THE KALMAN FILTER: OPTIMAL STATE ESTIMATION IN THE PRESENCE OF NOISE — lectures 1 and 2

John M. Finn

9th February 2004

UQWG seminar, Feb 5, 2004

http://public.lanl.gov/kmh/uncertainty/meetings
Outline

- **Counting statistics** with equal $\sigma_i$ by least squares approach. Minimum variance. Recursive nature.

- **Counting statistics** with unequal $\sigma_i$. Least squares, minimum variance approach. Recursive nature.

- Linear process with measurement noise only – estimating initial vs. current state.

- **Random walk** with zero measurement noise. Estimating the initial position.

- **Random walk**, estimating the current position.

- **Random walk** with measurement noise, estimating the current state.

- Preview of next lecture.
Counting statistics; sample mean and variance — equal $\sigma_i^2$

\[ x_i = x_0 + \xi_i, \quad y_i = x_i, \]
\[ < \xi_i > = 0, \quad < \xi_i \xi_j > = \sigma_i^2 \delta_{ij} = \delta_{ij} \text{ and gaussian distribution by Bayes' theorem} \]

\[
f(x_0(n)|y_1, ..., y_n) \propto f(y_1, ..., y_n|x_0(n)) \times \frac{\text{prior}}{\text{normalization}}
\]
\[ \sim \prod_{i=1}^{n} e^{-\frac{1}{2\sigma_i^2}[x_i-x_0(n)]^2} = e^{-\sum_{i=1}^{n} \frac{1}{2\sigma_i^2}[x_i-x_0(n)]^2} \]

Maximum likelihood

\[ \chi^2(n) = -\ln f = \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{[x_i - x_0(n)]^2}{2}. \]
\[ \partial \chi^2 / x_0(n) = 0 \Rightarrow \text{state estimate} \]

\[ x_0(n) = \frac{1}{n} \sum_{i=1}^{n} x_i, \]

The variance of the estimate at this stage is (uncorrelated)

\[ V(n) = \sigma^2(n) = \sum_{i=1}^{n} \sigma_i^2 \left( \frac{\partial x_0(n)}{\partial x_i} \right)^2 = \frac{1}{n}, \]

y and x0 vs time
y and x0 vs time
Minimum variance approach

\[ x_0(n) = \sum_{i=1}^{n} \rho_i x_i, \]

with \( \sum_{i=1}^{n} \rho_i = 1, \quad \sigma^2 = 1 \)

\[ V(n) = \sum_{i=1}^{n} \rho_i^2 \quad V^*(n) = \sum_{i=1}^{n} \rho_i^2 - \lambda \sum_{i=1}^{n} \rho_i \]

\[ \frac{\partial V(n)}{\partial \rho_k} = 0 \quad \Rightarrow \]

\[ \rho_k = \frac{\lambda}{2}, \]

or \( \rho_k = \frac{1}{n} \) for all \( k \)

\[ x_0(n) = \frac{1}{n} \sum_{i=1}^{n} x_i \quad V(n) = \frac{1}{n} \]
Recursive Kalman filter form

\[(n + 1)x_0(n + 1) = \sum_{i=1}^{n} x_i + x_{n+1}\]

\[x_0(n + 1) = \frac{n}{n + 1} x_0(n) + \frac{1}{n + 1} x_{n+1}\]

or

\[x_0(n + 1) = x_0(n) + K_n [x_{n+1} - x_0(n)],\]

Kalman gain \(K_n\)

\[K_n = \frac{1}{n + 1}.\]

\[\frac{1}{K_n} = \frac{1}{K_{n-1}} + 1\]

or

\[K_n = \frac{K_{n-1}}{1 + K_{n-1}}\]

\[K_n = V(n + 1),\]
Counting statistics for unequal $\sigma_i^2$

Uncorrelated but different confidence: $< \xi_i \xi_j > = \sigma_i^2 \delta_{ij}$

$$\chi^2 = \sum_{i=1}^{n} \frac{[x_i - x_0(n)]^2}{2\sigma_i^2}$$

$$\partial W/\partial x_0(n) = 0 \implies$$

$$x_0(n) = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}.$$

Example: $(x_1, x_2, x_3), x_4$

Take $z_1 = (x_1 + x_2 + x_3)/3$, $z_2 = x_4$. $\sigma_1^2 = 1/3$, $\sigma_2^2 = 1$

Then $x_0(4) = (z_1/\sigma_1^2 + z_2/\sigma_2^2) / (1/\sigma_1^2 + 1/\sigma_2^2)$

$$= (x_1 + x_2 + x_3 + x_4)/4$$
\[ V(n) = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}, \]

Again, take

\[ x_0(n) = \sum_{i=1}^{n} \rho_i x_i, \]

with \( \sum_{i=1}^{n} \rho_i = 1 \)

\[ V^*(n) = \sum_{i=1}^{n} \sigma_i^2 \rho_i^2 - \lambda \sum_{i=1}^{n} \rho_i \]

\[ \rho_k = \frac{\lambda}{2\sigma_k^2} = \frac{1/\sigma_k^2}{\sum_i 1/\sigma_i^2}. \]

Same.
Recursive Kalman filter form for unequal $\sigma^2_i$

\[ x_0(n + 1) = x_0(n) + K_n[x_{n+1} - x_0(n)] \]

with

\[ K_n = \frac{1}{\sigma^2_{n+1} \left( \sum_{i=1}^{n} \frac{1}{\sigma^2_i} + \frac{1}{\sigma^2_{n+1}} \right)} = \frac{1}{\sigma^2_{n+1} \sum_{i=1}^{n} \frac{1}{\sigma^2_i} + 1}. \]

\[ K_n = V(n + 1)/\sigma^2_{n+1}, \text{ and} \]

\[ \frac{1}{K_n} = \left( \frac{\sigma^2_{n+1}}{\sigma^2_n} \right) \frac{1}{K_{n-1}} + 1 \quad \text{or} \quad K_n = \frac{K_{n-1}}{K_{n-1} + \sigma^2_{n+1}/\sigma^2_n}, \]

The recursion in terms of the variance

\[ \frac{1}{V(n + 1)} = \frac{1}{V(n)} + \frac{1}{\sigma^2_{n+1}}. \]

with $K_n = V(n + 1)/\sigma^2_{n+1}$ $K_n$ tends to decrease with $n$ (more data)

If $\sigma^2_{n+1} < \sigma^2_n$, then $K_n$ will be larger than if $\sigma^2_{n+1} > \sigma^2_n$
One dimensional example of estimating the initial state and the current state

Simple stochastic system with measurement noise

\[ x_{k+1} = \gamma x_k, \]

\[ y_k = x_k + \eta_k. \]

\[ \langle \eta_k \eta_l \rangle = \delta_{kl} \]

\[ \chi^2 = \frac{1}{2} \sum_{k=1}^{n} (\gamma^k x_0 - y_k)^2, \]

\[ \frac{\partial \chi^2}{\partial x_0} = 0 \] gives

\[ x_0(n) = \frac{\sum_{k=1}^{n} \gamma^k y_k}{\sum_{k=1}^{n} \gamma^{2k}}. \]

\[ \gamma > 1 \ldots \text{weighted toward recent results, } \gamma < 1 \ldots \text{weighted toward initial results. } \]

Recursive form

\[ x_0(n+1) = x_0(n) + \frac{\gamma^{n+1}}{\sum_{k=1}^{n} \gamma^{2k} + \gamma^{2n+2}} (y_{n+1} - \gamma^{n+1} x_0(n)). \]
An estimate of $x_n$ rather than $x_0$.

$$
\chi^2 = \frac{1}{2} \sum_{k=1}^{n} (\gamma^{k-n} x_n - y_k)^2,
$$

$$
x_n(n) = \frac{\sum_{k=1}^{n} \gamma^{k-n} y_k}{\sum_{k=1}^{n} \gamma^{2k-2n}} = \gamma^n x_0(n),
$$

Exactly what you might guess. Recursive form:

$$
x_{n+1}(n+1) = \gamma x_n(n) + \frac{1}{\sum_{k=1}^{n} \gamma^{2k-2n-2} + 1} (y_{n+1} - \gamma x_n(n)).
$$

Notice $K_n x_n(n) = \gamma^{n+1} K_n x_0(n)$.
Random walk with zero measurement error – estimating the initial position

Counting statistics, with only measurement noise, is:

\[ x_{k+1} = x_k, \]

\[ y_k = x_k + \eta_k \]

Random walk problem (Wiener process, Brownian motion), with only dynamical noise:

\[ x_{k+1} = x_k + \xi_k, \]

\[ y_k = x_k. \]

\[ < \xi_k > = 0,\ < \xi_k \xi_k > = \sigma_0^2 \delta_{kl}. \] To estimate the initial position. Ship wrecks at \( x_0 \) – to find the ship.

\[ y_k = x_0 + \sum_{i=0}^{k-1} \xi_i = x_0 + \zeta_k. \]

\( \zeta_k \) has \( < \zeta_k > = 0 \)
And \( n \times n \) covariance matrix

\[
C_{kl} = \langle \zeta_k \zeta_l \rangle = C_{kl} = \langle \zeta_k \zeta_l \rangle = \sum_{i=0}^{k} \sum_{j=0}^{l} \langle \xi_i \xi_j \rangle = \sigma_0^2 \min(k, l),
\]

i.e.

\[
C = \sigma_0^2 \begin{bmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & 2 & 2 & 2 & \cdots & 2 \\
1 & 2 & 3 & 3 & \cdots & 3 \\
1 & 2 & 3 & 4 & \cdots & 4 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 2 & 3 & 4 & \cdots & n
\end{bmatrix}.
\]

Least squares in terms of the inverse of the covariance matrix \( D = C^{-1} \)

\[
\chi^2 = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} \zeta_k D_{kl} \zeta_l = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} (y_k - x_0) D_{kl} (y_l - x_0).
\]
\[ \frac{\partial \chi^2}{\partial x_0} = 0 \quad \Rightarrow \]

\[
x_0(n) = \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} D_{kl} y_l}{\sum_{k=1}^{n} \sum_{l=1}^{n} D_{kl}},
\]

\[
D = \sigma_0^{-2} \begin{bmatrix}
2 & -1 & 0 & 0 & \cdots \\
-1 & 2 & -1 & 0 & \cdots \\
0 & -1 & 2 & -1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
0 & 0 & 0 & -1 & 1
\end{bmatrix}.
\]

In the estimate, the value of \( \sigma_0^2 \) cancels.

\[
\sum_{kl} D_{kl} = 1, \quad \sum_{kl} D_{kl} y_l = y_1 \quad x_0(n) = y_1
\]

\[
V(n) = \sum_{ij} \frac{\partial x_0(n)}{\partial y_i} C_{ij} \frac{\partial x_0(n)}{\partial y_j} = C_{11} = 1.
\]

Recursive \( x_0(n + 1) = x_0(n) + K_n [y_{n+1} - x_0(n)] \) with \( K_n = 0 \).
Try minimum variance again

\[
x_0(n) = \sum_{i=1}^{n} \rho_i y_i
\]

\[
V^*(n) = \sum_{ij} C_{ij} \rho_i \rho_j - \lambda \sum_i \rho_i;
\]

\[
\partial V(n) / \partial \rho_k = 0 \quad \Rightarrow \quad \sum_j C_{kj} \rho_j = \lambda / 2
\]

\[
\rho_i = \frac{\lambda}{2} \sum_j D_{ij} \quad \text{or} \quad \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{pmatrix} = \frac{\lambda}{2} D \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.
\]

\[
\frac{\lambda}{2} = \frac{1}{\sum_{ij} D_{ij}}; \quad \rho_i = \frac{\sum_j D_{ij}}{\sum_{ij} D_{ij}}.
\]

\[
\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},
\]

\[
x_0(n) = y_1
\]

\[
V(n) = \sum_{ij} C_{ij} \rho_i \rho_j = C_{11} = 1. \quad \text{A third approach – next lecture.}
\]
Random walk with zero measurement noise — estimating the current position

\[ x_{k+1} = x_k + \xi_k, \]
\[ y_k = x_k. \]

Ship wrecks at \( x_0 \), but we wish to find the position of the survivor.

\[ y_k = x_n - \sum_{i=k}^{n-1} \xi_i = x_n(n) - \zeta_k, \]
\[ \langle \xi_i \rangle = 0, \quad \langle \xi_i \xi_j \rangle = \sigma_i^2 \delta_{ij} \quad \zeta_k^{new} = \zeta_n^{old} - \zeta_k^{old}. \]

\[ \chi^2 = \frac{1}{2} \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} \zeta_k D_{kl} \zeta_l = \frac{1}{2} \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} (x_n - y_k) D_{kl} (x_n - y_l). \]
\[ C_{kl} = \langle \zeta_k \zeta_l \rangle = \sum_{i=k}^{n-1} \sum_{j=l}^{n-1} \langle \xi_i \xi_j \rangle \]

\[ = \sigma_0^2 \sum_{i=k}^{n-1} \sum_{j=l}^{n-1} \delta_{ij} = \sigma_0^2 [n - \max(k, l)], \]

\[ C = \sigma_0^2 \begin{bmatrix}
  n - 1 & n - 2 & n - 3 & \cdots & 1 \\
  n - 2 & n - 2 & n - 3 & \cdots & 1 \\
  n - 3 & n - 3 & n - 3 & \cdots & 1 \\
  \vdots & \vdots & \vdots & \ddots & 1 \\
  1 & 1 & 1 & \cdots & 1
\end{bmatrix}. \]

\[ D = \sigma_0^{-2} \begin{bmatrix}
  1 & -1 & 0 & 0 & \cdots \\
  -1 & 2 & -1 & 0 & \cdots \\
  0 & -1 & 2 & -1 & \cdots \\
  \vdots & \vdots & \vdots & \ddots & \cdots \\
  0 & 0 & 0 & -1 & 2
\end{bmatrix}, \]

\[ x_n(n) = y_{n-1}, \]

\[ V(n) = 1. \]

Recursive \[ x_0(n+1) = x_0(n) + K_n[y_n - x_0(n)] \] with \( K_n = 1 \) now.
Random walk with measurement noise – estimating the current position

Estimating the current position of the shipwreck survivor

\[ x_{k+1} = x_k + \xi_k, \]

\[ y_k = x_k + \eta_k. \]

Solve for \( y_k \) in terms of \( x_n \):

\[ y_k = x_n - \sum_{i=k}^{n-1} \xi_i + \eta_k = x_n - \zeta_k + \eta_k. \]

\[ \chi^2 = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} (x_n - y_k) D_{kl} (x_n - y_l), \]

\( D = C^{-1} \), with \( C_{kl} = \langle (-\zeta_k + \eta_k)(-\zeta_l + \eta_l) \rangle \). Again using \( \langle \xi_k \xi_l \rangle = \sigma_0^2 \delta_{kl}, \langle \eta_k \eta_l \rangle = \sigma_1^2 \delta_{kl}, \langle \zeta_k \eta_l \rangle = 0 \) we have, for \( k = 1, \ldots, m \)
\[ C_{kl}^{(n)} = \sigma_0^2 [n - \max(k, l)] + \sigma_1^2 \delta_{kl}, \]

or

\[
C^{(n)} = \sigma_0^2 \begin{bmatrix}
  n - 1 & n - 2 & n - 3 & \cdots & 0 \\
  n - 2 & n - 2 & n - 3 & \cdots & 0 \\
  n - 3 & n - 3 & n - 3 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & 0 & \cdots & 0
\end{bmatrix}
\]

\[ + \sigma_1^2 \begin{bmatrix}
  1 & 0 & 0 & \cdots & 0 \\
  0 & 1 & 0 & \cdots & 0 \\
  0 & 0 & 1 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & 0 & \cdots & 1
\end{bmatrix} \]

\[ = \sigma_0^2 C^{(0,n)} + \sigma_1^2 C^{(1,n)} = C^{(n)}. \]
NEXT LECTURE

- Probabilistic (Bayesian) approach

- Application to higher dimension, with dynamical and measurement noise

- Application to control theory

- Application to nonlinear problems – the extended Kalman filter
Outline - Second Lecture

• Review of previous lecture

• Estimation of $M$ correlated variables. Alternate method based on the trace of the covariance matrix.

• Alternate method for the random walk with zero measurement noise. Estimating the initial or current position

• Probabilistic (Bayesian) approach

• Alternate method, for random walk with measurement noise added

• Higher dimensional stochastic process with measurement noise

• Application to control theory – the separation theorem

• Nonlinear stochastic systems – the Extended Kalman Filter
REVIEW: ESTIMATING A SCALAR VARIABLE

Measurement of a scalar – measurement noise but no dynamical noise

\[ \chi^2 = \sum_{i=1}^{n} \frac{[x_i - x_0(n)]^2}{2\sigma_i^2} \]

\[ x_0(n) = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} \]

Minimum variance form \( x_0(n) = \sum_{i=1}^{n} \rho_i x_i \), with

\[ V^*(n) = \sum_{i=1}^{n} \sigma_i^2 \rho_i^2 - \lambda \sum_{i=1}^{n} \rho_i \]

Recursive form: \( x_0(n+1) = x_0(n) + K_n [x_{n+1} - x_0(n)] \)

Innovation, Kalman gain (matrix)
ESTIMATION OF A CORRELATED HIGHER DIMENSIONAL VARIABLE

\[ x^i = x_0 + \sum \xi^i \]

with \[< \xi^i > = 0 < \xi^i \xi^j > = \delta_{ij} C_{kl} \]

\[ \chi^2 = \frac{1}{2} \sum_{i=1}^{n} (\xi^i, D^i \xi^i) \]

\[ = \frac{1}{2} \sum_{i=1}^{n} ( (x^i - x_0), D^i (x^i - x_0)) \]

where \( D^i = (C^i)^{-1} \).

\[ x_0 = \left( \sum_{i=1}^{n} D^i \right)^{-1} \sum_{i=1}^{n} D^i x^i \quad (1) \]

Note: this gives sample mean if all \( D_i \) are equal.

Also, if \( D_i \) are diagonal, this gives the weighted sample mean.
MINIMUM VARIANCE ALTERNATIVE
– TRACE OF THE COVARIANCE
MATRIX

\[ x_0(n) = \sum_{i=1}^{n} A^i x^i \]
\[ \sum_{i=1}^{n} A^i = 1 \]

\[ C(\vec{x}_0)_{kl} = \langle \delta x_{0,k} \delta x_{0,l} \rangle. \text{ Its trace is} \]

\[ T = \langle \sum_k \delta x_{0,k} \delta x_{0,k} \rangle = \langle |\delta x_0|^2 \rangle \]

\[ T = \sum_{ijkmn} A^i_{km} A^j_{kn} \langle \xi^i_m \xi^j_n \rangle = \sum_{ikmn} A^i_{km} A^i_{kn} C^i_{mn} \]

\[ T = \text{trace} \sum_i \left( ACA^T \right)^i. \text{ Minimize } T \]
\[ T^* = \sum_{ikmn} A^i_{km} A^i_{kn} C^i_{mn} - \sum_{mn} \lambda_{mn} \left( \sum_i A^i_{mn} - \delta_{mn} \right) \]

Differentiating with respect to \( A^i_{ab} \)
\[ \frac{\partial T^*}{\partial A^i_{ab}} = 0 \]

\[ 2A^i_{an} C^i_{nb} = \lambda_{ab} \quad A = \frac{1}{2} LD \quad L = 2 \left( \sum_i D^i \right)^{-1} \]

\[ A^i = \left( \sum_i D^i \right)^{-1} D^i \]

same as before

\[ x_0 = \left( \sum_{i=1}^{n} D^i \right)^{-1} \sum_{i=1}^{n} D^i x^i \quad (2) \]
**Random walk, estimating the current state – another alternate**

Recall $y_k = x_n - \sum_{i=k}^{n-1} \xi_i = x_n - \zeta_k$ ...

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} (x_n - y_k) D_{kl}(x_n - y_l) \quad (3)$$

$$C = \sigma_0^2 \begin{bmatrix} n - 1 & n - 2 & n - 3 & \cdots & 1 \\ n - 2 & n - 2 & n - 3 & \cdots & 1 \\ n - 3 & n - 3 & n - 3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \quad (4)$$

Alternatively,

$$\chi^2 = \frac{1}{2\sigma_0^2} \sum_{k=1}^{n-1} \xi_i^2$$

$$= \frac{1}{2\sigma_0^2} \left[ (y_2 - y_1)^2 + \cdots + (y_{n-1} - y_{n-2})^2 + (x_n - y_{n-1})^2 \right]$$

Obviously gives $x_n(n) = y_{n-1}$. $(\xi_0, \ldots, \xi_{n-1}) \rightarrow (\zeta_0, \ldots, \zeta_{n-1})$ – change of variable.
PROBABILISTIC (BAYESIAN) APPROACH – counting statistics

Bayes’

\[ f(x_0|y_1) \propto f(y_1|x_0) \propto e^{-\frac{(y_1-x_0)^2}{2\sigma_1^2}} \]

\[ f(x_0|y_1, y_2) \propto f(y_2|x_0, y_1) f(x_0|y_1) \]
\[ f(y_2|x_0, y_1) = f(y_2|x_0) \]
\[ f(x_0|y_1) \propto f(y_2|x_0) f(y_1|x_0) \]

Similarly \( f(x_0|y_1, y_2, \ldots, y_n) \propto f(y_n|x_0) \cdots f(y_2|x_0) f(y_1|x_0) \)

\[ \propto e^{-\frac{(y_n-x_0)^2}{2\sigma_n^2}} \cdots e^{-\frac{(y_2-x_0)^2}{2\sigma_2^2}} e^{-\frac{(y_1-x_0)^2}{2\sigma_1^2}} \]

Likelihood \( \chi^2 = -\ln f \propto \sum_{k=1}^{n} \frac{(y_k-x_0)^2}{2\sigma_k^2} \ldots \text{SAME} \)
RANDOM WALK

\[ f(x_0 | y_1) \propto f(y_1 | x_0) \propto e^{-\frac{(y_1 - x_0)^2}{2\sigma_1^2}} \]

\[ f(x_0 | y_1, y_2) \propto f(y_2 | x_0, y_1) f(x_0 | y_1) \]

\[ \propto f(y_2 | y_1) f(x_0 | y_1) \ (Markov) \propto f(y_2 | y_1) f(y_1 | x_0) \]

\[ f(x_0 | y_1, y_2, \ldots, y_n) \propto f(y_n | y_{n-1}) \cdots f(y_2 | y_1) f(x_0 | y_1) \]

\[ f \propto e^{-\frac{(y_n - y_{n-1})^2}{2\sigma_{n-1}^2}} \cdots e^{-\frac{(y_2 - y_1)^2}{2\sigma_1^2}} e^{-\frac{(y_1 - x_0)^2}{2\sigma_0^2}} \]

\[ \chi^2 = -\ln f \propto \sum_{k=1}^{n-1} \frac{(y_{k+1} - y_k)^2}{2\sigma_k^2} + \frac{(y_1 - x_0)^2}{2\sigma_0^2} \]
Random walk with measurement noise added – alternate approach

\[ x_{k+1} = x_k + \xi_k, \]

\[ y_k = x_k + \eta_k. \]

Then \( y_{k+1} - y_k = \xi_k + \eta_{k+1} - \eta_k \) and

\[ \chi^2 = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} (y_{k+1} - y_k) D_{kl} (y_{l+1} - y_l) \quad y_n \rightarrow x_n(n) \]

with \( C_{kl} = \langle (\xi_k + \eta_{k+1} - \eta_k) (\xi_l + \eta_{l+1} - \eta_l) \rangle > \text{tridiagonal} \)

\[
\sigma^2 \eta
= \begin{bmatrix}
2 & -1 & 0 & \cdots & 0 \\
-1 & 2 & -1 & 0 & \cdots \\
0 & -1 & 2 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\cdots & 0 & 0 & -1 & 2
\end{bmatrix}
\]

\[ + \sigma_{\xi^2} \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix} \]
Minimize with respect to $x_n(n)$:

$$(x_n(n) - y_{n-1}) D_{n-1,n-1} + \sum_{k=1}^{n-2} D_{n-1,k} (y_{k+1} - y_k) = 0$$

Limit 1: no measurement noise $\sigma^2_\eta = 0$ ...  $C = \sigma^2_\xi I$

or $D = \sigma^{-2}_\xi I$ .... $x_n(n) = y_{n-1}$

Limit 2: no dynamical noise $\sigma^2_\eta = 0$ ...

$$C = \sigma^2_\eta \begin{bmatrix}
  2 & -1 & 0 & \cdots & 0 \\
  -1 & 2 & -1 & 0 & \cdots \\
  0 & -1 & 2 & 1 & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \ddots \\
  \cdots & 0 & 0 & -1 & 2
\end{bmatrix}$$

gives sample mean

$$x_n(n) = \frac{1}{n-1} \sum_{k=1}^{n-1} y_k$$
Recall one dimensional system with measurement noise

Recall 1D system with measurement noise \( \langle \eta_k \eta_l \rangle = \delta_{kl} \)

\[
x_{k+1} = \gamma x_k,
\]

\[
y_k = x_k + \eta_k.
\]

\[
x_0(n+1) = x_0(n) + \frac{\gamma^{n+1}}{\sum_{k=1}^{n} \gamma^{2k} + \gamma^{2n+2}} (y_{n+1} - \gamma^{n+1} x_0(n)).
\]

\[
x_{n+1}(n+1) = \gamma x_n(n) + \frac{1}{\sum_{k=1}^{n} \gamma^{2k-2n-2} + 1} (y_{n+1} - \gamma x_n(n)).
\]

Estimates \( x_0(n) \) and \( x_n(n) = \gamma^n x_0(n) \) ... \( K_n x_n(n) = \gamma^{n+1} K^n x_0(n) \).

Also, \( V(x_0) = \frac{1}{\sum_{k=1}^{n} \gamma^{2k}} \) \( V(x_n) = \frac{\gamma^{2n}}{\sum_{k=1}^{n} \gamma^{2k}} \) = \( \gamma^{2n} V(x_0(n)) \) ... Recursive
Higher dimensional system with measurement noise - est. for $x_0(n)$

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

(5)

with measurement $\langle \eta^i_k, \eta^j_k \rangle = \delta_{ij}\delta_{kl}$

$$y_k = M_k x_k + \bar{\eta}_k.$$  

(6)

$$x_k = U_{k,0} x_0 = A_{k-1} A_{k-2} \ldots A_0 x_0$$

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n} \| \bar{\eta}_k \|^2 = \frac{1}{2} \sum_{k=1}^{n} \| M_k U_{k,0} x_0 - y_k \|^2,$$

(7)

$$x_0(n) = \left[ \sum_{k=1}^{n} N_{k,0}^T N_{k,0} \right]^{-1} \sum_{k=1}^{n} N_{k,0}^T y_k$$

$$N_{k,0} \equiv M_k U_{k,0}.$$  

(8)
\[ x_0(n+1) = x_0(n) + K_n [y_{n+1} - M_{n+1} U_{n+1,0} x_0(n)] \] (9)

\[ P^{-1}_{n+1} = P^{-1}_n + N^T_{n+1,0} N_{n+1,0} \]

\[ K_n = P_{n+1} N^T_{n+1,0} \quad C(x_0(n)) = P_n \]

\( x_0(n) \) propagated \( n \rightarrow n + 1 \) by \( U_{n+1,0} \)

Measurement applied \( M_{n+1} U_{n+1,0} x_0(n) \) is best guess for \( y_{n+1} \) before measurement
Higher dimensional system with measurement noise - est. for $x_n(n)$

$$
\chi^2 = \frac{1}{2} \sum_{k=1}^{n} \eta_k^2 = \frac{1}{2} \sum_{k=1}^{n} \|M_k U_{k,n} x_n - y_k\|^2
$$

$$
x_n(n) = \left[ \sum_{k=1}^{n} U_{0,n}^T N_{k,n}^T N_{k,n} U_{0,n} \right]^{-1} \sum_{k=1}^{n} U_{0,n}^T N_{k,n}^T y_k,
$$

or

$$
x_n(n) = U_{n,0} x_0(n) \quad \tilde{K}_n = U_{n+1,0} K_n
$$

$$
x_{n+1}(n+1) = A_n x_n(n) + \tilde{K}_n [y_{n+1} - M_{n+1} A_n x_n(n)]
$$

$x_n(n)$ is advanced in time $x_n(n) \rightarrow A_n x_n(n)$ and the measurement operation $M_{n+1}$ is done. This is the best guess for $y_{n+1}$ before $y_{n+1}$ is measured.
Continuous time advance, discrete time measurement formulation

\[
\begin{align*}
\frac{d\mathbf{x}}{dt} &= A(t)\mathbf{x} + \mathbf{\xi}(t) \quad <\xi_i \xi_j> = C_0 \\
\mathbf{y}_k &= M\mathbf{x}_k + \mathbf{\eta}_k \quad <\eta_i \eta_j> = C_1
\end{align*}
\]

1. Time advance of estimate and covariance between measurements

\[
\frac{d\hat{\mathbf{x}}}{dt} = A(t)\hat{\mathbf{x}} \quad \frac{dC}{dt} = AC + CA^T + C_0
\]

2. Adjust estimate and covariance at new measurement

\[
K_k = C^{(-)}(t_k)M^T [MC^{-(t_k)}M^T + C_1]^{-1} \\
C(t_k) = [I - K_k M] C^{(-)}(t_k) \\
\hat{\mathbf{x}}_k = \hat{\mathbf{x}}^{(-)}_k + K_k \left( \mathbf{y}_k - M\hat{\mathbf{x}}^{(-)}_k \right)
\]

\(\hat{\mathbf{x}}^{(-)}_k\) is the best guess for \(\mathbf{y}_k\) at \(t_k\) before its measurement; \(C^{(-)}(t_k)\) is the covariance matrix at \(t_k\) before measurement of \(\mathbf{y}_k\).
Application to control theory – separation theorem

\[ x_{k+1} = A_k x_k + \vec{\xi}_k + u_k \]  \hspace{1cm} (12)

\[ y_k = M_k x_k + \vec{\eta}_k. \]  \hspace{1cm} (13)

Continuum model...

\[ \frac{dx}{dt} = A(t)x + \vec{\xi}(t) + u(t) \]

\[ y(t) = x(t) = M(t)x(t) + \vec{\eta}(t) \text{ special case} \]

Optimal control \( \Rightarrow \) minimizing for example

\[ J = \int_0^T \{ (x(t), Q(t)x(t)) + (u(t), R(t)u(t)) \} \, dt \]
Minimizing $J$ determines $u[x]$ optimally for $\dot{\xi}(t) = 0$. Degree of control vs. cost.

For $\dot{\xi}(t) \neq 0$ do the following:

- Find optimal control $u[x(t), t]$ for $\dot{\xi}(t) = 0$

- Use Kalman filter on $y(t)$ to determine the optimal estimate $\hat{x}(t)$

- Add control $u(\hat{x}(t), t)$ based on estimate to equation $dx/dt = ...$

  ❌ Allows one to design controller and estimator independently

  ❌ More general form with measurement noise exists too

  ❌ A similar formulation exists for the discrete system
Extended Kalman Filter – for nonlinear systems

Most real problems (systems and measurements) are nonlinear

\[
\frac{dx}{dt} = a(x, t) + \xi(t)
\]

\[
y_k = h(x_k) + \eta(t)
\]

- Advance the estimate between measurements by the nonlinear dynamics

\[
d\hat{x}/dt = a(\hat{x}, t)
\]

- Advance the covariance between measurements by

\[
dC/dt = A(\hat{x}, t)C + CA^T(\hat{x}, t) + C_0
\]

with \( A_{ij} = \partial a_i/\partial x_j \) LINEARIZE with respect to \( x \)
• Kalman gain

\[
K_k = C(-)(t_k)M^T(\hat{x}_k^{-})
\times \left[ M(\hat{x}_k^{-})C(-)(t_k)M^T(\hat{x}_k^{-}) + C_1 \right]^{-1}
\]

where \( M_{i,j} = \partial h_i / \partial x_j \ldots \) \( \text{LINEARIZE} \) with respect to \( \mathbf{x} \). Covariance similarly

• Update estimate after new data: \( \hat{x}_k = \hat{x}_k^{-} + K_k \left( y_k - h(\hat{x}_k^{-}) \right) \)

• Caveat: \( d\hat{x}/dt = a(\mathbf{x}, t) = a(\hat{x}, t) + (\mathbf{x} - \hat{x}) \cdot \nabla a(\hat{x}, t) + \ldots \)

• Caveat: \textit{gaussian statistics} remains gaussian only if \( C \) remains small –if \textit{linearizations} hold over the range specified by \( C \)

• Caveat: what if the model \( [\text{i.e. } a(\mathbf{x}, t)] \) is known poorly? Model errors
SUMMARY

- **Least squares approach**

- **Recursive** least squares. Kalman gain ↔ covariance matrix; 'innovation'

- Minimum variance - minimum trace of the covariance matrix

- Estimating the **initial state or the current state**

  Only measurement noise – initial and current state estimates are related by the dynamics

  Only dynamical noise – initial and current state estimates are dominated by nearby data

- Bayesian approach and **maximum likelihood → least squares**

- Higher dimension – principles the same (recursion for estimate and covariance matrix; relation with Kalman gain)

- Control theory and the **separation theorem**
• Extended Kalman Filter – advance estimate nonlinearly, covariance matrix by linearized system. Caveats:

1) small covariance for linearization to be accurate ... otherwise not gaussian

2) systematic errors – model errors