Basic Kalman Filter Theory

Technical Note

Document TBD
Version: Draft

Authors: Mark Pedley and Michael Stanley

Date: September 2014

Table of Contents

Glossary

$\boldsymbol{A}_k$	The linear prediction or state matrix at sample k. $\boldsymbol{x}_k = \boldsymbol{A}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k$ $\hat{\boldsymbol{x}}_k^- = \boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1}^+$
$\boldsymbol{C}_k$	The measurement matrix relating $\boldsymbol{x}_k$ to $\boldsymbol{z}_k$ at sample k. $\boldsymbol{z}_k = \boldsymbol{C}_k \boldsymbol{x}_k + \boldsymbol{v}_k$
$E[]$	Expectation operator
$\boldsymbol{K}_k$	The Kalman filter gain at sample k
$\boldsymbol{P}_k^-$	The *a priori* covariance matrix of the linear prediction (*a priori*) error $\hat{\boldsymbol{x}}_{\varepsilon,k}^-$ at sample k. $\boldsymbol{P}_k^- = E\left[\hat{\boldsymbol{x}}_{\varepsilon,k}^- \hat{\boldsymbol{x}}_{\varepsilon,k}^{-\ T}\right]$
$\boldsymbol{P}_k^+$	The *a posteriori* covariance matrix of the Kalman (*a posteriori*) error $\hat{\boldsymbol{x}}_{\varepsilon,k}^+$ at sample k. $\boldsymbol{P}_k^+ = E\left[\hat{\boldsymbol{x}}_{\varepsilon,k}^+ \hat{\boldsymbol{x}}_{\varepsilon,k}^{+\ T}\right]$
$\boldsymbol{Q}_{w,k}$	The covariance matrix of the additive noise $\boldsymbol{w}_k$ on the process $\boldsymbol{x}_k$ $\boldsymbol{Q}_{w,k} = E[\boldsymbol{w}_k \boldsymbol{w}_k^T]$
$\boldsymbol{Q}_{v,k}$	The covariance matrix of the additive noise $\boldsymbol{v}_k$ on the measured process $\boldsymbol{z}_k$ $\boldsymbol{Q}_{v,k} = E[\boldsymbol{v}_k \boldsymbol{v}_k^T]$
$V[]$	Variance operator
$\boldsymbol{v}_k$	The additive noise on the measured process $\boldsymbol{z}_k$ at sample k
$\boldsymbol{w}_k$	The additive noise on the process of interest $\boldsymbol{x}_k$ at sample k
$\boldsymbol{x}_k$	The state vector at time sample k of the process of interest $\boldsymbol{x}_k$ $\boldsymbol{x}_k = \boldsymbol{A}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k$
$\hat{\boldsymbol{x}}_k^-$	The linear prediction (*a priori*) estimate of the process $\boldsymbol{x}_k$ at sample k. $\hat{\boldsymbol{x}}_k^- = \boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1}^+$
$\hat{\boldsymbol{x}}_k^+$	The Kalman filter (*a posteriori*) estimate of the process $\boldsymbol{x}_k$ at sample k. $\hat{\boldsymbol{x}}_k^+ = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k)\hat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k \boldsymbol{z}_k = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k)\boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1}^+ + \boldsymbol{K}_k \boldsymbol{z}_k$
$\hat{\boldsymbol{x}}_{\varepsilon,k}^-$	The error in the linear prediction (a priori) estimate of the process $\boldsymbol{x}_k$. $\hat{\boldsymbol{x}}_{\varepsilon,k}^- = \hat{\boldsymbol{x}}_k^- - \boldsymbol{x}_k$
$\hat{\boldsymbol{x}}_{\varepsilon,k}^+$	The error in the *a posteriori* Kalman filter estimate of the process $\boldsymbol{x}_k$. $\hat{\boldsymbol{x}}_{\varepsilon,k}^+ = \hat{\boldsymbol{x}}_k^+ - \boldsymbol{x}_k$
$\boldsymbol{z}_k$	The measured process at sample k. $\boldsymbol{z}_k = \boldsymbol{C}_k \boldsymbol{x}_k + \boldsymbol{v}_k$
$\delta_{k,j}$	The Kronecker delta function. $\delta_{k,j} = 1$ for $k = j$ and zero otherwise.

1 Introduction

This document describes the assumptions underlying the basic Kalman filter and derives the standard Kalman equations. It is intended as a primer that should be read before tackling the documentation for the more specialized indirect complementary Kalman filter used for the fusion of accelerometer, magnetometer and gyroscope data.

Section 2 derives some mathematical results used in the derivation. The derivation itself is in section 3.

2 Mathematical Lemmas

2.1 Lemma 1

The trace of the sum of two matrices equals the sum of the individual traces.

Proof

$$tr(\boldsymbol{A}+\boldsymbol{B}) = \sum_{i=0}^{N-1} A_{ii}+B_{ii} = \sum_{i=0}^{N-1} A_{ii} + \sum_{i=0}^{N-1} B_{ii} = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$$

Eq 2.1.1

2.2 Lemma 2

The derivative with respect to $\boldsymbol{A}$ of the trace of the matrix product $\boldsymbol{C} = \boldsymbol{AB}$ equals $\boldsymbol{B}^T$.

Proof

$$\frac{\partial\{tr(\boldsymbol{C})\}}{\partial \boldsymbol{A}} = \frac{\partial\{tr(\boldsymbol{AB})\}}{\partial \boldsymbol{A}} = \begin{pmatrix} \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{0,0}}\right) & \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{0,1}}\right) & \cdots & \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{0,N-1}}\right) \\ \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{1,0}}\right) & \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{1,1}}\right) & \cdots & \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{1,N-1}}\right) \\ \cdots & \cdots & \cdots & \cdots \\ \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{M-1,0}}\right) & \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{M-1,1}}\right) & \cdots & \left(\frac{\partial tr(\boldsymbol{AB})}{\partial A_{M-1,N-1}}\right) \end{pmatrix}$$

Eq 2.2.1

Assuming that the matrix $\boldsymbol{A}$ has dimensions MxN and the matrix $\boldsymbol{B}$ has dimensions NxM, then $\boldsymbol{C} = \boldsymbol{AB}$ has dimensions MxM.

The element C_{ij} of matrix $\boldsymbol{C}$ has value:

$$C_{ij} = \sum_{k=0}^{N-1} A_{ik}B_{kj} \Rightarrow tr(\boldsymbol{C}) = tr(\boldsymbol{AB}) = \sum_{i=0}^{M-1} C_{ii} = \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}$$

Eq 2.2.2

Substituting gives:

$$\frac{\partial\{tr(\boldsymbol{AB})\}}{\partial \boldsymbol{A}} = \begin{pmatrix} \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{0,0}}\right) & \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{0,1}}\right) & \cdots & \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{0,N-1}}\right) \\ \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{1,0}}\right) & \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{1,1}}\right) & \cdots & \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{1,N-1}}\right) \\ \cdots & \cdots & \cdots & \cdots \\ \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{M-1,0}}\right) & \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{M-1,1}}\right) & \cdots & \left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{M-1,N-1}}\right) \end{pmatrix}$$

Eq 2.2.3

By inspection:

$$\left(\frac{\partial \sum_{i=0}^{M-1}\sum_{k=0}^{N-1} A_{ik}B_{ki}}{\partial A_{lm}}\right) = B_{ml}$$

Eq 2.2.4

Substituting back gives:

$$\frac{\partial\{tr(\boldsymbol{AB})\}}{\partial \boldsymbol{A}} = \begin{pmatrix} B_{0,0} & B_{1,0} & \dots & B_{N-1,0} \\ B_{0,1} & B_{1,1} & \dots & B_{N-1,1} \\ \dots & \dots & \dots & \dots \\ B_{0,M-1} & B_{1,M-1} & \dots & B_{N-1,M-1} \end{pmatrix} = \boldsymbol{B}^T$$

Eq 2.2.5

2.3 Lemma 3

The derivative with respect to $\boldsymbol{A}$ of the trace of the matrix product $\boldsymbol{ABA}^T$ equals $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{B}^T)$.

Proof

$$\frac{\partial\{tr(\boldsymbol{ABA}^T)\}}{\partial \boldsymbol{A}} = \begin{pmatrix} \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{0,0}}\right) & \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{0,1}}\right) & \dots & \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{0,N-1}}\right) \\ \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{1,0}}\right) & \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{1,1}}\right) & \dots & \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{1,N-1}}\right) \\ \dots & \dots & \dots & \dots \\ \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{M-1,0}}\right) & \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{M-1,1}}\right) & \dots & \left(\frac{\partial tr(\boldsymbol{ABA}^T)}{\partial A_{M-1,N-1}}\right) \end{pmatrix}$$

Eq 2.3.1

If the matrix $\boldsymbol{A}$ has dimensions MxN then the matrix $\boldsymbol{B}$ must be square with dimensions NxN for the product $\boldsymbol{ABA}^T$ to exist. The product $\boldsymbol{ABA}^T$ is always square with dimensions MxM.

The element C_{ij} of the matrix $\boldsymbol{C}=\boldsymbol{AB}$ has value:

$$C_{ij} = \sum_{k=0}^{N-1} A_{ik}B_{kj}$$

Eq 2.3.2

The element D_{il} of matrix $\boldsymbol{D}=\boldsymbol{ABA}^T=\boldsymbol{CA}^T$ has value:

$$D_{il} = \sum_{j=0}^{N-1} C_{ij}A_{lj} = \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} A_{ik}B_{kj}A_{lj}$$

Eq 2.3.3

The trace of matrix $\boldsymbol{D}$ has value:

$$tr(\boldsymbol{D}) = \sum_{i=0}^{N-1} D_{ii} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1} A_{ik}B_{kj}A_{ij}$$

Eq 2.3.4

The derivative of $tr(\boldsymbol{D})$ with respect to A_{lm} is then:

$$\left(\frac{\partial tr(\boldsymbol{D})}{\partial A_{lm}}\right) = \left(\frac{\partial \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1} A_{ik}B_{kj}A_{ij}}{\partial A_{lm}}\right) = \left(\frac{\partial \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} A_{lk}B_{kj}A_{lj}}{\partial A_{lm}}\right)$$

Eq 2.3.5

$$= \sum_{j=0}^{N-1} A_{lj}B_{mj} + \sum_{j=0}^{N-1} A_{lj}B_{jm} = (\boldsymbol{AB}^T)_{lm} + (\boldsymbol{AB})_{lm}$$

Eq 2.3.6

$$\Rightarrow \frac{\partial\{tr(\boldsymbol{ABA}^T)\}}{\partial \boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{B}+\boldsymbol{B}^T)$$

Eq 2.3.7

If $\boldsymbol{B}$ is also symmetric then:

$$\frac{\partial\{tr(\boldsymbol{ABA}^T)\}}{\partial\boldsymbol{A}} = 2\boldsymbol{AB}\ if\ \boldsymbol{B} = \boldsymbol{B}^T$$

Eq 2.3.8

3 Kalman Filter Derivation

3.1 Process Model

The Kalman filter models the vector process of interest $\boldsymbol{x}_k$ as linear and recursive:

$$\boldsymbol{x}_k = \boldsymbol{A}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k \qquad \text{Eq 3.1.1}$$

If $\boldsymbol{x}_k$ has N degrees of freedom then $\boldsymbol{A}_k$ is an NxN linear prediction matrix (possibly time varying but assumed known) and $\boldsymbol{w}_k$ is an $Nx1$ noise vector.

The process $\boldsymbol{x}_k$ is assumed to be not directly measurable and must be estimated from a process $\boldsymbol{z}_k$ which can be measured. $\boldsymbol{z}_k$ is modeled as being linearly related to $\boldsymbol{x}_k$ with additive noise $\boldsymbol{v}_k$.

$$\boldsymbol{z}_k = \boldsymbol{C}_k \boldsymbol{x}_k + \boldsymbol{v}_k \qquad \text{Eq 3.1.2}$$

$\boldsymbol{z}_k$ is an $Nx1$ vector, $\boldsymbol{C}_k$ is an NxN matrix (possibly time varying but assumed known) and $\boldsymbol{v}_k$ is an $Nx1$ noise vector.

The noise vectors $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are assumed to be zero mean white processes:

$$E[\boldsymbol{w}_k] = \boldsymbol{0} \qquad \text{Eq 3.1.3}$$

$$E[\boldsymbol{v}_k] = \boldsymbol{0} \qquad \text{Eq 3.1.4}$$

$$cov\{\boldsymbol{w}_k, \boldsymbol{w}_j\} = E\left[\boldsymbol{w}_k {\boldsymbol{w}_j}^T\right] = \boldsymbol{Q}_{w,k}\delta_{kj} \qquad \text{Eq 3.1.5}$$

$$cov\{\boldsymbol{v}_k, \boldsymbol{v}_j\} = E\left[\boldsymbol{v}_k {\boldsymbol{v}_j}^T\right] = \boldsymbol{Q}_{v,k}\delta_{kj} \qquad \text{Eq 3.1.6}$$

By definition, covariance matrices are symmetric.

$${\boldsymbol{Q}_{w,k}}^T = \{E[\boldsymbol{w}_k {\boldsymbol{w}_k}^T]\}^T = E[(\boldsymbol{w}_k {\boldsymbol{w}_k}^T)^T] = E\left[\boldsymbol{w}_k {\boldsymbol{w}_j}^T\right] = \boldsymbol{Q}_{w,k} \qquad \text{Eq 3.1.7}$$

3.2 Derivation

The objective of the Kalman filter is to compute an unbiased *a posterori* estimate $\hat{\boldsymbol{x}}_k^+$ of the underlying process $\boldsymbol{x}_k$ from i) extrapolation from the previous iteration's *a posteriori* estimate $\hat{\boldsymbol{x}}_{k-1}^+$ and ii) from the current measurement $\boldsymbol{z}_k$:

$$\hat{\boldsymbol{x}}_k^+ = \boldsymbol{K}'_k \hat{\boldsymbol{x}}_{k-1}^+ + \boldsymbol{K}_k \boldsymbol{z}_k \qquad \text{Eq 3.2.1}$$

The time-varying Kalman gain matrices $\boldsymbol{K}'_k$ and $\boldsymbol{K}_k$ define the relative weightings given to the previous iteration's Kalman filter estimate $\hat{\boldsymbol{x}}_{k-1}^+$ and to the current measurement $\boldsymbol{z}_k$. If the measurements $\boldsymbol{z}_k$ have low noise then a higher weighting will be given to the term $\boldsymbol{K}_k \boldsymbol{z}_k$ compared to the extrapolated component $\boldsymbol{K}'_k \hat{\boldsymbol{x}}_{k-1}^+$ and vice versa. The Kalman filter is therefore a time varying recursive filter.

Unbiased estimate constraint (determines $\boldsymbol{K}'_k$)

For $\hat{\boldsymbol{x}}_k^+$ to be an unbiased estimate of $\boldsymbol{x}_k$, the expectation value of the *a posteriori* Kalman filter error $\hat{\boldsymbol{x}}_{\varepsilon,k}^+$ must be zero:

$$E\left[\hat{\boldsymbol{x}}_{\varepsilon,k}^+\right] = E[\hat{\boldsymbol{x}}_k^+ - \boldsymbol{x}_k] = \boldsymbol{0} \qquad \text{Eq 3.2.2}$$

Subtracting $\boldsymbol{x}_k$ from equation 3.2.1 gives:

$$\hat{\boldsymbol{x}}_{\varepsilon,k}^{+} = \hat{\boldsymbol{x}}_k^{+} - \boldsymbol{x}_k = \boldsymbol{K}'_k \hat{\boldsymbol{x}}_{k-1}^{+} + \boldsymbol{K}_k \boldsymbol{z}_k - \boldsymbol{x}_k \qquad \text{Eq 3.2.3}$$

Substituting equation 3.1.2 for $\boldsymbol{z}_k$ gives:

$$\hat{\boldsymbol{x}}_{\varepsilon,k}^{+} = \boldsymbol{K}'_k \hat{\boldsymbol{x}}_{k-1}^{+} + \boldsymbol{K}_k (\boldsymbol{C}_k \boldsymbol{x}_k + \boldsymbol{v}_k) - \boldsymbol{x}_k \qquad \text{Eq 3.2.4}$$

Substituting for $\boldsymbol{x}_k$ from equation 3.1.1 and re-arranging gives:

$$\hat{\boldsymbol{x}}_{\varepsilon,k}^{+} = \boldsymbol{K}'_k \left(\hat{\boldsymbol{x}}_{\varepsilon,k-1}^{+} + \boldsymbol{x}_{k-1} \right) + \boldsymbol{K}_k \{ \boldsymbol{C}_k (\boldsymbol{A}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k) + \boldsymbol{v}_k \} - (\boldsymbol{A}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k) \qquad \text{Eq 3.2.5}$$

$$= \boldsymbol{K}'_k \hat{\boldsymbol{x}}_{\varepsilon,k-1}^{+} + (\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{A}_k - \boldsymbol{A}_k + \boldsymbol{K}'_k) \boldsymbol{x}_{k-1} + (\boldsymbol{K}_k \boldsymbol{C}_k - \boldsymbol{I}) \boldsymbol{w}_k + \boldsymbol{K}_k \boldsymbol{v}_k \qquad \text{Eq 3.2.6}$$

Taking the expectation value of equation 3.2.6 and applying the unbiased estimate constraint gives:

$$E\left[\hat{\boldsymbol{x}}_{\varepsilon,k}^{+}\right] = E\left[\boldsymbol{K}'_k \hat{\boldsymbol{x}}_{\varepsilon,k-1}^{+}\right] + E[(\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{A}_k - \boldsymbol{A}_k + \boldsymbol{K}'_k) \boldsymbol{x}_{k-1}] + E[(\boldsymbol{K}_k \boldsymbol{C}_k - \boldsymbol{I}) \boldsymbol{w}_k] + E[\boldsymbol{K}_k \boldsymbol{v}_k] = \boldsymbol{0} \qquad \text{Eq 3.2.7}$$

Since the noise vectors $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are zero mean and uncorrelated with the Kalman matrices for the same iteration, it follows that:

$$E[(\boldsymbol{K}_k \boldsymbol{C}_k - \boldsymbol{I}) \boldsymbol{w}_k] = E[\boldsymbol{K}_k \boldsymbol{v}_k] = \boldsymbol{0} \qquad \text{Eq 3.2.8}$$

With the additional assumption that the process $\boldsymbol{x}_{k-1}$ is independent of the Kalman matrices at iteration k:

$$E[(\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{A}_k - \boldsymbol{A}_k + \boldsymbol{K}'_k) \boldsymbol{x}_{k-1}] = (\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{A}_k - \boldsymbol{A}_k + \boldsymbol{K}'_k) E[\boldsymbol{x}_{k-1}] = \boldsymbol{0} \qquad \text{Eq 3.2.9}$$

Since $\boldsymbol{x}_k$ is not, in general, a zero mean process:

$$\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{A}_k - \boldsymbol{A}_k + \boldsymbol{K}'_k = \boldsymbol{0} \Rightarrow \boldsymbol{K}'_k = \boldsymbol{A}_k - \boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{A}_k = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k) \boldsymbol{A}_k \qquad \text{Eq 3.2.10}$$

Eliminating $\boldsymbol{K}'_k$ in equation 3.2.1 gives:

$$\hat{\boldsymbol{x}}_k^{+} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k) \boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1}^{+} + \boldsymbol{K}_k \boldsymbol{z}_k \qquad \text{Eq 3.2.11}$$

A priori estimate

The *a priori* Kalman filter estimate $\hat{\boldsymbol{x}}_k^{-}$ is defined as resulting from the application of the linear prediction matrix $\boldsymbol{A}_k$ to the previous iteration's *a posteriori* estimate $\hat{\boldsymbol{x}}_{k-1}^{+}$:

$$\hat{\boldsymbol{x}}_k^{-} = \boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1}^{+} \qquad \textbf{Kalman equation 1} \qquad \text{Eq 3.2.12}$$

Definition of a posteriori estimate

Substituting the a priori estimate $\hat{\boldsymbol{x}}_k^{-}$ into equation 3.2.11 gives:

$$\hat{\boldsymbol{x}}_k^{+} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k) \hat{\boldsymbol{x}}_k^{-} + \boldsymbol{K}_k \boldsymbol{z}_k \qquad \textbf{Kalman equation 4} \qquad \text{Eq 3.2.13}$$

An equivalent form is:

$$\hat{\boldsymbol{x}}_k^{+} = \hat{\boldsymbol{x}}_k^{-} + \boldsymbol{K}_k (\boldsymbol{z}_k - \boldsymbol{C}_k \hat{\boldsymbol{x}}_k^{-}) \qquad \text{Eq 3.2.14}$$

$\boldsymbol{P}_k^{-}$ as a function of $\boldsymbol{P}_{k-1}^{+}$

The *a priori and a posteriori* error covariance matrices $\boldsymbol{P}_k^-$ and $\boldsymbol{P}_k^+$ are defined as:

$$\boldsymbol{P}_k^- = cov\{\widehat{\boldsymbol{x}}_{\varepsilon,k}^-, \widehat{\boldsymbol{x}}_{\varepsilon,k}^-\} = E\left[\widehat{\boldsymbol{x}}_{\varepsilon,k}^- {\widehat{\boldsymbol{x}}_{\varepsilon,k}^-}^T\right] = E[(\widehat{\boldsymbol{x}}_k^- - \boldsymbol{x}_k)(\widehat{\boldsymbol{x}}_k^- - \boldsymbol{x}_k)^T] \quad \text{Eq 3.2.15}$$

$$\boldsymbol{P}_k^+ = cov\{\widehat{\boldsymbol{x}}_{\varepsilon,k}^+, \widehat{\boldsymbol{x}}_{\varepsilon,k}^+\} = E\left[\widehat{\boldsymbol{x}}_{\varepsilon,k}^+ {\widehat{\boldsymbol{x}}_{\varepsilon,k}^+}^T\right] = E[(\widehat{\boldsymbol{x}}_k^+ - \boldsymbol{x}_k)(\widehat{\boldsymbol{x}}_k^+ - \boldsymbol{x}_k)^T] \quad \text{Eq 3.2.16}$$

Substituting the definitions of $\widehat{\boldsymbol{x}}_k^-$ and $\boldsymbol{x}_k$ into equation 3.2.15 gives:

$$\boldsymbol{P}_k^- = E[(\boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k-1}^+ - \boldsymbol{A}_k\boldsymbol{x}_{k-1} - \boldsymbol{w}_k)(\boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k-1}^+ - \boldsymbol{A}_k\boldsymbol{x}_{k-1} - \boldsymbol{w}_k)^T] \quad \text{Eq 3.2.17}$$

$$= E[\{\boldsymbol{A}_k(\widehat{\boldsymbol{x}}_{k-1}^+ - \boldsymbol{x}_{k-1}) - \boldsymbol{w}_k\}\{\boldsymbol{A}_k(\widehat{\boldsymbol{x}}_{k-1}^+ - \boldsymbol{x}_{k-1}) - \boldsymbol{w}_k\}^T] \quad \text{Eq 3.2.18}$$

$$= \boldsymbol{A}_k E[(\widehat{\boldsymbol{x}}_{k-1}^+ - \boldsymbol{x}_{k-1})(\widehat{\boldsymbol{x}}_{k-1}^+ - \boldsymbol{x}_{k-1})^T]{\boldsymbol{A}_k}^T + \boldsymbol{Q}_{w,k} \quad \text{Eq 3.2.19}$$

$$\Rightarrow \boldsymbol{P}_k^- = \boldsymbol{A}_k\boldsymbol{P}_{k-1}^+{\boldsymbol{A}_k}^T + \boldsymbol{Q}_{w,k} \qquad \textbf{Kalman equation 2} \qquad \text{Eq 3.2.20}$$

Minimum error covariance constraint (determines K_k)

The Kalman gain matrix $\boldsymbol{K}_k$ minimizes the *a posteriori* error $\widehat{\boldsymbol{x}}_{\varepsilon,k}^+$ variance via the trace of the *a posteriori* error covariance matrix $\boldsymbol{P}_k^+$:

$$E\left[{\widehat{\boldsymbol{x}}_{\varepsilon,k}^+}^T \widehat{\boldsymbol{x}}_{\varepsilon,k}^+\right] = tr(\boldsymbol{P}_k^+) \quad \text{Eq 3.2.21}$$

Substituting equation 2.1.2 for $\boldsymbol{z}_k$ into equation 3.2.11 gives a relation between the *a posteriori* and *a priori* errors:

$$\widehat{\boldsymbol{x}}_k^+ = \widehat{\boldsymbol{x}}_{\varepsilon,k}^+ + \boldsymbol{x}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\widehat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k\boldsymbol{z}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\left(\widehat{\boldsymbol{x}}_{\varepsilon,k}^- + \boldsymbol{x}_k\right) + \boldsymbol{K}_k(\boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{v}_k) \quad \text{Eq 3.2.22}$$

$$\Rightarrow \widehat{\boldsymbol{x}}_{\varepsilon,k}^+ + \boldsymbol{x}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\widehat{\boldsymbol{x}}_{\varepsilon,k}^- + \boldsymbol{x}_k - \boldsymbol{K}_k\boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{K}_k(\boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{v}_k) \quad \text{Eq 3.2.23}$$

$$\Rightarrow \widehat{\boldsymbol{x}}_{\varepsilon,k}^+ = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\widehat{\boldsymbol{x}}_{\varepsilon,k}^- + \boldsymbol{K}_k\boldsymbol{v}_k \quad \text{Eq 3.2.24}$$

Substituting this result into the definition of the *a posteriori* covariance matrix $\boldsymbol{P}_k^+$ gives:

$$\boldsymbol{P}_k^+ = E\left[\{(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\widehat{\boldsymbol{x}}_{\varepsilon,k}^- + \boldsymbol{K}_k\boldsymbol{v}_k\}\{(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\widehat{\boldsymbol{x}}_{\varepsilon,k}^- + \boldsymbol{K}_k\boldsymbol{v}_k\}^T\right] \quad \text{Eq 3.2.25}$$

$$= (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)E\left[\widehat{\boldsymbol{x}}_{\varepsilon,k}^- {\widehat{\boldsymbol{x}}_{\varepsilon,k}^-}^T\right](\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)^T + \boldsymbol{K}_k E[\boldsymbol{v}_k{\boldsymbol{v}_k}^T]{\boldsymbol{K}_k}^T \quad \text{Eq 3.2.26}$$

$$= (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\boldsymbol{P}_k^-(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)^T + \boldsymbol{K}_k\boldsymbol{Q}_{v,k}{\boldsymbol{K}_k}^T \quad \text{Eq 3.2.27}$$

$$= \boldsymbol{P}_k^- - \boldsymbol{P}_k^-{\boldsymbol{C}_k}^T{\boldsymbol{K}_k}^T - \boldsymbol{K}_k\boldsymbol{C}_k\boldsymbol{P}_k^- + \boldsymbol{K}_k\boldsymbol{C}_k\boldsymbol{P}_k^-{\boldsymbol{C}_k}^T{\boldsymbol{K}_k}^T + \boldsymbol{K}_k\boldsymbol{Q}_{v,k}{\boldsymbol{K}_k}^T \quad \text{Eq 3.2.28}$$

The Kalman filter gain $\boldsymbol{K}_k$ is that which minimizes the trace of the *a posteriori* error covariance matrix $\boldsymbol{P}_k^+$:

$$\frac{\partial}{\partial \boldsymbol{K}_k} tr(\boldsymbol{P}_k^+) = \frac{\partial}{\partial \boldsymbol{K}_k}\left\{tr(\boldsymbol{P}_k^-) - tr\left(\boldsymbol{P}_k^- \boldsymbol{C}_k{}^T \boldsymbol{K}_k{}^T\right) - tr(\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^-) + tr\left(\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T \boldsymbol{K}_k{}^T\right) + tr\left(\boldsymbol{K}_k \boldsymbol{Q}_{v,k} \boldsymbol{K}_k{}^T\right)\right\} = 0$$

Eq 3.2.29

The term $tr(\boldsymbol{P}_k^-)$ has no dependence on $\boldsymbol{K}_k$ giving:

$$\frac{\partial\{tr(\boldsymbol{P}_k^-)\}}{\partial \boldsymbol{K}_k} = \frac{\partial\left\{tr(\boldsymbol{A}_k \boldsymbol{P}_{k-1}^+ \boldsymbol{A}_k{}^T + \boldsymbol{Q}_{w,k})\right\}}{\partial \boldsymbol{K}_k} = 0$$

Eq 3.2.30

Since the trace of a transposed matrix equals the trace of the original matrix and using equation 2.2.5 gives:

$$\frac{\partial\left\{tr(\boldsymbol{P}_k^- \boldsymbol{C}_k{}^T \boldsymbol{K}_k{}^T)\right\}}{\partial \boldsymbol{K}_k} = \frac{\partial\{tr(\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^-)\}}{\partial \boldsymbol{K}_k} = (\boldsymbol{C}_k \boldsymbol{P}_k^-)^T = \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T$$

Eq 3.2.31

The third term can be simplified using equations 2.3.7 and 2.3.8 exploiting the fact that the covariance matrix is symmetric:

$$\frac{\partial\left\{tr(\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T \boldsymbol{K}_k{}^T)\right\}}{\partial \boldsymbol{K}_k} = \boldsymbol{K}_k \left\{\boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T + \left(\boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T\right)^T\right\} = 2\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T$$

Eq 3.2.32

The final term can be simplified also using equations 2.3.7 and 2.3.8 to give:

$$\frac{\partial\left\{tr(\boldsymbol{K}_k \boldsymbol{Q}_{v,k} \boldsymbol{K}_k{}^T)\right\}}{\partial \boldsymbol{K}_k} = 2\boldsymbol{K}_k \boldsymbol{Q}_{v,k}$$

Eq 3.2.33

Substituting back into equation 2.2.29 gives the optimal Kalman filter gain matrix $\boldsymbol{K}_k$:

$$-2\boldsymbol{P}_k^- \boldsymbol{C}_k{}^T + 2\boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T + 2\boldsymbol{K}_k \boldsymbol{Q}_{v,k} = \boldsymbol{0}$$ Eq 3.2.34

$$\Rightarrow \boldsymbol{K}_k\left(\boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T + \boldsymbol{Q}_{v,k}\right) = \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T$$ Eq 3.2.35

$$\Rightarrow \boldsymbol{K}_k = \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T\left(\boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T + \boldsymbol{Q}_{v,k}\right)^{-1}$$ **Kalman equation 3** Eq 3.2.36

$\boldsymbol{P}_k^+$ as a function of $\boldsymbol{P}_k^-$

Rearranging equation 3.2.35 gives:

$$\boldsymbol{K}_k \boldsymbol{Q}_{v,k} = \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T - \boldsymbol{K}_k \boldsymbol{C}_k \boldsymbol{P}_k^- \boldsymbol{C}_k{}^T$$ Eq 3.2.37

Substituting equation 3.2.37 into equation 3.2.27 gives:

$$\boldsymbol{P}_k^+ = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k)\boldsymbol{P}_k^-\left(\boldsymbol{I} - \boldsymbol{C}_k{}^T \boldsymbol{K}_k{}^T\right) + (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k)\boldsymbol{P}_k^- \boldsymbol{C}_k{}^T \boldsymbol{K}_k{}^T$$ Eq 3.2.38

$$\Rightarrow \boldsymbol{P}_k^+ = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k)\boldsymbol{P}_k^-$$ **Kalman equation 5** Eq 3.2.39

This completes the derivation of the standard Kalman filter equations.

3.3 Standard Kalman Equations

Kalman equation 1

The linear prediction (*a priori*) estimate $\hat{\boldsymbol{x}}_k^-$ is made by applying the linear prediction matrix $\boldsymbol{A}_k$ to the previous sample's Kalman (*a posteriori*) filter estimate $\hat{\boldsymbol{x}}_{k-1}^+$.

$$\hat{\boldsymbol{x}}_k^- = \boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1}^+ \qquad \text{Eq 3.3.1}$$

Kalman equation 2

The *a priori* (linear extrapolation) error covariance matrix $\boldsymbol{P}_k^-$ is then updated using the model matrix $\boldsymbol{A}_k$ and the noise matrix $\boldsymbol{Q}_{w,k}$.

$$\boldsymbol{P}_k^- = \boldsymbol{A}_k \boldsymbol{P}_{k-1}^+ {\boldsymbol{A}_k}^T + \boldsymbol{Q}_{w,k} \qquad \text{Eq 3.3.2}$$

Kalman equations 2 and 5 can be combined to give a recursive update of $\boldsymbol{P}_k^-$ without explicit calculation of the *a posteriori* error covariance matrix $\boldsymbol{P}_k^+$ in Kalman equation 5:

$$\boldsymbol{P}_k^- = \boldsymbol{A}_k (\boldsymbol{I} - \boldsymbol{K}_{k-1} \boldsymbol{C}_{k-1}) \boldsymbol{P}_{k-1}^- {\boldsymbol{A}_k}^T + \boldsymbol{Q}_{w,k} \qquad \text{Eq 3.3.3}$$

Kalman equation 3

The Kalman filter gain matrix $\boldsymbol{K}_k$ is updated:

$$\boldsymbol{K}_k = \boldsymbol{P}_k^- {\boldsymbol{C}_k}^T \left(\boldsymbol{C}_k \boldsymbol{P}_k^- {\boldsymbol{C}_k}^T + \boldsymbol{Q}_{v,k} \right)^{-1} \qquad \text{Eq 3.3.4}$$

Kalman equation 4

The Kalman filter (*a posteriori*) estimate $\hat{\boldsymbol{x}}_k^+$ is computed from the current *a priori* estimate $\hat{\boldsymbol{x}}_k^-$ and the current measurement $\boldsymbol{z}_k$:

$$\hat{\boldsymbol{x}}_k^+ = \hat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k (\boldsymbol{z}_k - \boldsymbol{C}_k \hat{\boldsymbol{x}}_k^-) = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k) \hat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k \boldsymbol{z}_k \qquad \text{Eq 3.3.5}$$

Kalman equation 5

The *a posteriori* Kalman error covariance matrix $\boldsymbol{P}_k^+$ is updated ready for the next iteration. This equation can be skipped if $\boldsymbol{P}_k^-$ is updated recursively in terms of itself as in equation 3.3.3.

$$\boldsymbol{P}_k^+ = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{C}_k) \boldsymbol{P}_k^- \qquad \text{Eq 3.3.6}$$