The Kalman Filter
An Adventure in Derivation

Robert S. Huston
Pinpoint Dynamics, LLC

August 31, 2023

Table of Contents

1. Introduction

Consider the linear system dynamical process model

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_{k+1|k}\mathbf{x}_k + \mathbf{u}_k + \mathbf{\Gamma}_{k+1|k}\mathbf{w}_k$$

and the linear observation model

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{x}$ is an $n \times 1$ state vector, $\mathbf{\Phi}$ is an $n \times n$ state transition matrix, $\mathbf{u}$ is an $n \times 1$ control input, $\mathbf{\Gamma}$ is an $n \times p$ transformation matrix, and $\mathbf{w}_k$ is a $p \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{w}_k\mathbf{w}_k^T\right\} = \mathbf{Q}_k$$

and where $\mathbf{z}$ is an $m \times 1$ measurement vector, $\mathbf{H}$ is an $m \times n$ transformation matrix, and $\mathbf{v}_k$ is an $m \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{v}_k\mathbf{v}_k^T\right\} = \mathbf{R}_k$$

and where process and measurement noise sequences are uncorrelated

$$E\left\{\mathbf{w}_j\mathbf{v}_k^T\right\} = \mathbf{0}, \text{ for all } j \text{ and } k$$

Given the measurements, $\mathbf{z}_k$, the modeling knowledge of $\mathbf{\Phi}_{k+1|k}$, $\mathbf{\Gamma}_{k+1|k}$, and $\mathbf{H}_k$, and the statistical characteristics of $\mathbf{w}_k$ and $\mathbf{v}_k$, the Kalman filter provides an optimal estimate of the state, $\mathbf{x}_k$, at time event k

$$\hat{\mathbf{x}}_k = E\left\{\mathbf{x}_k\right\}$$

with a state estimation error

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$$

and where the covariance of the state estimation error is

$$\begin{aligned}\mathbf{P}_k &= E\left\{\mathbf{e}_k\mathbf{e}_k^T\right\} \\ &= E\left\{\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right]\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right]^T\right\}\end{aligned}$$

The well-known Kalman filter recursion update cycle is comprised of two steps: a “projection” or *a priori* update, and a “correction” or *a posteriori* update.

I. Projection (*a priori*) update:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{u}_{k-1}$$
$$\mathbf{P}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\mathbf{P}_{k-1}\mathbf{\Phi}_{k|k-1}^{T} + \mathbf{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\mathbf{\Gamma}_{k|k-1}^{T}$$

II. Correction (*a posteriori*) update:

$$\hat{\mathbf{z}}_k = \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$$
$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k$$
$$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{T} + \mathbf{R}_k$$
$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{T}\mathbf{S}_k^{-1}$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\tilde{\mathbf{z}}_k$$
$$\mathbf{P}_k = \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1}$$

Unfortunately, the vast literature for the Kalman filter, and its varieties, can be overwhelmingly confounding. Different variables are used. Different notations are used. Different terms are used. The complexity of some of the mathematical proofs and theorems can obscure the more important matters. And then there are questions … many questions. What's with all these equations? What makes them so special? Where do they come from? What is meant by "optimal" gain? Why can't I just use least-squares estimation? Isn't the Kalman filter just a specialized least-squares filter anyway? What does it mean when it is said that "$\mathbf{S}_k$" represents the covariance of the measurement error? Why are there so many different forms of the filter, and which one should I use? And did Jeffrey Epstein really kill himself?

This document is an assorted collection of derivations and notes pertaining to the Kalman filter. They are targeted to the practicing engineer who has been tasked with developing and implementing a Kalman filtering solution. The mathematical derivations are therefore intended to eliminate some of the mysteries of how and why the Kalman filter works, as well as what the various versions and implementation strategies have to offer.

Some assumptions are made about the reader:

- The reader understands basic matrix algebra
- The reader understands basic calculus
- The reader understands basic probability and statistics

Lastly, this document is written the the way I wished I could have experienced when I was first learning the Kalman filter. Hopefully, it is helpful to the reader as is. If not, well then too bad.

1. Conventions and Notations

When I began my journey into studying the Kalman filter, my primary reference was the "Robert Grover Brown" book (see the "References" section at the end of this document), and much of my choice of conventions and notations comes from that book. I also heavily relied on Sorenson's material which was consistent with the R.G. Brown book. As I accumulated other references over the years, I discovered that the Brown/Sorenson conventions and notations were dominant. Granted there are other conventions and notations in the Kalman filtering literature, but I feel most at home when using the Brown/Sorenson conventions and notations, and those are the ones I've chosen to adopt in this document.

In particular, $\mathbf{x}_k$ represents a system state vector at time event "k", $\mathbf{\Phi}_{k+1|k}$ represents a linear state transition matrix that transitions the state from time events "k" to "$k+1$", $\mathbf{z}_k$ represents a measurement observation vector that observes the system output at time event "k", and $\mathbf{H}_k$ represents a linear observation transformation from $\mathbf{x}_k$ to $\mathbf{z}_k$. Additionally, $\mathbf{w}_k$ and $\mathbf{v}_k$ represent random sequences that model the system dynamics variations and measurement variations, respectively, with covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively. Lastly, the system state estimation error covariance is represented by $\mathbf{P}_k$.

In addition to the many different choices of variables in the technical literature, there are also many styles. For instance, instead of using the "k" and "$k+1\,|\,k$"subscripts, some references use "(k)" and "$(k+1\,|\,k)$" function forms instead, as in $\mathbf{x}(k)$ and $\mathbf{\Phi}(k+1\,|\,k)$. Others use "$+$" and "$-$" indicators to designate *a priori* variables from *a posteriori* variables, e.g., $\mathbf{P}_k^-$ instead of $\mathbf{P}_{k|k-1}$, and $\mathbf{P}_k^+$ instead of $\mathbf{P}_k$. The notations adopted in this document strive to make the equations clear (or as clear as they can be) without being "too busy" in the variable descriptors; $\mathbf{\Phi}_{k+1|k}$ is more readable in an expression than $\mathbf{\Phi}(k+1\,|\,k)$.

There are many references that do not distinguish visually between a scalar quantity and a matrix quantity. While this may have been a necessity in the days of typewriters and early word processors, it also does the reader an extreme disservice. Knowing that a variable is a scalar or a matrix is fundamental in the understanding of an equation or derivation, and given today's technology in word processing and electronic typesetting, there is no excuse for not adhering to the accepted conventions where an italic variable, s, is a scalar quantity, and a boldface variable, $\mathbf{M}$, is a matrix quantity.

I've often wondered why this particular choice of variables was picked by certain authors. For instance, why use **z** for the observation vector when the state vector is **x** and the natural choice would be to use **y** instead? After all, that's what all the classical linear system theory resources use. And why use **H** as the linear observation transformation instead of something more consistent with classical linear system theory? My uneducated guess is that using **z** allows for the use of **y** to represent additional state values (e.g., hidden states) in augmented system formulations, and using **H** is simply a natural alphabetic progression after the use of **F** and **G** are

assigned for system functional transformations. For instance, consider the classical continuous-time linear system description

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

where $\mathbf{x}$ is the system state, $\mathbf{u}$ is the control input, and $\mathbf{y}$ is the output observation. In transforming to a discrete-time description, a natural choice of variables would be

$$\mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{G}_k\mathbf{u}_k$$
$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{J}_k\mathbf{u}_k$$

We should stay away from using $\mathbf{E}$ because it could easily be confused with the expectation operation $E\{\cdots\}$, and we can't use $\mathbf{I}$ because it's the identity matrix (although I've seen some authors use $\mathbf{I}_k$ to represent an information matrix value at time event "k", causing me horrible confusion), so these choices seem to make logical sense. Similarly, the choices for covariance values appear to be alphabetically assigned. If $\mathbf{P}$ is our first-assigned covariance, then using $\mathbf{Q}$ and $\mathbf{R}$ make sense for the next two assignments. Of course, this reasoning behind the assignments is pure conjecture on my part, but it allows me to find order in the conventions that I've chosen from R.G. Brown and others.

The following list summarizes the general conventions and notations used in this document:

- We use an italic variable, s, to represent a scalar quantity (the variable can be of any case)
- We use a boldface lower-case variable, $\mathbf{v}$, to represent a vector matrix quantity
- We use a boldface upper-case variable, $\mathbf{M}$, to represent a block matrix quantity
- We use $\mathbf{x}_k$ to represent a system state vector at time event "k"
- We use $\hat{\mathbf{x}}_k$ to represent an estimate of the system state vector, $\mathbf{x}_k$, at time event "k"
- We use $\mathbf{P}_k$ to represent the covariance matrix of the state estimation error, $\mathbf{x}_k - \hat{\mathbf{x}}_k$
- We use $\mathbf{\Phi}_{k+1|k}$ to represent a linear state transition matrix that transitions $\mathbf{x}_k$ from time events "k" to "$k + 1$"
- We use $\mathbf{u}_k$ to represent a system control input vector at time event "k"
- We use $\mathbf{w}_k$ to represent a system process random variation vector at time event "k"
- We use $\mathbf{Q}_k$ to represent the covariance matrix of the system process random variation, $\mathbf{w}_k$
- We use $\mathbf{\Gamma}_{k+1|k}$ to represent a linear matrix that transforms $\mathbf{w}_k$ from time events "k" to "$k + 1$"
- We use $\mathbf{z}_k$ to represent a measurement observation vector that observes the system output at time event "k"
- We use $\hat{\mathbf{z}}_k$ to represent an estimate of the measurement observation vector, $\mathbf{z}_k$, at time event "k"
- We use $\mathbf{S}_k$ to represent the covariance matrix of the measurement estimation error, $\mathbf{z}_k - \hat{\mathbf{z}}_k$
- We use$\mathbf{H}_k$ to represent a linear observation transformation matrix from $\mathbf{x}_k$ to $\mathbf{z}_k$
- We use $\mathbf{v}_k$ to represent a measurement random variation vector at time event "k"
- We use $\mathbf{R}_k$ to represent the covariance matrix of the measurement random variation, $\mathbf{v}_k$
- We use $\mathbf{K}_k$ to represent the state estimation filter gain matrix

2. Weighted Least-Squares Estimation

Consider the overdetermined measurement system

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

where $\mathbf{z}$ is an $m \times 1$ measurement vector, $\mathbf{x}$ is an $n \times 1$ state vector, $\mathbf{H}$ is an $m \times n$ observation matrix, and $\mathbf{v}$ is an $m \times 1$ random vector of measurement errors with covariance

$$E\left\{\mathbf{v}\mathbf{v}^T\right\} = \mathbf{R}$$

As stated previously, $m \geq n$, i.e., there can be more measurements than there are states. So $\mathbf{H}$ is not necessarily square:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2n} \\ h_{31} & h_{32} & h_{33} & \dots & h_{3n} \\ h_{41} & h_{42} & h_{43} & \dots & h_{4n} \\ h_{51} & h_{52} & h_{53} & \dots & h_{5n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & h_{m3} & \dots & h_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ \vdots \\ v_m \end{bmatrix}$$

Assume that we already have a state estimate, $\hat{\mathbf{x}}$. We can then determine the corresponding measurement estimate, $\hat{\mathbf{z}}$, from

$$\hat{\mathbf{z}} = \mathbf{H}\,\hat{\mathbf{x}}$$

We form the measurement residual, i.e., the measurement error

$$\tilde{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{z}}$$

We can then form the weighted squared error term

$$\epsilon^2 = \tilde{\mathbf{z}}^T \mathbf{W} \tilde{\mathbf{z}}$$

where $\mathbf{W}$ is an arbitrary symmetric weighting matrix.

The objective is to determine the "best" value for $\hat{\mathbf{x}}$ that fits the measurement data $\mathbf{z}$. We will define "best" to be the value of $\hat{\mathbf{x}}$ that minimizes ϵ^2. In other words, we want to find the "least-squares" estimate of $\mathbf{x}$.

Expanding the terms of ϵ^2 gives

$$
\begin{aligned}
\epsilon^2 &= \tilde{\mathbf{z}}^T \mathbf{W} \tilde{\mathbf{z}} \\
&= \left[\mathbf{z} - \hat{\mathbf{z}}\right]^T \mathbf{W} \left[\mathbf{z} - \hat{\mathbf{z}}\right] \\
&= \left[\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\right]^T \mathbf{W} \left[\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\right] \\
&= \left[\mathbf{z}^T - \hat{\mathbf{x}}^T \mathbf{H}^T\right] \mathbf{W} \left[\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\right] \\
&= \mathbf{z}^T \mathbf{W} \mathbf{z} - \hat{\mathbf{x}}^T \mathbf{H}^T \mathbf{W} \mathbf{z} - \mathbf{z}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}} + \hat{\mathbf{x}}^T \mathbf{H}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}}
\end{aligned}
$$

To find the value of $\hat{\mathbf{x}}$ that minimizes ϵ^2, we differentiate ϵ^2 with respect to $\hat{\mathbf{x}}$ and set the resulting expression to zero. We can make use of the following differentiation formulas:

$$
\frac{d\left(\mathbf{a}^T \mathbf{x}\right)}{d\mathbf{x}} = \frac{d\left(\mathbf{x}^T \mathbf{a}\right)}{d\mathbf{x}} = \mathbf{a}
$$

$$
\frac{d\left(\mathbf{x}^T \mathbf{A} \mathbf{x}\right)}{d\mathbf{x}} = 2\mathbf{A}\mathbf{x}\,, \quad (\mathbf{A} \text{ is symmetric})
$$

We use the differential operation where, if s is a scalar value and $\mathbf{x}$ is a column vector value, then

$$
\frac{ds}{d\mathbf{x}} \triangleq \begin{bmatrix} \dfrac{\partial s}{\partial x_1} \\ \dfrac{\partial s}{\partial x_2} \\ \dfrac{\partial s}{\partial x_3} \\ \vdots \end{bmatrix}
$$

Then

$$
\begin{aligned}
\frac{d\left(\epsilon^2\right)}{d\hat{\mathbf{x}}} &= -\mathbf{H}^T \mathbf{W} \mathbf{z} - \left(\mathbf{z}^T \mathbf{W} \mathbf{H}\right)^T + 2\mathbf{H}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}} \\
&= \mathbf{0}
\end{aligned}
$$

and so, observing that $\mathbf{H}^T \mathbf{W} \mathbf{H}$ is invertible, we see that

$$
\begin{aligned}
-\mathbf{H}^T \mathbf{W} \mathbf{z} - \left(\mathbf{z}^T \mathbf{W} \mathbf{H}\right)^T + 2\mathbf{H}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}} &= \mathbf{0} \\
-\mathbf{H}^T \mathbf{W} \mathbf{z} - \mathbf{H}^T \mathbf{W} \mathbf{z} + 2\mathbf{H}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}} &= \mathbf{0} \\
-2\mathbf{H}^T \mathbf{W} \mathbf{z} + 2\mathbf{H}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}} &= \mathbf{0} \\
\mathbf{H}^T \mathbf{W} \mathbf{H} \hat{\mathbf{x}} &= \mathbf{H}^T \mathbf{W} \mathbf{z} \\
\hat{\mathbf{x}} &= \left(\mathbf{H}^T \mathbf{W} \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{W} \mathbf{z}
\end{aligned}
$$

Hence the "best" value of $\hat{\mathbf{x}}$ that fits the measurement data $\mathbf{z}$ in the least-squares sense is

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}$$

To find the optimal value of $\mathbf{W}$, we form the state estimation error

$$\begin{aligned}
\mathbf{e} &= \mathbf{x} - \hat{\mathbf{x}} \\
&= \mathbf{x} - \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z} \\
&= \mathbf{x} - \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\left[\mathbf{H}\mathbf{x} + \mathbf{v}\right] \\
&= \mathbf{x} - \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{H}\mathbf{x} - \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v} \\
&= \mathbf{x} - \mathbf{x} - \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v} \\
&= -\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v}
\end{aligned}$$

Then the covariance of the state estimation error is

$$\begin{aligned}
\mathbf{P} &= E\left\{\mathbf{e}\,\mathbf{e}^T\right\} \\
&= E\left\{\left[-\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v}\right]\left[-\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v}\right]^T\right\} \\
&= E\left\{\left[-\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v}\right]\left[-\mathbf{v}^T\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\right]\right\} \\
&= E\left\{\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{v}\mathbf{v}^T\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\right\} \\
&= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\;E\left\{\mathbf{v}\mathbf{v}^T\right\}\;\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{R}\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}
\end{aligned}$$

In order to simplify this expression, let $\mathbf{R} = \mathbf{W}^{-1}$. Then

$$\begin{aligned}
\mathbf{P} &= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{R}\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{W}^{-1}\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}
\end{aligned}$$

Hence, the optimal value for $\mathbf{W}$ is

$$\mathbf{W} = \mathbf{R}^{-1}$$

Note that the least-squares solution only uses information from the measurement description. It has no knowledge of the system state dynamics.

Let us now consider the unweighted ($\mathbf{W} = \mathbf{I}$) least-squares solution

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{z}$$

Define

$$\mathbf{A} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$$

so that

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{z}$$

If, for the moment, we treat the vector $\mathbf{z}$ as a zero-mean Gaussian random vector with unit noise covariance,

$$E\left\{\mathbf{z}\,\mathbf{z}^T\right\} = \mathbf{I}$$

then, because the solution for $\hat{\mathbf{x}}$ is a linear transformation on $\mathbf{z}$, the covariance of $\hat{\mathbf{x}}$ is then

$$\begin{aligned} E\left\{\hat{\mathbf{x}}\,\hat{\mathbf{x}}^T\right\} &= E\left\{\left[\mathbf{A}\mathbf{z}\right]\left[\mathbf{A}\mathbf{z}\right]^T\right\} \\ &= E\left\{\left[\mathbf{A}\mathbf{z}\right]\left[\mathbf{z}^T\mathbf{A}^T\right]\right\} \\ &= E\left\{\mathbf{A}\mathbf{z}\,\mathbf{z}^T\mathbf{A}^T\right\} \\ &= \mathbf{A}E\left\{\mathbf{z}\,\mathbf{z}^T\right\}\mathbf{A}^T \\ &= \mathbf{A}\mathbf{A}^T \end{aligned}$$

Substituting the expansion for $\mathbf{A}$ into $E\left\{\hat{\mathbf{x}}\,\hat{\mathbf{x}}^T\right\}$ gives

$$
\begin{aligned}
E\left\{\hat{\mathbf{x}}\,\hat{\mathbf{x}}^T\right\} &= \mathbf{A}\mathbf{A}^T \\
&= \left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]^T \\
&= \left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\left[\mathbf{H}\left(\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right)^T\right] \\
&= \left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\left[\mathbf{H}\left(\left(\mathbf{H}^T\mathbf{H}\right)^T\right)^{-1}\right] \\
&= \left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\left[\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right] \\
&= \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\left(\mathbf{H}^T\mathbf{H}\right)\left(\mathbf{H}^T\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{H}\right)^{-1}
\end{aligned}
$$

Note that this is identical to the unweighted ($\mathbf{W} = \mathbf{I}$) expression for the state estimation error covariance, $\mathbf{P}$, with $\mathbf{R} = \mathbf{I}$:

$$
\begin{aligned}
\mathbf{P} &= E\left\{\mathbf{e}\,\mathbf{e}^T\right\} \\
&= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{R}\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1} \\
&= \left(\mathbf{H}^T\mathbf{H}\right)^{-1}
\end{aligned}
$$

In navigation applications, $\mathbf{H}$ is a geometric transformation from navigation coordinates, $\mathbf{x}$, to their observation coordinates, $\mathbf{z}$. A parameter known as the Geometric Dilution of Precision (GDOP) is an indicator of the goodness of fit of the navigation solution, $\hat{\mathbf{x}}$. The GDOP is based on the above covariance relation for $E\left\{\hat{\mathbf{x}}\,\hat{\mathbf{x}}^T\right\}$, since smaller covariance values indicate a solution that exhibits less dispersion. The GDOP is defined as

$$
\begin{aligned}
\text{GDOP} &= \sqrt{\text{tr}\left(E\left\{\hat{\mathbf{x}}\,\hat{\mathbf{x}}^T\right\}\right)} \\
&= \sqrt{\text{tr}\left(\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right)}
\end{aligned}
$$

The GDOP is a measure of the precision of the navigation solution, $\hat{\mathbf{x}}$. A low GDOP value indicates a high confidence in the precision of the solution, and vice versa. Also, observe that the GDOP is based only on $\mathbf{H}$; it does not require the existence of measurements, $\mathbf{z}$, or the determination of a solution, $\hat{\mathbf{x}}$. Hence, one can compute the GDOP of different geometric conditions as a way to optimize the selection of signaling sources to be used for navigating.

3. The Kalman Filter

We begin with the unforced linear system dynamical process model

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_{k+1|k}\mathbf{x}_k + \mathbf{\Gamma}_{k+1|k}\mathbf{w}_k$$

where $\mathbf{x}$ is an $n \times 1$ state vector, $\mathbf{\Phi}$ is an $n \times n$ state transition matrix, $\mathbf{\Gamma}$ is an $n \times p$ transformation matrix, and $\mathbf{w}_k$ is a $p \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{w}_k\mathbf{w}_k^T\right\} = \mathbf{Q}_k$$

We also have a linear observation model

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{z}$ is an $m \times 1$ measurement vector, $\mathbf{H}$ is an $m \times n$ transformation matrix, and $\mathbf{v}_k$ is an $m \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{v}_k\mathbf{v}_k^T\right\} = \mathbf{R}_k$$

The process and measurement noise sequences are uncorrelated

$$E\left\{\mathbf{w}_j\mathbf{v}_k^T\right\} = \mathbf{0}, \quad \text{for all } j \text{ and } k$$

We wish to develop an optimal state estimator so that

$$\hat{\mathbf{x}}_k = E\left\{\mathbf{x}_k\right\}$$

where the state estimation error is

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$$

and the covariance of the state estimation error is

$$\begin{aligned}\mathbf{P}_k &= E\left\{\mathbf{e}_k\mathbf{e}_k^T\right\} \\ &= E\left\{\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right]\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right]^T\right\}\end{aligned}$$

The estimation error sequence is uncorrelated with the process and observation noise sequences

$$E\left\{\mathbf{e}_j\mathbf{w}_k^T\right\} = \mathbf{0}, \quad \text{for all } j \text{ and } k$$

$$E\left\{\mathbf{e}_j\mathbf{v}_k^T\right\} = \mathbf{0}\,, \text{ for all } j \text{ and } k$$

Observe that the system dynamical model tells us that the state for time event k is given by

$$\mathbf{x}_k = \mathbf{\Phi}_{k|k-1}\mathbf{x}_{k-1} + \mathbf{\Gamma}_{k|k-1}\mathbf{w}_{k-1}$$

Then, given a previously determined state estimate, $\hat{\mathbf{x}}_{k-1}$, the *a priori* state projection estimate at time event k is

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\,\hat{\mathbf{x}}_{k-1}$$

with *a priori* state estimation error

$$\mathbf{e}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$$

and *a priori* state estimation error covariance

$$\begin{aligned}\mathbf{P}_{k|k-1} &= E\left\{\mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^T\right\} \\ &= E\left\{\left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right]\left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right]^T\right\}\end{aligned}$$

Substituting the expressions for $\mathbf{x}_k$ and $\hat{\mathbf{x}}_{k|k-1}$ into $\mathbf{e}_{k|k-1}$ gives

$$\begin{aligned}\mathbf{e}_{k|k-1} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \\ &= \mathbf{\Phi}_{k|k-1}\mathbf{x}_{k-1} + \mathbf{\Gamma}_{k|k-1}\mathbf{w}_{k-1} - \mathbf{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1} \\ &= \mathbf{\Phi}_{k|k-1}\left[\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right] + \mathbf{\Gamma}_{k|k-1}\mathbf{w}_{k-1} \\ &= \mathbf{\Phi}_{k|k-1}\mathbf{e}_{k-1} + \mathbf{\Gamma}_{k|k-1}\mathbf{w}_{k-1}\end{aligned}$$

At this point, it will be helpful to introduce two useful lemmas for equations of this form.

Lemma 1:

Let $\mathbf{a}_k$ and $\mathbf{b}_k$ both be uncorrelated random sequences, $E\left\{\mathbf{a}_k\mathbf{b}_k^T\right\} = \mathbf{0}$, with covariances $\mathbf{A}_k = E\left\{\mathbf{a}_k\mathbf{a}_k^T\right\}$ and $\mathbf{B}_k = E\left\{\mathbf{b}_k\mathbf{b}_k^T\right\}$, respectively. Furthermore, let $\mathbf{C}$ and $\mathbf{D}$ be constant matrices such that $\mathbf{Ca}_k$ and $\mathbf{Db}_k$ have the same dimension: $\dim\left(\mathbf{Ca}_k\right) = \dim\left(\mathbf{Db}_k\right)$. Then

$$E\left\{\left[\mathbf{Ca}_k + \mathbf{Db}_k\right]\left[\mathbf{Ca}_k + \mathbf{Db}_k\right]^T\right\} = \mathbf{CA}_k\mathbf{C}^T + \mathbf{DB}_k\mathbf{D}^T$$

Proof:

$$
\begin{aligned}
&E\left\{\left[\mathbf{C}\mathbf{a}_k+\mathbf{D}\mathbf{b}_k\right]\left[\mathbf{C}\mathbf{a}_k+\mathbf{D}\mathbf{b}_k\right]^T\right\}\\
&\quad=E\left\{\left[\mathbf{C}\mathbf{a}_k+\mathbf{D}\mathbf{b}_k\right]\left[\mathbf{a}_k^T\mathbf{C}^T+\mathbf{b}_k^T\mathbf{D}^T\right]\right\}\\
&\quad=E\left\{\left[\mathbf{C}\mathbf{a}_k+\mathbf{D}\mathbf{b}_k\right]\mathbf{a}_k^T\mathbf{C}^T+\left[\mathbf{C}\mathbf{a}_k+\mathbf{D}\mathbf{b}_k\right]\mathbf{b}_k^T\mathbf{D}^T\right\}\\
&\quad=E\left\{\mathbf{C}\mathbf{a}_k\mathbf{a}_k^T\mathbf{C}^T+\mathbf{D}\mathbf{b}_k\mathbf{a}_k^T\mathbf{C}^T+\mathbf{C}\mathbf{a}_k\mathbf{b}_k^T\mathbf{D}^T+\mathbf{D}\mathbf{b}_k\mathbf{b}_k^T\mathbf{D}^T\right\}\\
&\quad=\mathbf{C}E\left\{\mathbf{a}_k\mathbf{a}_k^T\right\}\mathbf{C}^T+\mathbf{D}E\left\{\mathbf{b}_k\mathbf{a}_k^T\right\}\mathbf{C}^T+\mathbf{C}E\left\{\mathbf{a}_k\mathbf{b}_k^T\right\}\mathbf{D}^T+\mathbf{D}E\left\{\mathbf{b}_k\mathbf{b}_k^T\right\}\mathbf{D}^T\\
&\quad=\mathbf{C}\mathbf{A}_k\mathbf{C}^T+\mathbf{0}+\mathbf{0}+\mathbf{D}\mathbf{B}_k\mathbf{D}^T\\
&\quad=\mathbf{C}\mathbf{A}_k\mathbf{C}^T+\mathbf{D}\mathbf{B}_k\mathbf{D}^T
\end{aligned}
$$

Lemma 2:

Let $\mathbf{a}_k$ and $\mathbf{b}_k$ both be uncorrelated random sequences, $E\left\{\mathbf{a}_k\mathbf{b}_k^T\right\}=\mathbf{0}$, with covariances $\mathbf{A}_k=E\left\{\mathbf{a}_k\mathbf{a}_k^T\right\}$ and $\mathbf{B}_k=E\left\{\mathbf{b}_k\mathbf{b}_k^T\right\}$, respectively. Furthermore, let $\mathbf{C}$ and $\mathbf{D}$ be constant matrices such that $\mathbf{C}\mathbf{a}_k$ and $\mathbf{D}\mathbf{b}_k$ have the same dimension: $\dim\left(\mathbf{C}\mathbf{a}_k\right)=\dim\left(\mathbf{D}\mathbf{b}_k\right)$. Then

$$E\left\{\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]^T\right\}=\mathbf{C}\mathbf{A}_k\mathbf{C}^T+\mathbf{D}\mathbf{B}_k\mathbf{D}^T$$

Proof:

$$
\begin{aligned}
&E\left\{\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]^T\right\}\\
&\quad=E\left\{\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]\left[\mathbf{a}_k^T\mathbf{C}^T-\mathbf{b}_k^T\mathbf{D}^T\right]\right\}\\
&\quad=E\left\{\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]\mathbf{a}_k^T\mathbf{C}^T-\left[\mathbf{C}\mathbf{a}_k-\mathbf{D}\mathbf{b}_k\right]\mathbf{b}_k^T\mathbf{D}^T\right\}\\
&\quad=E\left\{\mathbf{C}\mathbf{a}_k\mathbf{a}_k^T\mathbf{C}^T-\mathbf{D}\mathbf{b}_k\mathbf{a}_k^T\mathbf{C}^T-\mathbf{C}\mathbf{a}_k\mathbf{b}_k^T\mathbf{D}^T+\mathbf{D}\mathbf{b}_k\mathbf{b}_k^T\mathbf{D}^T\right\}\\
&\quad=\mathbf{C}E\left\{\mathbf{a}_k\mathbf{a}_k^T\right\}\mathbf{C}^T-\mathbf{D}E\left\{\mathbf{b}_k\mathbf{a}_k^T\right\}\mathbf{C}^T-\mathbf{C}E\left\{\mathbf{a}_k\mathbf{b}_k^T\right\}\mathbf{D}^T+\mathbf{D}E\left\{\mathbf{b}_k\mathbf{b}_k^T\right\}\mathbf{D}^T\\
&\quad=\mathbf{C}\mathbf{A}_k\mathbf{C}^T-\mathbf{0}-\mathbf{0}+\mathbf{D}\mathbf{B}_k\mathbf{D}^T\\
&\quad=\mathbf{C}\mathbf{A}_k\mathbf{C}^T+\mathbf{D}\mathbf{B}_k\mathbf{D}^T
\end{aligned}
$$

Now, noting that

$$E\left\{\mathbf{e}_{k-1}\mathbf{w}_{k-1}^T\right\}=\mathbf{0}$$

and using **Lemma 1**, the expansion of the *a priori* state projection estimate error covariance gives

$$
\begin{aligned}
\mathbf{P}_{k|k-1} &= E\left\{\mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^T\right\} \\
&= E\left\{\left[\boldsymbol{\Phi}_{k|k-1}\mathbf{e}_{k-1} + \boldsymbol{\Gamma}_{k|k-1}\mathbf{w}_{k-1}\right]\left[\boldsymbol{\Phi}_{k|k-1}\mathbf{e}_{k-1} + \boldsymbol{\Gamma}_{k|k-1}\mathbf{w}_{k-1}\right]^T\right\} \\
&= \boldsymbol{\Phi}_{k|k-1}\mathbf{P}_{k-1}\boldsymbol{\Phi}_{k|k-1}^T + \boldsymbol{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\boldsymbol{\Gamma}_{k|k-1}^T
\end{aligned}
$$

and so the *a priori* state projection estimate error covariance is

$$\mathbf{P}_{k|k-1} = \boldsymbol{\Phi}_{k|k-1}\mathbf{P}_{k-1}\boldsymbol{\Phi}_{k|k-1}^T + \boldsymbol{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\boldsymbol{\Gamma}_{k|k-1}^T$$

In addition, observe that, because $\mathbf{P}_{k|k-1}$ is a covariance matrix,

$$\mathbf{P}_{k|k-1}^T = \mathbf{P}_{k|k-1}$$

Now, given our *a priori* state projection estimate, $\hat{\mathbf{x}}_{k|k-1}$, we can estimate the corresponding projection measurement estimate from

$$\hat{\mathbf{z}}_k = \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$$

We then form the measurement residual, i.e., the measurement estimation error

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k$$

Substituting the expressions for $\mathbf{z}_k$ and $\hat{\mathbf{z}}_k$ into $\tilde{\mathbf{z}}_k$ gives

$$
\begin{aligned}
\tilde{\mathbf{z}}_k &= \mathbf{z}_k - \hat{\mathbf{z}}_k \\
&= \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \\
&= \mathbf{H}_k\left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right] + \mathbf{v}_k \\
&= \mathbf{H}_k\mathbf{e}_{k|k-1} + \mathbf{v}_k
\end{aligned}
$$

Noting that

$$E\left\{\mathbf{e}_{k|k-1}\mathbf{v}_k^T\right\} = \mathbf{0}$$

and using **Lemma 1**, we then see that the covariance of the measurement residual is

$$E\left\{\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T\right\} = E\left\{\left[\mathbf{H}_k\mathbf{e}_{k|k-1} + \mathbf{v}_k\right]\left[\mathbf{H}_k\mathbf{e}_{k|k-1} + \mathbf{v}_k\right]^T\right\}$$
$$= \mathbf{H}_k\,\mathbf{P}_{k|k-1}\,\mathbf{H}_k^T + \mathbf{R}_k$$

Define

$$\mathbf{S}_k = \mathbf{H}_k\,\mathbf{P}_{k|k-1}\,\mathbf{H}_k^T + \mathbf{R}_k$$

Therefore $\mathbf{S}_k$ is the covariance of the measurement residual

$$\mathbf{S}_k = E\left\{\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T\right\}$$

Observe that $\mathbf{S}_k$ is a square, symmetric matrix, and so

$$\mathbf{S}_k^T = \mathbf{S}_k$$

Given $\hat{\mathbf{x}}_{k|k-1}$, $\mathbf{P}_{k|k-1}$, $\tilde{\mathbf{z}}_k$, and $\mathbf{S}_k$, our objective is to determine an optimal gain matrix, $\mathbf{K}_k$, such that the *a posteriori* state correction estimate is

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\tilde{\mathbf{z}}_k$$

with *a posteriori* state correction estimation error

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$$

and *a posteriori* state correction estimation error covariance

$$\mathbf{P}_k = E\left\{\mathbf{e}_k\mathbf{e}_k^T\right\}$$
$$= E\left\{\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right]\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right]^T\right\}$$

The gain, $\mathbf{K}_k$, is considered optimal if the diagonal elements of $\mathbf{P}_k$ are minimal, i.e., the *a posteriori* state estimation error variances are minimal.

Substituting the *a posteriori* state correction estimate expression for $\hat{\mathbf{x}}_k$ into $\mathbf{e}_k$ gives

$$
\begin{aligned}
\mathbf{e}_k &= \mathbf{x}_k - \hat{\mathbf{x}}_k \\
&= \mathbf{x}_k - \left(\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left[\mathbf{z}_k - \hat{\mathbf{z}}_k \right] \right) \\
&= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \left[\mathbf{z}_k - \hat{\mathbf{z}}_k \right] \\
&= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \mathbf{z}_k + \mathbf{K}_k \hat{\mathbf{z}}_k \\
&= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \left[\mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \right] + \mathbf{K}_k \left[\mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \right] \\
&= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \mathbf{H}_k \mathbf{x}_k - \mathbf{K}_k \mathbf{v}_k + \mathbf{K}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \\
&= \mathbf{x}_k - \mathbf{K}_k \mathbf{H}_k \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \mathbf{v}_k \\
&= \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{x}_k - \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \mathbf{v}_k \\
&= \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \right] - \mathbf{K}_k \mathbf{v}_k \\
&= \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{e}_{k|k-1} - \mathbf{K}_k \mathbf{v}_k
\end{aligned}
$$

Noting that

$$
E \left\{ \mathbf{e}_{k|k-1} \mathbf{v}_k^T \right\} = \mathbf{0}
$$

and using **Lemma 2**, the expansion of the *a posteriori* state correction estimate error covariance gives

$$
\begin{aligned}
\mathbf{P}_k &= E \left\{ \mathbf{e}_k \mathbf{e}_k^T \right\} \\
&= E \left\{ \left[\left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{e}_{k|k-1} - \mathbf{K}_k \mathbf{v}_k \right] \left[\left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{e}_{k|k-1} - \mathbf{K}_k \mathbf{v}_k \right]^T \right\} \\
&= \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{P}_{k|k-1} \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right]^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T
\end{aligned}
$$

and so the *a posteriori* state correction estimate error covariance is

$$
\mathbf{P}_k = \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{P}_{k|k-1} \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right]^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T
$$

This form of the *a posteriori* state correction error covariance is known as the Joseph form of the *a posteriori* covariance update (named after Peter D. Joseph). It is worth noting that it does not rely on $\mathbf{K}_k$ being optimal; it holds for any $\mathbf{K}_k$. Also, because each of the terms is a quadratic term of the form $\mathbf{ABA}^T$, the Joseph form is numerically suited for ill-conditioned systems.

Continuing our expansion and refactoring of $\mathbf{P}_k$ gives

$$
\begin{aligned}
\mathbf{P}_k &= \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1}\left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1}\left[\mathbf{I} - \mathbf{H}_k^T\mathbf{K}_k^T\right] + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \left[\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\right]\left[\mathbf{I} - \mathbf{H}_k^T\mathbf{K}_k^T\right] + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \left[\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\right] - \left[\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\right]\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \left[\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\right] - \left[\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T\right] + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{R}_k\,\mathbf{K}_k^T \\
&= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]\mathbf{K}_k^T
\end{aligned}
$$

Recall that

$$\mathbf{S}_k = \mathbf{H}_k\,\mathbf{P}_{k|k-1}\,\mathbf{H}_k^T + \mathbf{R}_k$$

Then

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T$$

What we wish to do is find the value of $\mathbf{K}_k$ that minimizes the diagonal elements of $\mathbf{P}_k$. In other words, we want the value of $\mathbf{K}_k$ that minimizes the state estimate error variances. To do this, we need to find the value of $\mathbf{K}_k$ that causes

$$\frac{d\left[\mathrm{tr}\left(\mathbf{P}_k\right)\right]}{d\,\mathbf{K}_k} = \mathbf{0}$$

We use the differential operation where, if s is a scalar value and $\mathbf{M}$ is a matrix value, then

$$\frac{d\,s}{d\,\mathbf{M}} \triangleq \begin{bmatrix} \frac{\partial s}{\partial m_{11}} & \frac{\partial s}{\partial m_{12}} & \frac{\partial s}{\partial m_{13}} & \cdots \\ \frac{\partial s}{\partial m_{21}} & \frac{\partial s}{\partial m_{22}} & \frac{\partial s}{\partial m_{23}} & \cdots \\ \frac{\partial s}{\partial m_{31}} & \frac{\partial s}{\partial m_{32}} & \frac{\partial s}{\partial m_{33}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

We can make use of the following differentiation formulas:

$$\frac{d\left[\text{tr}\left(\mathbf{AB}\right)\right]}{d\mathbf{A}} = \mathbf{B}^T, \quad (\mathbf{AB} \text{ must be square})$$

$$\frac{d\left[\text{tr}\left(\mathbf{ACA}^T\right)\right]}{d\mathbf{A}} = 2\mathbf{AC}, \quad (\mathbf{C} \text{ must be symmetric})$$

Note that $\text{tr}\left(\mathbf{B}^T\mathbf{A}^T\right) = \text{tr}\left(\mathbf{AB}\right)$, and so

$$\frac{d\left[\text{tr}\left(\mathbf{B}^T\mathbf{A}^T\right)\right]}{d\mathbf{A}} = \mathbf{B}^T, \quad (\mathbf{B}^T\mathbf{A}^T \text{ must be square})$$

Then

$$\begin{aligned}
\frac{d\left[\text{tr}(\mathbf{P}_k)\right]}{d\mathbf{K}_k} &= \frac{d\left[\text{tr}\left(\mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T\right)\right]}{d\,\mathbf{K}_k} \\
&= \frac{d\left[\text{tr}\left(\mathbf{P}_{k|k-1}\right)\right]}{d\,\mathbf{K}_k} - \frac{d\left[\text{tr}\left(\mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}\right)\right]}{d\,\mathbf{K}_k} - \frac{d\left[\text{tr}\left(\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T\right)\right]}{d\,\mathbf{K}_k} + \frac{d\left[\text{tr}\left(\mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T\right)\right]}{d\,\mathbf{K}_k} \\
&= \mathbf{0} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T - \mathbf{P}_{k|k-1}\mathbf{H}_k^T + 2\mathbf{K}_k\,\mathbf{S}_k \\
&= -2\mathbf{P}_{k|k-1}\mathbf{H}_k^T + 2\mathbf{K}_k\,\mathbf{S}_k
\end{aligned}$$

Setting

$$\frac{d\left[\text{tr}\left(\mathbf{P}_k\right)\right]}{d\mathbf{K}_k} = \mathbf{0}$$

gives

$$\begin{aligned}
\mathbf{0} &= -2\mathbf{P}_{k|k-1}\mathbf{H}_k^T + 2\mathbf{K}_k\mathbf{S}_k \\
\mathbf{K}_k\mathbf{S}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T \\
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}
\end{aligned}$$

and so the optimal gain, that is, the Kalman gain, is

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}$$

At this point, it is worth noting that we can also determine $\mathbf{K}_k$ using purely algebraic methods. Noting that $\mathbf{S}_k$ is a symmetric matrix, it is well-known that there exists a matrix, $\mathbf{M}$, such that

$$\mathbf{S}_k = \mathbf{M}\mathbf{M}^T$$

Then

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T$$

becomes

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{M}\mathbf{M}^T\,\mathbf{K}_k^T$$

Then, separating the non-$\mathbf{K}_k$ terms from the $\mathbf{K}_k$ terms gives

$$\mathbf{P}_k - \mathbf{P}_{k|k-1} = -\mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{M}\mathbf{M}^T\,\mathbf{K}_k^T$$

Note that the $-\mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T$ term is linear in $\mathbf{K}_k$ and the $\mathbf{K}_k\,\mathbf{M}\mathbf{M}^T\,\mathbf{K}_k^T$ term is quadratic in $\mathbf{K}_k$. If we expand a quadratic expression using the factor $\mathbf{A} - \mathbf{K}_k\mathbf{M}$, where $\mathbf{A}$ is yet to be determined, we have

$$\begin{aligned}\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]^T &= \left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\left[\mathbf{A}^T - \mathbf{M}^T\mathbf{K}_k^T\right] \\ &= \left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\mathbf{A}^T - \left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\mathbf{M}^T\mathbf{K}_k^T \\ &= \mathbf{A}\mathbf{A}^T - \mathbf{K}_k\mathbf{M}\mathbf{A}^T - \mathbf{A}\mathbf{M}^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{M}\mathbf{M}^T\mathbf{K}_k^T\end{aligned}$$

and so

$$\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]^T - \mathbf{A}\mathbf{A}^T = -\mathbf{K}_k\mathbf{M}\mathbf{A}^T - \mathbf{A}\mathbf{M}^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{M}\mathbf{M}^T\mathbf{K}_k^T$$

What we can now do is use this to complete the square of the right-hand side of the $\mathbf{P}_k - \mathbf{P}_{k|k-1}$ expression

$$\begin{aligned}\mathbf{P}_k - \mathbf{P}_{k|k-1} &= \left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]^T - \mathbf{A}\mathbf{A}^T \\ &= -\mathbf{K}_k\mathbf{M}\mathbf{A}^T - \mathbf{A}\mathbf{M}^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{M}\mathbf{M}^T\mathbf{K}_k^T\end{aligned}$$

Equating this expansion of $\mathbf{P}_k - \mathbf{P}_{k|k-1}$ with its previous expansion gives

$$\begin{aligned}-\mathbf{K}_k\mathbf{M}\mathbf{A}^T - \mathbf{A}\mathbf{M}^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{M}\mathbf{M}^T\mathbf{K}_k^T &= -\mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{M}\mathbf{M}^T\,\mathbf{K}_k^T \\ -\mathbf{K}_k\mathbf{M}\mathbf{A}^T - \mathbf{A}\mathbf{M}^T\mathbf{K}_k^T &= -\mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T \\ \mathbf{K}_k\mathbf{M}\mathbf{A}^T + \mathbf{A}\mathbf{M}^T\mathbf{K}_k^T &= \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T\end{aligned}$$

and so

$$\mathbf{A}\mathbf{M}^T = \mathbf{P}_{k|k-1}\mathbf{H}_k^T$$
$$\mathbf{A} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{M}^T\right)^{-1}$$

Now, recall that $\mathbf{P}_k - \mathbf{P}_{k|k-1}$ is

$$\mathbf{P}_k - \mathbf{P}_{k|k-1} = \left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]^T - \mathbf{A}\mathbf{A}^T$$

Then $\mathbf{P}_k$ is

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} + \left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]\left[\mathbf{A} - \mathbf{K}_k\mathbf{M}\right]^T - \mathbf{A}\mathbf{A}^T$$

Because the terms of this expression are square, symmetric matrices, and because $\mathbf{A}$ does not depend on $\mathbf{K}_k$, the only way to minimize the diagonal elements of $\mathbf{P}_k$ is to force the $\mathbf{A} - \mathbf{K}_k\,\mathbf{M}$ term to zero

$$\mathbf{A} - \mathbf{K}_k\,\mathbf{M} = \mathbf{0}$$

and so

$$\mathbf{A} = \mathbf{K}_k\,\mathbf{M}$$

Equating the two expressions for $\mathbf{A}$ then gives

$$\mathbf{K}_k\,\mathbf{M} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{M}^T\right)^{-1}$$
$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{M}^T\right)^{-1}(\mathbf{M})^{-1}$$
$$= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{M}\mathbf{M}^T\right)^{-1}$$
$$= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{S}_k\right)^{-1}$$

This, of course, agrees with the result we obtained from the differential calculus approach, although it took us significantly more steps to get here.

It should be noted that our optimization criteria of minimizing the diagonal elements of $\mathbf{P}_k$ constitutes a least-squares optimization criteria (the diagonal elements are the variances of the states). So $\mathbf{K}_k$ is said to be optimal in the least-squares sense. However, the Kalman filter is not a specialized least-squares estimator; it is its own formulation.

Now that we have an expression for $\mathbf{K}_k$, the *a posteriori* state correction estimate error covariance $\mathbf{P}_k$ becomes

$$
\begin{aligned}
\mathbf{P}_k &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T \\
&= \mathbf{P}_{k|k-1} - \left[\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\right]\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left[\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\right]^T \\
&\quad + \left[\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\right]\mathbf{S}_k\left[\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\right]^T \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&\quad + \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{S}_k\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&\quad + \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \mathbf{P}_{k|k-1} - 2\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1}
\end{aligned}
$$

and so the *a posteriori* state correction estimate error covariance is

$$\mathbf{P}_k = \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1}$$

Although this expression for $\mathbf{P}_k$ is typically the form presented in Kalman filtering literature, its direct implementation is discouraged due to numerical roundoff jeopardizing the covariance matrix symmetry, and its use is suitable mainly for academic formulations.

Note that, by remembering that $\mathbf{P}_{k|k-1}^T = \mathbf{P}_{k|k-1}$ and that $\mathbf{S}_k^T = \mathbf{S}_k$, the Kalman gain expression can be rearranged as follows

$$
\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\,\mathbf{H}_k^T\,\mathbf{S}_k^{-1} \\
\mathbf{K}_k\,\mathbf{S}_k &= \mathbf{P}_{k|k-1}\,\mathbf{H}_k^T\,\mathbf{S}_k^{-1}\,\mathbf{S}_k \\
\mathbf{K}_k\,\mathbf{S}_k &= \mathbf{P}_{k|k-1}\,\mathbf{H}_k^T \\
\left(\mathbf{K}_k\,\mathbf{S}_k\right)^T &= \left(\mathbf{P}_{k|k-1}\,\mathbf{H}_k^T\right)^T \\
\mathbf{S}_k\,\mathbf{K}_k^T &= \mathbf{H}_k\,\mathbf{P}_{k|k-1}
\end{aligned}
$$

We can then express $\mathbf{P}_k$ as

$$\begin{aligned}\mathbf{P}_k &= \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1} \\ &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{H}_k\,\mathbf{P}_{k|k-1} \\ &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T\end{aligned}$$

and so the *a posteriori* state correction estimate error covariance can also be expressed as

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T$$

This expression is quadratic in $\mathbf{K}_k$, and so, like the Joseph form, it is well-suited numerically for ill-conditioned systems. However, the difference between this form and the Joseph form is that this form is valid only when $\mathbf{K}_k$ is optimal according to the above derivations, whereas the Joseph form is valid for any $\mathbf{K}_k$. Also, because of the differencing operation, care must be taken so that $\mathbf{P}_k$ remains positive definite. Lastly, this expression does not rely on $\mathbf{H}_k$.

Given existing values for $\hat{\mathbf{x}}_{k-1}$ and $\mathbf{P}_{k-1}$, we now have established how to determine $\hat{\mathbf{x}}_{k|k-1}$, $\mathbf{P}_{k|k-1}$, $\hat{\mathbf{z}}_k$, $\tilde{\mathbf{z}}_k$, $\mathbf{S}_k$, $\mathbf{K}_k$, $\hat{\mathbf{x}}_k$, and $\mathbf{P}_k$, and so we can therefore specify the Kalman filter algorithm.

The Kalman filter update cycle for an unforced system (i.e., no control input) is

I. Projection (*a priori*) update:

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= \mathbf{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1} \\ \mathbf{P}_{k|k-1} &= \mathbf{\Phi}_{k|k-1}\,\mathbf{P}_{k-1}\,\mathbf{\Phi}_{k|k-1}^T + \mathbf{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\mathbf{\Gamma}_{k|k-1}^T\end{aligned}$$

II. Correction (*a posteriori*) update:

$$\begin{aligned}\hat{\mathbf{z}}_k &= \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \\ \tilde{\mathbf{z}}_k &= \mathbf{z}_k - \hat{\mathbf{z}}_k \\ \mathbf{S}_k &= \mathbf{H}_k\,\mathbf{P}_{k|k-1}\,\mathbf{H}_k^T + \mathbf{R}_k \\ \mathbf{K}_k &= \mathbf{P}_{k|k-1}\,\mathbf{H}_k^T\,\mathbf{S}_k^{-1} \\ \hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\,\tilde{\mathbf{z}}_k \\ \mathbf{P}_k &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T\end{aligned}$$

For a forced input system, the system dynamical and observation modeling equations are

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{\Phi}_{k+1|k}\mathbf{x}_k + \mathbf{u}_k + \mathbf{\Gamma}_{k+1|k}\mathbf{w}_k \\ \mathbf{z}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k\end{aligned}$$

where $\mathbf{u}_k$ is the $n \times 1$ control input sequence, and the only change to the Kalman filter update cycle is that the *a priori* state projection estimate becomes

$$\hat{\mathbf{x}}_{k|k-1} = \boldsymbol{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{u}_{k-1}$$

and so the Kalman filter update cycle for a forced input system is

I. Projection (*a priori*) update:

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \boldsymbol{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{u}_{k-1} \\
\mathbf{P}_{k|k-1} &= \boldsymbol{\Phi}_{k|k-1}\,\mathbf{P}_{k-1}\,\boldsymbol{\Phi}_{k|k-1}^{T} + \boldsymbol{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\boldsymbol{\Gamma}_{k|k-1}^{T}
\end{aligned}$$

II. Correction (*a posteriori*) update:

$$\begin{aligned}
\hat{\mathbf{z}}_k &= \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \\
\tilde{\mathbf{z}}_k &= \mathbf{z}_k - \hat{\mathbf{z}}_k \\
\mathbf{S}_k &= \mathbf{H}_k\,\mathbf{P}_{k|k-1}\,\mathbf{H}_k^{T} + \mathbf{R}_k \\
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\,\mathbf{H}_k^{T}\,\mathbf{S}_k^{-1} \\
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\,\tilde{\mathbf{z}}_k \\
\mathbf{P}_k &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^{T}
\end{aligned}$$

Before concluding this section, it is worth introducing two additional covariance matrices, particularly since they are relevant to the unscented Kalman filter formulation. These two covariances are the measurement propagation covariance, $\mathbf{P}_{\mathbf{zz},k}$, and the measurement-to-state transformation cross covariance, $\mathbf{P}_{\mathbf{xz},k}$.

Let $\mathbf{z}_{NF,k}$ be the noise-free measurement term (i.e., the $\mathbf{v}_k$ sequence is not included)

$$\mathbf{z}_{NF,k} = \mathbf{H}_k\mathbf{x}_k$$

The noise-free measurement residual is

$$\tilde{\mathbf{z}}_{NF,k} = \mathbf{z}_{NF,k} - \hat{\mathbf{z}}_k$$

Then the measurement propagation covariance is determined by

$$
\begin{aligned}
\mathbf{P}_{\mathbf{zz},k} &= E\left\{ \tilde{\mathbf{z}}_{NF,k}\ \tilde{\mathbf{z}}_{NF,k}^{T} \right\} \\
&= E\left\{ \left[\mathbf{z}_{NF,k} - \hat{\mathbf{z}}_{k}\right] \left[\mathbf{z}_{NF,k} - \hat{\mathbf{z}}_{k}\right]^{T} \right\} \\
&= E\left\{ \left[\mathbf{H}_{k}\mathbf{x}_{k} - \mathbf{H}_{k}\hat{\mathbf{x}}_{k|k-1}\right] \left[\mathbf{H}_{k}\mathbf{x}_{k} - \mathbf{H}_{k}\hat{\mathbf{x}}_{k|k-1}\right]^{T} \right\} \\
&= E\left\{ \left[\mathbf{H}_{k}\mathbf{e}_{k|k-1}\right] \left[\mathbf{H}_{k}\mathbf{e}_{k|k-1}\right]^{T} \right\} \\
&= E\left\{ \left[\mathbf{H}_{k}\mathbf{e}_{k|k-1}\right] \left[\mathbf{e}_{k|k-1}^{T}\mathbf{H}_{k}^{T}\right] \right\} \\
&= \mathbf{H}_{k} E\left\{ \mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^{T} \right\} \mathbf{H}_{k}^{T} \\
&= \mathbf{H}_{k}\mathbf{P}_{k|k-1}\mathbf{H}_{k}^{T}
\end{aligned}
$$

and so the measurement propagation covariance is

$$\mathbf{P}_{\mathbf{zz},k} = \mathbf{H}_{k}\mathbf{P}_{k|k-1}\mathbf{H}_{k}^{T}$$

As can be seen, then, the covariance of the measurement residual can be expressed as

$$\mathbf{S}_{k} = \mathbf{P}_{\mathbf{zz},k} + \mathbf{R}_{k}$$

In other words, $\mathbf{S}_{k}$ is simply the direct addition of the measurement propagation covariance and the measurement error covariance.

The measurement-to-state transformation cross covariance is determined by

$$
\begin{aligned}
\mathbf{P}_{\mathbf{xz},k} &= E\left\{\mathbf{e}_{k|k-1}\tilde{\mathbf{z}}_k^T\right\} \\
&= E\left\{\mathbf{e}_{k|k-1}\left[\mathbf{z}_k - \hat{\mathbf{z}}_k\right]^T\right\} \\
&= E\left\{\mathbf{e}_{k|k-1}\left[\mathbf{H}_k\mathbf{e}_{k|k-1} + \mathbf{v}_k\right]^T\right\} \\
&= E\left\{\mathbf{e}_{k|k-1}\left[\mathbf{e}_{k|k-1}^T\mathbf{H}_k^T + \mathbf{v}_k^T\right]\right\} \\
&= E\left\{\mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^T\mathbf{H}_k^T + \mathbf{e}_{k|k-1}\mathbf{v}_k^T\right\} \\
&= E\left\{\mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^T\mathbf{H}_k^T\right\} + E\left\{\mathbf{e}_{k|k-1}\mathbf{v}_k^T\right\} \\
&= E\left\{\mathbf{e}_{k|k-1}\mathbf{e}_{k|k-1}^T\right\}\mathbf{H}_k^T + \mathbf{0} \\
&= \mathbf{P}_{k|k-1}\,\mathbf{H}_k^T
\end{aligned}
$$

and so the measurement-to-state transformation cross covariance is

$$\mathbf{P}_{\mathbf{xz},k} = \mathbf{P}_{k|k-1}\,\mathbf{H}_k^T$$

Note that, whereas $\mathbf{P}_{\mathbf{zz},k}$ is square (and symmetric), $\mathbf{P}_{\mathbf{xz},k}$ is not square.

As can be seen, then, the Kalman filter gain can be expressed as

$$\mathbf{K}_k = \mathbf{P}_{\mathbf{xz},k}\,\mathbf{S}_k^{-1}$$

In other words, the Kalman filter gain is simply the measurement-to-state transformation cross covariance “divided by” the measurement residual covariance. Another viewpoint is that the filter gain is simply the cross covariance weighted by the inverse of the residual covariance, which is essentially a variance normalization operation.

Observe that, for the linear Kalman filter, the measurement estimate, $\hat{\mathbf{z}}_k$, measurement residual covariance, $\mathbf{S}_k$, and the filter correction gain, $\mathbf{K}_k$, are the only calculations that rely on $\mathbf{H}_k$. As we’ve just discovered, $\mathbf{P}_{\mathbf{zz},k}$ and $\mathbf{P}_{\mathbf{xz},k}$ are the components of $\mathbf{S}_k$, and $\mathbf{K}_k$, respectively, that are reliant on $\mathbf{H}_k$. As will be seen shortly, the unscented Kalman filter determines $\hat{\mathbf{z}}_k$, $\mathbf{P}_{\mathbf{zz},k}$, and $\mathbf{P}_{\mathbf{xz},k}$ without the necessity of $\mathbf{H}_k$.

4. The Information Form of the Kalman Filter

The inverse of a covariance matrix is known as an information matrix. Whereas a covariance matrix is an indicator of dispersion, an information matrix is an indicator of precision. The smaller the covariance is, the higher the precision is, and vice versa. An information matrix of $\mathbf{0}$ value implies infinite covariance, meaning that no statistical information can be inferred. The Kalman filter can be reformulated to use the inverse of $\mathbf{P}$, although this form of the Kalman filter is rarely used for implementation. However, it does have useful theoretical purposes.

Recall the Kalman filter *a posteriori* covariance expression

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \, \mathbf{S}_k \, \mathbf{K}_k^T$$

Remembering that $\mathbf{P}$ and $\mathbf{S}$ are symmetric, inserting the expressions for $\mathbf{K}_k$ and $\mathbf{S}_k$ gives

$$\begin{aligned}
\mathbf{P}_k &= \mathbf{P}_{k|k-1} - \left(\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\right)\mathbf{S}_k\left(\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\right)^T \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{S}_k\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}
\end{aligned}$$

We now need to introduce some useful matrix lemmas.

Lemma 3:

$$\mathbf{P}\left(\mathbf{I} + \mathbf{QP}\right) = \left(\mathbf{I} + \mathbf{PQ}\right)\mathbf{P}$$

Proof:

Note that $\mathbf{P} + \mathbf{PQP}$ can be expanded as

$$\mathbf{P} + \mathbf{PQP} = \mathbf{P}\left(\mathbf{I} + \mathbf{QP}\right)$$

But $\mathbf{P} + \mathbf{PQP}$ can also be expanded as

$$\mathbf{P} + \mathbf{PQP} = \left(\mathbf{I} + \mathbf{PQ}\right)\mathbf{P}$$

Therefore

$$\mathbf{P}\left(\mathbf{I} + \mathbf{QP}\right) = \left(\mathbf{I} + \mathbf{PQ}\right)\mathbf{P}$$

Lemma 4:

$$\mathbf{P}\left(\mathbf{I}+\mathbf{QP}\right)^{-1}=\left(\mathbf{I}+\mathbf{PQ}\right)^{-1}\mathbf{P}$$

Proof:

From **Lemma 3**, we see that

$$\begin{aligned}\mathbf{P}\left(\mathbf{I}+\mathbf{QP}\right)&=\left(\mathbf{I}+\mathbf{PQ}\right)\mathbf{P}\\ \left[\mathbf{P}\left(\mathbf{I}+\mathbf{QP}\right)\right]^{-1}&=\left[\left(\mathbf{I}+\mathbf{PQ}\right)\mathbf{P}\right]^{-1}\\ \left(\mathbf{I}+\mathbf{QP}\right)^{-1}\mathbf{P}^{-1}&=\mathbf{P}^{-1}\left(\mathbf{I}+\mathbf{PQ}\right)^{-1}\\ \mathbf{P}\left(\mathbf{I}+\mathbf{QP}\right)^{-1}\mathbf{P}^{-1}&=\left(\mathbf{I}+\mathbf{PQ}\right)^{-1}\\ \mathbf{P}\left(\mathbf{I}+\mathbf{QP}\right)^{-1}&=\left(\mathbf{I}+\mathbf{PQ}\right)^{-1}\mathbf{P}\end{aligned}$$

Lemma 5:

$$(\mathbf{I}+\mathbf{P})^{-1}=\mathbf{I}-(\mathbf{I}+\mathbf{P})^{-1}\mathbf{P}$$

Proof:

Noting that multiplying by $\mathbf{I}+\mathbf{P}-\mathbf{P}$ does not alter the value of $(\mathbf{I}+\mathbf{P})^{-1}$, we see that

$$\begin{aligned}(\mathbf{I}+\mathbf{P})^{-1}&=(\mathbf{I}+\mathbf{P})^{-1}(\mathbf{I}+\mathbf{P}-\mathbf{P})\\ &=(\mathbf{I}+\mathbf{P})^{-1}(\mathbf{I}+\mathbf{P})-(\mathbf{I}+\mathbf{P})^{-1}(\mathbf{P})\\ &=\mathbf{I}-(\mathbf{I}+\mathbf{P})^{-1}\mathbf{P}\end{aligned}$$

Lemma 6:

$$\left(\mathbf{A}+\mathbf{BCD}\right)^{-1}=\mathbf{A}^{-1}-\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right)^{-1}\mathbf{A}^{-1}\mathbf{BCDA}^{-1}$$

Proof:

Using **Lemma 5**, we have that

$$
\begin{aligned}
\left(\mathbf{A}+\mathbf{BCD}\right)^{-1} &= \left(\mathbf{A}+\mathbf{AA}^{-1}\mathbf{BCD}\right)^{-1} \\
&= \left(\mathbf{A}\left[\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right]\right)^{-1} \\
&= \left[\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right]^{-1}\mathbf{A}^{-1} \\
&= \left[\mathbf{I}-\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right)^{-1}\mathbf{A}^{-1}\mathbf{BCD}\right]\mathbf{A}^{-1} \\
&= \mathbf{A}^{-1}-\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right)^{-1}\mathbf{A}^{-1}\mathbf{BCDA}^{-1}
\end{aligned}
$$

Lemma 7:

$$
\left(\mathbf{A}+\mathbf{BCD}\right)^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{I}+\mathbf{CDA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CDA}^{-1}
$$

Proof:

By repeatedly applying **Lemma 4** to **Lemma 6**, we have the following equivalence relations

$$
\begin{aligned}
\left(\mathbf{A}+\mathbf{BCD}\right)^{-1} &= \mathbf{A}^{-1}-\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right)^{-1}\mathbf{A}^{-1}\mathbf{BCDA}^{-1} && (1) \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\left(\mathbf{I}+\mathbf{BCDA}^{-1}\right)^{-1}\mathbf{BCDA}^{-1} && (2) \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{I}+\mathbf{CDA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CDA}^{-1} && (3) \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{BC}\left(\mathbf{I}+\mathbf{DA}^{-1}\mathbf{BC}\right)^{-1}\mathbf{DA}^{-1} && (4) \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{BCD}\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right)^{-1}\mathbf{A}^{-1} && (5) \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{BCDA}^{-1}\left(\mathbf{I}+\mathbf{BCDA}^{-1}\right)^{-1} && (6)
\end{aligned}
$$

The third equivalence relation establishes the lemma.

Lemma 8:

If $\mathbf{C}^{-1}$ exists, then

$$
\left(\mathbf{A}+\mathbf{BCD}\right)^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}
$$

Proof:

If $\mathbf{C}^{-1}$ exists, then **Lemma 7** becomes

$$
\begin{aligned}
\left(\mathbf{A}+\mathbf{BCD}\right)^{-1} &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{I}+\mathbf{CDA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CDA}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{CC}^{-1}+\mathbf{CDA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CDA}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}\left[\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right]\right)^{-1}\mathbf{CDA}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}^{-1}\mathbf{CDA}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}
\end{aligned}
$$

Lemma 9:

If $\mathbf{C}^{-1}$ exists, then

$$\left(\mathbf{A}-\mathbf{BCD}\right)^{-1}=\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}-\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$$

Proof:

Let $\mathbf{F}=-\mathbf{B}$. If $\mathbf{C}^{-1}$ exists, then, by **Lemma 8** we have

$$\left(\mathbf{A}+\mathbf{FCD}\right)^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{F}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{F}\right)^{-1}\mathbf{DA}^{-1}$$
$$\left(\mathbf{A}-\mathbf{BCD}\right)^{-1}=\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}-\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$$

Lemma 10:

If $\mathbf{C}^{-1}$ exists, then

$$\left(\mathbf{A}-\mathbf{B}^{T}\mathbf{CB}\right)^{-1}=\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}^{T}\left(\mathbf{C}^{-1}-\mathbf{BA}^{-1}\mathbf{B}^{T}\right)^{-1}\mathbf{BA}^{-1}$$

Proof:

Let $\mathbf{F}=\mathbf{B}^{T}$ and $\mathbf{G}=\mathbf{B}$. If $\mathbf{C}^{-1}$ exists, then, by **Lemma 9** we have

$$\left(\mathbf{A}-\mathbf{FCG}\right)^{-1}=\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{F}\left(\mathbf{C}^{-1}-\mathbf{GA}^{-1}\mathbf{F}\right)^{-1}\mathbf{GA}^{-1}$$
$$\left(\mathbf{A}-\mathbf{B}^{T}\mathbf{CB}\right)^{-1}=\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}^{T}\left(\mathbf{C}^{-1}-\mathbf{BA}^{-1}\mathbf{B}^{T}\right)^{-1}\mathbf{BA}^{-1}$$

Now that we've established **Lemma 10**, recall our expanded *a posteriori* covariance expression

$$\mathbf{P}_{k}=\mathbf{P}_{k|k-1}-\mathbf{P}_{k|k-1}\mathbf{H}_{k}^{T}\left[\mathbf{H}_{k}\mathbf{P}_{k|k-1}\mathbf{H}_{k}^{T}+\mathbf{R}_{k}\right]^{-1}\mathbf{H}_{k}\mathbf{P}_{k|k-1}$$

Define

$$\mathbf{A} = \mathbf{P}_{k|k-1}$$
$$\mathbf{B} = \mathbf{H}_k\mathbf{P}_{k|k-1}$$
$$\mathbf{C} = \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1}$$

so that

$$\mathbf{B}^T = \mathbf{P}_{k|k-1}\mathbf{H}_k^T$$
$$\mathbf{C}^{-1} = \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]$$

which allows us to express $\mathbf{P}_k$ as

$$\mathbf{P}_k = \mathbf{A} - \mathbf{B}^T\mathbf{C}\mathbf{B}$$

Then by using **Lemma 10** we have

$$\begin{aligned}
\mathbf{P}_k^{-1} &= \left(\mathbf{A} - \mathbf{B}^T\mathbf{C}\mathbf{B}\right)^{-1} \\
&= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}^T\left(\mathbf{C}^{-1} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{A}^{-1} \\
&= \mathbf{P}_{k|k-1}^{-1} + \mathbf{P}_{k|k-1}^{-1}\mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right] - \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{P}_{k|k-1}^{-1}\mathbf{P}_{k|k-1}\mathbf{H}_k^T\right)^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{P}_{k|k-1}^{-1} \\
&= \mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k - \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\right)^{-1}\mathbf{H}_k \\
&= \mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k
\end{aligned}$$

and so the *a posteriori* correction update of the Kalman filter information matrix, $\mathbf{P}_k^{-1}$, is

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k$$

Next, recall the expression of the Kalman filter gain with $\mathbf{S}_k$ expanded

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1} \\
&= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1}
\end{aligned}$$

Strategically inserting $\mathbf{R}_k^{-1}\mathbf{R}_k$ does not alter the gain, and so

$$
\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1} \\
&= \mathbf{P}_{k|k-1}\mathbf{H}_k^T \left(\mathbf{R}_k^{-1}\mathbf{R}_k\right) \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1} \\
&= \mathbf{P}_{k|k-1}\mathbf{H}_k^T \left(\mathbf{R}_k^{-1}\right) \left[\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right)\mathbf{R}_k^{-1}\right]^{-1} \\
&= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1}
\end{aligned}
$$

Likewise, strategically inserting $\mathbf{P}_k\mathbf{P}_k^{-1}$ does not alter the gain, and so

$$
\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \left(\mathbf{P}_k\mathbf{P}_k^{-1}\right)\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \mathbf{P}_k\left(\mathbf{P}_k^{-1}\right)\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \mathbf{P}_k\left[\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\right]\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \mathbf{P}_k\left[\mathbf{I} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}\right]\mathbf{H}_k^T\mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1}
\end{aligned}
$$

Using **Lemma 3**, we see that the $\left[\mathbf{I} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}\right]\mathbf{H}_k^T\mathbf{R}_k^{-1}$ term becomes

$$
\left[\mathbf{I} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}\right]\mathbf{H}_k^T\mathbf{R}_k^{-1} = \mathbf{H}_k^T\mathbf{R}_k^{-1}\left[\mathbf{I} + \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1}\right]
$$

and so

$$
\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_k\left(\left[\mathbf{I} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}\right]\mathbf{H}_k^T\mathbf{R}_k^{-1}\right) \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \mathbf{P}_k\left(\mathbf{H}_k^T\mathbf{R}_k^{-1}\left[\mathbf{I} + \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1}\right]\right) \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \mathbf{P}_k\mathbf{H}_k^T\mathbf{R}_k^{-1}\left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right] \left[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{R}_k^{-1} + \mathbf{I}\right]^{-1} \\
&= \mathbf{P}_k\mathbf{H}_k^T\mathbf{R}_k^{-1}
\end{aligned}
$$

and so the Kalman gain can be expressed as

$$
\mathbf{K}_k = \mathbf{P}_k\mathbf{H}_k^T\mathbf{R}_k^{-1}
$$

It should be noted that this form of the Kalman gain relies on the *a posteriori* covariance, $\mathbf{P}_k$. In comparison, the other form of the Kalman gain, $\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}$, relies on the *a priori* covariance, $\mathbf{P}_{k|k-1}$. Hence, the use of the *a posteriori* form of the Kalman gain requires that the *a posteriori* state update must be done after the *a posteriori* state covariance update and the associated gain evaluation.

The information form of the Kalman filter update cycle for a forced input system is

I. Projection (*a priori*) update:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{u}_{k-1}$$
$$\mathbf{P}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\mathbf{P}_{k-1}\mathbf{\Phi}_{k|k-1}^T + \mathbf{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\mathbf{\Gamma}_{k|k-1}^T$$

II. Correction (*a posteriori*) update:

$$\hat{\mathbf{z}}_k = \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$$
$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k$$
$$\mathbf{P}_k^{-1} = \mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k$$
$$\mathbf{K}_k = \mathbf{P}_k\mathbf{H}_k^T\mathbf{R}_k^{-1}$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\,\tilde{\mathbf{z}}_k$$

Note that this requires performing an inversion of $\mathbf{P}_{k|k-1}$ and another inversion of $\mathbf{P}_k^{-1}$. Also, instead of needing to form and compute $\mathbf{S}_k^{-1}$, we compute $\mathbf{R}_k^{-1}$. While the size of $\mathbf{R}_k$ and $\mathbf{S}_k$ are both $m \times m$, typically $\mathbf{R}_k$ is static across successive measurement updates and so $\mathbf{R}_k^{-1}$ needs to computed only once. In addition, if $\mathbf{R}_k$ is diagonal, evaluation of $\mathbf{R}_k^{-1}$ is trivial.

In general, this form of the Kalman filter is not implemented unless the equations can be coded such that they are computationally feasible and can numerically preserve the structure of $\mathbf{P}_k$.

5. Relationship to Least-Squares Estimation

As stated previously, the Kalman filter is not a specialized least-squares estimator; it is its own formulation. However, there is an interesting relationship between the Kalman filter and least-squares estimation. Recall that the least-squares estimation problem relies only on the measurement relation and does not involve the progression of time. The system description is static, and so the unforced system process model is

$$\mathbf{x}_{k+1} = \mathbf{I}\,\mathbf{x}_k + \mathbf{0}$$

Since the progression of time is not applicable, we assume that all measurements occur simultaneously at $k = 1$, and so the measurement equation is an overdetermined relationship of the form

$$\mathbf{z}_1 = \mathbf{H}_1\mathbf{x}_1 + \mathbf{v}_1$$

where $E\left\{\mathbf{v}_1\mathbf{v}_1^T\right\} = \mathbf{R}_1$. In addition, we have no *a priori* knowledge of $\mathbf{x}$ and so $\hat{\mathbf{x}}_0 = \mathbf{0}$, and the associated covariance is $\mathbf{P}_0 = \infty$. Therefore, using the information form of the Kalman filter, the state error covariance becomes

$$\begin{aligned}\mathbf{P}_1^{-1} &= \mathbf{P}_0^{-1} + \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 \\ &= (\infty)^{-1} + \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 \\ &= \mathbf{0} + \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 \\ &= \mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1\end{aligned}$$

and the Kalman gain becomes

$$\begin{aligned}\mathbf{K}_1 &= \mathbf{P}_1\mathbf{H}_1^T\mathbf{R}_1^{-1} \\ &= \left(\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1\right)^{-1}\mathbf{H}_1^T\mathbf{R}_1^{-1}\end{aligned}$$

and the Kalman *a posteriori* update of the state estimate becomes a least-squares solution with the optimal weight $\mathbf{W} = \mathbf{R}_1^{-1}$:

$$\begin{aligned}\hat{\mathbf{x}}_1 &= \hat{\mathbf{x}}_0 + \mathbf{K}_1\left[\mathbf{z}_1 - \hat{\mathbf{x}}_0\right] \\ &= \mathbf{0} + \mathbf{K}_1\left[\mathbf{z}_1 - \mathbf{0}\right] \\ &= \mathbf{K}_1\mathbf{z}_1 \\ &= \left(\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1\right)^{-1}\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{z}_1\end{aligned}$$

Hence, when applied to a static system description, the Kalman filter reduces to a least-squares solution. So I've got that going for me, which is nice.

6. Quadratic Factors

One of the critical computation objectives is the preservation of the symmetric and positive definite characteristics of the covariance matrices. In particular, there are three matrices of interest: $\mathbf{P}_{k|k-1}$, $\mathbf{S}_k$, and $\mathbf{P}_k$. Recalling the expressions for each of these, we have

$$\mathbf{P}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\,\mathbf{P}_{k-1}\,\mathbf{\Phi}_{k|k-1}^T + \mathbf{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\mathbf{\Gamma}_{k|k-1}^T$$
$$\mathbf{S}_k = \mathbf{H}_k\,\mathbf{P}_{k|k-1}\,\mathbf{H}_k^T + \mathbf{R}_k$$
$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T$$

Observe that each of these contains one or more quadratic terms of the form $\mathbf{C} = \mathbf{ABA}^T$, where $\mathbf{B}$ is a square, symmetric matrix. Hence, $\mathbf{C}$ is also a square, symmetric matrix. Let $\mathbf{A}$ be of dimension $m \times n$ and $\mathbf{B}$ be of dimension $n \times n$. Then $\mathbf{C}$ will be of dimension $m \times m$.

An algorithm that preserves the symmetry while computing $\mathbf{C} = \mathbf{ABA}^T$ is as follows:

```
FOR i = 1 TO m DO
    σ = 0.0
    FOR j = 1 TO n DO
        FOR k = 1 TO n DO
            σ = σ + A_ij * A_ik * B_jk
        END FOR
    END FOR
    C_ii = σ
    FOR l = i + 1 TO m DO
        σ = 0.0
        FOR j = 1 TO n DO
            FOR k = 1 TO n DO
                σ = σ + A_ij * A_lk * B_jk
            END FOR
        END FOR
        C_il = σ
        C_li = σ
    END FOR
END FOR
```

7. Uncorrelated Measurements

If m is the dimension of $\mathbf{z}_k$, then the size of $\mathbf{S}_k$ is $m \times m$. When the number of measurements is large, e.g., the pseudorange measurements from a 12-channel GPS receiver, the evaluation of $\mathbf{S}_k^{-1}$ is computationally expensive and numerically challenging in maintaining its symmetric and positive definite characteristics.

If the measurements $\mathbf{z}_k$ are uncorrelated, then $\mathbf{R}_k$ is diagonal. In this case, the correction operations can be structured as a series of sequential scalar operations.

Let us denote the element-wise components of $\mathbf{z}_k$, $\mathbf{H}_k$, and $\mathbf{R}_k$ as

$$\mathbf{z}_k = \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(m)} \end{bmatrix}_k, \qquad \mathbf{H}_k = \begin{bmatrix} \mathbf{h}^{(1)} \\ \mathbf{h}^{(2)} \\ \vdots \\ \mathbf{h}^{(m)} \end{bmatrix}_k, \qquad \mathbf{R}_k = \begin{bmatrix} r^{(1)} & 0 & \cdots & 0 \\ 0 & r^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r^{(m)} \end{bmatrix}_k$$

where m is the measurement vector dimension. Observe that $\mathbf{h}^{(i)}$ is a $1 \times n$ row vector

$$\mathbf{h}^{(i)} = \begin{bmatrix} h_{i1} & h_{i2} & \cdots & h_{in} \end{bmatrix}$$

where n is the state vector dimension. Then the sequential correction algorithm is

$$\begin{aligned}
&\mathbf{x}^{(0)} = \hat{\mathbf{x}}_{k|k-1} \\
&\mathbf{P}^{(0)} = \mathbf{P}_{k|k-1} \\
&\text{FOR } i = 1 \text{ TO } m \text{ DO} \\
&\quad \tilde{z}^{(i)} = z^{(i)} - \mathbf{h}^{(i)}\mathbf{x}^{(i-1)} \\
&\quad s^{(i)} = \mathbf{h}^{(i)}\mathbf{P}^{(i-1)}\mathbf{h}^{(i)T} + r^{(i)} \\
&\quad \mathbf{k}^{(i)} = \mathbf{P}^{(i-1)}\mathbf{h}^{(i)T}\left(\frac{1}{s^{(i)}}\right) \\
&\quad \mathbf{x}^{(i)} = \mathbf{x}^{(i-1)} + \mathbf{k}^{(i)}\tilde{z}^{(i)} \\
&\quad \mathbf{P}^{(i)} = \mathbf{P}^{(i-1)} - \mathbf{k}^{(i)}s^{(i)}\mathbf{k}^{(i)T} \\
&\text{END FOR} \\
&\hat{\mathbf{x}}_k = \mathbf{x}^{(m)} \\
&\mathbf{P}_k = \mathbf{P}^{(m)}
\end{aligned}$$

The structure of the sequential correction algorithm is such that each i^{th} step can be treated as a "mini-update" that uses the results of the preceding step as its *a priori* input.

If $\mathbf{R}_k$ is not diagonal, i.e., the measurements are correlated, then the measurement relation can be transformed to a new uncorrelated form using U-D factorization. It is well-known that any $n \times n$ symmetric matrix, $\mathbf{M}$, can be factorized into a unit upper triangular matrix, $\mathbf{U}$, and a diagonal matrix, $\mathbf{D}$, such that

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{U}^T$$

where $\mathbf{U}$ and $\mathbf{D}$ are of the forms

$$\mathbf{U} = \begin{bmatrix} 1 & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & 1 & u_{23} & \cdots & u_{2n} \\ 0 & 0 & 1 & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} d_{11} & 0 & 0 & \cdots & 0 \\ 0 & d_{22} & 0 & \cdots & 0 \\ 0 & 0 & d_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_{nn} \end{bmatrix}$$

There are two useful properties to observe about unit triangular matrices:

- The determinant of a unit triangular matrix is 1. Therefore, $\mathbf{U}^{-1}$ exists.
- The inverse of a unit triangular matrix is also a unit triangular matrix, and it has the same structure as the non-inverted matrix. So $\mathbf{U}^{-1}$ is also a unit upper triangular matrix.

Now, let us suppose that our observation equation

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

has correlated measurement errors and, therefore, $\mathbf{R}_k$ is not diagonal. We then use the factorization:

$$\mathbf{R}_k = \mathbf{U}_R\mathbf{D}_R\mathbf{U}_R^T$$

Because $\mathbf{U}_R$ is a triangular matrix, we can use back-substitution methods for solving linear equations to transform $\mathbf{z}_k$ into a new value, $\acute{\mathbf{z}}_k$:

$$\mathbf{U}_R\acute{\mathbf{z}}_k = \mathbf{z}_k$$

Recall that we know that $\mathbf{U}_R^{-1}$ exists. Then

$$
\begin{aligned}
\acute{\mathbf{z}}_k &= \mathbf{U}_R^{-1}\mathbf{z}_k \\
&= \mathbf{U}_R^{-1}\left[\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k\right] \\
&= \mathbf{U}_R^{-1}\mathbf{H}_k\mathbf{x}_k + \mathbf{U}_R^{-1}\mathbf{v}_k \\
&= \acute{\mathbf{H}}_k\mathbf{x}_k + \acute{\mathbf{v}}_k
\end{aligned}
$$

The covariance of the transformed noise sequence $\acute{\mathbf{v}}_k$ is

$$
\begin{aligned}
\acute{\mathbf{R}}_k &= E\left\{\acute{\mathbf{v}}_k\acute{\mathbf{v}}_k^T\right\} \\
&= E\left\{\left[\mathbf{U}_R^{-1}\mathbf{v}_k\right]\left[\mathbf{U}_R^{-1}\mathbf{v}_k\right]^T\right\} \\
&= E\left\{\left[\mathbf{U}_R^{-1}\mathbf{v}_k\right]\left[\mathbf{v}_k^T\mathbf{U}_R^{-T}\right]\right\} \\
&= \mathbf{U}_R^{-1}E\left\{\mathbf{v}_k\mathbf{v}_k^T\right\}\mathbf{U}_R^{-T} \\
&= \mathbf{U}_R^{-1}\mathbf{R}_k\mathbf{U}_R^{-T} \\
&= \mathbf{U}_R^{-1}\left[\mathbf{U}_R\mathbf{D}_R\mathbf{U}_R^T\right]\mathbf{U}_R^{-T} \\
&= \mathbf{D}_R
\end{aligned}
$$

So $\acute{\mathbf{R}}_k$ is indeed diagonal. Also, observe that we do not need to compute $\mathbf{U}_R^{-1}$ in order to determine $\acute{\mathbf{z}}_k$ and $\acute{\mathbf{H}}_k$. Instead, we simply use back-substitution to solve

$$\mathbf{U}_R\acute{\mathbf{z}}_k = \mathbf{z}_k$$

and

$$\mathbf{U}_R\acute{\mathbf{H}}_k = \mathbf{H}_k$$

The algorithm that factorizes $\mathbf{M}$ into its $\mathbf{U}$ and $\mathbf{D}$ factors is known as the modified Cholesky decomposition algorithm (named after André-Louis Cholesky). The algorithm procedure is specified below:

FOR $j = n$ TO 1 BY -1 DO
    FOR $i = j$ TO 1 BY -1 DO
        $\sigma = \mathbf{M}_{ij}$
        FOR $k = j + 1$ TO n DO
            $\sigma = \sigma - \mathbf{U}_{ik} * \mathbf{D}_{kk} * \mathbf{U}_{jk}$
        END FOR
        IF $i = j$ THEN
            $\mathbf{D}_{jj} = \sigma$
            $\mathbf{U}_{jj} = 1.0$
        ELSE
            $\mathbf{U}_{ij} = \sigma / \mathbf{D}_{jj}$
            $\mathbf{U}_{ji} = 0.0$
        END IF
    END FOR
END FOR

8. The U-D Factorized Kalman Filter

As should be apparent, preservation of the integrity of the Kalman filter projection and correction computations is essential, especially when it comes to the **P** matrix. While using a quadratic matrix computation algorithm is clearly an appropriate implementation strategy, there are times when that is not enough. For instance, if the numerical range of the numbers challenges the precision range of double-precision floating-point storage, the behavior of the implemented filter can still be compromised.

One strategy of addressing the numerical range restrictions and floating-point roundoff errors is to maintain a matrix square root of the **P** matrix rather than **P** itself. Basically this compands the dynamic range of **P** into a more manageable range (the square root of very large numbers become smaller numbers, and the square root of very small numbers become larger numbers). It also guarantees that **P** remains symmetric and positive definite.

In practice, instead of using a matrix square root for **P**, a factorized form of **P** is used. Factorized algorithms provide better numerical stability, and they avoid the need to compute square root values. One of the more popular approaches is to use U-D factorization.

We begin the U-D Kalman filter algorithm description by identifying the items to be maintained:

$\hat{\mathbf{x}}_k$ state estimate at time event k
$\mathbf{U}_{P,k}$ U factor of $\mathbf{P}_k = \mathbf{U}_{P,k}\mathbf{D}_{P,k}\mathbf{U}_{P,k}^T$ at time event k
$\mathbf{D}_{P,k}$ D factor of $\mathbf{P}_k = \mathbf{U}_{P,k}\mathbf{D}_{P,k}\mathbf{U}_{P,k}^T$ at time event k

The algorithm relies on two subalgorithms: *UDProject* and *UDCorrect*. Both of these subalgorithms are specified below. In addition, the U-D Kalman filter algorithm relies on $\mathbf{R}_k$ being diagonal, so the above diagonalization procedure should be used if the measurements are correlated.

I. Projection (*a priori*) update:

Compute $\hat{\mathbf{x}}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\hat{\mathbf{x}}_{k-1}$
Factor $\mathbf{Q}_{k-1}$ into $\mathbf{U}_{Q,k-1}$ and $\mathbf{D}_{Q,k-1}$ such that $\mathbf{Q}_{k-1} = \mathbf{U}_{Q,k-1}\mathbf{D}_{Q,k-1}\mathbf{U}_{Q,k-1}^T$
(this can be done just once if $\mathbf{\Gamma}$ and $\mathbf{Q}$ are static)
Form $\mathbf{F}_k = \mathbf{\Phi}_{k|k-1}\mathbf{U}_{P,k-1}$
Form $\mathbf{G}_k = \mathbf{\Gamma}_{k|k-1}\mathbf{U}_{Q,k-1}$
Use *UDProject* to propagate $\mathbf{U}_{P,k-1}$ and $\mathbf{D}_{P,k-1}$ to $\mathbf{U}_{P,k|k-1}$ and $\mathbf{D}_{P,k|k-1}$

II. Correction (*a posteriori*) update:

$$\mathbf{x}^{(0)} = \hat{\mathbf{x}}_{k|k-1}$$
$$\mathbf{U}_P^{(0)} = \mathbf{U}_{P,k|k-1}$$
$$\mathbf{D}_P^{(0)} = \mathbf{D}_{P,k|k-1}$$

FOR $i = 1$ TO m DO

Form $\delta = z^{(i)} - \mathbf{h}^{(i)}\mathbf{x}^{(i-1)}$
Use *UDCorrect* to step update $\mathbf{x}^{(i-1)}$, $\mathbf{U}_P^{(i-1)}$, and $\mathbf{D}_P^{(i-1)}$ to $\mathbf{x}^{(i)}$, $\mathbf{U}_P^{(i)}$, and $\mathbf{D}_P^{(i)}$

END FOR

$$\hat{\mathbf{x}}_k = \mathbf{x}^{(m)}$$
$$\mathbf{U}_{P,k} = \mathbf{P}_P^{(m)}$$
$$\mathbf{D}_{P,k} = \mathbf{D}_P^{(m)}$$

The *UDProject* subalgorithm is based on Thornton's Modified Weighted Gram-Schmidt (MWGS) Orthogonalization procedure (named after Catherine Thornton, an engineer at NASA JPL). Recall the Kalman covariance projection update:

$$\mathbf{P}_{k|k-1} = \mathbf{\Phi}_{k|k-1}\mathbf{P}_{k-1}\mathbf{\Phi}_{k|k-1}^T + \mathbf{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\mathbf{\Gamma}_{k|k-1}^T$$

Substituting the U-D factors for $\mathbf{P}$ and $\mathbf{Q}$ gives

$$\mathbf{U}_{P,k|k-1}\mathbf{D}_{P,k|k-1}\mathbf{U}_{P,k|k-1}^T = \mathbf{\Phi}_{k|k-1}\mathbf{U}_{P,k-1}\mathbf{D}_{P,k-1}\mathbf{U}_{P,k-1}^T\mathbf{\Phi}_{k|k-1}^T + \mathbf{\Gamma}_{k|k-1}\mathbf{U}_{Q,k-1}\mathbf{D}_{Q,k-1}\mathbf{U}_{Q,k-1}^T\mathbf{\Gamma}_{k|k-1}^T$$

Define

$$\mathbf{F} = \mathbf{\Phi}_{k|k-1}\mathbf{U}_{P,k-1}$$
$$\mathbf{G} = \mathbf{\Gamma}_{k|k-1}\mathbf{U}_{Q,k-1}$$

Then

$$\mathbf{U}_{P,k|k-1}\mathbf{D}_{P,k|k-1}\mathbf{U}_{P,k|k-1}^T = \mathbf{F}\mathbf{D}_{P,k-1}\mathbf{F}^T + \mathbf{G}\mathbf{D}_{Q,k-1}\mathbf{G}^T$$

Define the block matrices

$$\mathbf{A} = [\mathbf{F} \quad \mathbf{G}]$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{D}_{P,k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{Q,k-1} \end{bmatrix}$$

Then

$$\mathbf{U}_{P,k|k-1}\mathbf{D}_{P,k|k-1}\mathbf{U}_{P,k|k-1}^{T} = \mathbf{A}\mathbf{W}\mathbf{A}^{T}$$

The Thornton MWGS algorithm performs a weighted Gram-Schmidt orthogonalization on the row vectors of $\mathbf{A}$ using $\mathbf{W}$ as the weighting matrix. The result is a unit lower triangular matrix, $\mathbf{L}$ and a diagonal coefficient matrix $\mathbf{D}_L$, such that

$$\mathbf{U}_{P,k|k-1} = \mathbf{L}^{T}$$
$$\mathbf{D}_{P,k|k-1} = \mathbf{D}_{L}$$

It should be noted that, although the primary purpose of Gram-Schmidt orthogonalization is to compute an optimal set of orthogonal basis vectors for a general set of basis vectors, it is the transformation matrices $\mathbf{L}$ and $\mathbf{D}_L$ that get produced as by-products of the procedure that we're interested in here.

The following *UDProject* subalgorithm implements the Thornton MWGS algorithm:

Use $\mathbf{d}$ as an $n \times 1$ scratchpad vector
FOR $i = n$ TO 1 BY -1 DO
    $\sigma = 0.0$
    FOR $j = 1$ TO n DO
        $\sigma = \sigma + \mathbf{F}_{ij} * \mathbf{D}_{P,jj} * \mathbf{F}_{ij}$
    END FOR
    FOR $j = 1$ TO p DO
        $\sigma = \sigma + \mathbf{G}_{ij} * \mathbf{D}_{Q,jj} * \mathbf{G}_{ij}$
    END FOR
    $\mathbf{d}_i = \sigma$
    $\mathbf{U}_{P,ii} = 1.0$
    FOR $j = 1$ TO $i - 1$ DO
        $\sigma = 0.0$
        FOR $k = 1$ TO n DO
            $\sigma = \sigma + \mathbf{F}_{ik} * \mathbf{D}_{P,kk} * \mathbf{F}_{jk}$
        END FOR
        FOR $k = 1$ TO p DO
            $\sigma = \sigma + \mathbf{G}_{ik} * \mathbf{D}_{Q,kk} * \mathbf{G}_{jk}$
        END FOR
        $\mathbf{U}_{P,ji} = \sigma / \mathbf{d}_i$
        FOR $k = 1$ TO n DO
            $\mathbf{F}_{jk} = \mathbf{F}_{jk} - \mathbf{U}_{P,ji} * \mathbf{F}_{ik}$
        END FOR
        FOR $k = 1$ TO p DO
            $\mathbf{G}_{jk} = \mathbf{G}_{jk} - \mathbf{U}_{P,ji} * \mathbf{G}_{ik}$
        END FOR
    END FOR
END FOR
FOR $i = 1$ TO n DO
    $\mathbf{D}_{P,ii} = \mathbf{d}_i$
END FOR

The inputs to the *UDProject* subalgorithm are

$\mathbf{F} = \mathbf{\Phi}_{k|k-1}\mathbf{U}_{P,k-1}$
$\mathbf{G} = \mathbf{\Gamma}_{k|k-1}\mathbf{U}_{Q,k-1}$
$\mathbf{D}_{P,k-1}$
$\mathbf{D}_{Q,k-1}$

and the outputs are:

$\mathbf{U}_{P,k|k-1}$
$\mathbf{D}_{P,k|k-1}$

The *UDCorrect* subalgorithm is based on Bierman's U-D factorization method for observational updates (named after Gerald Bierman, an engineer at NASA JPL). Bierman's method factorizes the covariance update of the i^{th} step of the sequential update algorithm:

$$\begin{aligned}\mathbf{P}^{(i)} &= \mathbf{P}^{(i-1)} - \mathbf{k}^{(i)}\mathbf{h}^{(i)}\mathbf{P}^{(i-1)} \\ &= \mathbf{P}^{(i-1)} - \frac{\mathbf{P}^{(i-1)}\mathbf{h}^{(i)^T}\mathbf{h}^{(i)}\mathbf{P}^{(i-1)}}{\mathbf{h}^{(i)}\mathbf{P}^{(i-1)}\mathbf{h}^{(i)^T} + r^{(i)}}\end{aligned}$$

Factoring this into U-D components gives

$$\mathbf{U}^{(i)}\mathbf{D}^{(i)}\mathbf{U}^{(i)^T} = \mathbf{U}^{(i-1)}\left[\mathbf{D}^{(i-1)} - \frac{\mathbf{D}^{(i-1)}\mathbf{v}^{(i)}\mathbf{v}^{(i)^T}\mathbf{D}^{(i-1)}}{\mathbf{v}^{(i)^T}\mathbf{D}^{(i-1)}\mathbf{v}^{(i)} + r^{(i)}}\right]\mathbf{U}^{(i-1)^T}$$

where $\mathbf{v}^{(i)}$ is the $n \times 1$ column vector

$$\mathbf{v}^{(i)} = \mathbf{U}^{(i-1)^T}\mathbf{h}^{(i)^T}$$

Bierman's method factorizes the inner expression into U-D factors

$$\mathbf{D}^{(i-1)} - \frac{\mathbf{D}^{(i-1)}\mathbf{v}^{(i)}\mathbf{v}^{(i)^T}\mathbf{D}^{(i-1)}}{\mathbf{v}^{(i)^T}\mathbf{D}^{(i-1)}\mathbf{v}^{(i)} + r^{(i)}} = \mathbf{B}^{(i)}\mathbf{D}^{(i)}\mathbf{B}^{(i)^T}$$

such that

$$\mathbf{U}^{(i)}\mathbf{D}^{(i)}\mathbf{U}^{(i)^T} = \mathbf{U}^{(i-1)}\mathbf{B}^{(i)}\mathbf{D}^{(i)}\mathbf{B}^{(i)^T}\mathbf{U}^{(i-1)^T}$$

Hence,

$$\mathbf{U}^{(i)} = \mathbf{U}^{(i-1)}\mathbf{B}^{(i)}$$

The following *UDCorrect* subalgorithm implements the U-D factorized correction procedure:

Use $\mathbf{k}$ as an $n \times 1$ scratchpad vector
FOR $j = 1$ TO n DO
    $\mathbf{k}_j = \mathbf{h}_j$
    FOR $i = 1$ TO $j - 1$ DO
        $\mathbf{k}_j = \mathbf{k}_j + \mathbf{U}_{P,ij} * \mathbf{h}_i$
    END FOR
END FOR
$\sigma = r$
FOR $j = 1$ TO n DO
    $v = \mathbf{k}_j$
    $\mathbf{k}_j = \mathbf{D}_{P,jj} * \mathbf{k}_j$
    $\omega = \mathbf{k}_j$
    FOR $i = 1$ TO $j - 1$ DO
        $\tau = \mathbf{U}_{P,ij} * \omega$
        $\mathbf{U}_{P,ij} = \mathbf{U}_{P,ij} - \mathbf{k}_i * v/\sigma$
        $\mathbf{k}_i = \mathbf{k}_i + \tau$
    END FOR
    $\mathbf{D}_{P,jj} = \mathbf{D}_{P,jj} * \sigma$
    $\sigma = \sigma + \omega * v$
    $\mathbf{D}_{P,jj} = \mathbf{D}_{P,jj}/\sigma$
END FOR
$\varepsilon = \delta/\sigma$
FOR $i = 1$ TO n DO
    $\mathbf{x}_i = \mathbf{x}_i + \mathbf{k}_i * \varepsilon$
END FOR

The inputs to the *UDCorrect* subalgorithm are

$\mathbf{x}^{(i-1)}$
$\mathbf{U}_P^{(i-1)}$
$\mathbf{D}_P^{(i-1)}$
$\mathbf{h}^{(i)}$
$\delta = z^{(i)} - \mathbf{h}^{(i)}\mathbf{x}^{(i-1)}$

and the outputs are:

$\mathbf{x}^{(i)}$
$\mathbf{U}_P^{(i)}$
$\mathbf{D}_P^{(i)}$

9. The Extended Kalman Filter

The Kalman filter can be extended to handle nonlinear systems. Consider the nonlinear system dynamical process model

$$\mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k\right) + \mathbf{\Gamma}_{k+1|k}\mathbf{w}_k$$

where $\mathbf{x}$ is an $n \times 1$ state vector, $\mathbf{f}()$ is an $n \times 1$ matrix function, $\mathbf{\Gamma}$ is an $n \times p$ transformation matrix, and $\mathbf{w}_k$ is a $p \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{w}_k\mathbf{w}_k^T\right\} = \mathbf{Q}_k$$

and the nonlinear observation model

$$\mathbf{z}_k = \mathbf{h}\left(\mathbf{x}_k\right) + \mathbf{v}_k$$

where $\mathbf{z}$ is an $m \times 1$ measurement vector, $\mathbf{h}()$ is an $m \times 1$ matrix function, and $\mathbf{v}_k$ is an $m \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{v}_k\mathbf{v}_k^T\right\} = \mathbf{R}_k$$

The process and measurement noise sequences are uncorrelated

$$E\left\{\mathbf{w}_j\mathbf{v}_k^T\right\} = \mathbf{0}, \text{ for all } j \text{ and } k$$

Given a previously determined state estimate, $\hat{\mathbf{x}}_{k-1}$, the *a priori* nonlinear projection estimate at time event k is

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}\left(\hat{\mathbf{x}}_{k-1}\right)$$

We can linearize $\mathbf{f}(\mathbf{x})$ at $\hat{\mathbf{x}}_{k|k-1}$ to generate a first-order Taylor approximation $\acute{\mathbf{\Phi}}_{k|k-1}$:

$$\acute{\mathbf{\Phi}}_{k|k-1} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}}$$

Given our *a priori* projection estimate, $\hat{\mathbf{x}}_{k|k-1}$, we can estimate the corresponding nonlinear projection measurement from

$$\hat{\mathbf{z}}_k = \mathbf{h}\left(\hat{\mathbf{x}}_{k|k-1}\right)$$

We can also linearize $\mathbf{h}(\mathbf{x})$ at $\hat{\mathbf{x}}_{k|k-1}$ to generate a first-order Taylor approximation $\acute{\mathbf{H}}_k$:

$$\acute{\mathbf{H}}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}}$$

The extended Kalman filter update cycle for a nonlinear system is then

I. Projection (*a priori*) update:

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}\left(\hat{\mathbf{x}}_{k-1}\right) \\
\acute{\mathbf{\Phi}}_{k|k-1} &= \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}} \\
\mathbf{P}_{k|k-1} &= \acute{\mathbf{\Phi}}_{k|k-1}\,\mathbf{P}_{k-1}\,\acute{\mathbf{\Phi}}_{k|k-1}^{T} + \mathbf{\Gamma}_{k|k-1}\mathbf{Q}_{k-1}\mathbf{\Gamma}_{k|k-1}^{T}
\end{aligned}$$

II. Correction (*a posteriori*) update:

$$\begin{aligned}
\hat{\mathbf{z}}_k &= \mathbf{h}\left(\hat{\mathbf{x}}_{k|k-1}\right) \\
\tilde{\mathbf{z}}_k &= \mathbf{z}_k - \hat{\mathbf{z}}_k \\
\acute{\mathbf{H}}_k &= \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}} \\
\mathbf{S}_k &= \acute{\mathbf{H}}_k\,\mathbf{P}_{k|k-1}\,\acute{\mathbf{H}}_k^{T} + \mathbf{R}_k \\
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\,\acute{\mathbf{H}}_k^{T}\,\mathbf{S}_k^{-1} \\
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\,\tilde{\mathbf{z}}_k \\
\mathbf{P}_k &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^{T}
\end{aligned}$$

10. The Unscented Kalman Filter

Consider the nonlinear system dynamical process model

$$\mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k\right) + \mathbf{\Gamma}_{k+1|k}\mathbf{w}_k$$

where $\mathbf{x}$ is an $n \times 1$ state vector, $\mathbf{f}()$ is an $n \times 1$ matrix function, $\mathbf{\Gamma}$ is an $n \times p$ transformation matrix, and $\mathbf{w}_k$ is a $p \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{w}_k\mathbf{w}_k^T\right\} = \mathbf{Q}_k$$

and the nonlinear observation model

$$\mathbf{z}_k = \mathbf{h}\left(\mathbf{x}_k\right) + \mathbf{v}_k$$

where $\mathbf{z}$ is an $m \times 1$ measurement vector, $\mathbf{h}()$ is an $m \times 1$ matrix function, and $\mathbf{v}_k$ is an $m \times 1$ white noise sequence with covariance

$$E\left\{\mathbf{v}_k\mathbf{v}_k^T\right\} = \mathbf{R}_k$$

The process and measurement noise sequences are uncorrelated

$$E\left\{\mathbf{w}_j\mathbf{v}_k^T\right\} = \mathbf{0}, \text{ for all } j \text{ and } k$$

Recall that the extended Kalman filter approximates the nonlinear transformations with first-order Taylor coefficient matrices. There are situations where these approximations can be inadequate in sufficiently modeling the system dynamics and measurement models, impacting the quality of the Kalman filter performance. The unscented Kalman filter attempts to address these shortcomings by relying on the notion that *it is better to approximate a probability distribution than to approximate a nonlinear transformation*. The basic approach is to use the standard deviations derived from the square root of the variances of the state estimation error covariance, $\mathbf{P}$, to generate a set of "sigma points" that approximate the probability distributions of the process and measurement operations. These sigma points are then used in weighted sum calculations to generate the necessary quantities needed for the Kalman filter update operations.

We begin with the set of $2n + 1$ weighting factors $\left\{W^{(0)}, W^{(1)}, \ldots, W^{(2n)}\right\}$, where, in order to provide unbiased estimates,

$$\sum_{j=0}^{2n} W^{(j)} = 1$$

We choose $W^{(0)}$ based on an empirical criteria. The remaining $2n$ weights are obtained from

$$W^{(j)} = \frac{1 - W^{(0)}}{2n}\ , j = 1,\ldots,2n$$

The $2n + 1$ sigma points, $\mathbf{X}^{(j)}$, for $\hat{\mathbf{x}}_{k-1}$ and $\mathbf{P}_{k-1}$ are determined from:

$$\mathbf{L}_{k-1} = \sqrt{\frac{n}{1 - W^{(0)}}\,\mathbf{P}_{k-1}}$$

$$\begin{aligned}
\mathbf{X}^{(0)} &= \hat{\mathbf{x}}_{k-1} \\
\mathbf{X}^{(i)} &= \hat{\mathbf{x}}_{k-1} + \left[\mathbf{L}_{k-1}\right]_i\ , i = 1,\ldots,n \\
\mathbf{X}^{(i+n)} &= \hat{\mathbf{x}}_{k-1} - \left[\mathbf{L}_{k-1}\right]_i\ , i = 1,\ldots,n
\end{aligned}$$

where $\sqrt{\mathbf{M}}$ is the lower triangular Cholesky decomposition of the symmetric matrix $\mathbf{M}$, and $\left[\mathbf{L}\right]_i$ is the i^{th} column of the lower triangular matrix $\mathbf{L}$. Basically, $\left[\mathbf{L}\right]_i$ represents the estimation error standard deviation matrix of the i^{th} component of $\hat{\mathbf{x}}_{k-1}$.

The primary weight, $W^{(0)}$, controls the spread of the sigma points. $W^{(0)} < 0$ moves the sigma points closer to the distribution origin, whereas $W^{(0)} > 0$ moves the sigma points further from the distribution origin.

One suggested approach by Julier and Uhlmann for choosing the weights is to set $W^{(0)}$ to

$$W^{(0)} = \frac{\kappa}{n + \kappa}$$

Then the remaining $2n$ weights become

$$W^{(j)} = \frac{1}{2(n + \kappa)}\ , j = 1,\ldots,2n$$

and $\mathbf{L}_{k-1}$ becomes

$$\mathbf{L}_{k-1} = \sqrt{(n + \kappa)\,\mathbf{P}_{k-1}}$$

The value κ is used to "fine tune" the higher order moments of the distribution approximation. When the distribution is assumed to be Gaussian, then κ should be chosen so that

$$n + \kappa = 3$$

Given a set of $2n + 1$ weighting factors, $W^{(j)}$, and a set of $2n + 1$ sigma points, $\mathbf{X}^{(j)}$, the unscented Kalman filter update cycle for a nonlinear system is

I. Projection (*a priori*) update:

$$\mathbf{x}^{(j)} = \mathbf{f}\left(\mathbf{X}^{(j)}\right) \ , j = 0,\ldots,2n$$

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{j=0}^{2n} W^{(j)} \mathbf{x}^{(j)}$$

$$\mathbf{P}_{k|k-1} = \sum_{j=0}^{2n} W^{(j)} \left[\mathbf{x}^{(j)} - \hat{\mathbf{x}}_{k|k-1}\right] \left[\mathbf{x}^{(j)} - \hat{\mathbf{x}}_{k|k-1}\right]^T + \mathbf{\Gamma}_{k|k-1} \mathbf{Q}_{k-1} \mathbf{\Gamma}_{k|k-1}^T$$

II. Correction (*a posteriori*) update:

$$\mathbf{z}^{(j)} = \mathbf{h}\left(\mathbf{x}^{(j)}\right) \ , j = 0,\ldots,2n$$

$$\hat{\mathbf{z}}_k = \sum_{j=0}^{2n} W^{(j)} \mathbf{z}^{(j)}$$

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k$$

$$\mathbf{P}_{\mathbf{zz},k} = \sum_{j=0}^{2n} W^{(j)} \left[\mathbf{z}^{(j)} - \hat{\mathbf{z}}_k\right] \left[\mathbf{z}^{(j)} - \hat{\mathbf{z}}_k\right]^T$$

$$\mathbf{P}_{\mathbf{xz},k} = \sum_{j=0}^{2n} W^{(j)} \left[\mathbf{x}^{(j)} - \hat{\mathbf{x}}_{k|k-1}\right] \left[\mathbf{z}^{(j)} - \hat{\mathbf{z}}_k\right]^T$$

$$\mathbf{S}_k = \mathbf{P}_{\mathbf{zz},k} + \mathbf{R}_k$$

$$\mathbf{K}_k = \mathbf{P}_{\mathbf{xz},k} \, \mathbf{S}_k^{-1}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \, \tilde{\mathbf{z}}_k$$

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \, \mathbf{S}_k \, \mathbf{K}_k^T$$

11. References

This following list comprises what I used for my references. They are not explicitly cited, but they do form a comprehensive list of the resources used for this document.

1. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Trans. ASME*, Journal of Basic Engineering, vol. 82, Series D, pp. 34-45, Mar. 1960
2. Thornton, C. L., *Triangular Covariance Factorizations for Kalman Filtering*, National Aeronautics and Space Administration, Technical Memorandum 33-798, Oct. 1976
3. Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, Inc., New York, NY, 1977
4. Brown, R. G., *Introduction to Random Signal Analysis and Kalman Filtering*, John Wiley & Sons, Inc., New York, NY, 1983
5. Sorenson, H. W. (Ed.), *Kalman Filtering: Theory and Application*, IEEE Press, New York, NY, 1985
6. Chui, C. K. and Chen, G., *Kalman Filtering with Real-Time Applications*, Springer-Verlag, New York, NY, 1987
7. Grewal, M. S., and Andrews, A. P., *Kalman Filtering: Theory and Practice*, Prentice-Hall, Englewood Cliffs, NJ, 1993
8. Julier, S. J., and Uhlmann, J. K., "A New Extension of the Kalman Filter to Nonlinear Systems," *Proc. SPIE 3068, Signal Processing, Sensor Fusion, and Target Recognition VI*, 28 July 1997; https://doi.org/10.1117/12.280797
9. Brookes, M., "The Matrix Reference Manual", [online] http://www.ee.imperial.ac.uk/hp/staff/dmb/matrix/intro.html, 2020