# AN INTRODUCTION TO THE KALMAN FILTER

by

Thomas Parrish

Under the Guidance of

Dr. Sebastian Pauli

A Thesis Submitted to
the Faculty of the
Lloyd International Honors College at
the University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for
Disciplinary Honors

Greensboro
2014

*This thesis is dedicated to Dr. Maya Chhetri, Dr. Sebastian Pauli, my family and friends, and the faculty of the University of North Carolina at Greensboro Mathematics Department, without whose encouragement, patience, and support I could not have succeeded.*

# Contents

# Introduction

The work on this thesis began research conducted through the UNCG math-bio program during the 2011-2012 and 2012-2013 school years. As part of a larger study examining vocal communication among wild deer mice (*Peromyscus* species) [2], infrared video was collected over 131 nights from dusk until dawn. The video was taken from a camera suspended in the tree canopy above the free-living mice on the forest floor. The volume of video recordings obtained in this study is a challenge to manually process. Computer vision techniques however, allowed us to detect and record the trajectories of moving objects from the video data without human intervention. As the result of this experience we were able to process the approximately 1500 hours of video and extract biologically meaningful data.

However, challenges presented themselves when occlusion occurred. The primary source of occlusion was objects moving beneath underbrush. In this case, objects will frequently disappear temporarily. Because of the simplicity of the tracking algorithm used, this causes objects to be considered "new" objects when they reappeared. In addition, because of the limited features available to distinguish objects, it was difficult to properly re-identify objects after they crossed paths. This lead us to investigate more advanced tracking algorithms. One commonly incorporated element of advanced tracking algorithms is the Kalman Filter, useful for its ability to filter noise from observed data and predict future data.

In this paper, we give an introduction to the Kalman filter targeted at an audience having minimal exposure to statistical methods and linear algebra. The Kalman filter a recursive linear filter which can be used to estimate the state of any system who's state at a time $t$ can be described by a linear transformation of the state at time $t-1$. We provide a brief introduction to recursive methods for calculating the mean and variance of a random variable in chapter one, and derive the Kalman filter for a single-dimensional space. We extend this derivation to an equal-dimensional filter in chapter 3. Finally, we consider the multi-dimensional case when the dimensions of the state space and measurement space disagree, and provide an alternate approach to the derivation of the general case Kalman Filter.

## Acknowledgments

The research preempting this thesis was supported by National Science Foundation (Grants IOB-0641530, IOB-1132419, DMS-0850465 and DBI-0926288). Thanks go to the Office of Undergraduate Research of UNCG and in particular Dr. Jan Rychtar and Dr. Mary Crowe.

We thank the students who worked on this project, including David Schuchart, Caitlin Bailey, Christian Bankester, and Benjamin Manifold, along with all the students who worked in the field collecting data. Special thanks go to Dr. Matina Kalcounis-Rueppell and Dr. Sebastian Pauli allowing me to be part of this research project, and to Dr. Pauli for his unending patience and support throughout the research project and this thesis work.

# Chapter 1

# An Overview of the Kalman Filter

In our object tracking application, foreground objects are differentiated from background objects by means of background subtraction. The background is the set of relatively unchanging pixels, or picture elements. In background subtraction, this background image is subtracted from the current image, and the resulting difference is considered to be the foreground. Because of the way background subtraction filters noise and how the location of objects is calculated (see [3]), there is an influential amount of error in the observed measurement of object locations within our tracking program. This problem, along with an inability to predict the behavior of objects that become hidden behind underbrush, prompted the implementation of the Kalman filter.

The Kalman filter is an algorithm that operates recursively on a stream of noisy input data, producing a statistically optimal estimate of the system's state; that is, the estimate with the least variance. The recursive nature of the algorithm allows it to predict the next state of the system using only information about the previous state, rather than necessitating analysis of all previous states at each new step. It is important to note, however, that the nature of the Kalman filter is that it estimates discrete time states, opposed to continuous time, and that standard Kalman filters assume that the state of the system is governed by a linear stochastic difference equation. The optimal estimate allows us to use measurements with some uncertainty, and estimate more certain values of unknowns. In addition, the filter is capable of estimating information that is unobservable, such as the velocity of moving objects in our case.

The three assumptions of the Kalman filter are that the state of a dynamic system changes over time, that the observer can describe these changes with a linear transformation, and that the state of the system is a random variable with Gaussian error. The linear transformation can be used to produce a forecast of the next system state. Since this forecast is a linear transformation on a Gaussian random variable, it will also have a Gaussian error. Then the random variable representing the state can be estimated by taking the joint probability distribution of the forecast and the measurement. This process can be summarized by the following algorithm.

**Algorithm 1** (Kalman Filter)**.**

Input: State transition matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$
Space translation matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$
Process error covariance $\mathbf{Q} \in \mathbb{R}^{n \times n}$
Measurement covariance $\mathbf{R} \in \mathbb{R}^{m \times m}$
Stream of measurements $(\mathbf{m}_t)_{t \in \mathbb{N}}$ with $\mathbf{m}_t \in \mathbb{R}^m$

Output: Stream of optimal state estimates $(\hat{\mathbf{x}}_t)_{t \in \mathbb{N}}$ with $\hat{x}_t \in \mathbb{R}^n$

- $\hat{\mathbf{x}}_0 \leftarrow \mathbf{m}_0$
- Repeat:
  - $\tilde{\mathbf{x}}_t \leftarrow \mathbf{A}\hat{\mathbf{x}}_{t-1}$ — Predict the current state
  - $\widetilde{\Sigma}_t \leftarrow \mathbf{A}\hat{\Sigma}_{t-1}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$ — Compute covariance of predicted state
  - $\mathbf{K}_t \leftarrow \widetilde{\Sigma}_t \mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\widetilde{\Sigma}_t \mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$ — Compute optimal measurement weight
  - $\hat{\mathbf{x}}_t \leftarrow \tilde{\mathbf{x}}_t + \mathbf{K}_t \left(\mathbf{m}_t - \mathbf{H}_t \tilde{\mathbf{x}}_t\right)$ — Correct prediction using current measurement
  - $\hat{\Sigma}_t \leftarrow (I - \mathbf{K}_t \mathbf{H})\widetilde{\Sigma}_t$ — Calculate covariance of the corrected estimate
- Until end of stream.

The principal assumption of the Kalman filter is that the state $\mathbf{x}$ of the system at time $t$ is

$$\mathbf{x} = \mathbf{A}\hat{\mathbf{x}}_{t-1} + \mathbf{v} \tag{1.1}$$

The state transition matrix $A$ describes the behavior, or relationship between the state at time $t$ and time $t-1$, which is assumed to be linear. The process error $\mathbf{v}$ describes how closely $\mathbf{A}$ represents the true behavior of the system. For our discussion of the Kalman filter, we assume that the behavior of the system does not change over time, and therefore $\mathbf{A}$ and $\mathbf{v}$ remain constant.

The optional state translation matrix allows for the dimension of the measurement space to differ from the dimension of the state space. This occurs when there are variables we wish to estimate that are unobservable.

In our application, we assumed the system describes motion at a constant velocity and our measurements consist of $x, y$ pairs. However, we define the state of our system to be $\mathbf{x}^{\mathrm{T}} = [x, y, v_x, v_y]$ where $v_x,$ and $v_y$ are the $x$ and $y$ velocities respectively.

$\mathbf{K}_t$ describes the weight placed on the measurement when the optimal estimate is produced. In the simple case when the dimension of the state agrees with the dimension of the measurements, $\mathbf{K}_t$ is obvious. In the generic derivation, $\mathbf{K}_t$ is unknown, and become the variable used to minimize the covariance of there state estimate.

The power of the Kalman Filter comes from the knowledge of the behavior of the state of the system. We assume that the system can be described by a linear transformation. This transformation allows us to predict the next (or current) state of the system. We assume this

transformation does not perfectly describe the state, but that the error in this transformation is centered at zero. Thus we assume that covariance of $\mathbf{v}$, $\mathbf{Q}$, is nonzero and $E[\mathbf{v}] = 0$. For these reasons, we can predict the state of the system at time $t$ as follows

$$\widetilde{\mathbf{x}}_t = \mathbf{A}\hat{\mathbf{x}}_{t-1}. \tag{1.2}$$

If the covariance matrix of the estimate of our random variable $x$ at time $t-1$ is $\widehat{\Sigma}_{t-1}$, then the covariance matrix of our forecast can be described by

$$\widetilde{\Sigma}_t = \mathbf{A}\widehat{\Sigma}_{t-1}\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \tag{1.3}$$

where $Q$ is the covariance matrix of the error term $v$.

Since we assume that $\hat{x}$ and $v$ are Gaussian random variables, and $A$ is a linear transformation, $\widetilde{x}_t$ will also be a Gaussian random variable. As we will show in the latter chapters, this forecast can be combined with a measurement $m_t$, as follows

$$\hat{\mathbf{x}}_t = \widetilde{\mathbf{x}}_t + K\left(\mathbf{m}_t - \widetilde{\mathbf{x}}_t\right) \tag{1.4}$$

where

$$K = \widetilde{\Sigma}_t\left(\widetilde{\Sigma}_t + \Sigma_{m_t}\right)^{-1}. \tag{1.5}$$

For our last step, we update $\widehat{\Sigma}$ as follows

$$\widehat{\Sigma}_t = (I - K)\,\widetilde{\Sigma}_t. \tag{1.6}$$

By combining this forecast with the observed measurements, we are able to filter noise attributable to measurement error. We reserve in depth discussion of the forecaster for the latter sections of the paper. Since the underlying principal of the Kalman Filter is recursively calculating the joint probability distribution of two random Gaussian variables, we will begin by provide some basic statistical definitions and methods for recursively calculating mean and variance, which will be helpful in our derivation of the Kalman Filter.

# Chapter 2

# Statistical Preliminaries

A *random variable* is a variable whose value is subject to variations due to chance. It consists of a discrete or continuous set of possible values, each with an associated probability. The distribution of possible values for a random variable can be partially characterized by two important quantities, the *mean* and *variance*. For a random variable $X$ with values $x_1, \ldots, x_n$, the *mean* of $X$, also called the expected value ( $E[X]$ ), is defined to be

$$\mu_X = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

This definition can be extended to a random variable $Y \in \mathbb{R}^n$, where $Y = [y_1, \ldots, y_n]^{\mathrm{T}}$, as follows

$$\mu_Y = [E(y_1), \ldots, E(y_n)]^{\mathrm{T}}.$$

The mean is associated with the most probable value of a random variable. The *standard deviation* of $X$ is defined to be

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)}.$$

The *variance* of $X$ is defined to be the square of the standard deviation ($\sigma^2$).

This notion of variance can be extended to random variables in $\mathbb{R}^n$ using a *covariance matrix*. Given two random variables $x_1$ and $x_2$, their *covariance* is defined as

$$\mathrm{Cov}[x_1, x_2] = E[(x_1 - E[x_1])(x_2 - E[x_2])] = E[x_1 x_2] - E[x_1]E[x_2].$$

The *covariance matrix* is a convenient way to summarize the covariances of all pairs of elements of a random variable in $\mathbb{R}^n$, and is defined as

$$\Sigma = \begin{bmatrix} \mathrm{Cov}[x_1, x_1] & \ldots & \mathrm{Cov}[x_1, x_n] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[x_n, x_1] & \ldots & \mathrm{Cov}[x_n, x_n] \end{bmatrix}.$$

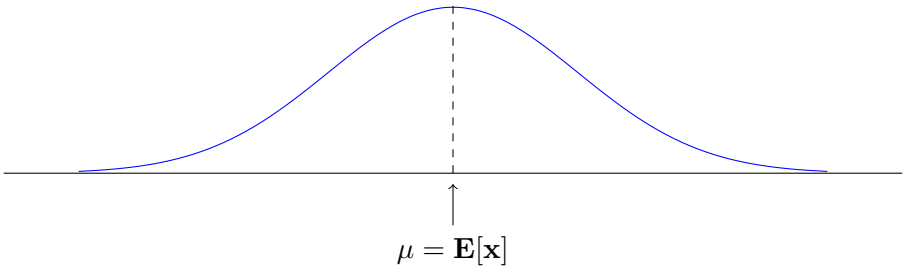


Figure 2.1: Gaussian distribution with expected value $E[x] = \mu$

Naturally, $\text{Cov}[x_i, x_i] = E[(x_i - E[x_i])^2] = \sigma_i^2$. While the standard deviation and variance are both measures of how far possible values of a random variable are spread out, covariance is measure of how much two random variables change together. A positive covariance between two random variables $x_1$ and $x_2$ means that large values of $x_1$ generally correspond to large values of $x_2$, while small values of $x_1$ generally correspond to small values of $x_2$. Thus, the covariance matrix measures both how wide the distribution of each component is, as well as how wide these spreads are relative to the other components.

Note that, given a random variable $X$ with variance $\sigma^2$ and a constant $a$, the variance of $aX$ is given by

$$\text{Var}(aX) = E[(aX - a\mu)^2] = E[a^2(X - \mu)^2] = a^2 E[(X - \mu)^2] = a^2 Var(X)$$

This result extends to higher dimensions as well. Given a vector $X \in \mathbb{R}^n$ and an $m \times n$ matrix $A$, it is not difficult to show that $\text{Cov}[AX] = A \cdot \text{Cov}[X] A^{\text{T}}$.

In the Kalman filter, we make the assumption that all random variables follow a Gaussian, or normal, distribution. A random variable $x$ is said to have a *Gaussian*, distribution if its corresponding probability density function is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

A Gaussian random variable is completely characterized by two parameters, its mean and variance.

Once again, this property extends to Gaussian random variables in $\mathbb{R}^n$. Given a random variable $x \in \mathbb{R}^n$ with mean $\mu$ and covariance matrix $\Sigma$, the probability density of $x$ is

$$p(x, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\left((x-\mu)^{\text{T}} \Sigma^{-1} (x-\mu)\right)}.$$

**Definition** (Time Series)**.** A *time series* is a sequence of data points $x_1, x_2, \ldots, x_n$, measured at successive points in time at uniform intervals.

Given a time series $x_1, x_2, \ldots, x_n$, a *filter* recomputes or corrects $x_{n+1}$, by taking into account all previous data.

## 2.1 Recursive Calculation of Mean and Variance

Recall that, for a set $X$ with values $\{x_1, \ldots, x_n\}$, the *mean* of $X$ is defined to be

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Suppose we observe a new measurement $x_{n+1}$. Using the definition of mean, we can recompute $\mu_{n+1}$ as $\mu_{n+1} = \frac{1}{n} \sum_{i=1}^{n+1} x_i$. However, it is possible, and often more desirable, to compute $\mu_{n+1}$ using only knowledge of $\mu_n$. We show that

**Lemma 1.**

$$\mu_{n+1} = \mu_n + \left( \frac{1}{n+1} \right) (x_{n+1} - \mu_n) \tag{2.1}$$

*Proof.* Let $\mu_n = \frac{1}{n} \sum_{i=1}^{n} x_i$. Solving for $\mu_{n+1}$ in terms of $\mu_n$, we see that

$$\begin{aligned}
\mu_{n+1} &= \frac{1}{n+1} \sum_{1}^{n+1} x_i = \frac{1}{n+1} \left( \sum_{1}^{n} x_i + x_{n+1} \right) \\
&= \frac{n}{n+1} \left( \frac{1}{n} \sum_{1}^{n} x_i + \frac{1}{n} x_{n+1} \right) \\
&= \frac{n}{n+1} \left( \mu_n + \frac{1}{n} x_{n+1} \right) = \frac{1}{n+1} \left( n\mu_n + x_{n+1} \right) \\
&= \frac{n+1}{n+1} \mu_n - \frac{n+1}{n+1} \mu_n + \frac{1}{n+1} \left( n\mu_n + x_{n+1} \right) \\
&= \mu_n + \frac{1}{n+1} \left( n\mu_n - (n+1)\mu_n + x_{n+1} \right) \\
&= \mu_n + \frac{1}{n+1} \left( x_{n+1} - \mu_n \right).
\end{aligned}$$

□

In applications related to the Kalman Filter, it is common to write equation 2.1 in the form of

$$\mu_{n+1} = \mu_n + K \left( x_{n+1} - \mu_n \right) \tag{2.2}$$

where $K = \frac{1}{n+1}$ is called the gain factor. This new mean is a weighted average of the old mean and the new data point, with the old mean being more heavily weighted (since we trust the previous mean $\mu_n$ more than our error-prone new data point.) Another way to interpret 2.1 is to say that the mean is corrected using the difference of $x_n + 1$ and $\mu_n$, with the gain $K$ determining how big the adjustment will be.

It is also useful to consider the variance of a set of data points. The variance can be described recursively as well, in a similar fashion to the mean. Recall that the variance of a set of data points $x_i, \ldots, x_n$ is defined by

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_n)^2 .$$

Then considering a new data point $x_{n+1}$, $\sigma_{n+1}^2$ can be calculated by

$$\sigma_{n+1}^2 = \frac{1}{n+1} \sum_{i=1}^{n+1} (x_i - \mu_{n+1})^2 .$$

Substituting in (1), we have

$$\begin{aligned}
\sigma_{n+1}^2 &= \frac{1}{n+1} \sum_{i=1}^{n+1} (x_i - \mu_n - K(x_{n+1} - \mu_n))^2 \\
&= \frac{1}{n+1} \sum_{1}^{n+1} \left[ (x_i - \mu_n)^2 - 2K(x_i - \mu_n)(x_{i+1} - \mu_n) + K^2 (x_{n+1} - \mu_n)^2 \right] \\
&= \frac{1}{n+1} \left[ \sum_{i=1}^{n+1} (x_i - \mu_n)^2 - 2K \sum_{1}^{n+1} (x_i - \mu_n)(x_{n+1} - \mu_n) + K^2 (n+1)(x_{n+1} - \mu_n)^2 \right] \\
&= \frac{1}{n+1} \left[ \sum_{i=1}^{n} (x_i - \mu_n)^2 - 2K \sum_{1}^{n} (x_i - \mu_n)(x_{n+1} - \mu_n) \right. \\
&\qquad \left. + (1 - 2K)(x_{n+1} - \mu_n)^2 + K^2 (n+1)(x_{n+1} - \mu_n)^2 \right] \\
&= \frac{1}{n+1} \left[ \sum_{i=1}^{n} (x_i - \mu_n)^2 - 2K \sum_{1}^{n} (x_i - \mu_n)(x_{n+1} - \mu_n) \right. \\
&\qquad \left. + (1 - 2K)(x_{n+1} - \mu_n)^2 + K^2 (x_{n+1} - \mu_n)^2 + nK^2 (x_{n+1} - \mu_n)^2 \right] \\
&= \frac{1}{n+1} \left[ \sum_{i=1}^{n} (x_i - \mu_n)^2 - 2K \sum_{1}^{n} (x_i - \mu_n)(x_{n+1} - \mu_n) \right. \\
&\qquad \left. + (1 - K)^2 (x_{n+1} - \mu_n)^2 + nK^2 (x_{n+1} - \mu_n)^2 \right] .
\end{aligned}$$

Note that $2K\sum_{i=1}^{i=n}(x_i-\mu_n)(x_{n+1}-\mu_n) = 2K(x_{n+1}-\mu_n)\sum_{i=1}^{n}(x_i-\mu_n) = 0$. This follows from

$$\sum_{i=1}^{n}(x_i-\mu_n) = \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\mu_n = \sum_{i=1}^{n}x_i - n\mu_n = \sum_{i=1}^{n}x_i - \frac{n}{n}\sum_{i=1}^{n}x_i = 0.$$

Then we can simplify $\sigma_{n+1}^2$ to

$$\begin{aligned}\sigma_{n+1}^2 &= \frac{1}{n+1}\left(\frac{n}{n}\sum_{i=1}^{n}(x_i-\mu_n)^2 + (1-K)^2(x_{n+1}-\mu_n)^2 + nK^2(x_{n+1}-\mu_n)^2\right)\\ &= \frac{1}{n+1}\left(n\sigma_n^2 + (1-K)^2(x_{n+1}-\mu_n)^2 + nK^2(x_{n+1}-\mu_n)^2\right)\\ &= \frac{1}{n+1}\left(n\sigma_n^2 + ((1-K)^2+nK^2)(x_{n+1}-\mu_n)^2\right).\end{aligned}$$

Now we can further simplify $\sigma_{n+1}^2$ to

$$\begin{aligned}\sigma_{n+1}^2 &= \frac{1}{n+1}\left(n\sigma_n^2 + nK(x_{n+1}-\mu_n)^2\right)\\ &= \frac{n}{n+1}\left(\sigma_n^2 + K(x_{n+1}-\mu_n)^2\right)\\ &= (1-K)\left(\sigma_n^2 + K(x_{n+1}-\mu_n)^2\right).\end{aligned}$$

If we allow $\sigma'^2_{n+1} = \sigma_n^2 + K(x_{n+1}-\mu_n)^2$, then we are left with

$$\sigma_{n+1}^2 = (1-K)\sigma'^2_{n+1}. \tag{2.3}$$

We now have a method for recursively updating the mean and variance of any time series. This recursive technique will become the basis for the derivation of the one dimensional Kalman Filter.

# Chapter 3

# The One Dimensional Kalman Filter

## 3.1 One-Dimensional Product of Gaussians

In the case when the state space and measurement space are one-dimensional, the derivation of the Kalman Filter can be reduced to producing the joint probability distribution of two independent variables. We begin by deriving generic formula for the mean and variance of this new distribution, and arranging them in a way that will become familiar throughout the remainder of the paper.

### Deriving the Mean and Variance of a Product of Gaussians

Suppose there exist two imprecise measurements for some quantity of interest $x$, say $x_1$ and $x_2$, where the error of $x_1$ and $x_2$ can be described by a *Gaussian*, distribution. These measurements are characterized by their mean $\mu_1$ and $\mu_2$, and variances $\sigma_1^2$ and $\sigma_2^2$, respectively. The associated with probability density functions can be given by

$$p_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\frac{(x-\mu_i)^2}{\sigma_i^2}}$$

for $i = 1, 2$ and $x \in (-\infty, \infty)$. Because we assume that $x_1$ and $x_2$ are independent, we can find the joint probability of $x$ given $x_1$ and $x_2$ by taking the product of their probability density functions, as follows

$$\begin{aligned} p(x) = p_1(x)p_2(x) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\frac{(x-x_1)^2}{\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\frac{(x-x_2)^2}{\sigma_2^2}} \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\left(\frac{(x-x_1)^2}{2\sigma_1^2} + \frac{(x-x_2)^2}{2\sigma_2^2}\right)}. \end{aligned}$$

If we take a close look at the exponential term

$$\begin{aligned}\alpha &= \left(\frac{(x-x_1)^2}{2\sigma_1^2} + \frac{(x-x_2)^2}{2\sigma_2^2}\right) \\ &= \frac{(\sigma_1^2+\sigma_2^2)x^2 - 2(x_1\sigma_2^2 + x_2\sigma_1^2)x + x_1^2\sigma_2^2 + x_2^2\sigma_1^2}{2\sigma_1^2\sigma_2^2} \\ &= \frac{x^2 - 2\frac{x_1\sigma_2^2+x_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}x + \frac{x_1\sigma_2^2+x_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\end{aligned}$$

we see that the term closely resembles the standard term in the standard Gaussian probability density function. All that is necessary is to complete the square and normalize the function. To complete the square, let $\mu = \frac{x_1\sigma_2^2+x_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$, $\sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$, and $\gamma = \frac{\mu-\mu^2}{2\sigma^2}$. Then we have

$$\begin{aligned}\alpha &= \frac{x^2 - 2\mu x + \mu}{2\sigma^2} \\ &= \frac{x^2 - 2\mu x + \mu + \mu^2 - \mu^2}{2\sigma^2} \\ &= \frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} + \frac{\mu - \mu^2}{2\sigma^2} \\ &= \frac{(x-\mu)^2}{2\sigma^2} + \gamma.\end{aligned}$$

Thus

$$p(x) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)-\gamma} = \frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)}.$$

In order to normalize $p(x)$, we can integrate $p$ over $\mathbb{R}$, as follows.

$$\int_{-\infty}^{\infty} \frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)} = \frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)}.$$

Recalling the standard result of

$$\int_{-\infty}^{\infty} e^{-ax^2} = \sqrt{\frac{\pi}{a}},$$

we see that

$$\frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)} = \frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2}\cdot\sqrt{\frac{\pi}{\frac{1}{2\sigma^2}}} = \frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2}\cdot\sqrt{2\pi\sigma^2}.$$

Finally, normalizing $p(x)$, we find that

$$\frac{p(x)}{\frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2}\cdot\sqrt{2\pi\sigma^2}} = \frac{\frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2}e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)}}{\frac{e^{-\gamma}}{2\pi\sigma_1\sigma_2}\cdot\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)}. \tag{3.1}$$

It is now clear that $\mu = \frac{x_1\sigma_2^2+x_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$ is the most probable value for the combined estimate of our measurements, and the combined variance is $\sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$. If we substitute a gain factor of $K = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$, then we can write the best estimate of $\hat{x}$ as

$$\hat{x} = x_1 + K(x_2 - x_1) \tag{3.2}$$

and the new variance as

$$\sigma^2 = (1-K)\sigma_1^2. \tag{3.3}$$

## 3.2 One Dimensional Kalman Filter

Note the similarities between equations 3.2 and 3.3 and the equations from the Kalman Filter in the introduction. In fact, in a single dimensional Kalman Filter, we are simply taking the joint probability distribution of a forecasted state estimate and the new measurement. A summary of the Kalman Filter in one dimension can be seen in the table below.

| Input | |
|---|---|
| State transition constant | $a$ |
| Measurement | $m$ |
| Measurement variance | $\sigma^2_{m_t}$ |
| Process variance | $q$ |
| Prediction Step | |
| Forecast (a priori) state estimate | $\tilde{x}_t = a\hat{x}_{t-1}$ |
| Forecast (a priori) estimate covariance | $\tilde{\sigma}_t = a^2\hat{\sigma}_{t-1} + q$ |
| Update Step | |
| Optimal Kalman gain | $K_t = \tilde{\sigma}_t^2 \left(\tilde{\sigma}_t^2 + \sigma^2_{m_t}\right)^{-1}$ |
| Updated (a posteriori) state estimate | $\hat{x}_t = \tilde{x}_t + K(m_t - \tilde{x}_t)$ |
| Updated (a posteriori) estimate covariance | $\hat{\sigma}^2 = (1 - K)\tilde{\sigma}_t^2$ |

Table 3.1: One Dimensional Kalman Filter Equations

A practical example of a one dimensional Kalman Filter can be seen as follows.

## Example

Suppose we wish to know the true temperature of a refrigerator, given a series of noisy measurements from an error prone thermometer. We assume that the temperature remains constant. In this case, our state transition will be represented by $A = 1$, and our forecast would be $\widetilde{x} = \hat{x} + v$. Because we assume that the temperature actually is constant, our process variance will be very small. Let us choose $Q = .0001$. At each step, $v$ is chosen as a random variable centered around 0 with variance $Q$. In this case, we expect our forecast to be very similar to our filtered estimate, because $v$ is negligible.

| Measurement | Forecast | Gain | Estimate | Forecast covariance | Estimate covariance |
|---|---|---|---|---|---|
| 3.231 | 3.000 | 0.909 | 3.000 | 1.000 | 0.091 |
| 3.209 | 3.210 | 0.476 | 3.210 | 0.091 | 0.048 |
| 2.963 | 3.209 | 0.323 | 3.210 | 0.048 | 0.032 |
| 2.311 | 3.130 | 0.245 | 3.130 | 0.032 | 0.024 |
| 2.772 | 2.929 | 0.197 | 2.929 | 0.025 | 0.020 |
| 2.640 | 2.898 | 0.165 | 2.898 | 0.020 | 0.017 |
| 3.018 | 2.856 | 0.143 | 2.856 | 0.017 | 0.014 |
| 2.731 | 2.879 | 0.126 | 2.879 | 0.014 | 0.013 |
| 2.485 | 2.860 | 0.112 | 2.860 | 0.013 | 0.011 |
| 3.195 | 2.818 | 0.102 | 2.818 | 0.011 | 0.010 |

Table 3.2: Kalman Filter values from $t = 0$ to $t = 10$

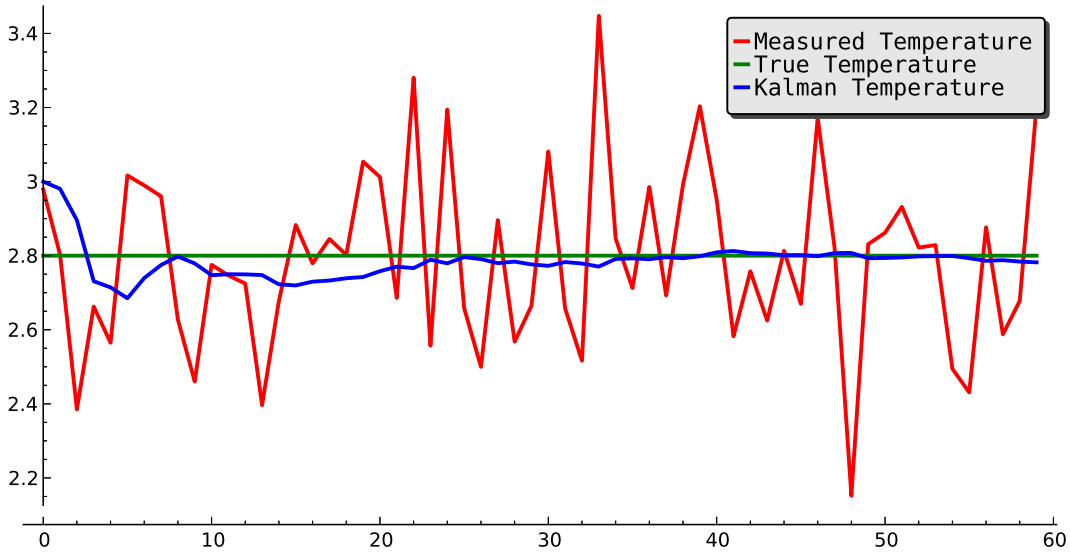


Figure 3.1: Temperature measurements and estimates from $t = 0$ to $t = 60$

# Chapter 4

# The Equal Dimensional Kalman Filter

This technique we used to derive the one dimensional Kalman Filter can be extended to the case of two multivariate Gaussian distributions, given some minor assumptions. Assume that we have two measurement vectors $\vec{x}_1 = [x_1 \dots x_n]^{\mathrm{T}}$ and $\vec{x}_2 = [x_2 \dots x_n]^{\mathrm{T}}$ that can be described by a multivariate Gaussian distribution with means $\mu_1, \mu_2 \in \mathbb{R}^n$ and covariance matrices $\Sigma_1, \Sigma_2 \in S^n_{++}$. Here $S^n_{++}$ is the space of symmetric positive definite square matrices of dimension $n^2$.

Recall that the definition of the covariance matrix of $\vec{x}_1$ is

$$\Sigma = \begin{bmatrix} \mathrm{Cov}[x_1, x_1] & \dots & \mathrm{Cov}[x_1, x_n] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[x_n, x_1] & \dots & \mathrm{Cov}[x_n, x_n] \end{bmatrix}.$$

Naturally, $\mathrm{Cov}[x_i, x_i] = E[(x_i - E[x_i])^2] = \sigma_i^2$. Note that $\Sigma$ is symmetric by definition, and in the case that each component of a vector $x$ is statistically independent, $\Sigma$ will be a diagonal matrix.

The general form for a multi-variate Gaussian probability density function is quite similar to the single dimensional case. For a vector $x$ in $\mathbb{R}^n$,

$$p(x, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\left((x-\mu)^{\mathrm{T}} \Sigma^{-1} (x-\mu)\right)}. \tag{4.1}$$

**Example 1.** As an example, let us consider a vector $x$ who's components are independent. Then we have

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

and

$$
\begin{aligned}
p(x, mu, \Sigma) &= \frac{1}{(2\pi)(\sigma_1^2\sigma_2^2 - 0)^{\frac{1}{2}}} e^{-\frac{1}{2}\left((x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right)} \\
&= \frac{1}{(2\pi)(\sigma_1^2\sigma_2^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}\right)} \\
&= \frac{1}{(2\pi)(\sigma_1^2\sigma_2^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \frac{(x_1-\mu_1)}{\sigma_1^2} \\ \frac{(x_2-\mu_2)}{\sigma_2^2} \end{bmatrix}\right)} \\
&= \frac{1}{(2\pi)(\sigma_1^2\sigma_2^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right)} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}
\end{aligned}
$$

which is precisely the product of two independent Gaussian probability density functions which we explored previously.

For the remainder of this section, we will drop the vector notation, and consider all variables to be vectors. Much like the univariate Gaussian distribution, the probability density function of an estimation $\hat{x}$, given two independent measurements $x_1$ and $x_2$ with covariance matrices $\Sigma_1$ and $\Sigma_2$, can be found by taking the product of their respective probability density functions and normalizing. That is,

$$
\begin{aligned}
p(\hat{x}, \mu, \Sigma) &= p(x_1, \mu_1, \Sigma_1) \cdot p(x_2, \mu_2, \Sigma_2) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_1|^{\frac{1}{2}}} e^{-\frac{1}{2}\left((\hat{x}-\mu_1)^{\mathrm{T}}\Sigma_1^{-1}(\hat{x}-\mu_1)\right)} \cdot \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_2|^{\frac{1}{2}}} e^{-\frac{1}{2}\left((\hat{x}-\mu_2)^{\mathrm{T}}\Sigma_2^{-1}(\hat{x}-\mu_2)\right)} \\
&= \frac{1}{(2\pi)^n|\Sigma_1|^{\frac{1}{2}}|\Sigma_2|^{\frac{1}{2}}} e^{-\frac{1}{2}\left((\hat{x}-\mu_1)^{\mathrm{T}}\Sigma_1^{-1}(\hat{x}-\mu_1) + (\hat{x}-\mu_2)^{\mathrm{T}}\Sigma_2^{-1}(\hat{x}-\mu_2)\right)}
\end{aligned}
$$

Like the single dimensional case, we can show that this represents a new Gaussian distribution, and the new mean and variance can be found by completing the square of the exponent again, and normalizing.

Let us consider only the exponent term, disregarding the $-\frac{1}{2}$, which we will re-assume later. Then we have

$$E = (\hat{x} - \mu_1)^{\mathrm{T}} \Sigma_1^{-1} (\hat{x} - \mu_1) + (\hat{x} - \mu_2)^{\mathrm{T}} \Sigma_2^{-1} (\hat{x} - \mu_2). \tag{4.2}$$

Given an $n \times n$ symmetric matrix $A$, and two vectors $x$ and $y$ of dimension $n$,

$$(x - y)^{\mathrm{T}} A (x - y) = x^{\mathrm{T}} A x + y^{\mathrm{T}} A y + 2 x^{\mathrm{T}} A y,$$

so we can expand (14) as

$$\begin{aligned} E = {} & \hat{x}^{\mathrm{T}} \Sigma_1^{-1} \hat{x} + {\mu_1}^{\mathrm{T}} \Sigma_1^{-1} \mu_1 - 2 \hat{x}^{\mathrm{T}} \Sigma_1^{-1} \mu_1 + \\ & \hat{x}^{\mathrm{T}} \Sigma_2^{-1} \hat{x} + {\mu_2}^{\mathrm{T}} \Sigma_2^{-1} \mu_2 - 2 \hat{x}^{\mathrm{T}} \Sigma_2^{-1} \mu_2 \\ = {} & \hat{x}^{\mathrm{T}} \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right) \hat{x} - 2 \hat{x}^{\mathrm{T}} \left( \Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 \right) + {\mu_1}^{\mathrm{T}} \Sigma_1^{-1} \mu_1 + {\mu_2}^{\mathrm{T}} \Sigma_2^{-1} \mu_2. \end{aligned}$$

Given ${\mu_1}^{\mathrm{T}} \Sigma_1^{-1} \mu_1 + {\mu_2}^{\mathrm{T}} \Sigma_2^{-1} \mu_2$ is constant, we can rewrite $E$ as

$$\begin{aligned} E &= \hat{x}^{\mathrm{T}} \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right) \hat{x} - 2 \hat{x}^{\mathrm{T}} \left( \Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 \right) + K \\ &= \hat{x}^{\mathrm{T}} \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right) \hat{x} - 2 \hat{x}^{\mathrm{T}} \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right) \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right)^{-1} \left( \Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 \right) + K \end{aligned}$$

If we let $\hat{\Sigma} = \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right)^{-1}$, then we have

$$E = \hat{x}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{x} - 2 \hat{x}^{\mathrm{T}} \hat{\Sigma}^{-1} \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right)^{-1} \left( \Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 \right) + K.$$

Again letting $\hat{\mu} = \left( \Sigma_1^{-1} + \Sigma_2^{-1} \right)^{-1} \left( \Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 \right)$, we find that

$$E = \hat{x}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{x} - 2 \hat{x}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{\mu} + K. \tag{4.3}$$

We now complete the square and see that

$$\begin{aligned} E &= \hat{x}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{x} - 2 \hat{x}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{\mu} + \hat{\mu}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{\mu} - \hat{\mu}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{\mu} + K \\ &= (\hat{x} - \hat{\mu})^{\mathrm{T}} \hat{\Sigma}^{-1} (\hat{x} - \hat{\mu}) - \hat{\mu}^{\mathrm{T}} \hat{\Sigma}^{-1} \hat{\mu} + K \\ &= (\hat{x} - \hat{\mu})^{\mathrm{T}} \hat{\Sigma}^{-1} (\hat{x} - \hat{\mu}) + K'. \end{aligned}$$

Then the probability density function of $\hat{x}$ given two measurements $\mu_1, \mu_2$ can be written as

$$p(\hat{x}, \mu, \Sigma) = \frac{Ne^{-\frac{1}{2}K'}}{(2\pi)^n |\Sigma_1|^{\frac{1}{2}} |\Sigma_2|^{\frac{1}{2}}} e^{-\frac{1}{2}(\hat{x}-\hat{\mu})^{\mathrm{T}} \hat{\Sigma}^{-1} (\hat{x}-\hat{\mu})} \tag{4.4}$$

where $N$ is the normalization constant found by integrating $p$ over $\mathbb{R}^n$. From this new Gaussian we see that our new mean is given by

$$\hat{\mu} = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1} \left(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right)$$

and our new covariance matrix is given by

$$\hat{\Sigma} = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1}$$

## 4.1 The Kalman Filter Equations

The equations we have previously derived for the combination of two multi-variate Gaussians are not the ones typically referenced in Kalman Filter literature. However, we will show that they are equivalent to the alternate form found in most literature. One reason for the popularity of the alternate form is that it is less expensive computationally, since it minimizes the number of computations of inverse matrices.

Recall that $\hat{\Sigma} = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1}$. If we invert $\hat{\Sigma}$, we see that

$$\begin{aligned}
\hat{\Sigma}^{-1} &= \Sigma_1^{-1} + \Sigma_2^{-1} \\
\hat{\Sigma}\hat{\Sigma}^{-1} &= \hat{\Sigma}\left(\Sigma_1^{-1} + \Sigma_2^{-1}\right) \\
I &= \hat{\Sigma}\left(\Sigma_1^{-1} + \Sigma_2^{-1}\right) \\
I &= \hat{\Sigma}\Sigma_1^{-1} + \hat{\Sigma}\Sigma_2^{-1} \\
I - \hat{\Sigma}\Sigma_2^{-1} &= \hat{\Sigma}\Sigma_1^{-1}.
\end{aligned}$$

If we allow $K = \hat{\Sigma}\Sigma_2^{-1}$, we can rewrite the equation as

$$\begin{aligned}
\hat{\Sigma}\Sigma_1^{-1} &= I - K \\
\hat{\Sigma} &= (I - K)\,\Sigma_1
\end{aligned} \tag{4.5}$$

We can see that $\hat{\Sigma}$ can now be written in terms of the gain factor $K$ and the covariance of the forecast $\Sigma_1$. All that remains is to simplify $K$, so it can be written strictly in terms of the known covariance matrices $\Sigma_1$ and $\Sigma_2$.

Given we defined $K$ to be $K = \hat{\Sigma}\Sigma_2^{-1}$,

$$
\begin{aligned}
K\Sigma_1 &= \hat{\Sigma} \\
\Sigma_1 + K\Sigma_1 &= \hat{\Sigma} + \Sigma_1 \\
\Sigma_1 - \hat{\Sigma} + K\Sigma_2 &= \Sigma_1 \\
\Sigma_1 - (I - K)\,\Sigma_1 + K\Sigma_2 &= \Sigma_1 \\
\Sigma_1 - I\Sigma_1 + K\Sigma_1 + K\Sigma_2 &= \Sigma_1 \\
K(\Sigma_1 + \Sigma_2) &= \Sigma_1 \\
K &= \Sigma_1 \left(\Sigma_1 + \Sigma_2\right)^{-1} \qquad (4.6)
\end{aligned}
$$

A similar process can be applied to our measurement update. Recall that

$$
\hat{\mu} = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1} \left(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right).
$$

Then

$$
\begin{aligned}
\hat{x} &= \hat{\Sigma}\left(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu 2\right) \\
\hat{\Sigma}^{-1}\hat{x} &= \Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2 \\
\hat{\Sigma}^{-1}\hat{x} &= \left(\Sigma_1^{-1} + \Sigma_2^{-1} - \Sigma_2^{-1}\right)\mu_1 + \Sigma_2^{-1}\mu_2 \\
\hat{\Sigma}^{-1}\hat{x} &= \left(\hat{\Sigma} - \Sigma_2^{-1}\right)\mu_1 + \Sigma_2^{-1}\mu_2 \\
\hat{\Sigma}^{-1}\hat{x} &= \hat{\Sigma}^{-1}\mu_1 - \Sigma_2^{-1}\mu_1 + \Sigma_2^{-1}\mu_2 \\
\hat{x} &= \mu_1 + \hat{\Sigma}\Sigma_2^{-1}\left(\mu_2 - \mu_1\right) \\
\hat{x} &= \mu_1 + K\left(\mu_2 - \mu_1\right). \qquad (4.7)
\end{aligned}
$$

Note that these equations for the Kalman gain and covariance estimates resemble the single dimensional equations 3.2 and 3.3. We can now describe and equal dimensional Kalman Filter by the equations in the following table.

| Prediction Step | |
|---|---|
| Forecast (a priori) state estimate | $\tilde{\mathbf{x}}_t = \mathbf{A}_t\hat{\mathbf{x}}_{t-1}$ |
| Forecast (a priori) estimate covariance | $\tilde{\Sigma}_t = \mathbf{A}_t\hat{\Sigma}_{t-1}\mathbf{A}_t^{\mathrm{T}} + \mathbf{Q}_t$ |
| **Update Step** | |
| Optimal Kalman gain | $\mathbf{K}_t = \tilde{\Sigma}_t^{\mathrm{T}}\left(\tilde{\Sigma}_t + \mathbf{R}_t\right)^{-1}$ |
| Updated (a posteriori) state estimate | $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t\left(\mathbf{m}_t - \tilde{\mathbf{x}}_t\right)$ |
| Updated (a posteriori) estimate covariance | $\hat{\Sigma}_t = (I - \mathbf{K}_t)\tilde{\Sigma}_t$ |

Table 4.1: Equal Dimensional Kalman Filter Equations

# Chapter 5

# The General Form Kalman Filter

If we consider a more general case of the Kalman Filter, in which the measurement space is a subspace of the state space, or $m \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^m$ with $n < m$, we see that our naive approach to the derivation of the filter no longer applies, as we cannot take the joint probability distribution of two vectors whose dimensions disagree.

In this case, we alter our initial assumption about the relation between $\mathbf{m}_t$ and $\mathbf{x}_t$ as follows

$$\mathbf{m}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t,$$

where $\mathbf{H}_t$ is the state translation matrix that relates the measurement space to the state space and $\mathbf{w}_t$ is the measurement error.

In this case, we choose a different approach, using matrix calculus. Similarly to our previous derivation, we assume that the optimal estimate can be derived from the a posteriori estimate as follows

$$\hat{\mathbf{x}}_t = \tilde{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{m}_t - \mathbf{H}_t\tilde{\mathbf{x}}_t). \tag{5.1}$$

$\mathbf{K}_t$ represents the weight of the newest measurement which minimizes the error.

One approach to minimizing the error is to minimize the variance of each component of $\mathbf{x}_t$. Recall that $[\Sigma_t]_{ii}$ is the variance of the $i$th component of $\mathbf{x}$. One way to minimize the variance of each component is to minimize the sum of the variances, or to minimize the trace of the covariance matrix $\Sigma_t$. Since the weight $\mathbf{K}_t$ is the only unknown, then our problem can be reduced to finding the value of $\mathbf{K}_t$ which minimizes the trace of $\Sigma_t$ (the sum of the elements along the diagonal of $\Sigma_t$).

Note that for the remainder of this section, we will suspend use of the time subscript $t$. In addition, we will use the familiar linear algebra notation, where $\sigma_{ij}$ represents the element in the $i$th row and $j$th column of the matrix $\Sigma$.

## 5.1 The Matrix Gradient

In order to minimize the error covariance matrix $\Sigma$ for each element of $\mathbf{K}$, given the solution form shown above (4.1), we need only set the derivative of the trace of $\Sigma, \mathrm{tr}(\Sigma)$ equal to 0. In order to simplify this calculation, we will introduce one form of the matrix derivative, the *matrix gradient.*

Let $F$ be a function operating on a vector space of dimension $n$. We define the gradient of $F$ as

$$\vec{\nabla} F(\mathbf{x}) = \sum_{i=1}^{n} \hat{x}_i \frac{\partial F}{\partial x_i} \tag{5.2}$$

$$= \begin{pmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{pmatrix} \tag{5.3}$$

where $\hat{x}_i$ represents a unit vector.

The gradient is a differential operator which maps the function $F$ to a vector whose components are the partial derivatives of $F$ with respect to each component of the variable $F$ operates on.

We can naturally extend this idea to a function $G$ which operates on an $m \times n$ vector space. The matrix gradient of $G$ with respect to an $m \times n$ matrix $\mathbf{A}$ can be defined as

$$\nabla_A G(\mathbf{A}) = \begin{pmatrix} \frac{\partial G}{\partial x_{11}} & \cdots & \frac{\partial G}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial G}{\partial x_{m1}} & \cdots & \frac{\partial G}{\partial x_{mn}} \end{pmatrix}. \tag{5.4}$$

We can simplify the calculation of the gradient of the trace of the covariance matrix by first developing some simple matrix-gradient rules for the trace of a square matrix. For this, we consider the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ to be the variable in question?, and the matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ to be constant.

First, we consider the gradient of the trace of a matrix with respect to itself. We can show that this results in the identity matrix $\mathbf{I}$.

**Lemma 2.** *Given a square matrix* $\mathbf{A}$ *of dimension* $n^2$*,*

$$\nabla_A tr(\mathbf{A}) = \mathbf{I}.$$

*Proof.* Recall that $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$. Then the $i, j$th component of $\nabla A \mathrm{tr}(\mathbf{A})$ with respect to $\mathbf{A}$

can be expressed component wise as

$$\begin{aligned}[\nabla_A \mathrm{tr}(\mathbf{A})]_{ij} =& \frac{\partial \mathrm{tr}(\mathbf{A})}{\partial a_{ij}} \\ &= \frac{\partial}{\partial a_{ij}} \sum_{k=0}^{n} a_{kk} \\ &= \sum_{k=0}^{n} \frac{\partial}{\partial a_{ij}} a_{kk} \\ &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}\end{aligned}$$

□

**Corollary 1.** *Given a square matrix* $\mathbf{A}$ *of dimension* $n^2$*,*

$$\nabla_A tr(\mathbf{A}^{\mathrm{T}}) = \mathbf{I}.$$

Next we consider the gradient of the trace of the product of a matrix with a constant matrix.

**Lemma 3.** *Given a square matrix* $\mathbf{A}$ *and constant matrix* $\mathbf{B}$ *of dimension* $n^2$*,*

$$\nabla_A \ tr\ (\mathbf{AB}) = \mathbf{B}^{\mathrm{T}}.$$

*Proof.* Recall that $[\mathbf{AB}]_{ij} = \sum_{k=1}^{n} \mathbf{A}_{ik}\mathbf{B}_{kj}$. Then we have

$$\mathrm{tr}(\mathbf{AB}) = \sum_{l=1}^{n}\sum_{k=1}^{n} a_{kl}b_{lk},$$

and

$$\begin{aligned}[\nabla_A \ \mathrm{tr}\ (\mathbf{AB})]_{ij} =& \frac{\partial \ \mathrm{tr}\ (\mathbf{AB})}{\partial a_{ij}} \\ &= \frac{\partial}{\partial a_{ij}} \sum_{l=1}^{n}\sum_{k=1}^{n} a_{kl}b_{lk} \\ &= \sum_{l=1}^{n}\sum_{k=1}^{n} \frac{\partial}{\partial a_{ij}} a_{kl}b_{lk}\end{aligned}$$

As above, we have

$$\frac{\partial}{\partial a_{ij}} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Thus

$$\begin{aligned} \sum_{l=1}^{n}\sum_{k=1}^{n} \frac{\partial}{\partial a_{ij}} a_{kl} b_{lk} &= \frac{\partial}{\partial a_{ij}} a_{ij} b_{ji} \\ &= b_{ji} \\ &= [\mathbf{B}^{\mathrm{T}}]_{ij} \end{aligned}$$

□

Similarly, we consider the gradient of the product of a constant matrix and a transposed matrix.

**Lemma 4.** *Given a square matrix* $\mathbf{A}$ *and a constant matrix* $B$ *of dimension* $n^2$*,*

$$\nabla_A \; tr\,(\mathbf{B}\mathbf{A}^{\mathrm{T}}) = \mathbf{B}.$$

*Proof.* Similarly to lemma 3, we have

$$\mathrm{tr}(\mathbf{B}\mathbf{A}^{\mathrm{T}}) = \sum_{l=1}^{n}\sum_{k=1}^{n} b_{kl} a_{kl},$$

and

$$\begin{aligned} [\nabla_A \; \mathrm{tr}\; (\mathbf{B}\mathbf{A}^{\mathrm{T}})]_{ij} =& \frac{\partial \; \mathrm{tr}\; (\mathbf{B}\mathbf{A}^{\mathrm{T}})}{\partial a_{ij}} \\ &= \frac{\partial}{\partial a_{ij}} \sum_{l=1}^{n}\sum_{k=1}^{n} b_{kl} a_{kl} \\ &= \sum_{l=1}^{n}\sum_{k=1}^{n} \frac{\partial}{\partial a_{ij}} b_{kl} a_{kl} \\ &= \frac{\partial}{\partial a_{ij}} b_{ij} a_{ij} \\ &= b_{ij} \\ &= [\mathbf{B}]_{ij} \end{aligned}$$

□

The final lemma we will use is the gradient of the quadratic form of a matrix.

**Lemma 5.** *Given a square matrix* $\mathbf{A}$ *and a constant matrix* $\mathbf{B}$ *of dimensions* $n^2$*,*

$$\nabla_A \; tr\left(\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A}\right) = \mathbf{B}\mathbf{A} + \mathbf{B}^{\mathrm{T}}\mathbf{A}.$$

*Proof.* First, note that

$$[\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A})]_{ij} = \sum_{l=1}^{n}\sum_{k=1}^{n} a_{ki}b_{kl}a_{lj}$$

and

$$\text{tr }(\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A})]_{ij} = \sum_{m=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n} a_{km}b_{kl}a_{lm}.$$

Then we have

$$\begin{aligned}
[\nabla_A \text{ tr } (\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A})]_{ij} =& \frac{\partial \text{ tr } (\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A})}{\partial a_{ij}} \\
&= \frac{\partial}{\partial a_{ij}} \sum_{m=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n} a_{km}b_{kl}a_{lm} \\
&= \sum_{m=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n} \frac{\partial}{\partial a_{ij}} a_{km}b_{kl}a_{lm}
\end{aligned}$$

Similarly to lemma 2, we find that

$$\frac{\partial}{\partial a_{ij}} a_{km}b_{kl}a_{lm} = \begin{cases} 2a_{ij}b_{ii} & \text{if } k = l = i, m = j \\ \sum\limits_{l=1, l\neq i}^{n} b_{il}a_{lj} & \text{if } k = i, m = j, l \neq i \\ \sum\limits_{k=1, k\neq i}^{n} a_{kj}b_{ki} & \text{if } l = i, m = j, k \neq i \\ 0 & else \end{cases}$$

By rearranging terms, we can see that

$$\begin{aligned}
\sum_{m=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n} \frac{\partial}{\partial a_{ij}} a_{km}b_{kl}a_{lm} &= \sum_{l=1, l\neq i}^{n} b_{il}a_{lj} + \sum_{k=1, k\neq i}^{n} a_{kj}b_{ki} + 2a_{ij}b_{ii} \\
&= \sum_{l=1}^{n} b_{il}a_{lj} + \sum_{k=1}^{n} a_{kj}b_{ki} \\
&= \sum_{l=1}^{n} b_{il}a_{lj} + \sum_{k=1}^{n} b_{ki}a_{kj} \\
&= (\mathbf{B}\mathbf{A})_{ij} + (\mathbf{B}^{\mathrm{T}}\mathbf{A})_{ij}.
\end{aligned}$$

□

**Corollary 2.** *Given a matrix* $\mathbf{A}$ *and a symmetric constant matrix* $\mathbf{B}$ *of dimension* $n^2$*,*

$$\nabla_A \; tr\,(\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A}) = 2\mathbf{B}\mathbf{A}.$$

## 5.2 Minimizing the Variance

As we described in equation 1.1 in the introduction, we make the assumption that state of the system at time $t$ is related to our estimate of the state at time $t-1$ by the following equation

$$\mathbf{x}_t = \mathbf{A}\hat{\mathbf{x}}_{t-1} + \mathbf{v}.$$

We also assume that the optimal estimate for the state of the system given both our forecast $\tilde{\mathbf{x}}_t$ and our measurement $\mathbf{m}_t$ has the same form as in the equal dimensional case, with one notable exception, the inclusion of the state translation matrix $\mathbf{H}_t$. Thus our estimate can be determined from

$$\hat{\mathbf{x}}_t = \tilde{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{m}_t - \mathbf{H}_t\tilde{\mathbf{x}}_t).$$

Let $\hat{\mathbf{e}}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$ and $\tilde{\mathbf{e}}_t = \mathbf{x}_t - \tilde{\mathbf{x}}_t$ be the estimate error and forecast error, respectively.

The definition of a covariance matrix of a random $\mathbf{x}$ is $\Sigma = E[\hat{\mathbf{e}}\hat{\mathbf{e}}^{\mathrm{T}}]$, where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. From above we see that

$$\begin{aligned}
\hat{\mathbf{e}}_t &= \mathbf{x}_t - \hat{\mathbf{x}}_t \\
&= \mathbf{x}_t - \tilde{\mathbf{x}}_t - \mathbf{K}_t(\mathbf{m}_t - \mathbf{H}_t\tilde{\mathbf{x}}_t) \\
&= \mathbf{x}_t - (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{x}}_t - \mathbf{K}_t\mathbf{m}_t \\
&= \mathbf{x}_t - (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{x}}_t - \mathbf{K}_t\mathbf{H}_t\mathbf{x}_t - \mathbf{K}_t\mathbf{w}_t \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)(\mathbf{x}_t - \tilde{\mathbf{x}}_t) - \mathbf{K}_t\mathbf{w}_t \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t - \mathbf{K}_t\mathbf{w}_t
\end{aligned}$$

And

$$\begin{aligned}
\hat{\mathbf{e}}_t\hat{\mathbf{e}}_t^{\mathrm{T}} &= ((\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t - \mathbf{K}_t\mathbf{w}_t)((\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}_t} - \mathbf{K}_t\mathbf{w}_t)^{\mathrm{T}} \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t^{\mathrm{T}}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_t\mathbf{w}_t\mathbf{w}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}} \\
&\quad - (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\mathbf{v}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}} - \mathbf{K}_t\mathbf{v}_t(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}_t}.
\end{aligned}$$

Since, given two independent random variables $a$ and $b$, $E[a+b] = E[a] + E[b]$, we find that

$$\begin{aligned}
\hat{\Sigma_t} &= E[\hat{\mathbf{e}}_t \hat{\mathbf{e}}_t^{\mathrm{T}}] \\
&= E[(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t^{\mathrm{T}}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_t\mathbf{w}_t\mathbf{w}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}} \\
&\quad - (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\mathbf{w}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}} - \mathbf{K}_t\mathbf{w}_t(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}_t}] \\
&= E[(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t^{\mathrm{T}}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}}] + E[\mathbf{K}_t\mathbf{w}_t\mathbf{w}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}}] \\
&\quad - E[(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\mathbf{w}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}}] - E[\mathbf{K}_t\mathbf{w}_t(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}_t}]
\end{aligned}$$

Recalling our assumption that the error is Gaussian with mean zero, we conclude that

$$\begin{aligned}
\hat{\Sigma_t} &= E[(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t^{\mathrm{T}}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}}] + E[\mathbf{K}_t\mathbf{w}_t\mathbf{w}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}}] \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)E[\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t^{\mathrm{T}}](\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_tE[\mathbf{w}_t\mathbf{w}_t^{\mathrm{T}}]\mathbf{K}_t^{\mathrm{T}} \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\tilde{\Sigma}_t(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^{\mathrm{T}} \\
&= \tilde{\Sigma}_t - \mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t - \tilde{\Sigma}_t(\mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}} + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^{\mathrm{T}} \\
&= \tilde{\Sigma}_t - \mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t - \tilde{\Sigma}_t(\mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_t(\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{R})K_t^{\mathrm{T}} \\
&= \tilde{\Sigma}_t - \mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t - \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}} + \mathbf{K}_t(\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{R})K_t^{\mathrm{T}} \qquad (5.5)
\end{aligned}$$

In order find the $\mathbf{K}_t$ that minimizes $\hat{\Sigma}_t$, we need only take the gradient of $\hat{\Sigma}_t$ with respect to $\mathbf{K}_t$ equal to zero and solve. We see that each term in equation 4.5 has the form of one of the gradient lemmas we derived, and so we can take the gradient as follows:

$$\begin{aligned}
\nabla_{\mathbf{K}_t}\,\mathrm{tr}\hat{\Sigma}_t &= 0 - \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} - \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + 2(\mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}) \\
&= \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} - \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + 2\mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + 2\mathbf{K}_t\mathbf{R} \qquad (5.6)
\end{aligned}$$

To find the minimum of the trace, we simply set equation 4.6 equal to zero and solve, and we find that

$$\begin{aligned}
0 &= -2\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + 2\mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + 2\mathbf{K}_t\mathbf{R} \\
&= -\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{K}_t\mathbf{R} \\
\mathbf{K}_t\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{K}_t\mathbf{R} &= \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} \\
\mathbf{K}_t &= \tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\tilde{\Sigma}_t\mathbf{H}_t^{\mathrm{T}} + \mathbf{R})^{-1} \qquad (5.7)
\end{aligned}$$

| Prediction Step | |
|---|---|
| Forecast (a priori) state estimate | $\tilde{\mathbf{x}}_t = \mathbf{A}_t \hat{\mathbf{x}}_{t-1}$ |
| Forecast (a priori) estimate covariance | $\tilde{\Sigma}_t = \mathbf{A}_t \hat{\Sigma}_{t-1} \mathbf{A}_t^{\mathrm{T}} + \mathbf{Q}_t$ |
| Update Step | |
| Optimal Kalman gain | $\mathbf{K}_t = \tilde{\Sigma}_t \mathbf{H}_t^{\mathrm{T}} \left( \mathbf{H}_t \tilde{\Sigma}_t \mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t \right)^{-1}$ |
| Updated (a posteriori) state estimate | $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t \left( \mathbf{m}_t - \mathbf{H}_t \tilde{\mathbf{x}}_t \right)$ |
| Updated (a posteriori) estimate covariance | $\hat{\Sigma}_t = (I - \mathbf{K}_t \mathbf{H}_t) \tilde{\Sigma}_t$ |

Table 5.1: Multi Dimensional Kalman Filter Equations

## 5.3 Example

Let us now consider the case of movement in a single dimension with constant velocity. In this case, we measure the current position with some error, although the state of our system will consist of both the current position and current velocity. The measurement will consist of a single value, $x$, while the state can be represented by a matrix which appears as

$$\hat{x} = \begin{vmatrix} x \\ v_x \end{vmatrix}.$$

Since we assume constant velocity, our state transition matrix will be

$$A = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}.$$

In this example, the dimensions of the state and measurement do not agree, so we introduce the observation matrix $H$, where

$$H = \begin{vmatrix} 1 & 0 \end{vmatrix}.$$

This matrix will transform our state matrix into a measurement vector.

Because our current position is statically independent from the velocity, our we choose the initial estimate of our process covariance to be a diagonal matrix. In addition, since we are confident that our system behaves with constant velocity, we choose variances for the transition of $x$ and $v_x$ close to zero. In this example, we have chosen

$$Q = \begin{vmatrix} 10^{-5} & 0 \\ 0 & 10^{-5} \end{vmatrix}.$$

Because we do not know what the initial position of $x$ is, we make an initial guess of $x_0 = 0$, and initialize the measurement error covariance (in this case variance) to be relatively large. Here we choose $\Sigma_{m_0} = |1|$.

Finally, since our initial estimate of the state of the system is random, we expect it to differ greatly from the true initial state. For this reason, we choose an initial estimate of the covariance of the system state to to be diagonal (since $x$ and $v_x$ are independent), with a value of

$$\widehat{\Sigma}_0 = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}$$

.

To generate data for this example, we choose a true velocity of $V_x = \frac{1}{2}$. At each step, we update the true position of $x$ by $x_t = x_{t-1} + \frac{1}{2}$. To generate our erroneous measurement, we adjust $x_t$ by adding a random Gaussian variable centered at 0 with a moderate variance of 1. Below is example output for 60 steps, with the first ten listed numerically.

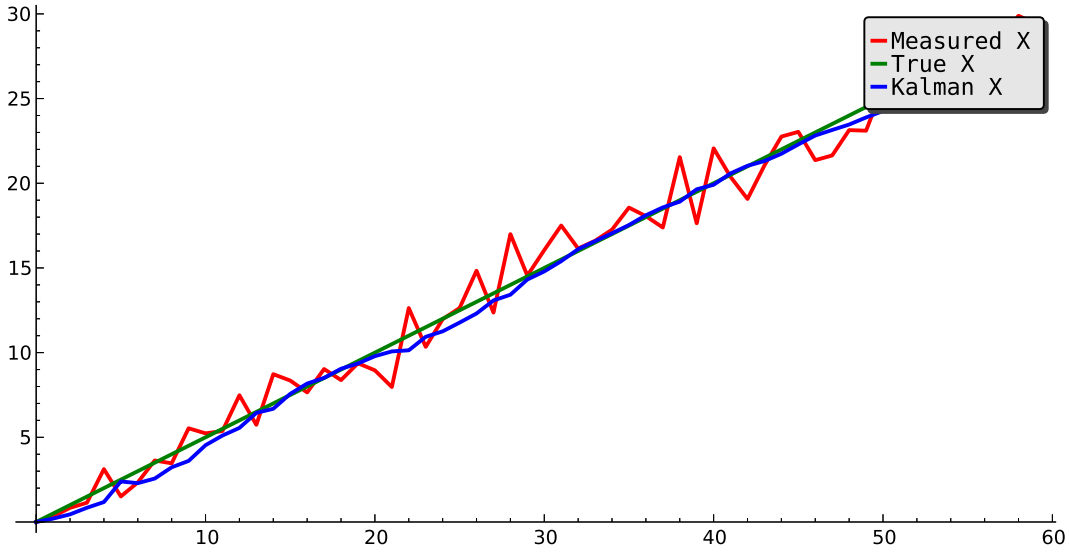


Figure 5.1: Simulated position and estimate from $t = 0$ to $t = 60$

| Measurement | Forecast | Estimate | Gain | Forecast covariance | Estimate covariance |
|---|---|---|---|---|---|
| 0.000 | $\begin{vmatrix} 1.000 \\ 1.000 \end{vmatrix}$ | $\begin{vmatrix} 0.000 \\ 1.000 \end{vmatrix}$ | $\begin{vmatrix} 0.800 \\ 0.400 \end{vmatrix}$ | $\begin{vmatrix} 4.000 & 2.000 \\ 2.000 & 2.000 \end{vmatrix}$ | $\begin{vmatrix} 0.800 & 0.400 \\ 0.400 & 1.200 \end{vmatrix}$ |
| 0.328 | $\begin{vmatrix} 0.800 \\ 0.600 \end{vmatrix}$ | $\begin{vmatrix} 0.200 \\ 0.600 \end{vmatrix}$ | $\begin{vmatrix} 0.737 \\ 0.421 \end{vmatrix}$ | $\begin{vmatrix} 2.800 & 1.600 \\ 1.600 & 1.200 \end{vmatrix}$ | $\begin{vmatrix} 0.737 & 0.421 \\ 0.421 & 0.526 \end{vmatrix}$ |
| 0.836 | $\begin{vmatrix} 0.853 \\ 0.401 \end{vmatrix}$ | $\begin{vmatrix} 0.452 \\ 0.401 \end{vmatrix}$ | $\begin{vmatrix} 0.678 \\ 0.305 \end{vmatrix}$ | $\begin{vmatrix} 2.105 & 0.947 \\ 0.947 & 0.526 \end{vmatrix}$ | $\begin{vmatrix} 0.678 & 0.305 \\ 0.305 & 0.237 \end{vmatrix}$ |
| 1.138 | $\begin{vmatrix} 1.237 \\ 0.396 \end{vmatrix}$ | $\begin{vmatrix} 0.841 \\ 0.396 \end{vmatrix}$ | $\begin{vmatrix} 0.604 \\ 0.215 \end{vmatrix}$ | $\begin{vmatrix} 1.525 & 0.542 \\ 0.542 & 0.237 \end{vmatrix}$ | $\begin{vmatrix} 0.604 & 0.215 \\ 0.215 & 0.121 \end{vmatrix}$ |
| 3.122 | $\begin{vmatrix} 1.552 \\ 0.375 \end{vmatrix}$ | $\begin{vmatrix} 1.178 \\ 0.375 \end{vmatrix}$ | $\begin{vmatrix} 0.536 \\ 0.156 \end{vmatrix}$ | $\begin{vmatrix} 1.154 & 0.336 \\ 0.336 & 0.121 \end{vmatrix}$ | $\begin{vmatrix} 0.536 & 0.156 \\ 0.156 & 0.069 \end{vmatrix}$ |
| 1.507 | $\begin{vmatrix} 3.013 \\ 0.619 \end{vmatrix}$ | $\begin{vmatrix} 2.394 \\ 0.619 \end{vmatrix}$ | $\begin{vmatrix} 0.478 \\ 0.117 \end{vmatrix}$ | $\begin{vmatrix} 0.916 & 0.224 \\ 0.224 & 0.069 \end{vmatrix}$ | $\begin{vmatrix} 0.478 & 0.117 \\ 0.117 & 0.042 \end{vmatrix}$ |
| 2.337 | $\begin{vmatrix} 2.736 \\ 0.443 \end{vmatrix}$ | $\begin{vmatrix} 2.293 \\ 0.443 \end{vmatrix}$ | $\begin{vmatrix} 0.430 \\ 0.091 \end{vmatrix}$ | $\begin{vmatrix} 0.755 & 0.159 \\ 0.159 & 0.042 \end{vmatrix}$ | $\begin{vmatrix} 0.430 & 0.091 \\ 0.091 & 0.028 \end{vmatrix}$ |
| 3.632 | $\begin{vmatrix} 2.971 \\ 0.407 \end{vmatrix}$ | $\begin{vmatrix} 2.565 \\ 0.407 \end{vmatrix}$ | $\begin{vmatrix} 0.390 \\ 0.072 \end{vmatrix}$ | $\begin{vmatrix} 0.640 & 0.119 \\ 0.119 & 0.028 \end{vmatrix}$ | $\begin{vmatrix} 0.390 & 0.072 \\ 0.072 & 0.019 \end{vmatrix}$ |
| 3.464 | $\begin{vmatrix} 3.684 \\ 0.455 \end{vmatrix}$ | $\begin{vmatrix} 3.229 \\ 0.455 \end{vmatrix}$ | $\begin{vmatrix} 0.357 \\ 0.059 \end{vmatrix}$ | $\begin{vmatrix} 0.554 & 0.092 \\ 0.092 & 0.019 \end{vmatrix}$ | $\begin{vmatrix} 0.357 & 0.059 \\ 0.059 & 0.014 \end{vmatrix}$ |
| 5.532 | $\begin{vmatrix} 4.047 \\ 0.442 \end{vmatrix}$ | $\begin{vmatrix} 3.605 \\ 0.442 \end{vmatrix}$ | $\begin{vmatrix} 0.328 \\ 0.049 \end{vmatrix}$ | $\begin{vmatrix} 0.488 & 0.073 \\ 0.073 & 0.014 \end{vmatrix}$ | $\begin{vmatrix} 0.328 & 0.049 \\ 0.049 & 0.010 \end{vmatrix}$ |

Table 5.2: Kalman Filter simulated values from $t = 0$ to $t = 10$

# Bibliography


[1] Kalman, Rudolph Emil. *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME - Journal of Basic Engineering, Volume 82, pp 35 - 45, 1960.

[2] Jessica R. Briggs and Matina C. Kalcounis-Rueppell. *Similar acoustic structure and behavioral context of vocalizations produced by male and female California mice in the wild*, Animal Behavior, 2011, `http://www.sciencedirect.com/science/article/pii/S0003347211003836`

[3] Matina Kalcounis-Rueppell, Thomas Parrish, and Sebastian Pauli. *Application of Object Tracking in Video Recordings to the Observation of Mice in the Wild*, 2013, Springer Proceedings in Mathematics & Statistics - Topics from the 8th Annual UNCG Regional Mathematics and Statistics Conference.

[4] Maybeck, Peter S. *Stochastic Models, Estimation, and Control, Volume 1*, 1979 , Academic Press, Inc.

[5] Welch, Greg and Bishop, Gary. 2006, *An Introduction to the Kalman Filter*, `http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf`

[6] Perreault, Brent. 2012, *Introduction to the Kalman Filter and its Derivation*, `http://www.academia.edu/1512888/Introduction_to_the_Kalman_Filter_and_its_Derivation`