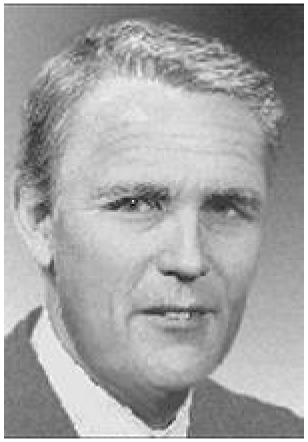
Least Squares Estimation and Kalman Filtering

Motivation: Kalman Filtering



Rudolf Kalman (Ph.D. Columbia 1957): “Published two highly influential mathematics papers – that helped man set foot on the moon in the late 1960s

ASME Journal of Basic Engineering

- New Approach to Linear Filtering and Prediction Problems
- New Results in Linear Filtering and Prediction Theory

Kalman filtering addresses an age-old question:

How do you get accurate information out of inaccurate data?

How do you update a "best" estimate for the state of a system as new, but still inaccurate, data pour in?

The answer lies in understanding least squares estimation

Motivational Example

Problem: Determine the normal acceleration (i.e. gravitational constant). Suppose a steel ball was dropped without initial velocity. The ball's position l was recorded at different times as follows:

Time [s]	Length of fall [m]
1	8.49
2	20.05
3	50.65
4	72.19
5	129.85
6	171.56

Solution: One approach: Calculate acceleration $g = \frac{2l}{t^2}$ and take average

$$g_1 = \frac{2(8.49)}{1^2} = 16.98$$

$$g_2 = \frac{2(20.05)}{2^2} = 10.025$$

$$\vdots$$

$$g_6 = \frac{2(171.56)}{6^2} = 9.53$$

Time [s]	Length of fall [m]	Calculated g $\left[m/s^2\right]$
1	8.49	16.980
2	20.05	10.025
3	50.65	11.256
4	72.19	9.0238
5	129.85	10.388
6	171.56	9.531

Taking average yields: $\bar{g} = \dfrac{16.980 + 10.025 + \cdots 9.531}{6} = 11.201 \left[\mathrm{m/s^2}\right]$

Averaging is not promising

Need another approach – recast as a least squares problem

$$l = \frac{gt^2}{2} + e$$

State model

$$J \triangleq \frac{1}{2}\sum_{i=1}^{N} e_i^2$$

Cost function to minimize

$$e_i = y_i - \hat{y}_i$$

Actual measurement

Predicted measurement

error

Solution: Least squares approach

If say: $e \square y_i - \hat{y}_i$ then $e_i = l_i - g\dfrac{t_i^2}{2}$ where g is what we want

$$
\begin{aligned}
e_1 &= l_1 - g\frac{t_1^2}{2} \\
e_2 &= l_2 - g\frac{t_2^2}{2} \\
&\;\vdots \\
\underbrace{e_6}_{\vec{e}} &= \underbrace{l_6}_{\vec{y}} - \underbrace{g\frac{t_6^2}{2}}_{\vec{\hat{y}} \,\square\, \Phi\Theta}
\end{aligned}
\qquad \text{where} \qquad
\underbrace{\begin{bmatrix} \frac{t_1^2}{2} \\ \frac{t_2^2}{2} \\ \vdots \\ \frac{t_6^2}{2} \end{bmatrix}}_{\Phi} \underbrace{g}_{\Theta}
$$

Thus if say $\vec{\hat{y}} \,\square\, \Phi\Theta$ can show that $\Theta = \left(\Phi^T\Phi\right)^{-1}\Phi^T\vec{y}$ (1)

Said that $\underbrace{\begin{bmatrix} \frac{t_1^2}{2} \\ \frac{t_2^2}{2} \\ \vdots \\ \frac{t_6^2}{2} \end{bmatrix}}_{\Phi}$ hence $\Phi^T \Phi = \begin{bmatrix} 0.5 & 2 & 4.5 & 8 & 12.5 & 18 \end{bmatrix} \begin{bmatrix} 0.5 \\ 2 \\ 4.5 \\ 8 \\ 12.5 \\ 18 \end{bmatrix} = 568.75$

From (1) have $\Theta = (568.75)^{-1} \begin{bmatrix} 0.5 & 2 & 4.5 & 8 & 12.5 & 18 \end{bmatrix} \begin{bmatrix} 8.49 \\ 20.05 \\ 50.65 \\ 72.19 \\ 129.85 \\ 171.56 \end{bmatrix}$

$\vec{y}$

Results in $\Theta = \frac{5.561 \times 10^3}{568.75} = 9.778\ m/s^2$

Least squares estimate of gravitational constant is quite good

Least squares looks very good – but consider the following:

Q1. What if all data is not immediately available?

Q2. What if the sensor is noisy?

Q3. What if the state model is questionable?

Answer 1: Use recursion to update your answers N = number of measurements

$$\hat{\theta}(N+1) = \hat{\theta}(N) + K(N)\left\{y_{N+1} - \phi^T(N+1)\hat{\theta}(N)\right\}$$

New = Current + Gain x (value – model predicted value)

$$K(N) = P(N)\phi(N+1)\left\{1+\phi^T(N+1)P(N)\phi(N+1)\right\}^{-1}$$

Kickoff value (Best Guess)

$$P(N+1) = \left\{I - K(N)\phi^T(N+1)\right\}P(N)$$

Demonstration: Suppose $P(0) \approx 1$ and $\hat{\theta}(0) = y_1 \approx 8.49\ m/s^2$

Iteration $N = 0$

$$K(0) = P(0)\phi(1)\left\{1+\phi^T(1)P(0)\phi(1)\right\}^{-1} = (1)(0.5)\left\{1+(0.5)(1)(0.5)\right\}^{-1} = 0.4$$

$$K(N) = P(N)\phi(N+1)\left\{1+\phi^T(N+1)P(N)\phi(N+1)\right\}^{-1}$$
$$\hat{\theta}(N+1) = \hat{\theta}(N) + K(N)\left\{y_{N+1} - \phi^T(N+1)\hat{\theta}(N)\right\}$$
$$P(N+1) = \left\{I - K(N)\phi^T(N+1)\right\}P(N)$$

Recall
$$\underbrace{\begin{bmatrix} \frac{t_1^2}{2} \\ \frac{t_2^2}{2} \\ \vdots \\ \frac{t_6^2}{2} \end{bmatrix}}_{\Phi} = \begin{array}{c|c} \begin{bmatrix} 0.5 \\ 2 \\ 4.5 \\ 8 \\ 12.5 \\ 18 \end{bmatrix} & \begin{array}{l} \leftarrow \phi(1) \\ \leftarrow \phi(2) \\ \leftarrow \phi(3) \\ \leftarrow \phi(4) \\ \leftarrow \phi(5) \\ \leftarrow \phi(6) \end{array} \end{array}$$

Demonstration:

Suppose $\underbrace{P(0) \square 1}_{\text{Best Guess}}$ and $\underbrace{\hat{\theta}(0) = y_1 \square 8.49\ m/s^2}_{\text{First data point}}$

Iteration $N = 0$ $\quad K(0) = P(0)\phi(1)\left\{1+\phi^T(1)P(0)\phi(1)\right\}^{-1} = (1)(0.5)\left\{1+(0.5)(1)(0.5)\right\}^{-1} = 0.4$

$$\hat{\theta}(1) = \hat{\theta}(0) + K(0)\left\{y_1 - \phi^T(1)\hat{\theta}(0)\right\} = 8.49 + 0.4\left\{8.49 - 0.5(8.49)\right\} = 10.188\ m/s^2$$ **Gravity constant**

$$P(1) = \left\{1 - K(0)\phi^T(1)\right\}P(0) = \left\{1-(0.4)(0.5)\right\}(1) = 0.8$$ **Best Guess has been updated**

Iteration $N=1$

$$K(1)=P(1)\phi(2)\{1+\phi^T(2)P(1)\phi(2)\}^{-1} = (0.8)(2)\{1+(2)(0.8)(2)\}^{-1}=0.381$$

$$\hat{\theta}(2)=\hat{\theta}(1)+K(1)\{y_2-\phi^T(2)\hat{\theta}(1)\} = 10.188+0.381\{20.05-2(10.188)\}=10.064\ m/s^2$$

$$P(2)=\{1-K(1)\phi^T(2)\}P(1) = \{1-(0.381)(2)\}(0.8)=0.1904$$

Iteration $N=2$

$$K(2)=P(2)\phi(3)\{1+\phi^T(3)P(2)\phi(3)\}^{-1}=(0.1904)(4.5)\{1+(4.5)(0.1904)(4.5)\}^{-1}=0.1765$$

$$\hat{\theta}(3)=\hat{\theta}(2)+K(2)\{y_3-\phi^T(3)\hat{\theta}(2)\} = 10.064+0.1765\{50.65-4.5(10.064)\}=11.010\ m/s^2$$

$$P(3)=\{1-K(2)\phi^T(3)\}P(2) = \{1-(0.1765)(4.5)\}(0.1904)=0.03917$$

Iteration $N=3$

$$K(3)=P(3)\phi(4)\{1+\phi^T(4)P(3)\phi(4)\}^{-1} = (0.03917)(8)\{1+(8)(0.03917)(8)\}^{-1}=0.0893$$

$$\hat{\theta}(4)=\hat{\theta}(3)+K(3)\{y_4-\phi^T(4)\hat{\theta}(3)\} = 11.010+0.0893\{72.19-8(11.010)\}=9.591\ m/s^2$$

$$P(4)=\{1-K(3)\phi^T(4)\}P(3) = \{1-(0.0893)(8)\}(0.03917)=0.0112$$

Note: gravity constant is updated as new data comes in

Iteration $N = 4$

$$K(4) = P(4)\phi(5)\left\{1+\phi^T(5)P(4)\phi(5)\right\}^{-1} = (0.0112)(12.5)\left\{1+(12.5)(0.0112)(12.5)\right\}^{-1} = 0.05095$$

$$\hat{\theta}(5) = \hat{\theta}(4) + K(4)\left\{y_5 - \phi^T(5)\hat{\theta}(4)\right\} = 9.591 + 0.05095\left\{129.85 - 12.5(9.591)\right\} = 10.0986\ m/s^2$$

$$P(5) = \left\{1 - K(4)\phi^T(5)\right\}P(4) = \left\{1-(0.05095)(12.5)\right\}(0.0112) = 4.067\times10^{-3}$$

Iteration $N = 5$

$$K(5) = P(5)\phi(6)\left\{1+\phi^T(6)P(5)\phi(6)\right\}^{-1} = \left(4.067\times10^{-3}\right)(18)\left\{1+(18)\left(4.067\times10^{-3}\right)(18)\right\}^{-1} = 0.0316$$

$$\hat{\theta}(6) = \hat{\theta}(5) + K(5)\left\{y_6 - \phi^T(6)\hat{\theta}(5)\right\} = 10.0986 + 0.0316\left\{171.56 - 18(10.0986)\right\} = 9.776\ m/s^2$$

$$P(6) = \left\{1 - K(5)\phi^T(6)\right\}P(5) = \left\{1-(0.0316)(18)\right\}\left(4.067\times10^{-3}\right) = 1.754\times10^{-3}$$

Gravity constant

Note: Value is almost the same as previously calculated in Slide 6

Take Home Message:

1. No such thing as “bad” data
2. Recursive method can be easily implemented as an algorithm
3. Rate to converge to “good” answer depends on “best guess” I.e. start up costs

Towards Kalman Filtering…

Least squares is a "special" case of Kalman Filtering

Recall that least squares says: $J \triangleq \frac{1}{2}\sum_{i=1}^{N} e_i^2$ Cost function to minimize where $e_i = y_i - \hat{y}_i = y_i - \Phi\Theta$

The general case: $J \triangleq \frac{1}{2}\sum_{i=1}^{N}\left(\vec{y} - \Phi\vec{\Theta}\right)^T W\left(\vec{y} - \Phi\vec{\Theta}\right)$

Weighting matrix. $W = I$ **yields least-squares form**

W values depend on:

- Accuracy of sensors quantified by a error co-variance matrix
- Accuracy of state model e.g. $\phi = t^2/2$ quantified by a state co-variance matrix

Kalman Filter: calculates the desired value <u>optimally</u> given Gaussian noise

Recommended Reading: See MEM 640 Web Page and
G.C. Dean, "An Introduction to Kalman Filters", *Measurement and Control*, Vol. 18, pp. 69-73,March 1986.