-1.2e+00	-4.5e-01	-4.8e+00	-3.7e+00	7.3e-01	2.3e-01	2.0e+00	-2.7e-01	8.6e-01
x0(3)	-7.0e-01	7.8e-01	-2.3e-01	-7.0e-01	8.7e-01	1.3e+00	-5.8e-01	2.3e-01	1.8e+00	-2.3e+00	-9.2e-02	-9.5e-02	-5.1e-01	-4.4e+00	4.0e+00	8.6e-01	-5.0e-01	-2.2e+00	4.7e-01	-5.6e-01
w0(1)	-3.9e-02	4.5e-01	-2.0e-01	9.7e-01	-2.0e+00	-7.1e-01	4.2e-01	1.2e+00	8.4e-01	-6.9e-01	-3.4e+00	-5.9e-01	-1.3e+00	-5.9e-01	2.3e+00	-3.4e-01	1.1e+00	-3.9e-02	2.1e+00	7.0e-01
w0(2)	-6.2e-01	-5.3e-01	1.6e-01	4.1e-01	2.2e-01	-9.7e-01	-2.1e+00	1.1e+00	-8.5e-01	1.1e+00	7.7e-01	1.4e+00	4.3e-01	1.3e-02	1.1e+00	4.0e-01	3.1e-01	-2.2e+00	1.0e+00	1.4e+00
w1(1)	1.9e-01	-8.8e-01	1.6e-02	-2.0e-01	8.5e-01	-3.1e-01	2.7e-01	-2.6e-01	4.1e-01	1.4e+00	-6.3e+00	-2.7e-01	1.4e+00	-4.2e-01	9.8e-01	5.6e-01	-8.9e-01	-3.6e-01	1.5e+00	-3.0e-01
w1(2)	-1.0e+00	-7.1e-01	-2.0e+00	3.0e-01	2.2e+00	3.2e-01	3.9e-01	7.8e-01	5.2e-01	-2.6e-01	-2.9e-01	-5.5e-01	9.8e-02	-7.6e-01	-4.9e+00	-7.9e-01	1.3e+00	-5.3e-01	5.8e-01	2.5e-01
w2(1)	-1.7e+00	6.8e-01	1.2e+00	-2.4e-01	5.2e-01	-1.8e-01	1.0e+00	1.6e+00	-9.9e-01	1.4e+00	1.0e+00	-1.5e+00	3.3e-01	3.1e+00	2.3e+00	-2.3e-01	-7.9e-01	-5.4e-01	-7.3e-01	1.5e-01
w2(2)	1.4e-01	4.4e-01	4.0e-01	-8.9e-01	-4.0e-01	-1.1e+00	-2.2e-01	1.5e+00	1.4e+00	-3.5e+00	-5.4e-01	5.8e-01	1.2e+00	-3.7e-01	-2.1e+00	-5.2e-01	-6.7e-02	1.1e+00	-2.5e+00	-7.1e-01
w3(1)	-3.9e-01	-1.3e+00	6.5e-01	1.2e+00	-1.3e+00	3.3e-01	7.9e-01	5.4e-01	-1.1e+00	-3.5e-01	7.5e-01	-5.3e-02	5.3e-01	-8.4e+00	-1.2e+00	7.3e-01	1.1e-01	-5.9e-01	-1.3e+00	-1.5e+00
w3(2)	-6.6e-02	-6.6e-01	-7.7e-01	-1.5e+00	-6.5e-01	-8.0e-01	1.1e+00	1.7e-02	2.7e-01	-3.7e-01	1.5e+00	-3.1e-01	-1.6e-01	-9.8e-01	-1.7e+00	2.1e+00	-4.4e-02	2.9e-02	6.3e-01	2.7e+00
w4(1)	1.5e+00	-4.8e-01	3.4e-01	9.2e-01	1.4e+00	-1.1e+00	4.6e-02	9.0e-01	1.2e+00	5.4e-01	1.7e+00	-5.3e-01	-1.1e+00	2.4e-01	-2.1e+00	3.7e-01	-1.4e+00	-8.6e-01	7.1e-01	-1.2e+00
w4(2)	-2.2e-02	-6.9e-01	-2.5e-01	-7.3e-01	2.9e-01	-2.6e-01	-5.5e-01	4.5e-01	-7.9e-01	-7.4e-01	1.2e+00	-5.0e-01	1.1e-01	2.7e+00	2.6e+00	1.1e+00	1.1e+00	1.8e+00	2.8e+00	-3.2e+00

mean of initial condition: $\mathbb{E}x_0 = 0_3$

variance of initial condition: $\mathbb{E}\left[x_0 x_0^T\right] = I_3$

mean of disturbance: $\mathbb{E}w_k = 0_2, \quad \forall k$

variance of disturbance: $\mathbb{E}w_k w_k^T = I_2, \quad \forall k$

uncorrelation across time: $\mathbb{E}w_k w_j^T = 0_{2\times 2}, \quad \text{for } k \neq j$

cross correlation: $\mathbb{E}x_0 w_k^T = 0_{3\times 2}, \quad \forall k$

So, this probability model for an initial condition, and (finite) sequence of disturbances exactly fits into the setup. Obviously there are "many" more such "tables" that would have different data, but the same $\boxed{\text{coarse}}$ properties.... And, we will see that the estimator only depends on these $\boxed{\text{listed}}$ coarse properties, not the details of the probability model.

With this in place, we begin the Kalman Filter (the estimator) derivation...

Linear system evolution: state

$$x_{k+1} = A_k x_k + E_k w_k$$

$$\mathcal{A}_{k,j} := A_{k-1} \cdots A_j, \quad j < k$$

(note: caligraphic $\mathcal{A}$ versus A)

If $w_k = 0$ for all k, then $x_k = \mathcal{A}_{k,0} x_0$

$$x_1 = A_0 x_0 + E_0 w_0$$

$$x_2 = A_1 A_0 x_0 + A_1 E_0 w_0 + E_1 w_1$$

$$x_3 = A_2 A_1 A_0 x_0 + A_2 A_1 E_0 w_0 + A_2 E_1 w_1 + E_2 w_2$$

$$x_4 = \underbrace{A_3 A_2 A_1 A_0}_{\mathcal{A}_{4,0}} x_0 + \underbrace{A_3 A_2 A_1}_{\mathcal{A}_{4,1}} E_0 w_0 + \underbrace{A_3 A_2}_{\mathcal{A}_{4,2}} E_1 w_1 + \underbrace{A_3}_{\mathcal{A}_{4,3}} E_2 w_2 + E_3 w_3$$

$$\boxed{x_k = \mathcal{A}_{k,0} x_0 + \mathcal{A}_{k,1} E_0 w_0 + \cdots + \mathcal{A}_{k,k-1} E_{k-2} w_{k-2} + E_{k-1} w_{k-1}}$$

Linear system evolution: state

$$x_{k+1} = A_k x_k + E_k w_k$$

$$\mathcal{A}_{k,j} := A_{k-1} \cdots A_j, \quad k > j \qquad \mathcal{A}_{k,k} := I$$

$$\boxed{x_k = \mathcal{A}_{k,0} x_0 + \mathcal{A}_{k,1} E_0 w_0 + \cdots + \mathcal{A}_{k,k-1} E_{k-2} w_{k-2} + E_{k-1} w_{k-1}}$$

$$x_k = \mathcal{A}_{k,0} x_0 + \underbrace{\begin{bmatrix} \mathcal{A}_{k,1} E_0 & \cdots & \mathcal{A}_{k,k-1} E_{k-2} & E_{k-1} \end{bmatrix}}_{=:\, \mathcal{B}_{A,E,k}} \overbrace{\begin{bmatrix} w_0 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}}^{\mathcal{W}_{k-1}}$$

$$\boxed{x_k = \mathcal{A}_{k,0} x_0 + \mathcal{B}_{A,E;k} \mathcal{W}_{k-1}} \qquad \boxed{\Rightarrow \ \Sigma_{x_k, w_k} = 0}$$

State x_k is a known linear combination of $x_0, w_0, \ldots, w_{k-1}$

Linear system evolution: collecting several time-steps

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{k-2} \\ x_{k-1} \end{bmatrix} = \underbrace{\begin{bmatrix} I \\ \mathcal{A}_{1,0} \\ \mathcal{A}_{2,0} \\ \vdots \\ \mathcal{A}_{k-2,0} \\ \mathcal{A}_{k-1,0} \end{bmatrix}}_{:= \Psi_{k-1}} x_0 + \underbrace{\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ E_0 & 0 & 0 & \cdots & 0 & 0 \\ \mathcal{A}_{2,1}E_0 & E_1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_{k-2,1}E_0 & \mathcal{A}_{k-2,2}E_1 & \mathcal{A}_{k-2,3}E_2 & \cdots & 0 & 0 \\ \mathcal{A}_{k-1,1}E_0 & \mathcal{A}_{k-1,2}E_1 & \mathcal{A}_{k-1,3}E_2 & \cdots & E_{k-2} & 0 \end{bmatrix}}_{:= \Gamma_{k-1}} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}$$

Next: express y as combinations of x_0 and all w

Linear system evolution: measured output

State x at time k is a known linear combination of $x_0, w_0, \ldots, w_{k-1}$

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{k-2} \\ x_{k-1} \end{bmatrix} = \begin{bmatrix} I \\ \mathcal{A}_{1,0} \\ \mathcal{A}_{2,0} \\ \vdots \\ \mathcal{A}_{k-2,0} \\ \mathcal{A}_{k-1,0} \end{bmatrix} x_0 + \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ E_0 & 0 & 0 & \cdots & 0 & 0 \\ \mathcal{A}_{2,1}E_0 & E_1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_{k-2,1}E_0 & \mathcal{A}_{k-2,2}E_1 & \mathcal{A}_{k-2,3}E_2 & \cdots & 0 & 0 \\ \mathcal{A}_{k-1,1}E_0 & \mathcal{A}_{k-1,2}E_1 & \mathcal{A}_{k-1,3}E_2 & \cdots & E_{k-2} & 0 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}$$

$$y_k = C_k x_k + F_k w_k$$

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{k-2} \\ y_{k-1} \end{bmatrix} = \begin{bmatrix} C_0 \\ C_1\mathcal{A}_{1,0} \\ C_2\mathcal{A}_{2,0} \\ \vdots \\ C_{k-2}\mathcal{A}_{k-2,0} \\ C_{k-1}\mathcal{A}_{k-1,0} \end{bmatrix} x_0 + \begin{bmatrix} F_0 & 0 & 0 & \cdots & 0 & 0 \\ C_1E_0 & F_1 & 0 & \cdots & 0 & 0 \\ C_2\mathcal{A}_{2,1}E_0 & C_2E_1 & F_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{k-2}\mathcal{A}_{k-2,1}E_0 & C_{k-2}\mathcal{A}_{k-2,2}E_1 & C_{k-2}\mathcal{A}_{k-2,3}E_2 & \cdots & F_{k-2} & 0 \\ C_{k-1}\mathcal{A}_{k-1,1}E_0 & C_{k-1}\mathcal{A}_{k-1,2}E_1 & C_{k-1}\mathcal{A}_{k-1,3}E_2 & \cdots & C_{k-1}E_{k-2} & F_{k-1} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}$$

Correlation: output and disturbance (fact #2)

$$\boxed{\Sigma_{y_k,w_k} = F_k W_k}$$

$$y_k = C_k x_k + F_k w_k \qquad\qquad \Sigma_{x_k,w_k} = 0$$

$$\begin{aligned}
\Rightarrow\ \Sigma_{y_k,w_k} &= \mathbb{E}\left[(C_k(x_k - m_k) + F_k w_k) w_k^T\right] \\
&= C_k \Sigma_{x_k,w_k} + F_k \Sigma_{w_k} \\
&= F_k W_k
\end{aligned}$$

Linear system evolution: simple variances and correlations

$$x_{k+1} = A_k x_k + E_k w_k \qquad y_k = C_k x_k + F_k w_k$$

Governing equations

$$\mathcal{W}_{k-1} := \begin{bmatrix} w_0 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}$$

$$x_k = \mathcal{A}_{k,0} x_0 + \mathcal{B}_{A,E;k} \mathcal{W}_{k-1}$$

$$y_k = C_k \mathcal{A}_{k,0} x_0 + C_k \mathcal{B}_{A,E;k} \mathcal{W}_{k-1} + F_k w_k$$

Derived evolution

$$\mathcal{Y}_{k-1} := \begin{bmatrix} y_0 \\ \vdots \\ y_{k-2} \\ y_{k-1} \end{bmatrix}$$

Given statistics

$$\mathbb{E} x_0 = m_0 \quad \mathbb{E}(x_0 - m_0)(x_0 - m_0)^T = \Sigma_0 \quad \mathbb{E}(x_0 - m_0) w_k^T = 0$$

$$\mathbb{E} w_k = 0 \quad \mathbb{E} w_k w_k^T = W_k \quad \mathbb{E} w_k w_j^T = 0 \text{ for } k \neq j$$

$$\mathbb{E} \mathcal{W}_k = 0 \qquad \mathbb{E} x_k = \mathcal{A}_{k,0} m_0$$

$$\Sigma_{x_k, w_k} = 0 \quad \mathbb{E} \mathcal{W}_{k-1} w_k^T = 0 \quad \Sigma_{y_k, w_k} = F_k W_k$$

Derived statistics (simple and fact #2)

Setup for "batch" estimation

$$x_k = \underbrace{\mathcal{A}_{k,0}}_{M_k} x_0 + \underbrace{\begin{bmatrix} \mathcal{A}_{k,1}E_0 & \mathcal{A}_{k,2}E_1 & \mathcal{A}_{k,3}E_2 & \cdots & \mathcal{A}_{k,k-1}E_{k-2} & E_{k-1} \end{bmatrix}}_{N_k (= \mathcal{B}_{A,E,k})} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}$$

$$Z := \quad x_k = M_k x_0 + N_k \mathcal{W}_{k-1}$$

$$P := \quad \mathcal{Y}_{k-1} = R_k x_0 + S_k \mathcal{W}_{k-1}$$

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{k-2} \\ y_{k-1} \end{bmatrix} = \underbrace{\begin{bmatrix} C_0 \\ C_1\mathcal{A}_{1,0} \\ C_2\mathcal{A}_{2,0} \\ \vdots \\ C_{k-2}\mathcal{A}_{k-2,0} \\ C_{k-1}\mathcal{A}_{k-1,0} \end{bmatrix}}_{R_k} x_0 + \underbrace{\begin{bmatrix} F_0 & 0 & 0 & \cdots & 0 & 0 \\ C_1E_0 & F_1 & 0 & \cdots & 0 & 0 \\ C_2\mathcal{A}_{2,1}E_0 & C_2E_1 & F_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{k-2}\mathcal{A}_{k-2,1}E_0 & C_{k-2}\mathcal{A}_{k-2,2}E_1 & C_{k-2}\mathcal{A}_{k-2,3}E_2 & \cdots & F_{k-2} & 0 \\ C_{k-1}\mathcal{A}_{k-1,1}E_0 & C_{k-1}\mathcal{A}_{k-1,2}E_1 & C_{k-1}\mathcal{A}_{k-1,3}E_2 & \cdots & C_{k-1}E_{k-2} & F_{k-1} \end{bmatrix}}_{S_k} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{k-2} \\ w_{k-1} \end{bmatrix}$$

“Batch” solution to optimal linear estimation problem (fact #3)

$$Z := \qquad x_k = M_k x_0 + N_k \mathcal{W}_{k-1}$$

$$P := \qquad \mathcal{Y}_{k-1} = R_k x_0 + S_k \mathcal{W}_{k-1}$$

$$\mathbb{E} x_0 = m_0$$

$$\mathbb{E} \mathcal{W}_{k-1} = 0$$

$$\mathbb{E}(x_0 - m_0)(x_0 - m_0)^T = \Sigma_0$$

$$\begin{aligned} \mathbb{E} \mathcal{W}_{k-1} \mathcal{W}_{k-1}^T &= \text{blkdiag}(W_0, \ldots, W_{k-1}) \\ &=: \bar{\Sigma}_W \end{aligned}$$

$$\mathbb{E}(x_0 - m_0) \mathcal{W}_{k-1}^T = 0$$

$$\boxed{\begin{aligned} \Sigma_{Z,P} &= M_k \Sigma_0 R_k^T + N_k \bar{\Sigma}_W S_k^T \\ \Sigma_P &= R_k \Sigma_0 R_k^T + S_k \bar{\Sigma}_W S_k^T \\ \mu_Z &= M_k m_0, \; \mu_P = R_k m_0 \end{aligned}}$$

$$\begin{aligned} \mathcal{L}(x_k | y_0, y_1, \ldots, y_{k-1}) &= \mathcal{L}(x_k | \mathcal{Y}_{k-1}) \\ &= \mathcal{L}(Z|P) \\ &\overset{1}{=} \Sigma_{Z,P} \Sigma_P^{-1} (P - \mu_P) + \mu_Z \\ &= (M_k \Sigma_0 R_k^T + N_k \bar{\Sigma}_W S_k^T)(R_k \Sigma_0 R_k^T + S_k \bar{\Sigma}_W S_k^T)^{-1} (\mathcal{Y}_{k-1} - R_k m_0) + M_k m_0 \end{aligned}$$

This is, indeed, an affine combination of $y_0, y_1, \ldots, y_{k-1}$

Properties of Batch solution

Benefits

- Solves problem
- Easy to understand
  - Optimal estimate at any time is one application of linear minimum variance estimate

$$\hat{x}_{k|k-1} = (M_k \Sigma_0 R_k^T + N_k \bar{\Sigma}_W S_k^T)(R_k \Sigma_0 R_k^T + S_k \bar{\Sigma}_W S_k^T)^{-1}(\mathcal{Y}_{k-1} - R_k m_0) + M_k m_0$$

Issues

- Matrices/Computation grow at each time step
- Much the work at time-step is discarded when moving to (k+1)

Challenge

- Can we do it recursively?