Kalman Filter Theory and Applications Equation Drilldown

https://github.com/musicarroll/kalman_course

Michael L. Carroll

June 17, 2023

©2023 by Michael L. Carroll



Part I

The Five Basic Kalman Equations Topics

Part I The Five Basic Kalman Equations Topics

• Understanding the Equations: Heuristic Introduction

Part I The Five Basic Kalman Equations Topics

- Understanding the Equations: Heuristic Introduction
- Equation Drilldown: Taking the Equations Apart

- Understanding the Equations: Heuristic Introduction
- Equation Drilldown: Taking the Equations Apart
- State Space Concepts

Part | The Five Basic Kalman Equations Topics

- Understanding the Equations: Heuristic Introduction
- Equation Drilldown: Taking the Equations Apart
- State Space Concepts

Mathematical Formulation of the Problem

- Mathematical Formulation of the Problem
- Drilldown on State Dynamics and Covariance Extrapolation Equations

- Mathematical Formulation of the Problem
- Drilldown on State Dynamics and Covariance Extrapolation Equations
- The Five Kalman Filter Equations

- Mathematical Formulation of the Problem
- Drilldown on State Dynamics and Covariance Extrapolation Equations
- The Five Kalman Filter Equations
- Examples

- Mathematical Formulation of the Problem
- Drilldown on State Dynamics and Covariance Extrapolation Equations
- The Five Kalman Filter Equations
- Examples
- Exercises

- Mathematical Formulation of the Problem
- Drilldown on State Dynamics and Covariance Extrapolation Equations
- The Five Kalman Filter Equations
- Examples
- Exercises

The Kalman Gain

• The heart of the solution is the **Kalman Gain**: K(k)

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

where R(k) is the **measurement noise covariance matrix** (to be defined later)

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

where R(k) is the **measurement noise covariance matrix** (to be defined later)

• R(k) governs the white measurement noise v(k) in the measurement model:

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

where R(k) is the **measurement noise covariance matrix** (to be defined later)

• R(k) governs the white measurement noise v(k) in the measurement model:

$$z(k) = H(k)x(k) + v(k)$$

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

where R(k) is the **measurement noise covariance matrix** (to be defined later)

• R(k) governs the white measurement noise v(k) in the measurement model:

$$z(k) = H(k)x(k) + v(k)$$

• But where did the Kalman gain Eq.(1) come from?

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

where R(k) is the **measurement noise covariance matrix** (to be defined later)

• R(k) governs the white measurement noise v(k) in the measurement model:

$$z(k) = H(k)x(k) + v(k)$$

• But where did the Kalman gain Eq.(1) come from?

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

The Kalman Gain

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

• We are not going to formally derive Eq.(1) right now.

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- We are not going to formally derive Eq.(1) right now.
- But we will take it apart and consider why it makes sense

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- We are not going to formally derive Eq.(1) right now.
- But we will take it apart and consider why it makes sense
- First consider that the expression on the right side has the form: AB^{-1}

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- We are not going to formally derive Eq.(1) right now.
- But we will take it apart and consider why it makes sense
- First consider that the expression on the right side has the form: AB^{-1}
- This is analogous to a scalar fraction: $\frac{a}{b}$, i.e., a ratio

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- We are not going to formally derive Eq.(1) right now.
- But we will take it apart and consider why it makes sense
- First consider that the expression on the right side has the form: AB^{-1}
- This is analogous to a scalar fraction: $\frac{a}{b}$, i.e., a ratio
- Now ignore the H matrix since it only serves to map the state space-related matrix P into measurement space; think of it as a scaling or projection factor

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- We are not going to formally derive Eq.(1) right now.
- But we will take it apart and consider why it makes sense
- First consider that the expression on the right side has the form: AB^{-1}
- This is analogous to a scalar fraction: $\frac{a}{b}$, i.e., a ratio
- Now ignore the H matrix since it only serves to map the state space-related matrix P into measurement space; think of it as a scaling or projection factor

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

¹Technically, covariance matrices only need to be positive semidefinite.

The Kalman Gain

$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

• Note also that the matrices P and R are positive definite¹

¹Technically, covariance matrices only need to be positive_semidefinite.



$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

- Note also that the matrices P and R are positive definite¹
- So, we are essentially considering the ratio of P to P+R

¹Technically, covariance matrices only need to be positive **semi**definite.



$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

- Note also that the matrices P and R are positive definite¹
- So, we are essentially considering the ratio of P to P + R
- But the P here is the extrapolated P that incorporates the previous P and Q (also positive definite)

¹Technically, covariance matrices only need to be positive semidefinite.



$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

- Note also that the matrices P and R are positive definite¹
- So, we are essentially considering the ratio of P to P + R
- But the P here is the extrapolated P that incorporates the previous P and Q (also positive definite)
- So, we effectively looking at a ratio of P + Q to P + Q + R

¹Technically, covariance matrices only need to be positive_semidefinite.



$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$
 (1)

- Note also that the matrices P and R are positive definite¹
- So, we are essentially considering the ratio of P to P + R
- But the P here is the extrapolated P that incorporates the previous P and Q (also positive definite)
- So, we effectively looking at a ratio of P + Q to P + Q + R
- If R were small (a near perfect sensor), then K would be near unity, and we would effectively ignore the prior state estimate and just trust the measurement; which makes sense, given a perfect sensor

¹Technically, covariance matrices only need to be positive **semi**definite.



$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- Note also that the matrices P and R are positive definite¹
- So, we are essentially considering the ratio of P to P + R
- But the P here is the extrapolated P that incorporates the previous P and Q (also positive definite)
- So, we effectively looking at a ratio of P + Q to P + Q + R
- If R were small (a near perfect sensor), then K would be near unity, and we would effectively ignore the prior state estimate and just trust the measurement; which makes sense, given a perfect sensor
- At the other extreme, if R were infinitely huge (a non-functional sensor), then K would be near 0 and we would effectively ignore the measurement in favor of the previously predicted estimate

¹Technically, covariance matrices only need to be positive semidefinite.



$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
 (1)

- Note also that the matrices P and R are positive definite¹
- So, we are essentially considering the ratio of P to P + R
- But the P here is the extrapolated P that incorporates the previous P and Q (also positive definite)
- So, we effectively looking at a ratio of P + Q to P + Q + R
- If R were small (a near perfect sensor), then K would be near unity, and we would effectively ignore the prior state estimate and just trust the measurement; which makes sense, given a perfect sensor
- At the other extreme, if R were infinitely huge (a non-functional sensor), then K would be near 0 and we would effectively ignore the measurement in favor of the previously predicted estimate

¹Technically, covariance matrices only need to be positive semidefinite.



State Update Equation

The Kalman Filter Solution State Update Equation

• Using the gain Eq. (1)

State Update Equation

• Using the gain Eq. (1)

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1},$$

• Using the gain Eq. (1)

 $K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$, the state update equation (which we've already seen) is

$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) \left[z(k) - H(k)\hat{x}^{-}(k) \right]$$
 (2)

where z(k) is the measurement and H(k) is the measurement matrix at time k.

• Using the gain Eq. (1)

$$K(k) = P^{-}(k) H(k)^{\top} [H(k)P^{-}(k) H(k)^{\top} + R(k)]^{-1}$$
, the state update equation (which we've already seen) is

$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) \left[z(k) - H(k)\hat{x}^{-}(k) \right]$$
 (2)

where z(k) is the measurement and H(k) is the measurement matrix at time k.

The Covariance Update Equation

The Covariance Update Equation

• Likewise, the error covariance P(k) is corrected using the gain:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$
(3)

The Covariance Update Equation

• Likewise, the error covariance P(k) is corrected using the gain:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$
(3)

• This is not a good form to use in numerical work!

The Covariance Update Equation

• Likewise, the error covariance P(k) is corrected using the gain:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$
(3)

- This is not a good form to use in numerical work!
- Better to use the Joseph form:

The Covariance Update Equation

• Likewise, the error covariance P(k) is corrected using the gain:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$
(3)

- This is not a good form to use in numerical work!
- Better to use the Joseph form:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)[I - K(k)H(k)]^{\top} + K(k)R(k)K(k)^{\top}$$

The Covariance Update Equation

• Likewise, the error covariance P(k) is corrected using the gain:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$
(3)

- This is not a good form to use in numerical work!
- Better to use the Joseph form:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)[I - K(k)H(k)]^{\top} + K(k)R(k)K(k)^{\top}$$

which better preserves symmetry of P
 (covariance matrices are always symmetric)

The Covariance Update Equation

• Likewise, the error covariance P(k) is corrected using the gain:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$
(3)

- This is not a good form to use in numerical work!
- Better to use the Joseph form:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)[I - K(k)H(k)]^{\top} + K(k)R(k)K(k)^{\top}$$

which better preserves symmetry of P
 (covariance matrices are always symmetric)



Taking Apart the Covariance Update Equation

• Again, ignore *H* in Eq.(3): $P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$

- Again, ignore *H* in Eq.(3): $P^{+}(k) = [I K(k)H(k)]P^{-}(k)$
- Thus, I K indicates, in a sense, how far we are from having no sensors at all

- Again, ignore *H* in Eq.(3): $P^{+}(k) = [I K(k)H(k)]P^{-}(k)$
- Thus, I K indicates, in a sense, how far we are from having no sensors at all
- If K = 0, (i.e., no sensors at all) there would be no reduction in uncertainty

- Again, ignore *H* in Eq.(3): $P^+(k) = [I K(k)H(k)]P^-(k)$
- Thus, I K indicates, in a sense, how far we are from having no sensors at all
- If K = 0, (i.e., no sensors at all) there would be no reduction in uncertainty
- If K = I, we would have perfect knowledge (thanks to perfect sensors), i.e., P = 0

- Again, ignore *H* in Eq.(3): $P^+(k) = [I K(k)H(k)]P^-(k)$
- Thus, I K indicates, in a sense, how far we are from having no sensors at all
- If K = 0, (i.e., no sensors at all) there would be no reduction in uncertainty
- If K = I, we would have perfect knowledge (thanks to perfect sensors), i.e., P = 0
- Thus, if 0 < K < I, this use of I − K will tend to reduce the uncertainty P

- Again, ignore *H* in Eq.(3): $P^+(k) = [I K(k)H(k)]P^-(k)$
- Thus, I K indicates, in a sense, how far we are from having no sensors at all
- If K = 0, (i.e., no sensors at all) there would be no reduction in uncertainty
- If K = I, we would have perfect knowledge (thanks to perfect sensors), i.e., P = 0
- Thus, if 0 < K < I, this use of I − K will tend to reduce the uncertainty P
- Later we will see how this form for K and its use in the update equations is in fact mandated by the requirement of optimal use of Φ , P, Q, R

- Again, ignore *H* in Eq.(3): $P^+(k) = [I K(k)H(k)]P^-(k)$
- Thus, I K indicates, in a sense, how far we are from having no sensors at all
- If K = 0, (i.e., no sensors at all) there would be no reduction in uncertainty
- If K = I, we would have perfect knowledge (thanks to perfect sensors), i.e., P = 0
- Thus, if 0 < K < I, this use of I − K will tend to reduce the uncertainty P
- Later we will see how this form for K and its use in the update equations is in fact mandated by the requirement of optimal use of Φ , P, Q, R



State Extrapolation: $\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$

State Extrapolation: $\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$

Covariance Extrapolation: $P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$

State Extrapolation: $\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$

Covariance Extrapolation:
$$P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$$

Kalman Gain:
$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$

State Extrapolation: $\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$

Covariance Extrapolation:
$$P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$$

Kalman Gain:
$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$

State Update:
$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) [z(k) - H(k)\hat{x}^{-}(k)]$$

State Extrapolation:
$$\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$$

Covariance Extrapolation:
$$P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$$

Kalman Gain:
$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$

State Update:
$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) [z(k) - H(k)\hat{x}^{-}(k)]$$

Covariance Update:
$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)$$

State Extrapolation:
$$\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$$

Covariance Extrapolation:
$$P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$$

Kalman Gain:
$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$

State Update:
$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) [z(k) - H(k)\hat{x}^{-}(k)]$$

Covariance Update: $P^+(k) = [I - K(k)H(k)]P^-(k)$ or the Joseph form for the Covariance Update:



State Extrapolation: $\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$

Covariance Extrapolation:
$$P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$$

Kalman Gain:
$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$

State Update:
$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) [z(k) - H(k)\hat{x}^{-}(k)]$$

Covariance Update: $P^+(k) = [I - K(k)H(k)]P^-(k)$ or the Joseph form for the Covariance Update:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)[I - K(k)H(k)]^{\top} + K(k)R(k)K(k)^{\top}$$



State Extrapolation: $\hat{x}^-(k) = \Phi(k)\hat{x}^+(k-1)$

Covariance Extrapolation:
$$P^{-}(k) = \Phi(k)P^{+}(k-1)\Phi(k)^{\top} + Q(k)$$

Kalman Gain:
$$K(k) = P^{-}(k) H(k)^{\top} \left[H(k) P^{-}(k) H(k)^{\top} + R(k) \right]^{-1}$$

State Update:
$$\hat{x}^{+}(k) = \hat{x}^{-}(k) + K(k) [z(k) - H(k)\hat{x}^{-}(k)]$$

Covariance Update: $P^+(k) = [I - K(k)H(k)]P^-(k)$ or the Joseph form for the Covariance Update:

$$P^{+}(k) = [I - K(k)H(k)]P^{-}(k)[I - K(k)H(k)]^{\top} + K(k)R(k)K(k)^{\top}$$



• $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)

- $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)
 - ullet where E[...] is statistical expectation (more on that later) and

- $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)
 - where E[...] is statistical expectation (more on that later) and
 - $\tilde{x}^+(k) = \hat{x}^+(k) x(k)$ is the estimation error

- $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)
 - where E[...] is statistical expectation (more on that later) and
 - $\tilde{x}^+(k) = \hat{x}^+(k) x(k)$ is the estimation error
 - P is often just referred to as the 'covariance matrix'

- $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)
 - ullet where E[...] is statistical expectation (more on that later) and
 - $\tilde{x}^+(k) = \hat{x}^+(k) x(k)$ is the estimation error
 - P is often just referred to as the 'covariance matrix'
- $Q(k) := E[w(k)w(k)^{\top}]$ is the process noise covariance matrix

- $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)
 - where E[...] is statistical expectation (more on that later) and
 - $\tilde{x}^+(k) = \hat{x}^+(k) x(k)$ is the estimation error
 - P is often just referred to as the 'covariance matrix'
- $Q(k) := E[w(k)w(k)^{\top}]$ is the process noise covariance matrix
- $R(k) := E[v(k)v(k)^{\top}]$ is the measurement noise covariance matrix

- $P^+(k) = E\left[\tilde{x}^+(k)\left[\tilde{x}^+(k)\right]^\top\right]$ is the estimation error covariance matrix (formal definition)
 - where E[...] is statistical expectation (more on that later) and
 - $\tilde{x}^+(k) = \hat{x}^+(k) x(k)$ is the estimation error
 - P is often just referred to as the 'covariance matrix'
- $Q(k) := E[w(k)w(k)^{\top}]$ is the process noise covariance matrix
- $R(k) := E[v(k)v(k)^{\top}]$ is the measurement noise covariance matrix



Summary

Vector Check

Summary

Vector Check

• Where are we?

Summary Vector Check

• Where are we?

• We've drilled down on the Kalman gain equation as a function of P, Q, R

Summary

Vector Check

- Where are we?
 - We've drilled down on the Kalman gain equation as a function of P, Q, R
 - Presented the Five Kalman equations

Summary

Vector Check

- Where are we?
 - We've drilled down on the Kalman gain equation as a function of P, Q, R
 - Presented the Five Kalman equations
- What's next?

- Where are we?
 - We've drilled down on the Kalman gain equation as a function of P, Q, R
 - Presented the Five Kalman equations
- What's next?
 - In the next video, we will consider 3 simple example applications with gradually increasing complexity

Summary Vector Check

- Where are we?
 - We've drilled down on the Kalman gain equation as a function of P, Q, R
 - Presented the Five Kalman equations
- What's next?
 - In the next video, we will consider 3 simple example applications with gradually increasing complexity