Covariance Intersection for Continuous Kalman Filters

by

Josephine Schikhof

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Wednesday November 20, 2024 at 10:00 AM.

Student number:	4944321	
Project duration:	January 4, 2024 – November 20, 2024	
Thesis committee:	Prof. dr. F. H. J. Redig	TU Delft
	Dr. R. C. Kraaij	TU Delft
	Dr. J. Söhl	TU Delft

TU Delft

Summary

The Kalman filter is a recursive algorithm that estimates the state of a dynamic system subject to measurement and model noise. If all noise terms affecting the system are white Gaussian noise with known mean and variance, and all noise terms are independent of each other, then the Kalman filter is the optimal estimator for the state variable. When measurements are collected from multiple sources, the covariance between these sources should be known or the sources should be independent to ensure that the estimate made by the Kalman filter is optimal. When the covariance between dependent measurement sources is not known, various methods exist which provide a solution to this problem. This thesis discusses two methods: the H_∞ filter and covariance intersection.

1. The H_∞ filter can be used instead of the Kalman filter, and is a generalisation of the Kalman filter. The H_∞ filter does not pose any constraints on the noise, therefore the correlation structure between the measurement sources does not need to be known. The H_∞ filter is not guaranteed to be optimal, but the estimation error is bounded.
2. Covariance intersection is used in addition to the Kalman filter and consists of two steps. First, each measurement source is used to calculate a partial estimate of the state. Second, these estimates and their covariances are fused using a linear combination. This method is consistent, which means that the estimated covariance matrix minus the error covariance matrix is positive semi-definite. This prevents overconfidence in the estimate of the state. Covariance intersection also converges as long as the constant controlling the linear combination is optimised for every time step.

The Kalman filter and the H_∞ filter can be formulated in both discrete and continuous time. However, covariance intersection is only suited for the discrete time Kalman filter, because the constant controlling the linear combination between the the partial estimates must be optimised for every time step to ensure convergence. The overall aim of this thesis is to formulate covariance intersection such that it can be applied to the continuous Kalman filter. This aim is partially achieved. To this end, three different topics are discussed in this thesis:

1. The proof showing the consistency of covariance intersection, first given by S. J. Julier and J. K. Uhlmann, is investigated. This proof makes the assumption that a Kalman filter step with information from only one measurement source is consistent, but this is neither trivial nor proven. This work highlights this assumption and explores it. Additionally, the proof given by S. J. Julier and J. K. Uhlmann does not use conditional expectations and conditional covariances, even though the estimates that are being fused are conditional expectations. This work contains a reformulation of the proof in the literature so that it does include conditional expectations and conditional covariances where appropriate.
2. Covariance intersection is reformulated to fuse the covariance matrices of the measurement noise instead of the partial estimates of the state and their covariances, without changing the result. This means that covariance intersection can be rewritten as a Kalman filter with slightly altered input. This alternative formulation therefore removes the need to calculate partial estimates. This reduces the number of Kalman filter applications from one per measurement source to one in total per time step when applying covariance intersection to the Kalman filter. Since a continuous Kalman filter exists, this fusion step can be included into the continuous Kalman filter. However, the constant that controls the linear combination still needs to be optimised at every time step to ensure that the algorithm converges. This work suggests a starting point for further research aimed at finding a method to optimise the constant in continuous time.
3. This work also includes a H_∞ generalisation of the Kalman filter with covariance intersection applied to it. This has no practical use because it imposes an additional constraint on the inputs of the H_∞ filter, limiting the opportunity for tuning the filter without improving the estimate. However, because the H_∞ filter can also be formulated in continuous time, it offers another avenue to make covariance intersection applicable to the continuous time Kalman filter.

Contents

1

Introduction

At the start of the Second World War, a military project was set up in the USA to develop an estimator that could predict the future position of fast-moving aircraft using noisy radar tracking data. The aircraft would drops bombs over targets while flying at high speed. To improve the hit rate, it was necessary to know the future position of the aircraft so that this could be taken into account when aiming. Norbert Wiener, as part of the programme, was tasked with developing an optimal estimator of the future position of aircraft. What he came up with was an estimator that minimised the mean squared prediction error using power spectral densities to characterise the statistical properties of the dynamic system. Andrei Kolomogorov developed the same theory around the same time, so this estimator became known as the Wiener-Kolomogorov filter. In the late 1950s, Rudolf Kalman built on this foundation. During a train journey, he came up with the idea of using state variables in the Wiener-Kolmogorov filter. State variables describe the dynamics of the system whose future state is to be estimated. The state variables describe the system without external forces, such as measurements. This idea led to the development of the Kalman filter [1]. The Kalman filter was first used in the Apollo space programme in the early 1960s [13]. Since then, the Kalman filter has been applied to other fields and is particularly widely used in the field of navigation [1].

In terms of terminology, the term 'filter' may seem unusual for an estimator. Historically, filters separated unwanted components of gas-liquid-solid mixtures and later referred to analogue circuits that filtered electronic signals by frequency. In the 1930s and 1940s, the concept was extended to the separation of signal from noise using power spectral densities developed by Kolmogorov and Wiener. With Kalman filtering, the term evolved to include the solution of inversion problems that estimate the independent variables as an inverted function of the measurements [1].

The Kalman filter is a recursive estimator that minimises the mean square prediction error and is used to estimate the (future) state of a dynamic system. The dynamic system consists of a state variable and measurements. The state variable is modelled with an additional white noise term, which accounts for undescribed model characteristics. The measurements are a linear function of the state variable perturbed by white noise. The recursive estimation process consists of two steps: first, a prediction is made based on the current state. Second, the state is updated to incorporate the latest measurement to correct the prediction. If the noise of the dynamic system and the measurements are Gaussian, then the Kalman filter is the optimal estimator. However, if the noise is not Gaussian, then the Kalman filter is the best *linear* estimator. Note that the noise of the dynamic system and the noise of the measurements are assumed to be independent of each other [2]. The Kalman filter also has a continuous time form, but this is rarely used in practice because modern measurement equipment produces discrete measurements. The continuous Kalman filter is still useful, however, because in some cases it is easier to analyse the effects of mathematical variations on the design of the filter in its continuous form than in its discrete one.

The Kalman filter can also incorporate multiple measurement sources, but the full covariance matrix between these sources should be known, or preferably the sources should be independent. For example, the acceleration of a vehicle can be measured using the Doppler effect. Another way of measuring its acceleration would be to carry a weight, suspended on springs, on the vehicle, and measuring the compression of the springs. Because these two measurements do not share the same error source, they are independent and can therefore be used in the Kalman filter.

In practice, errors in measurements from different sources are often correlated, and the full correlation structure is usually not known. An example is a GPS location calculation by two different devices that are located in the same place, using an overlapping set of satellites. If the data from one of the shared satellites is flawed because of, for example, atmospheric interference, the location estimate produced by both devices will have a correlated error. Many solutions of varying rigour and efficiency have been proposed. I will highlight three methods of dealing with measurement correlations: tuning of the covariance matrix, the H_∞ filter, and covariance intersection.

1. The most rudimentary solution of these three is tuning of the covariance matrix of the estimate. It does so by treating the measurements as independent when applying the Kalman filter. Having dependent measurements means that part of one measurement is 'contained' in the other measurement. This means that having dependent measurements generates less information than independent measurements. Thus assuming that the measurements are independent while they are actually dependent fictitiously increases the amount of information available. This makes the covariance matrix of the estimate smaller than it should be. The underestimation of the covariance matrix can be remedied by tuning the covariance matrix, which will ensure consistency of the algorithm under certain conditions. Consistency means than the estimated covariance matrix minus the error covariance matrix is positive definite. Thus ensuring that the confidence in the estimate of the state is not unreasonably large. But the filter can no longer be considered optimal and does not guarantee that the estimate is consistent in general [8].
2. Another solution to the problem of unknown correlated measurements is the use of the H_∞ filter. This filter can be seen as a generalisation of the Kalman filter. It makes no assumptions about the noise and therefore it does not matter whether the noise is correlated or not. The downside to this relaxation of assumptions is that the filter is no longer guaranteed to be optimal. Instead, the maximum error of the estimate is bounded. The H_∞ filter contains a trade-off. It allows the user to consider whether it is better to reduce the covariance matrix of the estimate by increasing the bound on the maximum error or to reduce the maximum error at the cost of a larger covariance matrix of the estimate.
3. The last solution discussed in this thesis that can be used to deal with unknown correlations between measurements is covariance intersection, introduced by Simon Julier and Jeffrey Uhlmann. Covariance intersection consists of two steps. First, partial estimates per dependent measurement source are calculated, using the Kalman filter. Second, these estimates are fused using a linear combination. This means that the number of applications of the Kalman filter scales with the number of dependent error sources. The constant controlling the linear combination needs to be optimised at each time step to ensure non-divergence [7]. The resulting estimate is consistent [9].

Fusing the partial estimates for every time step, as in covariance intersection, clearly requires the existence of discrete time steps. Covariance intersection in its current form can therefore not be applied to the continuous Kalman filter. The overall aim of this thesis is to formulate covariance intersection in such a way that it can be applied to the continuous Kalman filter. To this end, three topics are explored. First, this thesis investigates the proof given by S. J. Julier and J. K. Uhlmann which proves that covariance intersection is consistent. This thesis points out two issues with this proof.

1. Julier and Uhlmann's proof does not use conditional expectations even though all estimates and covariance matrices in the Kalman filter are conditional expectations. This thesis contains a rewritten proof that does include conditional expectations where appropriate.
2. Julier and Uhlmann's proof assumes that the partial estimates are all consistent. This assumption is neither trivial nor proven. This thesis highlights this assumption and shows why this assumption does not hold in general. An alternative assumption is proposed: that the partial estimates are consistent in expectation over many time steps, rather than for each time step individually.

The second topic this thesis addresses is a reformulation of covariance intersection. This reformulation fuses the covariance matrices of the measurements noise rather than the covariance matrices of the estimates. This means that the Kalman filter only has to be applied once to this fused covariance matrix instead of as often as there are dependent measurement sources, while the result remains the same. This reformulation allows covariance intersection - except for the optimising of the constant that controls the linear combination - to be written into the algorithm of the Kalman filter. This means that the fusion of the partial estimates can be translated to the continuous time Kalman filter if a differential equation of the constant is provided. However, the optimisation problem that needs to be solved for every time step has not yet been reformulated such that it is possible to find a differential equation for this constant. A starting point for further research is suggested to also solve this last barrier to applying covariance intersection to the continuous time Kalman filter.

The last topic discussed in this thesis is how the H_∞ filter can be adjusted such that it becomes a generalisation of the Kalman filter with covariance intersection applied to it. This adjustment impairs the performance of the H_∞, because additional requirements are posed on the input matrices of the H_∞ filter. However, this adjustment provides a different way of arriving at a continuous Kalman filter with covariance intersection applied to it, as there also exists a continuous time version of the H_∞ filter.

The outline of this thesis is as follows. Chapter 2 provides the mathematical background of both the discrete and continuous time Kalman filter. Section 2.1 derives the discrete Kalman filter. This section also contains several frequently used formulations of the Kalman filter. Section 2.2 derives the continuous time Kalman filter as a limit of the discrete time Kalman filter. Chapter 3 introduces the H_∞ filter; the discrete form can be found in Section 3.1 and the continuous form in Section 3.2. The Kalman filter and the H_∞ filter are compared in Section 3.3. Section 3.4 investigates whether the H_∞ filter can be seen as a Kalman filter with non zero noise terms. This does not turn out to be possible. Chapter 4 focusses on the subject of covariance intersection, starting with an introduction to the concept in Section 4.1. Section 4.2 gives the definition of consistency and investigates the proof of the consistency of covariance intersection, given by S. J. Julier and J. K. Uhlmann. This section also highlights and discusses the assumption that the partial estimates used in covariance intersection are consistent. Section 4.3 formulates a method which fuses the covariance matrices of the measurement noise instead of the covariance matrices of the estimates, as covariance intersection does. It is shown that this formulation leads to the same estimate as covariance intersection, provided that the constant which controls the linear combination is the same. Chapter 5 adjusts the H_∞ filter such that the H_∞ filter becomes a generalisation of the Kalman filter with covariance intersection applied to it. This provides a different path to make covariance intersection applicable to the continuous time Kalman filter. The conclusion of this work is contained in Chapter 6. Section 6.1 contains recommendations for future work.

2
Kalman filter

This chapter contains a mathematical derivation of both the discrete and the continuous Kalman filter. The discrete Kalman filter will be derived using the Bayes' method combined with moment generating functions. The discrete Kalman filter is a recursive algorithm consisting of two steps. The first step is the prediction step, which is derived in Section 2.1. The second step is the update step. To derive this step, some background on moment generating functions is needed which is provided in Subsection 2.1.1. Subsection 2.1.2 derives the update step of the Kalman filter. Subsection 2.1.3 contains several reformulations of the solution of the Kalman filter, which will be used throughout this thesis. The continuous Kalman filter is derived in Section 2.2 as a limit of a discrete Kalman filter.

2.1. Derivation of discrete Kalman filter

There are many ways to derive the discrete Kalman filter. This section will the derive the discrete Kalman filter by using Bayes' theorem and moment generating functions, which is based on a paper written by H. M. Masnadi-Shirazi et al. [12]. We consider the following dynamic system:

$$\begin{aligned} X_k &= F_{k-1}X_{k-1} + W_{k-1}, \\ Y_{k-1} &= H_{k-1}X_{k-1} + V_{k-1}, \end{aligned} \tag{2.1}$$

where $X_k \in \mathbb{R}^n$ is the state vector, this vector is a random variable; $F_k \in \mathbb{R}^n \times \mathbb{R}^n$ is the state transition matrix, which can describe the physics underlying the process that is modelled; $W_k \in \mathbb{R}^n$ is Gaussian white noise with covariance matrix $Q_k \in \mathbb{R}^n \times \mathbb{R}^n$. Therefore, $W_k \sim N(0, Q_k)$. W_k can be used to compensate for unmodelled factors. $Y_k \in \mathbb{R}^m$ represents the measurement at time k, $H_k \in \mathbb{R}^m \times \mathbb{R}^n$ provides a linear connection between the state vector and the measurement vector. $V_k \in \mathbb{R}^m$ is Gaussian white measurement noise with covariance matrix $R_k \in \mathbb{R}^m \times \mathbb{R}^m$. Thus $V_k \sim N(0, R_k)$. W_k and V_k are assumed to be independent. If V_k and W_k are Gaussian, the Kalman filter is the optimal estimator, otherwise the Kalman filter is the best linear estimator. The derivation of the discrete filter in this section uses that these two noise terms are Gaussian. Different derivations exist that do not need this assumption.

An example where the Kalman filter can be applied is in modelling the course of a ship. In this case, X_k represents the vector describing the ship's coordinates and velocity. Y_k could then represent GPS measurements of these coordinates and velocity, while H_k serves as a projection onto X_k.

Let us define the sigma algebra $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, which is generated by the observations up until and including time n. We will denote the density of X_k given the sigma algebra $\mathcal{F}_n$ as $f_{X_k|\mathcal{F}_n}$.

The Kalman filter minimises the the mean square error between the true state and its estimate. The mean square error is conditioned on the sigma algebra $\mathcal{F}_n$ to account for the measurements up until and including time n. We thus need to find the following estimator:

$$\hat{x}_{k|n} = \arg\min_{\hat{x}} \mathbb{E}[(X_k - \hat{x})^T (X_k - \hat{x}) | \mathcal{F}_n]. \tag{2.2}$$

The solution to this estimation problem is given in the following lemma.

Theorem 2.1.1. *[2] Assume that $\hat{x}^T\hat{x}$ is integrable. Then the solution to the minimisation problem of Equation* (2.2) *is the conditional mean $\mathbb{E}[X_k|\mathcal{F}_n]$.*

This theorem is proved later in this section.

From now on, the estimate of X_k given the observations up until and including time n will be denoted by $\hat{x}_{k|n} = \mathbb{E}[X_k|\mathcal{F}_n]$. The error covariance matrix of this estimate is denoted by $P_{k|n} = \mathbb{E}[(X_k - \hat{x}_{k|n})(X_k - \hat{x}_{k|n})^T|\mathcal{F}_n]$. Note that $\hat{x}_{k|n}$ and $P_{k|n}$ are the mean and variance of $f_{X_k|\mathcal{F}_n}$. Typically, when using the Kalman filter, we are interested in finding $\hat{x}_{k|k}$ and $P_{k|k}$. To this end, we need to find $f_{X_k|\mathcal{F}_k}$ which reflects a belief in the state X_k given $\mathcal{F}_k$. The following theorem shows how $f_{X_k|\mathcal{F}_k}$ can be found.

Theorem 2.1.2. *Given the dynamic system in Equation* (2.1)*. Then:*

$$\begin{aligned}
f_{X_k|\mathcal{F}_{k-1}}(x) &= \int f_{X_k|X_{k-1}}(x) f_{X_{k-1}|\mathcal{F}_{k-1}}(w) \ \mathrm{d}w, \\
f_{X_k|\mathcal{F}_k}(x) &= \frac{f_{Y_k|X_k}(y) f_{X_k|\mathcal{F}_{k-1}}(x)}{\int f_{Y_k|X_k}(y) f_{X_k|\mathcal{F}_{k-1}}(x) \ \mathrm{d}x}.
\end{aligned}$$

Theorem 2.1.2 is proven below.

Remark 2.1.3. *Theorem 2.1.2 shows that it is possible to derive the density of $f_{X_k|\mathcal{F}_k}$, called the posterior from $f_{X_{k-1}|\mathcal{F}_{k-1}}$ via $f_{X_k|\mathcal{F}_{k-1}}$, which is called the prior. This means that $f_{X_k|\mathcal{F}_k}$ can be calculated recursively, by first finding the prior and then the posterior. The calculation of the mean and the covariance matrix of the prior and the posterior are called the prediction step and the update step, respectively.*

To find the estimate of the states $\hat{x}_{k|k-1}$ and $\hat{x}_{k|k}$ the means of the prior and posterior need to be found. The error covariance matrices $P_{k|k-1}$ and $P_{k|k}$ are found by calculating the covariance matrices of the prior and posterior, respectively. We start with deriving the mean and covariance matrix of the prior.

Theorem 2.1.4 (Prediction step of Kalman filter)**.** *Given the dynamic system in Equation* (2.1)*, we find:*

$$\begin{aligned}
\hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1}, \\
P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}.
\end{aligned}$$

This theorem is proved below. The next theorem gives the update step, by deriving the mean and covariance matrix of the posterior.

Theorem 2.1.5 (Update step of Kalman filter)**.** *Consider the dynamic system of Equation* (2.1)*. Then:*

$$\begin{aligned}
\hat{x}_{k|k} &= (I - K_{X_k,Y_k}H_k)x_{k|k-1} + K_{X_k,Y_k}y_k \\
&= x_{k|k-1} + K_{X_k,Y_k}(y_k - H_k x_{k|k-1}), \\
P_{k|k} &= (I - K_{X_k,Y_k}H_k)P_{k|k-1}(I - K_{X_k,Y_k}H_k)^T + K_{X_k,Y_k}R_k K_{X_k,Y_k}^T,
\end{aligned}$$

where

$$K_{X_k,Y_k} = P_{k|k-1}H_k^T P_{Y_k}^{-1}, \qquad P_{Y_k} = H_k P_{k|k-1}H^T + R_k.$$

Some background on moment generating function and auxiliary results are needed to prove this theorem. The background on moment generating functions can be found in Subsection 2.1.1. The auxiliary results and proof can be found in Subsection 2.1.2.

In conclusion, we find the following algorithm to estimate the state of the dynamic system of Equation (2.1). The algorithm starts with $\hat{x}_{0|0}$ and $P_{0|0}$, where $P_{0|0}$ should reflect the certainty with which $\hat{x}_{0|0}$ was chosen. The exact starting point does not matter as long as $P_{0|0}$ is not the zero matrix. Note that from now on, we will denote K_{X_k,Y_k} as K_k for brevity. As mentioned before, first a prediction step is made, which was derived in Theorem 2.1.4:

$$\begin{aligned}\hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1},\\ P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}.\end{aligned}$$

After which the update step takes place, which was derived in Theorem 2.1.5:

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}),\\ P_{k|k} &= (I - K_kH_k)P_{k|k-1}(I - K_kH_k)^T + K_kR_kK_k^T,\\ K_k &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}.\end{aligned}$$

K_k is often called the Kalman gain and $I_k = y_k - H_k\hat{x}_{k|k-1}$ is called the innovation, which is the difference between the observation and the a priori estimate of the measurements, $\hat{y}_{k|k-1} = \mathbb{E}[Y_k|\mathcal{F}_{k-1}] = H_k\hat{x}_{k|k-1}$. The solution of the Kalman filter can be rewritten into multiple forms. These reformulations of the solution of the Kalman filter are derived in Subsection 2.1.3.

We now prove the above results, starting with Theorem 2.1.1.

Proof of Theorem 2.1.1. We prove this theorem by setting the gradient of the mean square error to zero. Before deriving the gradient of the mean squared error, we find a bound for $|(X_k - \hat{x})^T(X_k - \hat{x})|$:

$$0 \leq |(X_k - \hat{x})^T(X_k - \hat{x})| = |X_k^TX_k - X_k^T\hat{x} - \hat{x}^TX_k + \hat{x}^T\hat{x}|.$$

Note that $X_k^T\hat{x} \in \mathbb{R}$, so it is equal to its transpose:

$$\begin{aligned}0 &\leq |X_k^TX_k - 2\hat{x}^TX_k + \hat{x}^T\hat{x}|\\ &= X_k^TX_k + 2|\hat{x}^TX_k| + \hat{x}^T\hat{x}.\end{aligned} \tag{2.3}$$

We bring $|\hat{x}^TX_k|$ to the left, to obtain:

$$|\hat{x}^TX_k| \leq \frac{1}{2}(X_k^TX_k + \hat{x}^T\hat{x}).$$

We use the above expression in Equation (2.3), to find:

$$\begin{aligned}|(X_k - \hat{x})^T(X_k - \hat{x})| &\leq X_k^TX_k + X_k^TX_k + \hat{x}^T\hat{x} + \hat{x}^T\hat{x}\\ &= 2X_k^TX_k + 2\hat{x}^T\hat{x}.\end{aligned}$$

We define $Z = 2X_k^TX_k + 2\hat{x}^T\hat{x}$. Since we assumed that $\hat{x}^T\hat{x}$ is integrable and since X_k has a normal distribution, we conclude that Z is an integrable random variable which bounds $|(X_k - \hat{x})^T(X_k - \hat{x})|$ from above. We continue to rewrite the gradient of the mean square error. By the definition of the gradient:

$$\nabla_{\hat{x}}\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n] = \begin{bmatrix}\frac{\partial}{\partial\hat{x}_1}\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n]\\ \vdots\\ \frac{\partial}{\partial\hat{x}_n}\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n]\end{bmatrix}.$$

To avoid long derivations in matrix notation, we show that it is possible to interchange the partial derivative with respect to x_i and the expectation. To this end, we define $e_i = [0, \ldots, 0, 1, 0, \ldots, 0]$ where the 1 is on the ith position. Then using the definition of partial derivatives:

$$\begin{aligned}&\frac{\partial}{\partial\hat{x}_i}\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n]\\ &\quad= \lim_{h\to0}\frac{1}{h}\left(\mathbb{E}[(X_k - (\hat{x} + he_i))^T(X_k - (\hat{x} + he_i))|\mathcal{F}_n] - \mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n]\right)\\ &\quad= \lim_{h\to0}\mathbb{E}\left[\frac{1}{h}\Big((X_k - (\hat{x} + he_i))^T(X_k - (\hat{x} + he_i)) - (X_k - \hat{x})^T(X_k - \hat{x})\Big)|\mathcal{F}_n\right].\end{aligned}$$

We have seen that $|\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n]| \leq Z$, thus using conditional dominated convergence, we are allowed to interchange the limit and the expectation:

$$\begin{aligned}
\frac{\partial}{\partial \hat{x}_i}&\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n] \\
&= \mathbb{E}\left[\lim_{h\to 0}\frac{1}{h}\left((X_k - (\hat{x} + he_i))^T\left(X_k - (\hat{x} + he_i)\right) - (X_k - \hat{x})^T(X_k - \hat{x})\right)|\mathcal{F}_n\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial \hat{x}_i}(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n\right].
\end{aligned}$$

Thus we conclude:

$$\begin{aligned}
\nabla_{\hat{x}}\mathbb{E}[(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n] &= \mathbb{E}[\nabla_{\hat{x}}(X_k - \hat{x})^T(X_k - \hat{x})|\mathcal{F}_n] \\
&= 2\mathbb{E}[X_k - \hat{x}|\mathcal{F}_n] \\
&= 2(\mathbb{E}[X_k|\mathcal{F}_n] - \hat{x}) = 0.
\end{aligned}$$

Clearly, this only holds when $\hat{x} = \mathbb{E}[X_k|\mathcal{F}_n]$. □

We need a preliminary result before we can prove Theorem 2.1.2.

Lemma 2.1.6 (Bayes Theorem)**.** *[6] Let $X, Y \in \mathbb{R}^d$ be two continuous random variables with probability densities $f_X(x)$ and $f_Y(y)$, respectively. We assume $f_Y(y) > 0$. Let $f_{X,Y}(x,y)$ be the joint probability function of the random vector (X,Y). Then,*

$$f_{X|Y}(x) = \frac{f_{Y|X}(y)f_X(x)}{f_Y(y)}.$$

This result is proven at the end of the section.

Proof of Theorem 2.1.2. We start with the derivation of the prior. Using the definition of conditional probability functions, we find:

$$\begin{aligned}
f_{X_k|\mathcal{F}_{k-1}}(x) = \frac{f_{X_k,\mathcal{F}_{k-1}}(x,y)}{f_{\mathcal{F}_{k-1}}(y)} &= \frac{\int f_{X_{k-1},X_k,\mathcal{F}_{k-1}}(w,x,y)\ \mathrm{d}w}{f_{\mathcal{F}_{k-1}}(y)} \\
&= \frac{\int f_{X_k|X_{k-1},\mathcal{F}_{k-1}}(x)f_{X_{k-1}|\mathcal{F}_{k-1}}(w)f_{\mathcal{F}_{k-1}}(y)\ \mathrm{d}w}{f_{\mathcal{F}_{k-1}}(y)} \\
&= \int f_{X_k|X_{k-1},\mathcal{F}_{k-1}}(x)f_{X_{k-1}|\mathcal{F}_{k-1}}(w)\ \mathrm{d}w.
\end{aligned}$$

Using that conditioning X_k on X_{k-1} and $\mathcal{F}_{k-1}$ is the same as only conditioning on X_{k-1}, because $\mathcal{F}_{k-1}$ has been used to calculate X_{k-1}, so $\mathcal{F}_{k-1}$ is redundant information, gives:

$$f_{X_k|\mathcal{F}_{k-1}}(x) = \int f_{X_k|X_{k-1}}(x)f_{X_{k-1}|\mathcal{F}_{k-1}}(w)\ \mathrm{d}w.$$

We derive the probability function of the posterior. Using Bayes theorem, see Lemma 2.1.6, we obtain:

$$\begin{aligned}
f_{X_k|\mathcal{F}_k}(x) &= f_{X_k|Y_k,\mathcal{F}_{k-1}}(x) \\
&= \frac{f_{Y_k|X_k,\mathcal{F}_{k-1}}(y)f_{X_k|\mathcal{F}_{k-1}}(x)}{f_{Y_k|\mathcal{F}_{k-1}}(y)}.
\end{aligned}$$

Again, using that conditioning on $\mathcal{F}_{k-1}$ is redundant when conditioning on X_k, gives:

$$f_{X_k|\mathcal{F}_k}(x) = \frac{f_{Y_k|X_k}(y)f_{X_k|\mathcal{F}_{k-1}}(x)}{f_{Y_k|\mathcal{F}_{k-1}}(y)}.$$

We use the definition of conditional probability to rewrite $f_{Y_k|\mathcal{F}_{k-1}}(y)$ as follows:

$$\begin{aligned} f_{Y_k|\mathcal{F}_{k-1}}(y) &= \int f_{X_k,Y_k|\mathcal{F}_{k-1}}(x,y)\ \mathrm{d}x \\ &= \int f_{Y_k|X_k,\mathcal{F}_{k-1}}(y) f_{X_k|\mathcal{F}_{k-1}}(x)\ \mathrm{d}x. \end{aligned}$$

Using that $\mathcal{F}_{k-1}$ is redundant when conditioning on X_k gives:

$$f_{Y_k|\mathcal{F}_{k-1}}(y) = \int f_{Y_k|X_k}(y) f_{X_k|\mathcal{F}_{k-1}}(x)\ \mathrm{d}x.$$

□

Proof of Theorem 2.1.4. It is possible to derive the mean and covariance matrix of $f_{X_k|\mathcal{F}_{k-1}}$ by writing out the integral of the prior in Theorem 2.1.2. This is, however, a tedious process and can be done more easily by using basic rules of conditional expectations. We use the latter method. We start with deriving the mean of the prior:

$$\begin{aligned} \hat{x}_{k|k-1} &= \mathbb{E}[X_k|\mathcal{F}_{k-1}] = \mathbb{E}[F_{k-1}X_{k-1} + W_{k-1}|\mathcal{F}_{k-1}] \\ &= F_{k-1}\mathbb{E}[X_{k-1}|\mathcal{F}_{k-1}] + \mathbb{E}[W_{k-1}|\mathcal{F}_{k-1}] \\ &= F_{k-1}\hat{x}_{k-1|k-1}. \end{aligned}$$

We continue to derive the covariance matrix of the prior:

$$\begin{aligned} P_{k|k-1} &= \mathrm{Var}(X_k|\mathcal{F}_{k-1}) = \mathrm{Var}(F_{k-1}X_{k-1} + W_{k-1}|\mathcal{F}_{k-1}) \\ &= F_{k-1}\mathrm{Var}(X_{k-1}|\mathcal{F}_{k-1})F_{k-1}^T + \mathrm{Var}(W_{k-1}|\mathcal{F}_{k-1}) + 2F_{k-1}\mathrm{Cov}(X_{k-1}, W_{k-1}|\mathcal{F}_{k-1}) \\ &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}. \end{aligned}$$

This concludes the claim. □

We proceed with the proof of Theorem 2.1.5, whose main idea is to derive the moment generating function of the posterior that we found in Theorem 2.1.2. Since the moment generating function determines the distribution of a random variable uniquely, this will then give us the distribution, mean and covariance matrix. The proof of Theorem 2.1.5 is split over two subsections. Subsection 2.1.1 will give the definition and some basic properties of moment generating functions. Subsection 2.1.2 will then use moment generating functions to derive the distribution of the posterior.

Before moving on to Subsection 2.1.1, we prove the preliminary result.

Proof of Lemma 2.1.6. We prove this lemma via the definition of conditional density functions:

$$\begin{aligned} f_{X|Y}(x) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{\frac{f_{X,Y}(x,y)}{f_X(x)} f_X(x)}{f_Y(y)} \\ &= \frac{f_{Y|X}(y) f_X(x)}{f_Y(y)}. \end{aligned}$$

This concludes the claim. □

2.1.1. Moment generating functions

This subsection contains the definition of moment generating functions and some basic properties. We start with the definition of a moment generating function.

Definition 2.1.7 (Moment generating function). *[11] Let X be a random vector on $\mathbb{R}^n$ and t a vector on $\mathbb{R}^n$. Then the moment generating function is defined as:*

$$M_X(t) = \mathbb{E}\Big[\mathrm{e}^{\langle t,X\rangle}\Big],$$

provided that the right-hand side exists on some open neighbourhood of 0.

Remark 2.1.8. *For the moment generating function to exist, there must be a $\delta \in \mathbb{R}$ such that if $\|t\| < \delta$ then $\mathbb{E}\Big[\mathrm{e}^{\langle t,X\rangle}\Big] < \infty$. The name of the moment generating function comes from the fact that if the moment generating function of a random variable exists on an open neighbourhood around 0, then all its moments are finite and can be computed by partial differentiation.*

Some basic properties of the moment generating function will be shown in Lemma 2.1.9.

Lemma 2.1.9 (Basic properties of moment generating functions). *Let X_1 and X_2 be two independent $\mathbb{R}^n$ valued random variables for which the moment generating functions M_{X_1} and M_{X_2} exist on an open neighbourhood of 0. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map and let $b \in \mathbb{R}^n$. Then for $t \in \mathbb{R}^n$ in the neighbourhood where M_{X_1} and M_{X_2} exist:*

(a) $M_{AX_1+b}(t) = \mathrm{e}^{\langle t,b\rangle} M_{X_1}(A^T t)$, *where A^T is the transpose of A,*
(b) $M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t)$.

Let $X = (X_1, X_2)$ and $t = (t_1, t_2)$ with $t_1 \in \mathbb{R}^n$ and $t_2 \in \mathbb{R}^n$ in the neighbourhood where M_{X_1} and M_{X_2} exist, respectively. Then:

(c) $M_X(t) = M_{X_1}(t_1) M_{X_2}(t_2)$.

This result is proven later in this section.

A very important and useful property of the moment generating function is that the moment generating function uniquely determines the distribution. We state the relevant theorem without proof.

Theorem 2.1.10 (Uniqueness theorem for moment generating functions). *[11] If for some $\delta > 0$, $M_X(t) = \mathbb{E}[e^{\langle t,X\rangle}] < \infty$ and $M_X(t) = M_Y(t)$ for all t such that $\|t\| < \delta$, then X and Y have the same distribution.*

In Theorem 2.1.2 we derived the probability functions for the prior and posterior. These probability functions are conditional probability functions and therefore, it is necessary to know the dynamics of conditional moment generating functions.

Lemma 2.1.11 (Conditional moment generating function). *Let (X, Y) be a random vector with moment generating function $M_{X,Y}(t_1, t_2) = \mathbb{E}[\mathrm{e}^{\langle t_1,X\rangle + \langle t_2,Y\rangle}]$ for $t_1 \in \mathbb{R}^n$ and $t_2 \in \mathbb{R}^m$ with $n = \dim(X)$ and $m = \dim(Y)$, such that the moment generating function $M_{X,Y}$ exits. We define the probability measure $\mathbb{P}_{t_1}$ for Y as*

$$\frac{\mathrm{d}\mathbb{P}_{t_1}}{\mathrm{d}\mathbb{P}}(y) := \frac{M_{X|Y=y(t_1)}}{M_X(t_1)}. \tag{2.4}$$

Then the moment generating function M_{Y,t_1} of Y under the measure $\mathbb{P}_{t_1}$ is given by:

$$M_{Y,t_1}(t_2) := \mathbb{E}_{t_1}[\mathrm{e}^{\langle t_2,Y\rangle}] = \frac{M_{X,Y}(t_1, t_2)}{M_X(t_1)}. \tag{2.5}$$

This lemma is proven below. To find the distribution of $X_k|\mathcal{F}_k$, we will use Equation (2.4) and Equation (2.5) and we therefore need the moment generating functions of the vector (X_k, Y_k) and of X_k. Because we assumed the noise to be Gaussian, X_k and Y_k have a normal distribution and we thus need the moment generating function of a normal distribution. The moment generating function of a normal distribution will be derived in Lemma 2.1.12.

Lemma 2.1.12 (Moment generating function of normal distribution)**.** *The following three statements are true:*

(a) *Let $X \sim N(0,1)$, then $M_X(t) = \mathsf{e}^{\frac{1}{2}t^2}$ for $t \in \mathbb{R}$.*
(b) *Let $X \sim N(0,I)$ be a vector of n independent standard normal random variables, then $M_X(t) = \mathsf{e}^{\frac{1}{2}\|t\|^2}$ for $t \in \mathbb{R}^n$.*
(c) *Let $X \sim N(\mu,\Sigma)$ on $\mathbb{R}^n$. Then for $t \in \mathbb{R}^n$,*

$$M_X(t) = \mathbb{E}\Big[\mathsf{e}^{\langle t,X\rangle}\Big] = \mathsf{e}^{\langle t,\mu\rangle + \frac{1}{2}\langle t,\Sigma t\rangle}.$$

This lemma is proven at the end of this section. Below we prove the results of this section.

Proof of Lemma 2.1.9. We prove the above statements as follows:

(a) $M_{AX_1+b}(t) = \mathbb{E}\Big[\mathsf{e}^{\langle t,AX_1+b\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle t,b\rangle}\mathsf{e}^{\langle A^Tt,X_1\rangle}\Big] = \mathsf{e}^{\langle t,b\rangle}\mathbb{E}\Big[\mathsf{e}^{\langle A^Tt,X_1\rangle}\Big] = \mathsf{e}^{\langle t,b\rangle}M_{X_1}(A^Tt)$,
(b) $M_{X_1+X_2}(t) = \mathbb{E}\Big[\mathsf{e}^{\langle t,X_1+X_2\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle t,X_1\rangle}\mathsf{e}^{\langle t,X_2\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle t,X_1\rangle}\Big]\mathbb{E}\Big[\mathsf{e}^{\langle t,X_2\rangle}\Big] = M_{X_1}(t)M_{X_2}(t)$,
(c) $M_X(t) = \mathbb{E}\Big[\mathsf{e}^{\langle t,X\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle (t_1,t_2),(X_1,X_2)\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle t_1,X_1\rangle}\mathsf{e}^{\langle t_2,X_2\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle t_1,X_1\rangle}\Big]\mathbb{E}\Big[\mathsf{e}^{\langle t_2,X_2\rangle}\Big]$
$= M_{X_1}(t_1)M_{X_2}(t_2)$.

□

Proof of Lemma 2.1.11. Fix t_1. By the law of total expectation,

$$\begin{aligned}
M_{X,Y}(t_1,t_2) &= \mathbb{E}\Big[\mathsf{e}^{\langle t_1,X\rangle+\langle t_2,Y\rangle}\Big] = \mathbb{E}\Big[\mathsf{e}^{\langle t_1,X\rangle}\mathsf{e}^{\langle t_2,Y\rangle}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\mathsf{e}^{\langle t_1,X\rangle}\mathsf{e}^{\langle t_2,Y\rangle}|Y\Big]\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\mathsf{e}^{\langle t_1,X\rangle}|Y\Big]\mathsf{e}^{\langle t_2,Y\rangle}\Big] \\
&= \mathbb{E}\Big[M_{X|Y}(t_1)\mathsf{e}^{\langle t_2,Y\rangle}\Big] = \frac{\mathbb{E}\Big[M_{X|Y}(t_1)\mathsf{e}^{\langle t_2,Y\rangle}\Big]}{M_X(t_1)}M_X(t_1) \\
&= \mathbb{E}_{t_1}\Big[\mathsf{e}^{\langle t_2,Y\rangle}\Big]M_X(t_1).
\end{aligned}$$

Diving both sides by $M_X(t_1)$ concludes the proof. □

Proof of Lemma 2.1.12. We prove the statements as follows:

(a) We use $tx - \frac{1}{2}x^2 = \frac{1}{2}t^2 - \frac{1}{2}(x-t)^2$ to obtain:

$$M_X(t) = \mathbb{E}\Big[\mathsf{e}^{\langle t,X\rangle}\Big] = \frac{1}{\sqrt{2\pi}}\int \mathsf{e}^{tx}\mathsf{e}^{-\frac{1}{2}x^2}\,\mathsf{d}x = \mathsf{e}^{\frac{1}{2}t^2}\frac{1}{\sqrt{2\pi}}\int \mathsf{e}^{-\frac{1}{2}(x-t)^2}\,\mathsf{d}x = \mathsf{e}^{\frac{1}{2}t^2}.$$

(b) This is an immediate consequence of Lemma 2.1.9*(c)* and Lemma 2.1.12*(a)*.
(c) By definition $X = AY + \mu$ with $Y \sim N(0,I)$ for some map A such that $\Sigma = AA^T$. Note that it is possible to write $\Sigma = AA^T$ because Σ is a covariance matrix. The result immediately follows from Lemma 2.1.9 *(a)* and Lemma 2.1.12 *(b)*.

□

This concludes the theory on moment generating functions necessary to continue the derivation of the distribution of the posterior. The following subsection will derive this distribution.

2.1.2. Distribution of posterior

This subsection uses the theory discussed in Subsection 2.1.1 to derive the distribution of the posterior. To this end, the following lemma will derive the distribution of a vector (X, Y) with similar properties to the vector (X_k, Y_k).

Lemma 2.1.13. *Let $X \in \mathbb{R}^n$ and let $V \in \mathbb{R}^m$ with $X \sim N(\mu_X, \Sigma_X)$ and $V \sim N(0, \Sigma_V)$ independent. Let $Y \in \mathbb{R}^m$ such that $Y = HX + V$ with $H \in \mathbb{R}^m \times \mathbb{R}^n$ a linear map. Then,*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ H\mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_X & \Sigma_X H^T \\ H\Sigma_X & H\Sigma_X H^T + \Sigma_V \end{pmatrix} \right).$$

This result is proven at the end of the section.

In the following lemma we will derive the distribution of a normal random variable conditioned on another normal random variable with a given covariance structure. This will be used as a stepping stone to derive the distribution of $X_k|\mathcal{F}_k$.

Lemma 2.1.14. *Let (X, Y) be a normal distribution with mean $\mu = (\mu_X, \mu_Y)$ and block-covariance matrix Σ given by*

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} \\ \Sigma_{Y,X} & \Sigma_Y \end{bmatrix}.$$

Let $K_{X,Y} = \Sigma_{X,Y}\Sigma_Y^{-1}$. Then $(X|Y = y) \sim N(\mu_X + K_{X,Y}(y - \mu_Y), \Sigma_X - K_{X,Y}\Sigma_Y K_{X,Y}^T)$.

This result is proven below.

The following lemma combines Lemma 2.1.13 with Lemma 2.1.14, to obtain the mean and covariance matrix of a conditional distribution that has the same properties as the distribution of $X_k|\mathcal{F}_k$.

Lemma 2.1.15. *Let $X \in \mathbb{R}^n$ be a random vector with $X \sim N(\mu_X, \Sigma_X)$. Let $Y \in \mathbb{R}^m$ with $Y = HX + V$ for some linear map $H \in \mathbb{R}^m \times \mathbb{R}^n$ and $V \in \mathbb{R}^m$ a normal random variable such that $V \sim N(0, \Sigma_V)$. Assume that X and V are independent. Then:*

$$\begin{aligned} \mu_{X|Y=y} &= (I - K_{X,Y}H)\mu_X + K_{X,Y}y, \\ \Sigma_{X|Y=y} &= (I - K_{X,Y}H)\Sigma_X(I - K_{X,Y}H)^T + K_{X,Y}\Sigma_V K_{X,Y}^T, \end{aligned}$$

where

$$K_{X,Y} = \Sigma_{X,Y}\Sigma_Y^{-1} = \Sigma_X H^T \Sigma_Y^{-1}.$$

The proof of this lemma can be found at the end of this section. Finally, enough background is established to prove Theorem 2.1.5.

Proof of Theorem 2.1.5. This result is directly obtained by applying Lemma 2.1.15 to the dynamic system of Equation (2.1). □

The auxiliary results of this section are proven below.

Proof of Lemma 2.1.13. The mean and covariance matrix of X are already given. We calculate the mean and variance of Y:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[HX + V] = H\mathbb{E}[X] + \mathbb{E}[V] = H\mu_X, \\ \mathsf{Var}(Y) &= \mathsf{Var}(HX + V) = \mathsf{Var}(HX) + \mathsf{Var}(V) + 2\mathsf{Cov}(HX, Y) \\ &= H\Sigma_X H^T + \Sigma_V. \end{aligned}$$

What remains is the covariance between X and Y:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])^T] \\
&= \mathbb{E}[XY^T] - \mathbb{E}[X\mathbb{E}[Y]^T] - \mathbb{E}[\mathbb{E}[X]Y^T] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]^T] \\
&= \mathbb{E}[XY^T] - \mathbb{E}[X]\mathbb{E}[Y]^T = \mathbb{E}[X(HX+V)^T] - \mu_X\mu_X^T H^T \\
&= \mathbb{E}[XX^TH^T + XV^T] - \mu_X\mu_X^TH^T = \mathbb{E}[XX^T]H^T + \mathbb{E}[X]\mathbb{E}[V] - \mu_X\mu_X^TH^T \\
&= \Sigma_X H^T + \mu_X\mu_X^TH^T - \mu_X\mu_X^TH^T = \Sigma_X H^T.
\end{aligned}$$

Note that Y has a normal distribution because it is a linear combination of normal random variables. This concludes the claim. □

Proof of Lemma 2.1.14. Let $n = \dim(X)$ and $m = \dim(Y)$. Let $t_1 \in \mathbb{R}^n$ and $t_2 \in \mathbb{R}^m$ such that the moment generating function $M_{X,Y}(t_1,t_2) = \mathbb{E}[\mathrm{e}^{\langle t_1,X\rangle + \langle t_2,Y\rangle}]$ exists. By Lemma 2.1.11, the moment generating function of Y under the measure $\mathbb{P}_{t_1}$ is given by:

$$\begin{aligned}
\mathbb{E}_{t_1}\left[\mathrm{e}^{\langle t_2,Y\rangle}\right] &= \frac{M_{X,Y}(t_1,t_2)}{M_X(t_1)} = \frac{\exp\left[\langle t,\mu\rangle + \frac{1}{2}\langle t,\Sigma t\rangle\right]}{\exp\left[\langle t_1,\mu_X\rangle + \frac{1}{2}\langle t_1,\Sigma_X t_1\rangle\right]} \\
&= \frac{\exp\left[\langle t_1,\mu_X\rangle + \langle t_2,\mu_Y\rangle + \frac{1}{2}\Big(\langle t_1,\Sigma_X t_1\rangle + \langle t_1,\Sigma_{X,Y} t_2\rangle + \langle t_2,\Sigma_{Y,X} t_1\rangle + \langle t_2,\Sigma_Y t_2\rangle\Big)\right]}{\exp\left[\langle t_1,\mu_X\rangle + \frac{1}{2}\langle t_1,\Sigma_X t_1\rangle\right]} \\
&= \exp\left[\langle t_2,\mu_Y\rangle + \frac{1}{2}\Big(\langle t_1,\Sigma_{X,Y} t_2\rangle + \langle t_2,\Sigma_{Y,X} t_1\rangle + \langle t_2,\Sigma_Y t_2\rangle\Big)\right].
\end{aligned}$$

Using that $\Sigma_{Y,X} = \Sigma_{X,Y}^T$ and thus $\langle t_1,\Sigma_{X,Y}t_2\rangle = \langle t_2,\Sigma_{Y,X}t_1\rangle$, gives:

$$\mathbb{E}_{t_1}\left[\mathrm{e}^{\langle t_2,Y\rangle}\right] = \exp\left[\langle t_2,\mu_Y + \Sigma_{Y,X}t_1\rangle + \frac{1}{2}\langle t_2,\Sigma_Y t_2\rangle\right].$$

In the last line we can recognise the moment generating function of the normal distribution. Thus by Theorem 2.1.10, Y under the measure $\mathbb{P}_{t_1}$ has a normal distribution with mean $\mu_Y + \Sigma_{Y,X}t_1$ and covariance Σ_Y. We write this distribution as $f_{Y,t_1}(y) = f_{\mu_Y+\Sigma_{Y,X}t_1,\Sigma_Y}(y)$. Equation (2.4) gives:

$$\begin{aligned}
M_{X|Y=y}(t_1) &= \frac{\mathrm{d}\mathbb{P}_{t_1}}{\mathrm{d}\mathbb{P}}(y)M_X(t_1) \\
&= \frac{f_{\mu_Y+\Sigma_{Y,X}t_1,\Sigma_Y}(y)}{f_{\mu_Y,\Sigma_Y}(y)}M_X(t_1).
\end{aligned}$$

Because all terms on the right are in terms of an exponential, we compute the factor of the exponential by taking the logarithm of $M_{X|Y=y}(t_1)$:

$$\begin{aligned}
\log(M_{X|Y=y}(t_1)) &= -\frac{1}{2}\langle y-\mu_Y-\Sigma_{Y,X}t_1, \Sigma_Y^{-1}(y-\mu_Y-\Sigma_{Y,X}t_1)\rangle + \frac{1}{2}\langle y-\mu_Y,\Sigma_Y^{-1}(y-\mu_Y)\rangle \\
&\quad + \langle t_1,\mu_X\rangle + \frac{1}{2}\langle t_1,\Sigma_X t_1\rangle \\
&= -\frac{1}{2}\langle \Sigma_{Y,X}t_1, \Sigma_Y^{-1}\Sigma_{Y,X}t_1\rangle + \langle \Sigma_{Y,X}t_1,\Sigma_Y^{-1}(y-\mu_Y)\rangle + \langle t_1,\mu_X\rangle + \frac{1}{2}\langle t_1,\Sigma_X t_1\rangle \\
&= \langle t_1, \mu_X + \Sigma_{X,Y}\Sigma_Y^{-1}(y-\mu_Y)\rangle + \frac{1}{2}\langle t_1,(\Sigma_X - \Sigma_{X,Y}\Sigma_Y^{-1}\Sigma_Y\Sigma_Y^{-1}\Sigma_{Y,X})t_1\rangle.
\end{aligned}$$

Using the definition of $K_{X,Y}$ and that $(\Sigma_{X,Y}\Sigma_Y^{-1})^T = \Sigma_Y^{-1}\Sigma_{Y,X}$, gives:

$$\log(M_{X|Y=y}(t_1)) = \langle t_1,\mu_X + K_{X,Y}(y-\mu_Y)\rangle + \frac{1}{2}\langle t_1,(\Sigma_X - K_{X,Y}\Sigma_Y K_{X,Y}^T)t_1\rangle.$$

In the last line we recognise the exponent of the moment generating function of a normal distribution with mean $\mu_X + K_{X,Y}(y-\mu_Y)$ and variance $\Sigma_X - K_{X,Y}\Sigma_Y K_{X,Y}^T$. This concludes the claim. □

Proof of Lemma 2.1.15. First we derive the mean of $X|Y=y$. We use that $\mu_Y = H\mu_X$, which is shown in Lemma 2.1.13. Applying Lemma 2.1.14, we find:

$$\mu_{X|Y=y} = \mu_X + K_{X,Y}(y-\mu_Y) = \mu_X - K_{X,Y}H\mu_X + K_{X,Y}y = (I - K_{X,Y}H)\mu_X + K_{X,Y}y.$$

We continue with the covariance. By Lemma 2.1.14, we find:

$$\Sigma_{X|Y=y} = \Sigma_X - K_{X,Y}\Sigma_Y K_{X,Y}^T.$$

Adding and subtracting $K_{X,Y}\Sigma_Y K_{X,Y}^T$ and using that $K_{X,Y}\Sigma_Y K_{X,Y}^T = K_{X,Y}\Sigma_{Y,X} = \Sigma_{X,Y}K_{X,Y}^T$ gives:

$$\Sigma_{X|Y=y} = \Sigma_X - K_{X,Y}\Sigma_{Y,X} - \Sigma_{X,Y}K_{X,Y}^T + K_{X,Y}\Sigma_Y K_{X,Y}^T.$$

Using $\Sigma_{Y,X} = H\Sigma_X$ and $\Sigma_Y = H\Sigma_X H^T + \Sigma_V$, which was derived in Lemma 2.1.13, gives:

$$\begin{aligned}\Sigma_{X|Y=y} &= \Sigma_X - K_{X,Y}H\Sigma_X - \Sigma_X H^T K_{X,Y}^T + K_{X,Y}H\Sigma_X H^T K_{X,Y}^T + K_{X,Y}\Sigma_V K_{X,Y}^T \\ &= (I - K_{X,Y}H)\Sigma_X(I - K_{X,Y}H)^T + K_{X,Y}\Sigma_V K_{X,Y}^T.\end{aligned}$$

This concludes the claim. □

2.1.3. Alternative formulations of the solution of the Kalman filter

Throughout this thesis, different formulations of the Kalman filter will be used. This subsection contains a derivation of several formulations. The following lemma states the reformulations commonly used.

Theorem 2.1.16 (Rewritten solution of Kalman filter). *The solution of the Kalman filter can be rewritten in the following way:*

Prediction step:

$$\begin{aligned}\hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1}, \\ P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}.\end{aligned}$$

Update step:

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}), \\ P_{k|k} &= (I - K_kH_k)P_{k|k-1}(I-K_kH_k)^T + K_kR_kK_k^T \\ &= (I-K_kH_k)P_{k|k-1} \\ &= (P_{k|k-1}^{-1} + H_k^TR_k^{-1}H_k)^{-1}, \\ K_k &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1} \\ &= P_{k|k}H_k^TR_k^{-1}.\end{aligned} \tag{2.6}$$

Note that when using the second definition of the Kalman gain, the third definition of the covariance matrix should be used. Otherwise, the Kalman gain is needed to calculate $P_{k|k}$, while $P_{k|k}$ is also needed to calculate the Kalman gain.

To prove Theorem 2.1.16, we need a preliminary result called the Woodbury matrix identity. It is also sometimes called the Woodbury formula, matrix inversion lemma or Sherman-Morrison-Woodbury formula.

Lemma 2.1.17 (Woodbury matrix identity). *The Woodbury matrix identity is given by:*

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1},$$

provided the above matrix products and inverses exist.

This result is proven below.

Proof of Theorem 2.1.16. We start with deriving the second formulation of the covariance matrix $P_{k|k}$, starting from the first:

$$\begin{aligned}P_{k|k} &= (I - K_kH_k)P_{k|k-1}(I - K_kH_k)^T + K_kR_kK_k^T \\ &= P_{k|k-1} - P_{k|k-1}H_k^TK_k^T - K_kH_kP_{k|k-1} + K_kH_kP_{k|k-1}H_k^TK_k^T + K_kR_kK_k^T \\ &= (I - K_kH_k)P_{k|k-1} - P_{k|k-1}H_k^TK_k^T + K_k(H_kP_{k|k-1}H_k^T + R_k)K_k^T.\end{aligned}$$

Using the first definition of the Kalman gain, gives:

$$\begin{aligned}P_{k|k} &= (I - K_kH_k)P_{k|k-1} - P_{k|k-1}H_k^TK_k^T + P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}(H_kP_{k|k-1}H_k^T + R_k)K_k^T \\ &= (I - K_kH_k)P_{k|k-1} - P_{k|k-1}H_k^TK_k^T + P_{k|k-1}H_k^TK_k^T \\ &= (I - K_kH_k)P_{k|k-1}.\end{aligned}$$

Next the third characterisation of the covariance matrix is derived, starting from the second formulation. Using the first definition of the Kalman gain, we find:

$$\begin{aligned}P_{k|k} &= (I - K_kH_k)P_{k|k-1} = (I - P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}H_k)P_{k|k-1} \\ &= P_{k|k-1} - P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}H_kP_{k|k-1}.\end{aligned}$$

Applying the Woodbury matrix identity, see Lemma 2.1.17, we find:

$$P_{k|k} = (P_{k|k-1}^{-1} + H_k^TR_k^{-1}H_k)^{-1}.$$

We shift our focus to the Kalman gain. This derivation is based on a similar derivation in the book written by Y. Bar-Shalom et al. [2]. To make notation easier we use $S_k = H_kP_{k|k-1}H_k^T + R_k$. Thus:

$$\begin{aligned}K_k &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1} \\ &= P_{k|k-1}H_k^TS_k^{-1} \\ &= P_{k|k-1}H_k^TS_k^{-1}R_kR_k^{-1}.\end{aligned}$$

Adding and subtracting $H_kP_{k|k-1}H_k^T$ to R_k, gives:

$$\begin{aligned}K_k &= P_{k|k-1}H_k^TS_k^{-1}(H_kP_{k|k-1}H_k^T + R_k - H_kP_{k|k-1}H_k^T)R_k^{-1} \\ &= P_{k|k-1}H_k^TS_k^{-1}(S_k - H_kP_{k|k-1}H_k^T)R_k^{-1} \\ &= (P_{k|k-1}H_k^T - P_{k|k-1}H_k^TS_k^{-1}H_kP_{k|k-1}H_k^T)R_k^{-1}.\end{aligned}$$

Using $K_k = P_{k|k-1}H_k^TS_k^{-1}$, we find:

$$\begin{aligned}K_k &= (P_{k|k-1}H_k^T - K_kH_kP_{k|k-1}H_k^T)R_k^{-1} \\ &= (I - K_kH_k)P_{k|k-1}H_k^TR_k^{-1}.\end{aligned}$$

Using the second definition of $P_{k|k}$, we obtain:

$$K_k = P_{k|k}H_k^TR_k^{-1}.$$

This concludes the claim. □

Proof of Lemma 2.1.17. We prove the formula directly by multiplying both sides by $A + UCV$.

$$\begin{aligned}&(A + UCV)(A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}) \\ &\quad = I - U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} + UCVA^{-1} - UCVA^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \\ &\quad = I + UCVA^{-1} - (U + UCVA^{-1}U)(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \\ &\quad = I + UCVA^{-1} - UC(C^{-1} + VA^{-1}U)(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \\ &\quad = I + UCVA^{-1} - UCIVA^{-1} \\ &\quad = I\end{aligned}$$

This concludes the claim. □

2.2. Derivation of continuous Kalman filter

This section contains a derivation of the continuous times Kalman filter. The continuous time Kalman filter will be derived as a limit of the discrete time Kalman filter.

The derivation in this section is based on the book 'Optimal and robust estimation' written by F. L. Lewis et al. [10] and the book 'Estimation with applications to tracking and navigation' written by Y. Bar-Shalom et al. [2]. We consider the following continuous dynamic system:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = A(t)X(t) + D(t)W(t), \tag{2.7a}$$

$$Y(t) = C(t)X(t) + V(t), \tag{2.7b}$$

where $W : \mathbb{R}^+ \to \mathbb{R}^n$ and $V : \mathbb{R}^+ \to \mathbb{R}^m$ are continuous Gaussian white noise and are independent of each other. The intensity of $W(t)$ is $Q : \mathbb{R}^+ \to \mathbb{R}^n \times \mathbb{R}^n$ and the intensity of $V(t)$ is $R(t)$, with $R : \mathbb{R} \to \mathbb{R}^m \times \mathbb{R}^m$. Engineers use the word 'intensity' for the variance of Brownian noise. From here on, 'intensity' will be used when the random variable is continuous and variance will be used when the random variable is discrete. Note that A, C and D are continuous linear functions for fixed t. $A : \mathbb{R}^+ \to \mathbb{R}^n \times \mathbb{R}^n$, $C : \mathbb{R}^+ \to \mathbb{R}^m \times \mathbb{R}^n$, $D : \mathbb{R}^+ \to \mathbb{R}^n \times \mathbb{R}^n$. $X : \mathbb{R}^+ \to \mathbb{R}^n$ and $Y : \mathbb{R}^+ \to \mathbb{R}^m$ are continuous random variables.

Theorem 2.2.1 derives the solution of the continuous time Kalman filter.

Theorem 2.2.1. *The solution to the dynamic system given in Equation* (2.7) *is:*

$$\begin{aligned}
\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} &= A(t)\hat{x}(t) + P(t)C(t)^T R(t)^{-1}[y(t) - C(t)\hat{x}(t)] \\
&= A(t)\hat{x}(t) + K(t)[y(t) - C(t)\hat{x}(t)], \\
\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - P(t)C(t)^T R(t)^{-1}C(t)P(t) \\
&= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - K(t)C(t)P(t) \\
&= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - K(t)R(t)K(t)^T, \\
K(t) &= P(t)C(t)^T R(t)^{-1}.
\end{aligned}$$

To prove this theorem, two auxiliary results are needed. Lemma 2.2.2 discretises the system given in Equation (2.7). Then Lemma 2.2.4 finds the discrete Kalman solution to the discretised system of Lemma 2.2.2. Theorem 2.2.1 is proven by taking the limit over the time step size of the solution of Lemma 2.2.4. It is also possible to derive the continuous Kalman filter directly, but this is not shown in this work.

Lemma 2.2.2. *Discretising the system with time step length* Δ *given by Equation* (2.7) *gives the following dynamic system:*

$$\begin{aligned}
X(t_k) &= (I + A(t_{k-1})\Delta) X(t_{k-1}) + \tilde{W}(t_{k-1}), \\
Y(t_{k-1}) &= C(t_{k-1})X(t_{k-1}) + \tilde{V}(t_{k-1}),
\end{aligned} \tag{2.8}$$

with $\tilde{W}(t_{k-1}) = \Delta D(t_{k-1})W(t_{k-1})$ *and* $\tilde{V}(t_{k-1}) = \frac{1}{\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta} V(\tau)\ \mathrm{d}\tau$. *The covariance of* $\tilde{W}$ *and* $\tilde{V}$ *are then given b:*

$$\begin{aligned}
\tilde{Q}(t_{k-1}) &= \mathit{Cov}(\tilde{W}(t_{k-1})) = \Delta D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T, \\
\tilde{R}(t_{k-1}) &= \mathit{Cov}(\tilde{V}(t_{k-1})) = \frac{1}{\Delta}R(t_{k-1}),
\end{aligned}$$

respectively.

This lemma will be proven at the end of this subsection.

Remark 2.2.3. *It may seem contradictory that Cov*$(\tilde{W}(t_{k-1})) = \Delta D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T$ *while Cov*$(\tilde{V}(t_{k-1})) = \frac{1}{\Delta}R(t_{k-1})$ *as these differ a factor* Δ^2*. It may also seem strange that as* Δ *gets smaller, the covariance of the noise* $\tilde{R}(t)$ *grows. Since* $V(t)$ *is discretised by averaging over an interval, the longer the interval, the closer the average will be to zero (since the noise is centred around zero). Therefore, by averaging over a smaller interval (*Δ *goes to zero), the average will be further from zero. This is reflected in an increasing variance, see Figure 2.1.*

$W(t)$ *is discretised by the Euler Maruyama method, which is the stochastic extension of the Euler method or the first order Taylor expansion.* $\tilde{W}(t_k)$ *is the accumulated noise of* $W(t)$ *over an interval of length* Δ*, because the noise has an accumulative effect on the state variable. By integrating over* $W(t_k)$*,* $\tilde{W}(t_k)$ *accounts for all the noise in the interval of length* Δ*. The integral over* $W(t)$ *is approximated by multiplying* $W(t_k)$ *with* Δ*. When* Δ *goes to zero, the error of this approximation goes to zero. When* Δ *goes to zero, the area of* $\Delta W(t_k)$ *also decreases. Thus* $\tilde{W}(t)$ *decreases, which means that the variance of* $\tilde{W}(t)$ *also decreases. For a visual interpretation, see Figure 2.1.*

Lemma 2.2.4. *When applying the Kalman filter to the system given in Equation* (2.8) *we find:*

Prediction step:

$$\begin{aligned}
\hat{x}(t_k|t_{k-1}) &= (I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1}),\\
P(t_k|t_{k-1}) &= (I + A(t_{k-1})\Delta)P(t_{k-1}|t_{k-1})(I + A(t_{k-1})\Delta)^T + \Delta D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T.
\end{aligned}$$

Update step:

$$\begin{aligned}
\hat{x}(t_k|t_k) &= \hat{x}(t_k|t_{k-1}) + K(t_k)[y(t_k) - C(t_k)\hat{x}(t_k|t_{k-1})],\\
P(t_k|t_k) &= (I - K(t_k)C(t_k))P(t_k|t_{k-1}),\\
K(t_k) &= P(t_k|t_{k-1})C(t_k)^T\left(C(t_k)P(t_k|t_{k-1})C(t_k)^T + \frac{R(t_k)}{\Delta}\right)^{-1}.
\end{aligned}$$

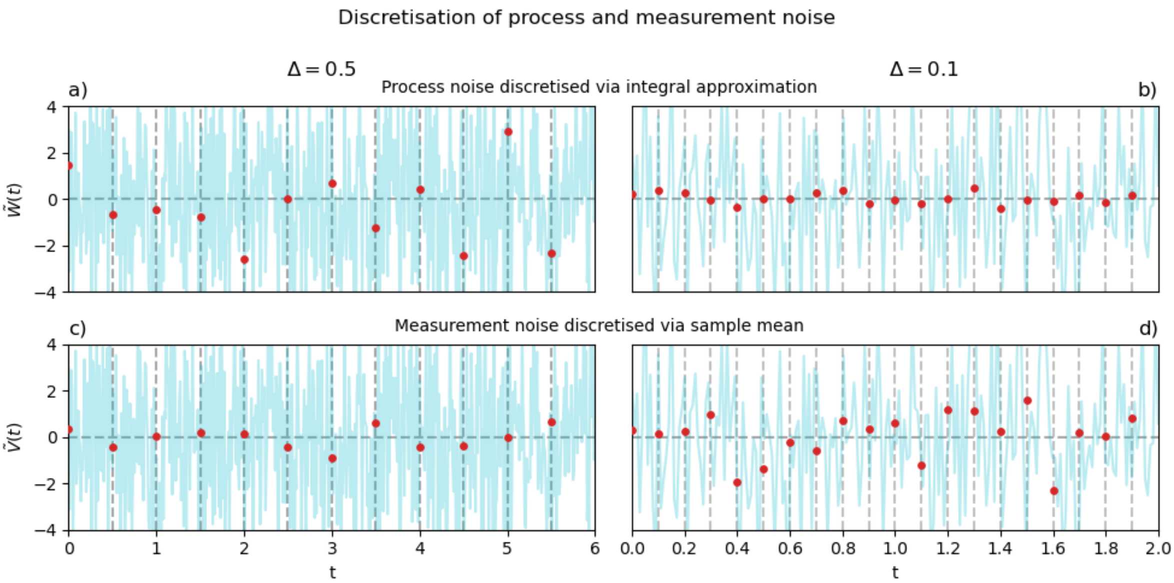

Figure 2.1: This figure shows the discretisation of the process and measurement noise. All figures show continuous time Gaussian white noise in blue in the background with the red dots representing the discrete approximation of the continuous noise. The process noise is discretised by approximating the integral over the the Euler Maruyama discretisation, which is shown in figure a) and b). Figure a) shows the discretisation with $\Delta = 0.5$ and figure b) uses $\Delta = 0.1$. It is clear that the variance of the discrete process noise decreases when Δ decreases. Figure c) and d) show the discretisation of the measurement noise with $\Delta = 0.5$ and $\Delta = 0.1$, respectively. The measurement noise is discretised by calculating the mean over intervals of length Δ. It is clear that the variance of the noise increases when Δ decreases.

This result is proven below. We have enough background to prove Theorem 2.2.1 by taking the limit over the time step size of the solution given in Lemma 2.2.4.

Proof of Theorem 2.2.1. We start by taking the limit of $\Delta \to 0$ of the Kalman gain of the discretised solution from Lemma 2.2.4:

$$\lim_{\Delta\to 0} K(t_k) = \lim_{\Delta\to 0} P(t_k|t_{k-1})C(t_{k-1})^T\left(C(t_{k-1})P(t_k|t_{k-1})C(t_{k-1})^T + \frac{R(t_{k-1})}{\Delta}\right)^{-1}.$$

Taking Δ out of the brackets gives:

$$\lim_{\Delta\to 0} K(t_k) = \lim_{\Delta\to 0} \Delta P(t_k|t_{k-1})C(t_{k-1})^T\left(\Delta C(t_{k-1})P(t_k|t_{k-1})C(t_{k-1})^T + R(t_{k-1})\right)^{-1} = 0.$$

To obtain the differential equation for the a priori covariance, the difference equation of the a priori covariance matrix is derived. We then take the limit of $\Delta \to 0$ of the difference equation to obtain the differential equation. To this end, we rewrite the expression for $P(t_k|t_{k-1})$ by working out the brackets in the definition of $P(t_k|t_{k-1})$:

$$\begin{aligned}P(t_k|t_{k-1}) = P(t_{k-1}|t_{k-1}) &+ \Delta[A(t_{k-1})P(t_{k-1}|t_{k-1}) + P(t_{k-1}|t_{k-1})A(t_{k-1})^T \\ &+ D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T + \Delta A(t_{k-1})P(t_{k-1}|t_{k-1})A(t_{k-1})^T].\end{aligned}$$

Using the expression for $P(t_{k-1}|t_{k-1})$, we obtain:

$$\begin{aligned}P(t_k|t_{k-1}) = (I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2}) &+ \Delta[A(t_{k-1})(I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2}) \\ &+ (I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2})A(t_{k-1})^T + \Delta D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T \\ &+ \Delta A(t_{k-1})(I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2})A(t_{k-1})].\end{aligned}$$

We subtract $P(t_{k-1}|t_{k-2})$ from both sides and divide both sides by Δ to find:

$$\begin{aligned}\frac{1}{\Delta}\left(P(t_k|t_{k-1}) - P(t_{k-1}|t_{k-2})\right) &= -\frac{1}{\Delta}K(t_{k-1})C(t_{k-1})P(t_{k-1}|t_{k-2}) \\ &+ A(t_{k-1})(I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2}) + (I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2})A(t_{k-1})^T \\ &+ D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T + \Delta A(t_{k-1})(I - K(t_{k-1})C(t_{k-1}))P(t_{k-1}|t_{k-2})A(t_{k-1})^T.\end{aligned}$$

Working out the brackets gives:

$$\begin{aligned}\frac{1}{\Delta}(P(t_k|t_{k-1}) - P(t_{k-1}|t_{k-2})) &= -\frac{1}{\Delta}K(t_{k-1})C(t_{k-1})P(t_{k-1}|t_{k-2}) + A(t_{k-1})P(t_{k-1}|t_{k-2}) \\ &- A(t_{k-1})K(t_{k-1})C(t_{k-1})P(t_{k-1}|t_{k-2}) + P(t_{k-1}|t_{k-2})A(t_{k-1})^T \\ &- K(t_{k-1})C(t_{k-1})P(t_{k-1}|t_{k-2})A(t_{k-1})^T + D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T.\end{aligned} \tag{2.9}$$

Before taking the limit of $\Delta \to 0$ of this difference equation, we calculate the limit of $\frac{1}{\Delta}K(t_k)$ when $\Delta \to 0$:

$$\lim_{\Delta\to 0}\frac{1}{\Delta}K(t_k) = \frac{1}{\Delta}P(t_k|t_{k-1})C(t_k)^T\left(C(t_k)P(t_k|t_{k-1})C(t_k)^T + \frac{R(t_k)}{\Delta}\right)^{-1}.$$

Taking Δ out of the brackets gives:

$$\begin{aligned}\lim_{\Delta\to 0}\frac{1}{\Delta}K(t_k) &= \lim_{\Delta\to 0} P(t_k|t_{k-1})C(t_k)^T(\Delta C(t_k)P(t_k|t_{k-1})C(t_k)^T + R(t_k))^{-1} \\ &= P(t_k|t_{k-1})C(t_k)^T R(t_k)^{-1}.\end{aligned}$$

Using $\lim_{\Delta\to 0} K(t_k) = 0$ and the above result in Equation (2.9) and taking the limit of Δ to zero, gives:

$$\begin{aligned}\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= \lim_{\Delta\to 0}\frac{1}{\Delta}(P(t_k|t_{k-1}) - P(t_{k-1}|t_{k-2})) \\ &= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - P(t)C(t)^T R(t)^{-1}C(t)P(t).\end{aligned}$$

Note that we can drop the subscripts k and $k+1$ of t because A, C, D, P, R and Q are all continuous functions. We left out notation to indicate that this is the differential equation for the a priori covariance because we will show that the differential equation for the a priori and a posteriori covariance matrices are equal. To do this we take a look at the difference between the a posteriori and the a priori covariance:

$$\begin{aligned}P(t_k|t_k) - P(t_k|t_{k-1}) &= (I - K(t_k)C(t_k))P(t_k|t_{k-1}) - P(t_k|t_{k-1}) \\ &= -K(t_k)C(t_k)P(t_k|t_{k-1}) \\ &= -K(t_k)C(t_k)\big((I + A(t_{k-1})\Delta)P(t_{k-1}|t_{k-1})(I + A(t_{k-1})\Delta)^T \\ &\quad + \Delta D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T\big) \\ &= -K(t_k)C(t_k)P(t_{k-1}|t_{k-1}) - \Delta K(t_k)C(t_k)P(t_{k-1}|t_{k-1})A(t_{k-1})^T \\ &\quad - \Delta K(t_k)C(t_k)A(t_{k-1})P(t_{k-1}|t_{k-1}) - \Delta^2 K(t_k)C(t_k)A(t_{k-1})P(t_{k-1}|t_{k-1})A(t_{k-1})^T \\ &\quad - \Delta K(t_k)C(t_k)D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T.\end{aligned}$$

We take the limit over this difference with $\Delta \to 0$. Because $\lim_{\Delta\to 0} K(t_k) = 0$ we find:

$$\lim_{\Delta\to 0} P(t_k|t_k) - P(t_k|t_{k-1}) = 0.$$

From this, we conclude that the differential equation for the a priori and the a posteriori covariance are the same. We can rewrite the differential equation of the covariance matrix to find the expression for $K(t)$:

$$\begin{aligned}\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - P(t)C(t)^T R(t)^{-1}C(t)P(t) \\ &= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - P(t)C(t)^T R(t)^{-1}R(t)R(t)^{-1}C(t)P(t) \\ &= A(t)P(t) + P(t)A(t)^T + D(t)Q(t)D(t)^T - K(t)R(t)K(t)^T,\end{aligned}$$

with

$$K(t) = P(t)C(t)^T R(t)^{-1}.$$

We continue the proof by finding the differential equation for the state estimate. We do this in a similar way to the covariance matrix. So we start with rewriting the formula for $\hat{x}(t_k|t_k)$ by working out the brackets:

$$\begin{aligned}\hat{x}(t_k|t_k) &= (I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1}) + K(t_k)[y(t_k) - C(t_k)(I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1})] \\ &= \hat{x}(t_{k-1}|t_{k-1}) + \Delta A(t_{k-1})\hat{x}(t_{k-1}|t_{k-1}) + K(t_k)[y(t_k) - C(t_k)\hat{x}(t_{k-1}|t_{k-1})] \\ &\quad - \Delta K(t_k)C(t_k)A(t_{k-1})\hat{x}(t_{k-1}|t_{k-1}).\end{aligned}$$

We subtract $\hat{x}(t_{k-1}|t_{k-1})$ from both sides and divide by Δ to obtain:

$$\begin{aligned}\frac{1}{\Delta}(\hat{x}(t_k|t_k) - \hat{x}(t_{k-1}|t_{k-1})) &= A(t_{k-1})\hat{x}(t_{k-1}|t_{k-1}) + \frac{1}{\Delta}K(t_k)[y(t_k) - C(t_k)\hat{x}(t_{k-1}|t_{k-1})] \\ &\quad - K(t_k)C(t_k)A(t_{k-1})\hat{x}(t_{k-1}|t_{k-1}).\end{aligned}$$

We take the limit $\Delta \to 0$ using $\lim_{\Delta\to 0} K(t_k) = 0$ and $\lim_{\Delta\to 0} \frac{1}{\Delta}K(t_k) = P(t)C(t)^T R(t)^{-1}$, we find:

$$\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} = A(t)\hat{x}(t) + P(t)C(t)^T R(t)^{-1}[y(t) - C(t)\hat{x}(t)] \tag{2.10}$$

We can again drop the subscripts of t because all functions involved are continuous. Finally, we show that the a priori and a posteriori differential equations of state are equal by calculating the difference between the a priori and a posteriori estimate:

$$\begin{aligned}\hat{x}(t_k|t_k) - \hat{x}(t_k|t_{k-1}) &= \hat{x}(t_k|t_{k-1}) + K(t_k)[y(t_k) - C(t_k)\hat{x}(t_k|t_{k-1})] - (I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1}) \\ &= (I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1}) + K(t_k)[y(t_k) - C(t_k)(I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1})] \\ &\quad - (I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1}) \\ &= K(t_k)[y(t_k) - C(t_k)(I + A(t_{k-1})\Delta)\hat{x}(t_{k-1}|t_{k-1})] \\ &= K(t_k)y(t_k) - K(t_k)C(t_k)\hat{x}(t_{k-1}|t_{k-1}) - \Delta K(t_k)C(t_k)A(t_{k-1})\hat{x}(t_{k-1}|t_{k-1}).\end{aligned}$$

Taking the limit of $\Delta \to 0$ to zero over this difference and using $\lim_{\Delta\to 0} K(t_k) = 0$ gives:

$$\lim_{\Delta\to 0}(\hat{x}(t_k|t_k) - \hat{x}(t_k|t_{k-1})) = 0.$$

This concludes the proof. □

Below we prove the auxiliary results.

Sketch of proof of Lemma 2.2.2. We will discretise Equation (2.7a) using the Euler-Maruyama method. This gives:

$$\begin{aligned} X(t_k) &= (I + A(t_{k-1})\Delta)X(t_{k-1}) + \tilde{W}(t_{k-1}), \\ Y(t_{k-1}) &= C(t_{k-1})X(t_{k-1}) + \tilde{V}(t_{k-1}). \end{aligned}$$

We derive $\tilde{W}(t_{k-1})$ and $\tilde{V}(t_{k-1})$ and their perspective covariances. $\tilde{W}(t_{k-1})$ is discretised by the Euler Maruyama method along with the rest of the state equation. So we find:

$$\tilde{W}(t_{k-1}) = \lim_{\Delta\to 0}\int_{t_{k-1}}^{t_{k-1}+\Delta} (I + A(\tau)\Delta)D(\tau)W(\tau)\ \mathrm{d}\tau.$$

When Δ is sufficiently small $I + A(\tau)\Delta$ tends to the identity matrix, so we find:

$$\begin{aligned} \tilde{W}(t_{k-1}) &\approx \lim_{\Delta\to 0}\int_{t_{k-1}}^{t_{k-1}+\Delta} D(\tau)W(\tau)\ \mathrm{d}\tau \\ &\approx \Delta D(t_{k-1})W(t_{k-1}). \end{aligned}$$

We continue to derive the covariance of $\tilde{W}(t_{k-1})$:

$$\begin{aligned} \tilde{Q}(t_{k-1}) &= \mathbb{E}[\tilde{W}(t_{k-1})\tilde{W}(t_{k-1})^T] \\ &= \mathbb{E}\left[\Big(\int_{t_{k-1}}^{t_{k-1}+\Delta} D(\tau_1)W(\tau_1)\ \mathrm{d}\tau_1\Big)\Big(\int_{t_{k-1}}^{t_{k-1}+\Delta} D(\tau_2)W(\tau_2)\ \mathrm{d}\tau_2\Big)^T\right]. \end{aligned}$$

Using Fubini's theorem, we find:

$$\begin{aligned} \tilde{Q}(t_{k-1}) &= \int_{t_{k-1}}^{t_{k-1}+\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta} \mathbb{E}[D(\tau_1)W(\tau_1)(D(\tau_2)W(\tau_2))^T]\ \mathrm{d}\tau_1\,\mathrm{d}\tau_2 \\ &= \int_{t_{k-1}}^{t_{k-1}+\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta} D(\tau_1)Q(\tau_1)D(\tau_2)^T\delta(\tau_1 - \tau_2)\ \mathrm{d}\tau_1\,\mathrm{d}\tau_2 \\ &= \int_{t_{k-1}}^{t_{k-1}+\Delta} D(\tau)Q(\tau)D(\tau)^T\,\mathrm{d}\tau \\ &\approx \Delta D(t_{k-1})Q(t_{k-1})D(t_{k-1})^T, \end{aligned}$$

where $\delta(t)$ is the Dirac-delta function. Note that $\tilde{Q}(t_{k-1})$ is the covariance, not the intensity because the dynamic system is discrete. Equation (2.7b) is discretised by taking the average over a small interval, so $\tilde{V}(t_{k-1})$ is discretised in the same manner. We thus obtain:

$$\tilde{V}(t_{k-1}) = \lim_{\Delta\to 0}\frac{1}{\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta} V(\tau)\ \mathrm{d}\tau.$$

We continue to derive the covariance of $\tilde{V}(t_{k-1})$:

$$\begin{aligned} \tilde{R}(t_{k-1}) &= \mathbb{E}[\tilde{V}(t_{k-1})\tilde{V}(t_{k-1})^T] \\ &= \mathbb{E}\left[\Big(\frac{1}{\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta} V(\tau_1)\ \mathrm{d}\tau_1\Big)\Big(\frac{1}{\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta} V(\tau_2)\ \mathrm{d}\tau_2\Big)^T\right]. \end{aligned}$$

Applying Fubini's theorem, we obtain:

$$\begin{aligned}
\tilde{R}(t_{k-1}) &= \frac{1}{\Delta^2}\int_{t_{k-1}}^{t_{k-1}+\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta}\mathbb{E}[V(\tau_1)V(\tau_2)^T]\ \mathsf{d}\tau_1\,\mathsf{d}\tau_2 \\
&= \frac{1}{\Delta^2}\int_{t_{k-1}}^{t_{k-1}+\Delta}\int_{t_{k-1}}^{t_{k-1}+\Delta}R(\tau_1)\delta(\tau_1-\tau_2)\ \mathsf{d}\tau_1\,\mathsf{d}\tau_2 \\
&= \frac{1}{\Delta^2}\int_{t_{k-1}}^{t_{k-1}+\Delta}R(\tau)\ \mathsf{d}\tau \approx \frac{1}{\Delta}R(t_{k-1}).
\end{aligned}$$

□

Proof of Lemma 2.2.4. The solution is directly obtained by applying the discrete Kalman filter to the dynamic system in Equation (2.8). □

3

H_∞ filter

This chapter introduces the H_∞ filter. Unlike the Kalman filter, the H_∞ filter does not require any statistical properties of the noise. The cost of this relaxation is that the H_∞ filter does not provide an optimal estimate. However, it does ensure that the maximum error is bounded. The H_∞ filter involves a trade-off: either the maximum error can be kept small, resulting in a large error covariance of the estimate, or the error covariance can be reduced at the expense of a larger maximum error. Since the H_∞ filter does not rely on knowledge of noise characteristics, it is particularly useful in situations where error covariances are unknown. One example is estimating a location using satellites, where the extent to which the satellite signal is disturbed by cosmic radiation is uncertain.

Section 3.1 explains the background of the discrete H_∞ filter through the perspective of game theory. Section 3.2 gives a brief introduction into the continuous H_∞ filter. Section 3.3 compares the H_∞ filter with the Kalman filter. In this section, it becomes apparent that the H_∞ filter can be viewed as a generalisation of the Kalman filter. Finally, in Section 3.4 the Kalman filter is derived for non zero mean noise in an attempt to view the H_∞ filter as a Kalman filter with the mean of the noise shifted. It becomes clear that it is not correct to interpret the H_∞ filter that way.

3.1. Discrete H_∞ filter

This section will explain the discrete H_∞ filter using game theory. The H_∞ filter requires no assumptions on the statistics of the noise terms. The H_∞ filter minimises the worst case estimation error, whereas the Kalman filter minimises the mean squared error. The consequence is that the H_∞ filter is not necessarily the optimal filter in the least square sense. We consider the following dynamic system:

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + w_{k-1}, \\ y_{k-1} &= H_{k-1}x_{k-1} + v_{k-1}, \end{aligned} \tag{3.1}$$

where $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are the noise terms. These noise terms do not need to satisfy any conditions. The noise terms can, for example, be deterministic, they may have a non-zero mean and the covariance matrices need not be known. Small letters are used instead of capital letters to indicate that these noise terms are not necessarily random variables. $x_k \in \mathbb{R}^n$ is the state variable. $y_k \in \mathbb{R}^m$ represents the measurements. $F_k \in \mathbb{R}^n \times \mathbb{R}^n$ describes the model dynamics. $H_k \in \mathbb{R}^m \times \mathbb{R}^n$ provides a linear connection between the state variable and the measurements. The H_∞ filter estimates a linear combination of the state, $z_k = L_k x_k$. $L_k \in \mathbb{R}^n \times \mathbb{R}^n$ can be any full rank matrix. It is also possible to choose $L_k = I$ to estimate the state itself. L_k can be used to estimate any linear combination of the state. This may be useful when it is already known that the variable of interest is a linear combination of the state. We will explain the H_∞ filter according to the game theory approach, as in [14]. To this end, we define the cost function:

$$J_1 = \frac{\sum_{k=0}^{N-1} \|z_k - \hat{z}_k\|_{S_k}^2}{\|x_0 - \hat{x}_0\|_{P_0^{-1}}^2 + \sum_{k=0}^{N-1} \left(\|w_k\|_{Q_k^{-1}}^2 + \|v_k\|_{R_k^{-1}}^2\right)}, \tag{3.2}$$

where $P_0,\, S_k,\, Q_k \in \mathbb{R}^n \times \mathbb{R}^n$ and $R_k \in \mathbb{R}^m \times \mathbb{R}^m$ are all symmetric positive definite matrices. The vector norm is defined as $\|x\|_A^2 = x^T A x$. The idea of the game with the above cost function is that our opponent chooses the starting point x_0 and the noise terms v_k and w_k, with the goal to maximise J_1 and therefore $z_k - \hat{z}_k$. By choosing a large v_k and w_k, $z_k - \hat{z}_k$ will be large. However, since v_k and w_k are in the denominator, our opponent can not choose v_k and w_k arbitrarily large. We try to minimise J_1 and therefore $z_k - \hat{z}_k$. Thus this game can be written down as the following minimax problem:

$$\min_{\hat{z}_k} \max_{w_k, v_k, x_0} J_1. \tag{3.3}$$

Note that P_0, S_k, Q_k and R_k are chosen in accordance with a specific problem. If information about the noise terms is known, this information can be incorporated into these matrices. If, for example, the covariance would be known and the noise is zero mean, then P_0, Q_k and R_k should be chosen to be the covariance matrices. It is unfortunately not tractable to minimise J_1 directly, therefore a performance bound is chosen such that $J_1 < \frac{1}{\theta}$, where $\theta > 0$ is the performance bound. Using this performance bound gives the following variant on Equation (3.2) and Equation (3.3):

$$J = -\frac{1}{\theta}\|x_0 - \hat{x}_0\|_{P_0^{-1}}^2 + \sum_{k=0}^{N-1} \left[\|z_k - \hat{z}_k\|_{S_k}^2 - \frac{1}{\theta}\left(\|w_k\|_{Q_k^{-1}}^2 + \|v_k\|_{R_k^{-1}}^2\right)\right] < 1, \tag{3.4}$$

the minimax problem then becomes:

$$J^\star = \min_{\hat{z}_k} \max_{w_k, v_k, x_0} J.$$

Since $y_k = H_k x_k + v_k$ we find $v_k = y_k - H_k x_k$. So the minimax problem can be rewritten as:

$$J^\star = \min_{\hat{z}_k} \max_{w_k, y_k, x_0} -\frac{1}{\theta}\|x_0 - \hat{x}_0\|_{P_0^{-1}}^2 + \sum_{k=0}^{N-1} \left[\|z_k - \hat{z}_k\|_{S_k}^2 - \frac{1}{\theta}\left(\|w_k\|_{Q_k^{-1}}^2 + \|y_k - H_k x_k\|_{R_k^{-1}}^2\right)\right]. \tag{3.5}$$

The solution to this minimax problem is the H_∞ filter. We will not derive the solution, but we will state it in the following theorem.

Theorem 3.1.1 (H_∞ filter)**.** *The solution to the dynamic system given in Equation* (3.1) *by the H_∞ filter is given by:*

Prediction step:

$$\begin{aligned}
\hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1}, \\
P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}.
\end{aligned}$$

Update step:

$$\begin{aligned}
\bar{S}_k &= L_k^T S_k L_k, \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}), \\
P_{k|k} &= P_{k|k-1}[I - \theta\bar{S}_k P_{k|k-1} + H_k^T R_k^{-1} H_k P_{k|k-1}]^{-1}, \\
K_k &= P_{k|k-1}[I - \theta\bar{S}_k P_{k|k-1} + H_k^T R_k^{-1} H_k P_{k|k-1}]^{-1} H_k^T R_k^{-1}.
\end{aligned} \tag{3.6}$$

provided that at each time step the following holds:

$$P_{k|k-1} - \theta\bar{S}_k + H_k^T R_k^{-1} H_k \succ 0. \tag{3.7}$$

Remark 3.1.2. *Even though the noise terms are allowed to be deterministic, the covariance matrices Q_k and R_k cannot be the zero matrix, because they need to be positive definite. So, even if the noise is deterministic, the noise is modelled with a small variance.*

The condition in Equation (3.7) *ensures that the solution found is a minimum for $\hat{z}_k$. In the proof of Theorem 3.1.1, see [14], another condition arises to ensure that the H_∞ filter is a maximum for y_k. This condition is:*

$$R_k - H_k P_{k|k} H_k^T \succ 0.$$

In the derivation given in [14], this condition is considered only relevant for academic purposes. Do note however, that when this additional condition is not satisfied, the H_∞ filter is strictly speaking not a solution to the optimisation problem in Equation (3.5).

In the H_∞ filter, it is not necessary to know the statistics of the noise. In case these statistics are known, they can be incorporated. The choice of Q_k and R_k influences the quality of the filter, even though the maximum error is bounded [16]. Q_k and R_k should therefore always be tuned before using the filter. This could in a sense be seen as estimating Q_k and R_k. Of course, the tightness of the bound on the error is dependent on the choice of θ. Choosing a large θ will lead to a tighter bound on the error, however, this will increase the variance of the estimate. Thus there is a trade-off between the tightness of the bound error of the estimate and the size of the variance of the estimate. Also θ can not be chosen infinitely large because $P_{k|k-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k \succ 0$ needs to hold for every time step. Choosing θ too large will violate this condition.

From a game theory perspective it is interesting to see whether is it possible to interchange the minimum and maximum. Being allowed to switch the order of the minimum and the maximum means that it does not matter whether our opponent or we choose our position in the game first. It is unfortunately not possible to use the minimax theorem from von Neumann since $\hat{z}_k, w_k, x_0 \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ and $\mathbb{R}^n$ and $\mathbb{R}^m$ are not compact spaces. There is however an generalisation of the minimax theorem, which can be found in the book 'Convex Analysis and Variational Problems' written by I. Ekeland and R. Téman [4]. This generalisation gives rise to the following proposition.

Proposition 3.1.3. *The minimax problem given in Equation* (3.5) *can be generalised in de following way:*

$$\begin{aligned}
&\min_{\hat{z}_k} \sup_{x_0, w_k, y_k} -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1}\left[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\left(\|w_k\|^2_{Q_k^{-1}} + \|y_k - H_k x_k\|^2_{R_k^{-1}}\right)\right] \\
&\quad = \max_{x_0, w_k, y_k} \inf_{\hat{z}_k} -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1}\left[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\left(\|w_k\|^2_{Q_k^{-1}} + \|y_k - H_k x_k\|^2_{R_k^{-1}}\right)\right].
\end{aligned}$$

To prove this proposition, some definitions and a theorem are needed.

Definition 3.1.4 (Convex set)**.** *[4] Let V be a vector space. If $u, v \in V$, u and v are called the end points of the line-segment denoted by $[u, v]$, where*

$$[u, v] = \{\lambda u + (1-\lambda)v \,|\, 0 \leq \lambda \leq 1\}.$$

A set $A \subset V$ is convex if and only if for every pair of elements (u, v) of A the line-segment $[u, v]$ is contained in A.

Definition 3.1.5 (Reflexive Banach space)**.** *[4] A Banach space is a complete normed vector space. A Banach space is reflexive if its unit ball is compact in the weak topology. This implies that every bounded sequence admits a weakly converging subsequence.*

Note that every finite dimensional normed space is a reflexive Banach space.

Definition 3.1.6 (Lower semi-continuous function)**.** *[4] A function $F : V \to \mathbb{R}$ is said to be lower semi-continuous on V if it satisfies the two equivalent conditions:*

(a) $\forall a \in \mathbb{R}, \{u \in V | F(u) \leq a\}$ *is closed,*
(b) $\forall \bar{u} \in V, \lim_{u \to \bar{u}} F(u) \geq F(\bar{u})$.

F is called upper semi-continuous if $-F$ is lower semi-continuous. A function is continuous if it is both upper and lower semi-continuous.

Theorem 3.1.7 (Generalisation of minimax theorem)**.** *[4] Let V and Z be two reflexive Banach spaces. Let $A \subset V$ and $B \subset Z$ both convex, closed and non-empty. If the function $L : A \times B \to \mathbb{R}$ satisfies:*

(a) $\forall u \in A, w \to L(u, w)$ *is concave and upper semi-continuous,*
(b) $\forall w \in B, u \to L(u, w)$ *is convex and lower semi-continuous,*

such that $\exists w_0 \in B$ such that:

$$\lim_{\substack{u \in A \\ \|u\| \to \infty}} L(u, w_0) = +\infty,$$

and $\exists u_0 \in A$ such that:

$$\lim_{\substack{w \in B \\ \|w\| \to \infty}} L(u_0, w) = -\infty.$$

Then L possesses at least one saddle point and

$$\min_{u \in A} \sup_{w \in B} L(u, w) = \max_{w \in B} \inf_{u \in A} L(u, w).$$

For the proof we refer to [4]. Enough background is established to prove Proposition 3.1.3.

Proof of Proposition 3.1.3. To apply Theorem 3.1.7 to the cost function of Equation (3.5) we define the vector $c_k \in \mathbb{R}^{(2n+m)}$ as $c_k = [x_0, w_k, y_k]^T$. We rewrite Equation (3.5) as follows:

$$\begin{aligned} J^\star &= \min_{\hat{z}_k} \max_{c_k} -\frac{1}{\theta} \|c_k - \hat{c}_k\|^2_{\tilde{P}_0^{-1}} + \sum_{k=0}^{N-1} \left[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta} \left(\|c_k\|^2_{\tilde{Q}_k^{-1}} + \|c_k - \tilde{H}_k \tilde{x}_k\|^2_{\tilde{R}_k^{-1}} \right) \right] \\ &= \min_{\hat{z}_k} \max_{c_k} F(\hat{z}_k, c_k). \end{aligned}$$

We define $d_m \in \mathbb{R}^m$ as a zero vector, we also define $d_n \in \mathbb{R}^n$ in the same fashion. $D_{nn} \in \mathbb{R}^n \times \mathbb{R}^n$ is a zero matrix, similarly for $D_{mm} \in \mathbb{R}^m \times \mathbb{R}^m$, $D_{nm} \in \mathbb{R}^n \times \mathbb{R}^m$ and $D_{mn} \in \mathbb{R}^m \times \mathbb{R}^n$. Then,

$$\hat{c}_k = [\hat{x}_0, d_n, d_m]^T, \qquad \tilde{P}_0 = \begin{bmatrix} P_0 & D_{nn} & D_{nm} \\ D_{nn} & D_{nn} & D_{nm} \\ D_{mn} & D_{mn} & D_{mm} \end{bmatrix},$$

$$\tilde{H}_k = [D_{nn}, D_{nn}, H_k]^T, \qquad \tilde{Q}_k = \begin{bmatrix} D_{nn} & D_{nn} & D_{nm} \\ D_{nn} & Q_k & D_{nm} \\ D_{mn} & D_{mn} & D_{mm} \end{bmatrix},$$

$$\tilde{R}_k = \begin{bmatrix} D_{nn} & D_{nn} & D_{nm} \\ D_{nn} & D_{nn} & D_{nm} \\ D_{mn} & D_{mn} & R_k \end{bmatrix}.$$

If we equip $\mathbb{R}^n$ and $\mathbb{R}^{(2n+m)}$ with any matrix norm, then $\mathbb{R}^n$ and $\mathbb{R}^{(2n+m)}$ are reflexive Banach spaces. Then clearly, $\forall c_k \in \mathbb{R}^{(2n+m)}$ $F(\hat{z}_k, c_k)$ is convex and continuous and therefore lower semi-continuous. Similarly, $\forall \hat{z}_k \in \mathbb{R}^n$ $F(\hat{z}_k, c_k)$ is concave and continuous and therefore upper semi-continuous. Note that this only holds if θ is indeed positive. Let $c_0 \in \mathbb{R}^{(2n+m)}$ be the zero vector. Then,

$$\lim_{\substack{\hat{z}_k \in \mathbb{R}^n \\ \|\hat{z}_k\| \to \infty}} F(\hat{z}_k, c_0) = +\infty,$$

and of course if $\hat{z}_0 \in \mathbb{R}^n$ is the zero vector, then:

$$\lim_{\substack{c_k \in \mathbb{R}^{(2n+m)} \\ \|c_k\| \to \infty}} F(\hat{z}_0, c_k) = -\infty.$$

Thus we conclude with Theorem 3.1.7 that:

$$\begin{aligned}
& \min_{\hat{z}_k} \sup_{c_k} F(\hat{z}_k, c_k) = \max_{c_k} \inf_{\hat{z}_k} F(\hat{z}_k, c_k), \\
\Leftrightarrow \quad & \min_{\hat{z}_k} \sup_{x_0, w_k, y_k} -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1} \Big[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\big(\|w_k\|^2_{Q_k^{-1}} + \|y_k - H_k x_k\|^2_{R_k^{-1}}\big)\Big] \\
& = \max_{x_0, w_k, y_k} \inf_{\hat{z}_k} -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1} \Big[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\big(\|w_k\|^2_{Q_k^{-1}} + \|y_k - H_k x_k\|^2_{R_k^{-1}}\big)\Big].
\end{aligned}$$

Note that it is not possible to replace the supremum with a maximum and the infimum with a minimum, because $\mathbb{R}^n$ and $\mathbb{R}^{(2n+m)}$ are not compact. □

3.2. Continuous H_∞ filter

There also exists a continuous version of the H_∞ filter. This section will briefly introduce the continuous H_∞ filter. Note that the H_∞ filter is not applied in its continuous form. It is, however, very useful to analyse the filter. We do not derive the continuous filter. The filter is cited from [14].

We consider the following continuous dynamic system:

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= A(t)x(t) + w(t), \\
y(t) &= C(t)x(t) + v(t), \\
z(t) &= L(t)x(t).
\end{aligned} \tag{3.8}$$

for $t \in [0, T]$, with T. Note again that we use small letters for $x : \mathbb{R} \to \mathbb{R}^n, w : \mathbb{R} \to \mathbb{R}^n, y : \mathbb{R} \to \mathbb{R}^m$ and $v : \mathbb{R} \to \mathbb{R}^m$, because these are not necessarily continuous random variables, and might be continuous functions instead. We again estimate a linear combination of the state, where $z : \mathbb{R} \to R^n$ with $L : \mathbb{R} \to R^n \times \mathbb{R}^n$. A and C are continuous linear functions. $A : \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^n$ and $C : \mathbb{R} \to \mathbb{R}^m \times \mathbb{R}^n$. We define a continuous version of the cost function:

$$J_1 = \frac{\int_0^T \|z(t) - \hat{z}(t)\|^2_{S(t)}\, \mathrm{d}t}{\|x(0) - \hat{x}(0)\|^2_{P_0^{-1}} + \int_0^T (\|w(t)\|^2_{Q(t)^{-1}} + \|v(t)\|^2_{R(t)^{-1}}\, \mathrm{d}t},$$

where $P_0 \in \mathbb{R}^n \times \mathbb{R}^n$, $S : \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^n$, $Q : \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^n$ and $R : \mathbb{R} \to \mathbb{R}^m \times \mathbb{R}^m$. The matrices $S(t)$, $Q(t)$ and $R(t)$ must be positive definite for every $t \in \mathbb{R}$.

In continuous time it is also not tractable to minimise J_1 directly, so we set $J_1 < \frac{1}{\theta}$. We can rewrite this to:

$$J = -\frac{1}{\theta}\|x(0) - \hat{x}(0)\|^2_{P_0^{-1}} + \int_0^T \|z(t) - \hat{z}(t)\|^2_{S(t)} - \frac{1}{\theta}\left(\|w(t)\|^2_{Q(t)^{-1}} + \|v(t)\|^2_{R(t)^{-1}}\right) \mathrm{d}t < 1.$$

The continuous H_∞ gives the solution to the following optimisation problem:

$$J^\star = \min_{\hat{z}(t)} \max_{\substack{w(t), v(t), \\ x_0}} J.$$

The H_∞ filter is given in the following theorem.

Theorem 3.2.1. *Given the dynamic system of Equation* (3.8)*. The H_∞ filter is given by:*

$$\begin{aligned}
\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} &= A(t)\hat{x}(t) + K(t)(y(t) - C(t)\hat{x}(t). \\
\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= A(t)P(t) + P(t)A(t)^T + Q(t) - K(t)C(t)P(t) + \theta P(t)L(t)^T S(t)L(t)P(t), \\
K(t) &= P(t)C(t)^T R(t)^{-1}, \\
\hat{z}(t) &= L(t)\hat{x}(t),
\end{aligned}$$

provided that $P(t) \succ 0$ for $t \in [0, T]$.

Remark 3.2.2. *Note that $P(t) \succ 0$ for $t \in [0, T]$ does not automatically hold, because $Q(t)$ and $R(t)$ are not necessarily the intensity of $w(t)$ and $v(t)$, respectively. So, $Q(t)$ and $R(t)$ can be chosen such that $P(t) \succ 0$ for $t \in [0, T]$ does not hold. In practice, for the H_∞ filter to behave as desired the covariance matrix calculated by the H_∞ filter will be larger than the covariance matrix calculated by the Kalman filter. This reflects the extra uncertainty added by not knowing the intensities $Q(t)$ and $R(t)$. Because the covariance matrix calculated by the Kalman filter is positive definite, in practice the covariance matrix calculated by the H_∞ filter is expected to be positive definite as well.*

3.3. Comparison Kalman filter and H_∞ filter

In this section we compare the Kalman filter with the H_∞ filter. We start comparing the discrete filters in Subsection 3.3.1 and then move on to the continuous filters in Subsection 3.3.2.

3.3.1. Discrete filters

This subsection compares the discrete Kalman filter with the discrete H_∞ filter. At first glance, it is not clear how these two filters are related. Therefore, we rewrite the solution of the H_∞ filter so that the differences become apparent. We do so in the following lemma.

Lemma 3.3.1 (Rewritten solution of the H_∞ filter)**.** *The solution of the H_∞ can be rewritten in the following way:*

Prediction step:

$$\begin{aligned}
\hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1}, \\
P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k+1}.
\end{aligned}$$

Update step:

$$\begin{aligned}
\bar{S}_k &= L_k^T S_k L_k, \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}), \\
P_{k|k} &= P_{k|k-1}(I - \theta\bar{S}_k P_{k|k-1} + H_k^T R_k^{-1} H_k P_{k|k-1})^{-1} && \text{(3.9a)} \\
&= (P_{k|k-1}^{-1} - \theta\bar{S}_k + H_k^T R_k^{-1} H_k)^{-1} && \text{(3.9b)} \\
&= (I - K_k H_k)(P_{k|k-1}^{-1} - \theta\bar{S}_k)^{-1}, && \text{(3.9c)} \\
K_k &= P_{k|k-1}[I - \theta\bar{S}_k P_{k|k-1} + H_k^T R_k^{-1} H_k P_{k|k-1}]^{-1} H_k^T R_k^{-1} && \text{(3.9d)} \\
&= P_{k|k} H_k^T R_k^{-1} && \text{(3.9e)} \\
&= (P_{k|k-1}^{-1} - \theta\bar{S}_k)^{-1} H_k^T [(H_k(P_{k|k-1}^{-1} - \theta\bar{S}_k)^{-1} H_k^T + R_k]^{-1}, && \text{(3.9f)}
\end{aligned}$$

provided that at each time step the following hold:

$$P_{k|k-1} - \theta\bar{S}_k + H_k^T R_k^{-1} H_k \succ 0.$$

Note that the second definition of $P_{k|k}$ only goes together with the second definition of K_k.

This lemma is proven later in this subsection.

Proposition 3.3.2. *Given the dynamic system in Equation* (3.1)*, with $w_k \sim N(0, Q_k)$ and $v_k \sim N(0, R_k)$, where Q_k and R_k are covariance matrices, sending $\theta \to 0$, makes the H_∞ filter equivalent to the Kalman filter.*

This proposition is proven later in this section.

Remark 3.3.3. *The Kalman filter does not include the matrix S_k like the H_∞ filter does. The role of this matrix is eliminated when sending $\theta \to 0$ however, the role of this matrix is different in the Kalman filter and H_∞ filter. In the H_∞ filter, the choice of S_k influences the outcome, but as $\theta \to 0$ all terms with S_k disappear, making the choice of S_k irrelevant in the limit. The Kalman filter minimises the S_k-weighted sum for any symmetric matrix S_k, as shown in Lemma 3.3.4, which explains why the choice of the matrix S_k is irrelevant in the limit.*

Note that sending $\theta \to 0$ essentially sets the bound on the maximum error to infinity. So the Kalman filter can be seen as a H_∞ filter where the covariance matrices of the noise are known and there is no guaranteed bound on the estimation error [16]. However, the Kalman filter is still the best (linear) estimator.

Lemma 3.3.4. *Let $S_k \in \mathbb{R}^n \times \mathbb{R}^n$ be a symmetric matrix. Assume that $\hat{x}^T S_k \hat{x}$ is integrable. Then the solution to the following optimisation problem:*

$$\hat{x}_{k|n} = \arg\min_{\hat{x}} \mathbb{E}[\|X_k - \hat{x}\|^2_{S_k} | \mathcal{F}_n] = \arg\min_{\hat{x}} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n]$$

is the conditional mean $\mathbb{E}[X_k|\mathcal{F}_n]$.

This lemma is proven below. We prove the results of this subsection, starting with the proof of Lemma 3.3.1.

Proof of Lemma 3.3.1. The reformulations are proven one by one, starting with Equation (3.9b).

Proof of Equation (3.9b)**:** Using the Woodbury matrix identity, see Lemma 2.1.17, we find:

$$\begin{aligned}
P_{k|k} &= P_{k|k-1}(I - \theta \bar{S}_k P_{k|k-1} + H_k^T R_k^{-1} H_k P_{k|k-1})^{-1} \\
&= P_{k|k-1}\Big((I - \theta \bar{S}_k P_{k|k-1})^{-1} \\
&\quad - (I - \theta P_{k|k-1})^{-1} H_k^T (R_k + H_k P_{k|k-1}(I - \theta P_{k|k-1})^{-1} H_k^T)^{-1} H_k P_{k|k-1}(I - \theta P_{k|k-1})^{-1}\Big) \\
&= P_{k|k-1}(I - \theta \bar{S}_k P_{k|k-1})^{-1} - P_{k|k-1}(I - \theta \bar{S}_k P_{k|k-1})^{-1} H_k^T \times \\
&\qquad (R_k + H_k P_{k|k-1}(I - \theta \bar{S}_k P_{k|k-1})^{-1} H_k^T)^{-1} H_k P_{k|k-1}(I - \theta \bar{S}_k P_{k|k-1})^{-1}.
\end{aligned}$$

Again, using the Woodbury matrix identity, see Lemma 2.1.17, gives:

$$\begin{aligned}
P_{k|k} &= ((P_{k|k-1}(I - \theta \bar{S}_k P_{k|k-1})^{-1})^{-1} + H_k^T R_k^{-1} H_k)^{-1} \\
&= ((I - \theta \bar{S}_k P_{k|k-1}) P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k)^{-1} \\
&= (P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k)^{-1}.
\end{aligned}$$

To derive Equation (3.9c), we need Equation (3.9f). So we continue to rewrite the Kalman gain.

Proof of Equation (3.9e)**:** Using Equation (3.9a), we obtain:

$$\begin{aligned}
K_k &= P_{k|k-1}[I - \theta \bar{S}_k P_{k|k-1} + H_k^T R_k^{-1} H_k P_{k|k-1}]^{-1} H_k^T R_k^{-1} \\
&= P_{k|k} H_k^T R_k^{-1}.
\end{aligned}$$

Proof of Equation (3.9f)**:** Using Equation (3.9b), we find:

$$K_k = (P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k)^{-1} H_k^T R_k^{-1}.$$

Applying the Woodbury matrix identity, see Lemma 2.1.17, gives:

$$\begin{aligned}
K_k &= \Big((P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} \\
&\quad - (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} \Big) H_k^T R_k^{-1} \\
&= \Big((P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T \\
&\quad - (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T \Big) R_k^{-1}.
\end{aligned}$$

We multiply $(P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T$ with $I = [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]$, to obtain:

$$\begin{aligned}
K_k &= \Big((P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T] \\
&\quad - (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T \Big) R_k^{-1}.
\end{aligned}$$

We take $(P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1}$ out of the brackets, to find:

$$\begin{aligned}
K_k &= \Big((P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T \\
&\quad - H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T] \Big) R_k^{-1} \\
&= \Big((P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1} R_k \Big) R_k^{-1} \\
&= (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T [R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T]^{-1}.
\end{aligned}$$

Proof of Equation (3.9c)**:** Using the Woodbury matrix identity, see Lemma 2.1.17, we obtain:

$$\begin{aligned}
&(P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k)^{-1} \\
&\quad = (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} - (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T (R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T)^{-1} H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} \\
&\quad = \Big(I - (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T (R_k + H_k (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1} H_k^T)^{-1} H_k \Big) (P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1}.
\end{aligned}$$

Using Equation (3.9f), we find:

$$(P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k)^{-1} = (I - K_k H_k)(P_{k|k-1}^{-1} - \theta \bar{S}_k)^{-1}.$$

This concludes the claim. □

Proof of Proposition 3.3.2. The reformulation of the H_∞ filter in Lemma 3.3.1 makes it clear that given the statistics of the noise are known and used in the H_∞ filter, there are two differences between the Kalman filter and the H_∞ filter. First, the term $\theta \bar{S}_k P_{k|k-1}$ in the calculation of the covariance matrix $P_{k|k}$, which is removed by sending $\theta \to 0$. Second, the condition $P_{k|k-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k \succ 0$. This condition is satisfied for every time step when sending $\theta \to 0$, because both $P_{k|k-1}$ and $H_k^T R_k^{-1} H_k$ are positive definite matrices. Thus sending $\theta \to 0$ eliminates the difference between the Kalman filter and the H_∞ filter. □

Proof of Lemma 3.3.4. We prove this theorem by setting the gradient of the mean square error to zero. Before deriving the gradient of the mean squared error, we find a bound for $|(X_k - \hat{x})^T (X_k - \hat{x})|$:

$$0 \le |(X_k - \hat{x})^T S_k (X_k - \hat{x})| = |X_k^T S_k X_k - X_k^T S_k \hat{x} - \hat{x}^T S_k X_k + \hat{x}^T S_k \hat{x}|.$$

Note that $X_k^T S_k \hat{x} \in \mathbb{R}$, so it is equal to its transpose:

$$\begin{aligned}
0 &\le |X_k^T S_k X_k - 2\hat{x}^T S_k X_k + \hat{x}^T S_k \hat{x}| \\
&= X_k^T S_k X_k + 2|\hat{x}^T S_k X_k| + \hat{x}^T S_k \hat{x}.
\end{aligned} \tag{3.10}$$

We bring $|\hat{x}^T S_k X_k|$ to the left, to obtain:

$$|\hat{x}^T S_k X_k| \leq \frac{1}{2}(X_k^T S_k X_k + \hat{x}^T S_k \hat{x}).$$

We use the above expression in Equation (3.10), to find:

$$\begin{aligned}|(X_k - \hat{x})^T S_k (X_k - \hat{x})| &\leq X_k^T S_k X_k + X_k^T S_k X_k + \hat{x}^T S_k \hat{x} + \hat{x}^T S_k \hat{x} \\ &= 2X_k^T S_k X_k + 2\hat{x}^T S_k \hat{x}.\end{aligned}$$

We define $Z = 2X_k^T S_k X_k + 2\hat{x}^T S_k \hat{x}$. Since we assumed that $\hat{x}^T S_k \hat{x}$ is integrable and X_k has a normal distribution, we conclude that Z is an integrable random variable which bounds $|(X_k - \hat{x})^T S_k (X_k - \hat{x})|$ from above. We continue to rewrite the gradient of the mean square error. By the definition of the gradient:

$$\nabla_{\hat{x}} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] = \begin{bmatrix} \frac{\partial}{\partial \hat{x}_1} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] \\ \vdots \\ \frac{\partial}{\partial \hat{x}_n} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] \end{bmatrix}.$$

To avoid long derivations in matrix notation, we show that it is possible to interchange the partial derivative with respect to x_i and the expectation. To this end, we define $e_i = [0, \ldots, 0, 1, 0, \ldots, 0]$ where the 1 is on the ith position. Then using the definition of partial derivatives:

$$\begin{aligned}&\frac{\partial}{\partial \hat{x}_i} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] \\ &\quad= \lim_{h \to 0} \frac{1}{h} \left(\mathbb{E}[(X_k - (\hat{x} + he_i))^T S_k (X_k - (\hat{x} + he_i)) | \mathcal{F}_n] - \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] \right) \\ &\quad= \lim_{h \to 0} \mathbb{E}\left[\frac{1}{h} \Big((X_k - (\hat{x} + he_i))^T S_k (X_k - (\hat{x} + he_i)) - (X_k - \hat{x})^T S_k (X_k - \hat{x}) \Big) | \mathcal{F}_n \right].\end{aligned}$$

We have seen that $|\mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n]| \leq Z$, thus using conditional dominated convergence, we are allowed to interchange the limit and the expectation:

$$\begin{aligned}&\frac{\partial}{\partial \hat{x}_i} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] \\ &\quad= \mathbb{E}\left[\lim_{h \to 0} \frac{1}{h} \Big((X_k - (\hat{x} + he_i))^T S_k (X_k - (\hat{x} + he_i)) - (X_k - \hat{x})^T S_k (X_k - \hat{x}) \Big) | \mathcal{F}_n \right] \\ &\quad= \mathbb{E}\left[\frac{\partial}{\partial \hat{x}_i} (X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n \right].\end{aligned}$$

Thus we conclude:

$$\begin{aligned}\nabla_{\hat{x}} \mathbb{E}[(X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] &= \mathbb{E}[\nabla_{\hat{x}} (X_k - \hat{x})^T S_k (X_k - \hat{x}) | \mathcal{F}_n] \\ &= 2S_k \mathbb{E}[X_k - \hat{x} | \mathcal{F}_n] \\ &= 2S_k (\mathbb{E}[X_k | \mathcal{F}_n] - \hat{x}) = 0.\end{aligned}$$

Clearly, this only holds when $\hat{x} = \mathbb{E}[X_k | \mathcal{F}_n]$. Thus, the Kalman filter minimizes the S_k weighted sum for any choice of a symmetric matrix S_k. □

3.3.2. Continuous filters

After comparing the discrete Kalman filter and the discrete H_∞ filter, this subsection compares the continuous version of these two filters. We used a slightly different dynamical system for the continuous Kalman filter in Section 2.2 than for the continuous H_∞ filter in Section 3.2. To compare the solutions of the filters, we will use the dynamical system given in Equation (3.8). So we use $D(t) = I$ in the dynamic system of Equation (2.7). The continuous Kalman filter then becomes:

$$\begin{aligned}\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} &= A(t)\hat{x}(t) + K(t)(y(t) - C(t)\hat{x}(t),\\ \frac{\mathrm{d}P(t)}{\mathrm{d}t} &= A(t)P(t) + P(t)A(t)^T + Q(t) - K(t)C(t)P(t),\\ K(t) &= P(t)C(t)^T R(t)^{-1}.\end{aligned}$$

It is now immediately clear that the difference between the continuous Kalman filter and the continuous H_∞ filter is the term $\theta P(t)L(t)^T S(t)L(t)P(t)$ in the differential equation for the covariance matrix. So we see again that sending $\theta \to 0$ in the continuous H_∞ filter gives the continuous Kalman filter provided the known statistics of the noise are used in the H_∞ filter.

Note that when the known statistics of the noise are used in the H_∞ filter, $P(t)$ is positive definite for every $t \in [0, T]$. This is because $P(t)$ is positive definite in the Kalman filter, and the corresponding $P(t)$ in the H_∞ is larger than or equal to its Kalman filter counterpart, making it positive definite as well.

3.4. Kalman filter with non zero mean noise

This section attempts to formulate a different interpretation of the H_∞ filter by viewing it as a Kalman filter with non zero mean noise. The idea for this interpretation arises because the terms containing the measurement noise in the cost function of the H_∞ filter looks like the relative entropy between two random normal variables. The relative entropy in this context can be seen as the error that arises from using a normal distribution with a shifted mean instead of a normal distribution with zero mean, with the same covariance matrix. What remains of the cost function is similar to the cost function used in the Kalman filter. This could mean that the H_∞ filter is a Kalman filter where the noise is allowed to have a shifted mean, for which the relative entropy compensates. However, this interpretation is shown to be incorrect. Thereto, the definition of relative entropy is given and the relative entropy of two Gaussian random variables is derived. Then the Kalman filter is derived for non zero mean noise. We start with the definition of relative entropy, also called the Kullback-Leibler divergence or distance.

Definition 3.4.1 (Relative entropy)**.** *[3] Let p and q be two probability distributions defined on the same support, with p absolutely continuous with respect to q, meaning $p(x) = 0$ whenever $q(x) = 0$. The relative entropy between distribution $p(x)$ and $q(x)$ is then defined as:*

$$D_{KL}(p\|q) = \mathbb{E}_p\left[\log \frac{p(x)}{q(x)}\right].$$

Intuitively, relative entropy can be seen as the error of using distribution q instead of distribution p, while the actual distribution is p. This will then be denoted by $D_{KL}(p\|q)$. Note that the relative entropy is always non-negative and only 0 if $p = q$. However, the relative entropy is not a metric, because it is not symmetric and it does not satisfy the triangle inequality.

To continue, we need the relative entropy between two Gaussian random variables. This will be derived in the following lemma.

Lemma 3.4.2 (Relative entropy for two Gaussian random variables)**.** *Suppose we have two normal distributions $X_1 \sim N(\mu_1, \Sigma_1)$ and $X_2 \sim N(\mu_2, \Sigma_2)$. With $X_1, X_2, \mu_1, \mu_2 \in \mathbb{R}^n$ and $\Sigma_1, \Sigma_2 \in \mathbb{R}^n \times \mathbb{R}^n$. The relative entropy between these two Gaussian distributions is:*

$$D_{KL}(X_1\|X_2) = \frac{1}{2}\left(tr(\Sigma_2^{-1}\Sigma_1) - n + (\mu_2 - \mu_1)^T\Sigma_2^{-1}(\mu_2 - \mu_1) + \log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right)\right).$$

If $\Sigma_1 = \Sigma_2$ then,

$$D_{KL}(X_1\|X_2) = \frac{1}{2}\|\mu_2 - \mu_1\|^2_{\Sigma_1^{-1}}.$$

This lemma is proven below.

Remember the cost function of the H_∞ filter, see Equation (3.4):

$$J = -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1}\left[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\left(\|w_k\|^2_{Q_k^{-1}} + \|v_k\|^2_{R_k^{-1}}\right)\right].$$

The last term $-\frac{1}{\theta}\left(\|w_k\|^2_{Q_k^{-1}} + \|v_k\|^2_{R_k^{-1}}\right)$ can remind us of the relative entropy between two Gaussian random variables with the same covariance matrix and different mean. The term $\frac{1}{\theta}\|w_k\|^2_{Q_k^{-1}}$ would then be the relative entropy between $Z_1 \sim N(0, Q_k)$ and $Z_2 \sim N(w_k, Q_k)$. $\frac{1}{\theta}\|v_k\|^2_{R_k^{-1}}$ would then be the relative entropy between $Z_3 \sim N(0, R_k)$ and $Z_4 \sim N(v_k, R_k)$. Note that w_k and v_k are not random variables in this context, but constants. The intuitive interpretation of this reformulation would be that the H_∞ filter is essentially a Kalman filter where the mean of the noise is shifted. The term $-\frac{1}{\theta}\left(\|w_k\|^2_{Q_k^{-1}} + \|v_k\|^2_{R_k^{-1}}\right)$ would then compensate for mistakenly using that the noise is zero mean as the Kalman filter assumes. θ would then determine the degree of compensation for the shifted mean of the noise. We will derive the Kalman filter where the noise has a shifted mean to check this interpretation.

Theorem 3.4.3. *Given the following dynamic system:*

$$\begin{aligned} X_k &= F_k X_{k-1} + W_{k-1}, \\ Y_{k-1} &= H_{k-1} X_{k-1} + V_{k-1}, \end{aligned}$$

where $X_k \in \mathbb{R}^n$ is the state vector; $F_k \in \mathbb{R}^n \times \mathbb{R}^n$ is the state transition matrix; $W_k \in \mathbb{R}^n$ is shifted Gaussian white noise with covariance matrix $Q_k \in \mathbb{R}^n \times \mathbb{R}^n$; $Y_k \in \mathbb{R}^m$ represents the measurement at time k; $H_k \in \mathbb{R}^m \times \mathbb{R}^n$ provides a linear connection between the state vector and the measurement vector; $V_k \in \mathbb{R}^m$ is shifted Gaussian white noise with covariance matrix $R_k \in \mathbb{R}^m \times \mathbb{R}^m$. W_k and V_k are assumed to be independent. Furthermore, $\mathbb{E}[W_k] = w_k$; $\mathbb{E}[V_k] = v_k$; $Q_k = \text{Var}(W_k) = \mathbb{E}[W_k W_k^T] - \mathbb{E}[W_k]\mathbb{E}[W_k]^T = \tilde{Q}_k - w_k w_k^T$ and $R_k = \text{Var}(V_k) = \mathbb{E}[V_k V_k^T] - \mathbb{E}[V_k]\mathbb{E}[V_k]^T = \tilde{R}_k - v_k v_k^T$, with $w_k \in \mathbb{R}^n$; $v_k \in \mathbb{R}^m$ and $\tilde{Q}_k \in \mathbb{R}^n \times \mathbb{R}^n$; $\tilde{R}_k \in \mathbb{R}^m \times \mathbb{R}^m$.

The Kalman filter is given by:

Prediction step:

$$\begin{aligned} \hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1} + v_k, \\ P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_k. \end{aligned}$$

Update step:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - (H_k\hat{x}_{k|k-1} + w_k)), \\ P_{k|k} &= [I - K_k H_k]P_{k|k-1}, \\ K_k &= P_{k|k-1}H_k^T[H_k P_{k|k-1}H_k^T + R_k]^{-1}. \end{aligned}$$

Remark 3.4.4. *Note that for this interpretation to work, the noise should have a normal distribution and the covariance matrices should be known. So this interpretation could only work for a specific case of the H_∞ filter. It is, however, immediately clear that this solution has a different structure than the solution of the H_∞ filter for any choice of θ. In the H_∞ filter, $P_{k|k-1}$ is replaced by $(P_{k|k-1}^{-1} - \theta S_k)^{-1}$. Shifting the mean of the noise in the Kalman filter does not result in the same change. In the adjusted Kalman filter, $P_{k|k-1}$ remains unchanged. The terms that are changed by shifting the mean of the noise in the Kalman filter are $\hat{x}_{k|k-1}$ and $\hat{x}_{k|k}$. We must therefore conclude that viewing $\frac{1}{\theta}\left(\|w_k\|^2_{Q_k^{-1}} + \|v_k\|^2_{R_k^{-1}}\right)$ as a relative entropy to compensate for using the Kalman filter without a shifted mean of the measurement noise it not correct.*

We prove the results in this section below, starting with the proof of Lemma 3.4.2.

Proof of Lemma 3.4.2. We write out $D_{KL}(X_1\|X_2)$ when $X_1 \sim N(\mu_1,\Sigma_1)$ and $X_2 \sim N(\mu_2,\Sigma_2)$:

$$\begin{aligned} D_{KL}(X_1\|X_2) &= \mathbb{E}_{X_1}\left[\log\frac{(2\pi)^{-n/2}\,\text{det}(\Sigma_1)^{-1/2}\,\text{exp}(-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1))}{(2\pi)^{-n/2}\,\text{det}(\Sigma_2)^{-1/2}\,\text{exp}(-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2))}\right] \\ &= \mathbb{E}_{X_1}\left[\frac{1}{2}\Big(\log\left(\frac{\text{det}(\Sigma_2)}{\text{det}(\Sigma_1)}\right) - (x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\Big)\right] \\ &= \frac{1}{2}\Big(\log\left(\frac{\text{det}(\Sigma_2)}{\text{det}(\Sigma_1)}\right) - \mathbb{E}_{X_1}[(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)] + \mathbb{E}_{X_1}[(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)]\Big). \end{aligned} \tag{3.11}$$

Let $X \sim N(\mu,\Sigma)$ with $X,\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^n \times \mathbb{R}^n$. For a matrix $A \in \mathbb{R}^n \times \mathbb{R}^n$ and a vector $b \in \mathbb{R}^n$ we have

$$\mathbb{E}[(X-b)^TA(X-b) = \mathbb{E}[X^TAX - X^TAb - b^TAX + b^TAb].$$

Since $X^TAb \in \mathbb{R}$ it is equal to its transpose. Thus:

$$\begin{aligned} \mathbb{E}[(X-b)^TA(X-b) &= \mathbb{E}[X^TAX - 2X^TAb + b^TAb] \\ &= \mathbb{E}[X^TAX] - \mathbb{E}[2X^T]Ab + b^TAb \\ &= \mu^TA\mu + \text{tr}(A\Sigma) - 2\mu^TAb + b^TAb \\ &= (\mu-b)^TA(\mu-b) + \text{tr}(A\Sigma). \end{aligned}$$

Using the above in Equation (3.11), we find:

$$\begin{aligned} D_{KL}(X_1\|X_2) = \frac{1}{2}\Big(\log&\left(\frac{\text{det}(\Sigma_2)}{\text{det}(\Sigma_1)}\right) - (\mu_1-\mu_1)^T\Sigma_1^{-1}(\mu_1-\mu_1) + \text{tr}(\Sigma_1^{-1}\Sigma_1) \\ &+ (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) + \text{tr}(\Sigma_2^{-1}\Sigma_1) \\ = \frac{1}{2}\Big(\log&\left(\frac{\text{det}(\Sigma_1)}{\text{det}(\Sigma_2)}\right) - \text{tr}(I) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) + \text{tr}(\Sigma_2^{-1}\Sigma_1)\Big) \\ = \frac{1}{2}\Big(&\text{tr}(\Sigma_2^{-1}\Sigma_1) - n + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) + \log\left(\frac{\det\Sigma_2}{\det\Sigma_1}\right)\Big). \end{aligned}$$

Setting $\Sigma_1 = \Sigma_2$, we find:

$$\begin{aligned} D_{KL}(X_1\|X_2) &= \frac{1}{2}\left(\text{tr}(I) - n + (\mu_2-\mu_1)^T\Sigma_1^{-1}(\mu_2-\mu_1) + \ln\left(\frac{\det\Sigma_1}{\det\Sigma_1}\right)\right) \\ &= \frac{1}{2}\left(n - n + (\mu_2-\mu_1)^T\Sigma_1^{-1}(\mu_2-\mu_1) + \ln(1)\right) \\ &= \frac{1}{2}(\mu_2-\mu_1)^T\Sigma_1^{-1}(\mu_2-\mu_1) \\ &= \frac{1}{2}\|\mu_2-\mu_1\|^2_{\Sigma_1^{-1}}. \end{aligned}$$

This concludes the proof. □

Proof of Theorem 3.4.3. This proof is a modified version of the method used in [2]. We start with deriving the expression for the conditional mean:

$$\begin{aligned} \hat{x}_{k|k-1} = \mathbb{E}[X_k|\mathcal{F}_{k-1}] &= \mathbb{E}[F_{k-1}X_{k-1} + W_{k-1}|\mathcal{F}_{k-1}] \\ &= F_{k-1}\mathbb{E}[X_{k-1}|\mathcal{F}_{k-1}] + \mathbb{E}[W_{k-1}|\mathcal{F}_{k-1}] \\ &= F_{k-1}\hat{x}_{k-1|k-1} + w_{k-1}. \end{aligned}$$

We derive the conditional error covariance $P_{k|k-1}$:

$$\begin{aligned}
P_{k|k-1} &= \mathbb{E}[(X_k - \hat{x}_{k|k-1})(X_k - \hat{x}_{k|k-1})^T|\mathcal{F}_{k-1}] = \mathbb{E}[(F_{k-1}(X_{k-1} - \hat{x}_{k-1|k-1}) + W_{k-1} - w_{k-1})\times \\
&\qquad (F_{k-1}(X_{k-1} - \hat{x}_{k-1|k-1}) + W_{k-1} - w_{k-1})^T|\mathcal{F}_{k-1}] \\
&= F_{k-1}\mathbb{E}[(X_{k-1} - \hat{x}_{k-1|k-1})(X_{k-1} - \hat{x}_{k-1|k-1})^T|\mathcal{F}_{k-1}]F_{k-1}^T \\
&\quad + F_{k-1}\mathbb{E}[X_{k-1} - \hat{x}_{k-1|k-1}|\mathcal{F}_{k-1}]\mathbb{E}[(W_{k-1} - w_{k-1})^T|\mathcal{F}_{k-1}] \\
&\quad + \mathbb{E}[W_{k-1} - w_{k-1}|\mathcal{F}_{k-1}]\mathbb{E}[(X_{k-1} - \hat{x}_{k-1|k-1})^T|\mathcal{F}_{k-1}]F_{k-1}^T \\
&\quad + \mathbb{E}[(W_{k-1} - w_{k-1})(W_{k-1} - w_{k-1})^T|\mathcal{F}_{k-1}].
\end{aligned}$$

Using that $\mathbb{E}[X_{k-1} - \hat{x}_{k-1|k-1}|\mathcal{F}_{k-1}] = 0$ gives:

$$\begin{aligned}
P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + \mathbb{E}[W_{k-1}W_{k-1}^T] - \mathbb{E}[W_{k-1}]w_{k-1}^T - w_{k-1}\mathbb{E}[W_{k-1}^T] + w_{k-1}w_{k-1}^T \\
&= F_{k-1}P_{k-1|k-1}F_{k-1}^T + \tilde{Q}_{k-1} - w_{k-1}w_{k-1}^T \\
&= F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}.
\end{aligned}$$

We derive the estimator for y_k conditioned on $\mathcal{F}_{k-1}$:

$$\begin{aligned}
\hat{y}_{k|k-1} = \mathbb{E}[Y_k|\mathcal{F}_{k-1}] &= \mathbb{E}[H_k X_k + V_k|\mathcal{F}_{k-1}] \\
&= H_k\hat{x}_{k|k-1} + v_k.
\end{aligned}$$

We derive the covariance of the error of the measurement estimation:

$$\begin{aligned}
B_k &= \mathbb{E}[(Y_k - \hat{y}_{k|k-1})(Y_k - \hat{y}_{k|k-1})^T|\mathcal{F}_{k-1}] \\
&= \mathbb{E}[(H_k(X_k - \hat{x}_{k|k-1}) + V_k - v_k)(H_k(X_k - \hat{x}_{k|k-1}) + V_k - v_k)^T|\mathcal{F}_{k-1}] \\
&= H_{k-1}\mathbb{E}[(X_k - \hat{x}_{k|k-1})(X_k - \hat{x}_{k|k-1})^T|\mathcal{F}_{k-1}]H_k^T \\
&\quad + H_k\mathbb{E}[(X_k - \hat{x}_{k|k-1})|\mathcal{F}_{k-1}]\mathbb{E}[(V_k - v_k)^T|\mathcal{F}_{k-1}] \\
&\quad + \mathbb{E}[V_k - v_k|\mathcal{F}_{k-1}]\mathbb{E}[(X_k - \hat{x}_{k|k-1})^T|\mathcal{F}_{k-1}]H_{k-1}^T + \mathbb{E}[(V_k - v_k)(V_k - v_k)^T|\mathcal{F}_{k-1}] \\
&= H_kP_{k|k-1}H_k^T + \mathbb{E}[V_kV_k^T] - \mathbb{E}[V_k]v_k^T - v_k\mathbb{E}[V_k^T] + v_kv_k^T \\
&= H_kP_{k|k-1}H_k^T + \tilde{R}_k - v_kv_k^T \\
&= H_kP_{k|k-1}H_k^T + R_k.
\end{aligned}$$

Finally, we derive the cross covariance between the error of the state estimation and the error of the measurement estimation:

$$\begin{aligned}
\mathbb{E}[(X_k - \hat{x}_{k|k-1})(Y_k - \hat{y}_{k|k-1})^T|\mathcal{F}_{k-1}] &= \mathbb{E}[(X_k - \hat{x}_{k|k-1})(H_k(X_k - \hat{x}_{k|k-1}) + V_k - v_k)^T|\mathcal{F}_{k-1}] \\
&= \mathbb{E}[(X_k - \hat{x}_{k|k-1})(X_k - \hat{x}_{k|k-1})^T|\mathcal{F}_{k-1}]H_k^T + \mathbb{E}[X_k - \hat{x}_{k|k-1}|\mathcal{F}_{k-1}]\mathbb{E}[(V_k - v_k)^T] \\
&= P_{k|k-1}H_k^T,
\end{aligned}$$

which leads to the following Kalman gain:

$$\begin{aligned}
K_k &= \mathbb{E}[(X_k - \hat{x}_{k|k-1})(Y_k - \hat{y}_{k|k-1})^T|\mathcal{F}_{k-1}]B_k^{-1} \\
&= P_{k|k-1}H_k^TB_k^{-1} \\
&= P_{k|k-1}H_k^T[H_kP_{k|k-1}H_k^T + R_k]^{-1}.
\end{aligned}$$

This concludes the proof. □

4

Covariance Intersection

Up until now, we have only included one measurement source in the dynamic systems. In this chapter, we add another measurement source. The Kalman filter can also be applied to multiple measurement sources provided that all correlations between these sources are known. However, in practice these correlations are often not known. There exist multiple methods that calculate a state estimate when the correlations between measurement sources are not known. This chapter describes one possible method, namely covariance intersection.

Section 4.1 provides an introduction to covariance intersection. Covariance intersection is essentially a linear combination between two partial estimates. Covariance intersection is consistent, which means that the method does not underestimate the covariance of the estimate. The formal definition of consistency is given in Section 4.2. The consistency of covariance intersection is proven by S. J. Julier and J. K. Uhlmann. This proof contains two issues. First, it does not use conditional expectations even though the estimates made by the Kalman filter are conditional expectations. Second, Julier and Uhlmann's proof assumes that the partial estimates are all consistent. This assumption is neither trivial nor proven. Section 4.2 rewrites the proof to include conditional expectations where appropriate and discusses the assumption made in this proof. A reformulation of covariance intersection is presented in Section 4.3. This reformulation makes it possible to translate the fusion step of covariance intersection into the continuous Kalman filter. In Chapter 5 it will be shown that this reformulation can easily be added to the structure of the H_∞ filter.

4.1. Introduction to covariance intersection

We consider the following dynamic system with two measurement sources A and B:

$$\begin{aligned} X_{k+1} &= F_k X_k + W_k, \\ Y_k &= H_k X_k + V_k, \end{aligned} \tag{4.1}$$

with

$$\begin{aligned} Y_k &= \begin{bmatrix} Y_k^A \\ Y_k^B \end{bmatrix}, & V_k &= \begin{bmatrix} V_k^A \\ V_k^B \end{bmatrix}, \\ H_k &= \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}, & R_k &= \begin{bmatrix} R_k^A & \Sigma_k^{AB} \\ \Sigma_k^{BA} & R_k^B \end{bmatrix}, \end{aligned} \tag{4.2}$$

where $X_k \in \mathbb{R}^n$ is the random state vector; $F_k \in \mathbb{R}^n \times \mathbb{R}^n$ is the state transition matrix, which can describe the physics underlying the process that is modelled; $W_k \in \mathbb{R}^n$ is white noise with the following distribution: $W_k \sim N(0, Q_k)$, with $Q_k \in \mathbb{R}^n \times \mathbb{R}^n$; W_k can be used to compensate for factors not included in the linear combination of the state variable. $Y_k^A \in \mathbb{R}^m$ represents the measurements from source A at time k; $H_k^A \in \mathbb{R}^m \times \mathbb{R}^n$ provides a linear connection between the state vector and the measurement vector for measurement source A; $V_k^A \in \mathbb{R}^m$ is Gaussian white measurement noise, so: $V_k^A \sim N(0, R_k^A)$, with $R_k^A \in \mathbb{R}^m \times \mathbb{R}^m$. Measurement source B is allowed to have different dimensions, so $Y_k^B \in \mathbb{R}^q$, $H_k^B \in \mathbb{R}^q \times \mathbb{R}^n$, $V_k^B \in \mathbb{R}^q$, V_k^B is also Gaussian white noise, so $V_k^B \sim N(0, R_k^B)$, with $R_k^B \in \mathbb{R}^q \times \mathbb{R}^q$. $\Sigma_k^{AB} = (\Sigma_k^{BA})^T \in \mathbb{R}^m \times \mathbb{R}^q$ represents the cross-covariance between V_k^A and V_k^B. W_k is assumed to be independent from both V_k^A and V_k^B. Note that it is possible to add as many sources of measurements as desired, but for the sake of simplicity we use two measurement sources.

If all matrices are known, the Kalman filter can be applied as seen in Chapter 2. Unfortunately, Σ_k^{AB} is often not known in practice. One possible solution is the H_∞ filter, which can be found in Chapter 3. Another solution to this problem is provided by covariance intersection, which was first suggested by J. K. Uhlmann in his PhD thesis [15]. Covariance intersection consists of two steps which calculate an alternative update step for the Kalman filter, without knowing the full covariance matrix between the measurement sources:

1. Partial update steps are calculated by applying the Kalman filter with a partial covariance matrix for each measurement source separately.
2. These estimates are fused together using a linear combination. This step requires the optimisation of the constant that controls the linear combination.

The details of covariance intersection are given in Definition 4.1.1. In the rest of this chapter, we will use the following sigma algebras:

$$\mathcal{F}_n = \sigma(Y_1^A, Y_1^B, \ldots, Y_n^A, Y_n^B), \tag{4.3a}$$
$$\mathcal{F}_n^A = \sigma(\mathcal{F}_{n-1}, Y_n^A), \tag{4.3b}$$
$$\mathcal{F}_n^B = \sigma(\mathcal{F}_{n-1}, Y_n^B). \tag{4.3c}$$

In words, $\mathcal{F}_n$ contains information from the measurement sources A and B up until and including time n. $\mathcal{F}_n^A$ contains the information of both measurement sources up until and including time $n-1$ and additionally contains information about the measurement from source A at time n. $\mathcal{F}_n^B$ has the same interpretation as $\mathcal{F}_n^A$ except that information from measurement source B is available at time n.

Definition 4.1.1 (Covariance intersection)**.** *Given the dynamic system of Equation* (4.1) *with unknown cross correlations $\Sigma_k^{AB} = (\Sigma_k^{BA})^T$. Let $(\hat{x}_{k|k-1}^{CI}, P_{k|k-1}^{CI})$ be an a priori estimate of the mean and variance of the random variable X_k. We define the following partial a posteriori estimate for measurement source A, using the sigma algebra $\mathcal{F}_n^A$ as described in Equation* (4.3b)*:*

$$\begin{aligned}
\hat{x}_{k|k}^A &= \mathbb{E}[X_k|\mathcal{F}_k^A] = \hat{x}_{k|k-1}^{CI} + K_k^A(y_k^A - H_k^A \hat{x}_{k|k-1}^{CI}),\\
P_{k|k}^A &= \mathbb{E}[(X_k - \hat{x}_{k|k}^A)(X_k - \hat{x}_{k|k}^A)^T|\mathcal{F}_k^A]\\
&= ((P_{k|k-1}^{CI})^{-1} + (H_k^A)^T(R_k^A)^{-1}H_k^A)^{-1},\\
K_k^A &= P_{k|k}^A(H_k^A)^T(R_k^A)^{-1}.
\end{aligned} \tag{4.4}$$

Note that this is the standard Kalman update step except that all information on measurement source B is left out. An analogous partial a posteriori estimate is calculated for measurement source B, using $\mathcal{F}_n^B$ as described in Equation (4.3c)*. Then these two partial a posteriori estimates are fused in the following way:*

$$(P_{k|k}^{CI})^{-1} = \omega(P_{k|k}^A)^{-1} + (1-\omega)(P_{k|k}^B)^{-1}, \tag{4.5a}$$
$$\hat{x}_{k|k}^{CI} = P_{k|k}^{CI}(\omega(P_{k|k}^A)^{-1}\hat{x}_{k|k}^A + (1-\omega)(P_{k|k}^B)^{-1}\hat{x}_{k|k}^B), \tag{4.5b}$$

with

$$\omega = \arg\min_\omega \phi(\chi(\omega, P_{k|k-1}^{CI}, \hat{x}_{k|k-1}, y_k, H_k, \tilde{R}_k)) \in [0,1], \tag{4.6}$$

where $\tilde{R}_k = [R_k^A, R_k^B]^T$, $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is some measure and χ maps its input to a matrix of size $\mathbb{R}^n \times \mathbb{R}^n$.

Figure 4.1 depicts a schematic representation of the covariance intersection algorithm.

Remark 4.1.2. *The literature chooses χ such that $\chi(\omega, P^{CI}_{k|k-1}, \hat{x}_{k|k-1}, y_k, H_k, \tilde{R}_k) = P^{CI}_{k|k}$. Remark 4.3.6 discusses the possibility of different choices for χ. There is no consensus in the literature on which measure for ϕ is best. [9] claims that the Kalman filter with covariance intersection applied to it converges as long as ω is optimised for every time step, ensuring that the updated estimate is smaller than or equal to the previous estimate. Define $C = \chi(\omega, P^{CI}_{k|k-1}, \hat{x}_{k|k-1}, y_k, H_k, \tilde{R}_k)$. Common choices for ϕ include:*

- $\phi(C) = \det(C)$ *[15],*
- $\phi(C) = tr(C)$ *[7],*
- $\phi(C) = \|C\|$ *[9].*

It is not clear which choice for ϕ should be preferred, except that the trace and matrix norms are convex functions with respect to ω guaranteeing a unique minimum [7]. The determinant is neither concave nor convex, but is still used.

4.2. Consistency of covariance intersection

Covariance intersection is claimed to be consistent, which means that the covariance matrix of the estimate is 'large enough', thereby mitigating the risk of overconfidence in the algorithm. This section gives the formal definition of consistency. The consistency of covariance intersection is proven by S. J. Julier and J. K. Uhlmann. This section contains a rewritten version of their proof so that it includes conditional expectations where appropriate. This section also highlights and discusses the assumptions made in this proof. The contents of this section are loosely based on [5] and [7].

Definition 4.2.1. *Given $x \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^n \times \mathbb{R}^n$ a positive definite matrix, and $\mathcal{F}$ a sigma algebra. The estimate $(\hat{x}, \Sigma)$ is called consistent for the random variable $X|\mathcal{F}$ if:*

$$\Sigma \succeq \mathbb{E}[(X - \hat{x})(X - \hat{x})^T | \mathcal{F}], \tag{4.7}$$

where $\succeq$ means that the left-hand side minus the right-hand side is a positive (semi-)definite matrix.

Remark 4.2.2. *According to bias-variance decomposition:*

$$\mathbb{E}[(X - \hat{x})(X - \hat{x})^T | \mathcal{F}] = \Sigma_X + (\hat{x} - \mathbb{E}[X|\mathcal{F}])(\hat{x} - \mathbb{E}[X|\mathcal{F}])^T,$$

where $\Sigma_X = \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])(X - \mathbb{E}[X|\mathcal{F}])^T | \mathcal{F}]$. Using this in Equation (4.7) *gives:*

$$\Sigma \succeq \Sigma_X + (\hat{x} - \mathbb{E}[X|\mathcal{F}])(\hat{x} - \mathbb{E}[X|\mathcal{F}])^T. \tag{4.8}$$

Remark 4.2.3. *Note that an estimate $(\hat{x}_{k|k}, P_{k|k-1})$ made by the Kalman filter, when all assumptions are satisfied, is always consistent for $X_k|\mathcal{F}_k$. This follows directly from $\hat{x}_{k|k} = \mathbb{E}[X_k|\mathcal{F}_k]$ and $P_{k|k} = \mathbb{E}[(X_k - \mathbb{E}[X_k|\mathcal{F}_k])(X_k - \mathbb{E}[X_k|\mathcal{F}_k])^T | \mathcal{F}_k]$, as seen in Chapter 2.*

The following theorem proves the consistency of covariance intersection, provided that the partial a priori estimates are consistent.

Theorem 4.2.4 (Consistency of covariance intersection)**.** *Given the partial a posteriori estimates $(\hat{x}^A_{k|k}, P^A_{k|k})$ and $(\hat{x}^B_{k|k}, P^B_{k|k})$, as in Equation* (4.4)*. Given an estimate by covariance intersection $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$, as described in Definition 4.1.1. Assume that $(\hat{x}^A_{k|k}, P^A_{k|k})$ and $(\hat{x}^B_{k|k}, P^B_{k|k})$ are consistent for the random variable $X_k|\mathcal{F}_k$. If this assumption is satisfied, then $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$ is consistent for $X_k|\mathcal{F}_k$ for every choice of $\omega \in [0, 1]$.*

This theorem is proved later in this section.

Remark 4.2.5 (On the consistency of the partial a posteriori estimates)**.** *Theorem 4.2.4 uses that the partial a posteriori estimates are consistent. Using the definition of $P^A_{k|k}$, see Equation* (4.5a)*, and the definition of consistency, see Equation* (4.7)*, we see that if $(\hat{x}^A_{k|k}, P^A_{k|k})$ is consistent for $X_k|\mathcal{F}_k$ following needs to hold:*

$$P^A_{k|k} = \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T | \mathcal{F}^A_k] \succeq \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T | \mathcal{F}_k]. \tag{4.9}$$

We define the following random variable:

$$D_k = \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}^A_k] - \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]. \tag{4.10}$$

If Equation (4.9) *is true, then D_k must be a non-negative random variable, which is not necessarily true. So Equation* (4.9) *can not hold in general. However, we can show that $\mathbb{E}[D_k|\mathcal{F}^A_k] = 0$. Using the tower rule we can rewrite $P^A_{k|k}$ as follows:*

$$P^A_{k|k} = \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}^A_k] = \mathbb{E}[\mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]|\mathcal{F}^A_k].$$

Using this expressing in Equation (4.10)*, we find:*

$$D_k = \mathbb{E}[\mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]|\mathcal{F}^A_k] - \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k].$$

If we take the expectation conditioned on $\mathcal{F}^A_k$ of D_k, then:

$$\begin{aligned}\mathbb{E}[D_k|\mathcal{F}^A_k] &= \mathbb{E}[\mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]|\mathcal{F}^A_k] - \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]|\mathcal{F}^A_k] \\ &= \mathbb{E}[\mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]|\mathcal{F}^A_k] - \mathbb{E}[\mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k]|\mathcal{F}^A_k] = 0.\end{aligned}$$

Based on this realisation, we formulate conjecture.

Conjecture 4.2.6. *Given the partial a posteriori estimates $(\hat{x}^A_{k|k}, P^A_{k|k})$ and $(\hat{x}^B_{k|k}, P^B_{k|k})$, as in Equation* (4.4)*. Given an estimate by covariance intersection $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$, as described in Definition 4.1.1. $(\hat{x}^A_{k|k}, P^A_{k|k})$ and $(\hat{x}^B_{k|k}, P^B_{k|k})$ are consistent in expectation for the random variable $X|\mathcal{F}_k$. This means that, on average, the terms preventing consistency for every time step approach zero. This implies that $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$ is consistent in expectation for the random variable $X_k|\mathcal{F}_k$.*

Remark 4.2.7. *If this conjecture turns out to be true, covariance intersection would not be consistent for every time step, but would be consistent over many time steps, creating consistent behaviour in practice.*

We need an auxiliary lemma to simplify the proof of Theorem 4.2.4.

Lemma 4.2.8. *Let $A, B \in \mathbb{R}^n \times \mathbb{R}^n$ be two invertible matrices. Let $\omega \in (0,1)$. We define:*

$$C = \left(\omega A^{-1} + (1-\omega)B^{-1}\right)^{-1}.$$

And we define:

$$\begin{aligned} L &= \omega C A^{-1}, \\ M &= (1-\omega)CB^{-1}. \end{aligned}$$

Then $L + M = I$.

Below we prove the results of this section.

Proof of Theorem 4.2.4. We first show consistency for $\omega = 0$ and $\omega = 1$. Let us fix $\omega = 0$. Using the definitions of $P^{CI}_{k|k}$ and $\hat{x}^{CI}_{k|k}$, given in Equations (4.5a) and Equation (4.5b) we find:

$$\begin{aligned} P^{CI}_{k|k} &= P^B_{k|k}, \\ \hat{x}^{CI}_{k|k} &= P^{CI}_{k|k}(P^B_{k|k})^{-1}\hat{x}^B_{k|k} \\ &= \hat{x}^B_{k|k}. \end{aligned}$$

By assumption, this estimate is consistent for the random variable $X_k|\mathcal{F}_k$. A similar result can be obtained for $\omega = 1$. We will now prove consistency for $\omega \in (0,1)$. Definition 4.2.1 gives us what we need to show to prove the consistency of covariance intersection:

$$P^{CI}_{k|k} \succeq \mathbb{E}[(X_k - \hat{x}^{CI}_{k|k})(X_k - \hat{x}^{CI}_{k|k})^T|\mathcal{F}_k].$$

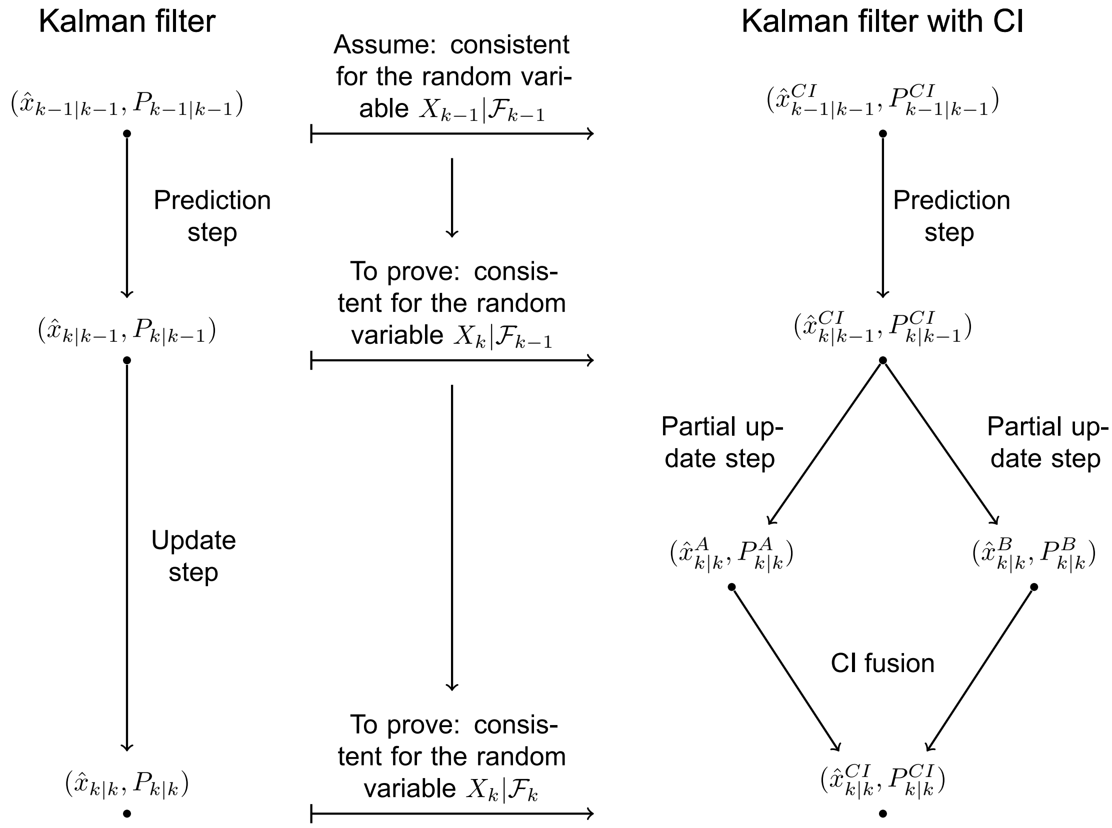

Figure 4.1: This figure shows a schematic representation of the Kalman filter on the left and covariance intersection applied to the Kalman filter on the right. The Kalman filter starts with an a posteriori estimate for time step $k-1$, $(\hat{x}_{k-1|k-1}, P_{k-1|k-1})$. Using a prediction step, an a priori estimate is calculated: $(\hat{x}_{k|k-1}, P_{k|k-1})$. Then with the update step, an a posteriori estimate is calculated for time step k: $(\hat{x}_{k|k}, P_{k|k})$. On the right it can be seen that, when applying covariance intersection to the Kalman filter, we start with an a posteriori estimate for the mean and variance for time step $k-1$: $(\hat{x}^{CI}_{k-1|k-1}, P^{CI}_{k-1|k-1})$. Using the prediction step of the Kalman filter, an a priori estimate is calculated: $(\hat{x}^{CI}_{k|k-1}, P^{CI}_{k|k-1})$. Then two partial a posteriori updates are calculated for time step k, both conditioned on one measurement source. This gives $(\hat{x}^{A}_{k|k}, P^{A}_{k|k})$ and $(\hat{x}^{B}_{k-1|k-1}, P^{B}_{k-1|k-1})$. Finally, these two a posteriori estimates are fused to one a posteriori estimate for time step k with the fusion rule given in Definition 4.1.1: $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$. The horizontal arrows indicate what is needed for covariance intersection to be consistent. This means that if $X_{k-1}|\mathcal{F}_{k-1}$ is consistent for the random variable $X_{k-1}|\mathcal{F}_{k-1}$, then both the prediction and the update step should be consistent for the relevant random variables. This work only proofs the consistency of $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$, provided that the partial estimates made by covariance intersection are consistent for the random variable $X_k|\mathcal{F}_k$. This is proven in Theorem 4.2.4.

Subtracting $\mathbb{E}[(X_k - \hat{x}^{CI}_{k|k})(X_k - \hat{x}^{CI}_{k|k})^T|\mathcal{F}_k]$ from both sides, we find:

$$P^{CI}_{k|k} - \mathbb{E}[(X_k - \hat{x}^{CI}_{k|k})(X_k - \hat{x}^{CI}_{k|k})^T|\mathcal{F}_k] \succeq 0. \tag{4.11}$$

For brevity we will use the following notation in this proof:

$$\begin{aligned}
\bar{P}^{CI}_{k|k} &= \mathbb{E}[(X_k - \hat{x}^{CI}_{k|k})(X_k - \hat{x}^{CI}_{k|k})^T|\mathcal{F}_k], \\
\bar{P}^{A}_{k|k} &= \mathbb{E}[(X_k - \hat{x}^{A}_{k|k})(X_k - \hat{x}^{A}_{k|k})^T|\mathcal{F}_k], \\
\bar{P}^{B}_{k|k} &= \mathbb{E}[(X_k - \hat{x}^{B}_{k|k})(X_k - \hat{x}^{B}_{k|k})^T|\mathcal{F}_k], \\
\bar{P}^{AB}_{k|k} &= \mathbb{E}[(X_k - \hat{x}^{A}_{k|k})(X_k - \hat{x}^{B}_{k|k})^T|\mathcal{F}_k], \\
\bar{P}^{BA}_{k|k} &= \mathbb{E}[(X_k - \hat{x}^{B}_{k|k})(X_k - \hat{x}^{A}_{k|k})^T|\mathcal{F}_k].
\end{aligned} \tag{4.12}$$

Using this notation we can rewrite Equation (4.11) as follows:

$$P^{CI}_{k|k} - \bar{P}^{CI}_{k|k} \succeq 0. \tag{4.13}$$

We will use Lemma 4.2.8 to rewrite $X_k - \hat{x}^{CI}_{k|k}$. To this end, define:

$$\begin{aligned} L_k^{CI} &= \omega P^{CI}_{k|k}(P^A_{k|k})^{-1}, \\ M_k^{CI} &= (1-\omega) P^{CI}_{k|k}(P^B_{k|k})^{-1}. \end{aligned} \tag{4.14}$$

By Lemma 4.2.8, we know $L_k^{CI} + M_k^{CI} = I$. Using the definition for $\hat{x}^{CI}_{k|k}$, see Equation (4.5b), we can rewrite $X_k - \hat{x}^{CI}_{k|k}$ as follows:

$$\begin{aligned} X_k - \hat{x}^{CI}_{k|k} &= (L_k^{CI} + M_k^{CI})X_k - P^{CI}_{k|k}(\omega(P^A_{k|k})^{-1}\hat{x}^A_{k|k} + (1-\omega)(P^B_{k|k})^{-1}\hat{x}^B_{k|k}) \\ &= (L_k^{CI} + M_k^{CI})X_k - (L_k^{CI}\hat{x}^A_{k|k} + M_k^{CI}\hat{x}^B_{k|k}) \\ &= L_k^{CI}(X_k - \hat{x}^A_{k|k}) + M_k^{CI}(X_k - \hat{x}^B_{k|k}). \end{aligned}$$

Using the above in the definition of $\bar{P}^{CI}_{k|k}$, see Equation (4.12), we find:

$$\begin{aligned} \bar{P}^{CI}_{k|k} &= \mathbb{E}[(X_k - \hat{x}^{CI}_{k|k})(X_k - \hat{x}^{CI}_{k|k})^T|\mathcal{F}_k] \\ &= \mathbb{E}[(L_k^{CI}(X_k - \hat{x}^A_{k|k}) + M_k^{CI}(X_k - \hat{x}^B_{k|k}))(L_k^{CI}(X_k - \hat{x}^A_{k|k}) + M_k^{CI}(X_k - \hat{x}^B_{k|k}))^T|\mathcal{F}_k]. \end{aligned}$$

By writing out the definitions of L_k^{CI} and M_k^{CI}, see Equation (4.14), we obtain:

$$\begin{aligned} \bar{P}^{CI}_{k|k} = \mathbb{E}\Big[&\Big(P^{CI}_{k|k}\big(\omega(P^A_{k|k})^{-1}(X_k - \hat{x}^A_{k|k}) + (1-\omega)(P^B_{k|k})^{-1}(X_k - \hat{x}^B_{k|k})\big)\Big)\times \\ &\Big(P^{CI}_{k|k}\big(\omega(P^A_{k|k})^{-1}(X_k - \hat{x}^A_{k|k}) + (1-\omega)(P^B_{k|k})^{-1}(X_k - \hat{x}^B_{k|k})\big)\Big)^T|\mathcal{F}_k\Big]. \end{aligned}$$

Using the notation in Equation (4.12) and that $P^A_{k|k}$ and $P^B_{k|k}$ are symmetric, we find:

$$\begin{aligned} \bar{P}^{CI}_{k|k} = P^{CI}_{k|k}\Big(&\omega^2(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} + \omega(1-\omega)(P^A_{k|k})^{-1}\bar{P}^{AB}_{k|k}(P^B_{k|k})^{-1} \\ &+ \omega(1-\omega)(P^B_{k|k})^{-1}\bar{P}^{BA}_{k|k}(P^A_{k|k})^{-1} + (1-\omega)^2(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1}\Big)P^{CI}_{k|k}. \end{aligned}$$

Then the left hand side of Equation (4.13) can be rewritten as:

$$\begin{aligned} P^{CI}_{k|k} - \bar{P}^{CI}_{k|k} = P^{CI}_{k|k} - \mathbb{E}[(X_k - \hat{x}^{CI}_{k|k})(X_k - \hat{x}^{CI}_{k|k})^T|\mathcal{F}_k] = P^{CI}_{k|k} - P^{CI}_{k|k}\Big(&\omega^2(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} \\ + \omega(1-\omega)(P^A_{k|k})^{-1}\bar{P}^{AB}_{k|k}(P^B_{k|k})^{-1} + \omega(1-\omega)(P^B_{k|k})^{-1}\bar{P}^{BA}_{k|k}(P^A_{k|k})^{-1}& \\ + (1-\omega)^2(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1}\Big)P^{CI}_{k|k}.& \end{aligned}$$

When multiplying both sides from the left and the right with $(P^{CI}_{k|k})^{-1}$, we find:

$$\begin{aligned} (P^{CI}_{k|k})^{-1} - (P^{CI}_{k|k})^{-1}\bar{P}^{CI}_{k|k}(P^{CI}_{k|k})^{-1} = (P^{CI}_{k|k})^{-1} - \omega^2(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} - \omega(1-\omega)(P^A_{k|k})^{-1}\bar{P}^{AB}_{k|k}(P^B_{k|k})^{-1} \\ - \omega(1-\omega)(P^B_{k|k})^{-1}\bar{P}^{BA}_{k|k}(P^A_{k|k})^{-1} - (1-\omega)^2(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1}. \end{aligned} \tag{4.15}$$

By assumption $(\hat{x}^A_{k|k}, P^A_{k|k})$ is a consistent estimate for $X_k|\mathcal{F}_k$. Note that this assumption is not proven, see Remark 4.2.5. If $(\hat{x}^A_{k|k}, P^A_{k|k})$ is a consistent estimate for $X_k|\mathcal{F}_k$ the following needs to hold:

$$P^A_{k|k} = \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}^A_k] \succeq \mathbb{E}[(X_k - \hat{x}^A_{k|k})(X_k - \hat{x}^A_{k|k})^T|\mathcal{F}_k].$$

Using the definition of $\bar{P}^{CI}_{k|k}$, see Equation (4.12), in Equation (4.9), we find:

$$P^A_{k|k} \succeq \bar{P}^A_{k|k}.$$

Multiplying from the left and the right with $(P^A_{k|k})^{-1}$ gives:

$$(P^A_{k|k})^{-1} \succeq (P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1}.$$

We find a similar expression for measurement source B: $(P^B_{k|k})^{-1} \succeq (P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1}$. These can be used to rewrite $(P^{CI}_{k|k})^{-1}$, as defined in Equation (4.5a):

$$\begin{aligned}(P^{CI}_{k|k})^{-1} &= \omega(P^A_{k|k})^{-1} + (1-\omega)(P^B_{k|k})^{-1} \\ &\succeq \omega(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} + (1-\omega)(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1}.\end{aligned}$$

We use the above to continue to rewrite Equation (4.15):

$$\begin{aligned}(P^{CI}_{k|k})^{-1} - (P^{CI}_{k|k})^{-1}\bar{P}^{CI}_{k|k}(P^{CI}_{k|k})^{-1} &\succeq \omega(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} + (1-\omega)(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1} \\ &\quad - \omega^2(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} - \omega(1-\omega)(P^A_{k|k})^{-1}\bar{P}^{AB}_{k|k}(P^B_{k|k})^{-1} \\ &\quad - \omega(1-\omega)(P^B_{k|k})^{-1}\bar{P}^{BA}_{k|k}(P^A_{k|k})^{-1} - (1-\omega)^2(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1} \\ &= \omega(1-\omega)(P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} + \omega(1-\omega)(P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1} \\ &\quad - \omega(1-\omega)(P^A_{k|k})^{-1}\bar{P}^{AB}_{k|k}(P^B_{k|k})^{-1} - \omega(1-\omega)(P^B_{k|k})^{-1}\bar{P}^{BA}_{k|k}(P^A_{k|k})^{-1} \\ &= \omega(1-\omega)\big((P^A_{k|k})^{-1}\bar{P}^A_{k|k}(P^A_{k|k})^{-1} - (P^A_{k|k})^{-1}\bar{P}^{AB}_{k|k}(P^B_{k|k})^{-1} - (P^B_{k|k})^{-1}\bar{P}^{BA}_{k|k}(P^A_{k|k})^{-1} \\ &\quad + (P^B_{k|k})^{-1}\bar{P}^B_{k|k}(P^B_{k|k})^{-1}\big).\end{aligned}$$

We use the definitions of $\bar{P}^A_{k|k}$, $\bar{P}^B_{k|k}$, $\bar{P}^{AB}_{k|k}$ and $\bar{P}^{BA}_{k|k}$, see Equation (4.12) to find:

$$\begin{aligned}(P^{CI}_{k|k})^{-1} - (P^{CI}_{k|k})^{-1}\bar{P}^{CI}_{k|k}(P^{CI}_{k|k})^{-1} &= \omega(1-\omega)\mathbb{E}\Big[((P^A_{k|k})^{-1}(X_k - \hat{x}^A_{k|k}) - (P^B_{k|k})^{-1}(X_k - \hat{x}^B_{k|k}))\times \\ &\quad ((P^A_{k|k})^{-1}(X_k - \hat{x}^A_{k|k}) - (P^B_{k|k})^{-1}(X_k - \hat{x}^B_{k|k}))^T|\mathcal{F}_k\Big] \succeq 0,\end{aligned}$$

because the expression on the right hand side is a square. Thus we conclude that covariance intersection is consistent for $X_k|\mathcal{F}_k$ every choice of $\omega \in [0,1]$, provided that the partial a posteriori estimates $(\hat{x}^A_{k|k}, P^A_{k|k})$ and $(\hat{x}^B_{k|k}, P^B_{k|k})$ are consistent for the random variable $X_k|\mathcal{F}_k$. □

Proof of Lemma 4.2.8. We rewrite L as follows:

$$\begin{aligned}L &= \omega CA^{-1} = \omega\left(\omega A^{-1} + (1-\omega)B^{-1}\right)^{-1}A^{-1} \\ &= \left(\frac{1}{\omega}A\left(\omega A^{-1} + (1-\omega)B^{-1}\right)\right)^{-1} = \left(I + \frac{1-\omega}{\omega}AB^{-1}\right)^{-1}.\end{aligned}$$

We use $I = \left(\frac{1}{1-\omega}B\right)\left(\frac{1}{1-\omega}B\right)^{-1}$ and take $\left(\frac{1}{1-\omega}B\right)^{-1}$ out of the brackets to obtain:

$$L = \left(\left(\frac{1}{1-\omega}B + \frac{1}{\omega}A\right)(1-\omega)B^{-1}\right)^{-1} = \frac{1}{1-\omega}B\left(\frac{1}{1-\omega}B + \frac{1}{\omega}A\right)^{-1}.$$

We can similarly rewrite M:

$$\begin{aligned}M &= (1-\omega)CB^{-1} = (1-\omega)\left(\omega A^{-1} + (1-\omega)B^{-1}\right)^{-1}B^{-1} \\ &= \left(\frac{1}{1-\omega}B\left(\omega A^{-1} + (1-\omega)B^{-1}\right)\right)^{-1} = \left(\frac{\omega}{1-\omega}BA^{-1} + I\right)^{-1}.\end{aligned}$$

We use $I = \left(\frac{1}{\omega}A\right)\left(\frac{1}{\omega}A\right)^{-1}$ and take $\left(\frac{1}{\omega}A\right)^{-1}$ out of the brackets to obtain:

$$\begin{aligned}M &= \left(\left(\frac{1}{1-\omega}B + \frac{1}{\omega}A\right)\omega A^{-1}\right)^{-1} \\ &= \frac{1}{\omega}A\left(\frac{1}{1-\omega}B + \frac{1}{\omega}A\right)^{-1},\end{aligned}$$

then

$$L + M = \frac{1}{1-\omega} B \left(\frac{1}{1-\omega} B + \frac{1}{\omega} A \right)^{-1} + \frac{1}{\omega} A \left(\frac{1}{1-\omega} B + \frac{1}{\omega} A \right)^{-1}$$
$$= \left(\frac{1}{1-\omega} B + \frac{1}{\omega} A \right) \left(\frac{1}{1-\omega} B + \frac{1}{\omega} A \right)^{-1} = I,$$

establishing the claim. □

4.3. Covariance intersection on covariance matrices of measurement noise

This section provides a reformulation of covariance intersection. Covariance intersection is essentially a linear combination between the partial estimates calculated by conditioning on part of the measurements. In this section we define a method that applies the linear combination that is used in covariance intersection to the covariance matrices of the noise. Then we prove that this method gives the same result as covariance intersection described in Section 4.1 for a fixed ω. The advantage of this reformulation is that the fusion step is done before the Kalman filter is applied, instead of breaking up the Kalman filter as in the standard approach to covariance intersection. This makes it possible to translate the fusion step and the Kalman filter separately to continuous time. The lemma below presents the reformulation of covariance intersection.

Lemma 4.3.1 (Reformulation of covariance intersection)**.** *Given the dynamic system of Equation* (4.1)*. For $\omega \in [0,1]$, define a matrix $(R_k^{RCI})^{-1}$ as follows:*

$$(R_k^{RCI})^{-1} = \begin{bmatrix} \omega (R_k^A)^{-1} & 0 \\ 0 & (1-\omega)(R_k^B)^{-1} \end{bmatrix}.$$

Applying the Kalman filter to the dynamic system of Equation (4.1)*, using matrix $(R_k^{RCI})^{-1}$ instead of matrix $(R_k)^{-1}$, gives an estimate for the state and its covariance, denoted by $(\hat{x}_{k|k}^{RCI}, P_{k|k}^{RCI})$. Using the definitions given in Equation* (4.2)*, this estimate is calculated by the following variation of the Kalman filter:*

Prediction step:

$$\hat{x}_{k|k-1}^{RCI} = F_{k-1} \hat{x}_{k-1|k-1}^{RCI},$$
$$P_{k|k-1}^{RCI} = F_{k-1} P_{k-1|k-1}^{RCI} F_{k-1}^T + Q_{k-1}.$$

Update step:

$$\begin{aligned}
\hat{x}_{k|k}^{RCI} &= \hat{x}_{k|k-1}^{RCI} + K_k^{RCI} (y_k - H_k \hat{x}_{k|k-1}^{RCI}) \\
&= \hat{x}_{k|k-1}^{RCI} + \omega P_{k|k}^{RCI} (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \hat{x}_{k|k-1}^{RCI}) + (1-\omega) P_{k|k}^{RCI} (H_k^B)^T (R_k^B)^{-1} (y_k^B - H_k^B \hat{x}_{k|k-1}^{RCI}), \\
P_{k|k}^{RCI} &= \left((P_{k|k-1}^{RCI})^{-1} + H_k^T (R_k^{RCI})^{-1} H_k \right)^{-1} \\
&= \left((P_{k|k}^{RCI})^{-1} + \omega (H_k^A)^T (R_k^A)^{-1} H_k^A + (1-\omega)(H_k^B)^T (R_k^B)^{-1} H_k^B \right)^{-1}, \\
K_k^{RCI} &= P_{k|k}^{RCI} H_k^T (R_k^{RCI})^{-1} \\
&= \begin{bmatrix} \omega P_{k|k}^{RCI} (H_k^A)^T (R_k^A)^{-1} \\ (1-\omega) P_{k|k}^{RCI} (H_k^B)^T (R_k^B)^{-1} \end{bmatrix}^T.
\end{aligned} \tag{4.16}$$

This lemma is proven below.

Remark 4.3.2. *R_k^{RCI} is defined such that:*

$$\begin{bmatrix} \omega(R_k^A)^{-1} & 0 \\ 0 & (1-\omega)(R_k^B)^{-1} \end{bmatrix} \preceq \begin{bmatrix} R_k^A & \Sigma_k^{AB} \\ \Sigma_k^{BA} & R_k^B \end{bmatrix}^{-1}.$$

To see this, we define the vectors $\tilde{V}_k^A = [V_k^A, 0]^T$ and $\tilde{V}_k^B = [0, V_k^B]^T$. Also define:

$$\tilde{R}_k^A = \begin{bmatrix} R_k^A & 0 \\ 0 & 0 \end{bmatrix} \qquad \tilde{R}_k^B = \begin{bmatrix} 0 & 0 \\ 0 & R_k^B \end{bmatrix}$$

Clearly, $(0, \tilde{R}_k^A)$ and $(0, \tilde{R}_k^B)$ are consistent estimates for $\tilde{V}_k^A$ and $\tilde{V}_k^B$, respectively. Then for $\omega \in (0,1)$, $R_k^{RCI} = \frac{1}{\omega}\tilde{R}_k^A + \frac{1}{1-\omega}\tilde{R}_k^B$. By Lemma 4.3.7, $(0, R_k^{RCI})$ is conservative for $\tilde{V}_k^A + \tilde{V}_k^B = [V_k^A, V_k^B]^T$ for $\omega \in (0,1)$. For $\omega \in \{0,1\}$, it is clear that $\tilde{R}_k^A \preceq R_k$ and $\tilde{R}_k^B \preceq R_k$. Thus we conclude for $\omega \in [0,1]$:

$$\begin{bmatrix} \omega(R_k^A)^{-1} & 0 \\ 0 & (1-\omega)(R_k^B)^{-1} \end{bmatrix} \preceq \begin{bmatrix} R_k^A & \Sigma_k^{AB} \\ \Sigma_k^{BA} & R_k^B \end{bmatrix}^{-1}.$$

Of course, changing the covariance matrix of the measurement noise like in Lemma 4.3.1 violates the conditions to apply the Kalman filter. This change therefore impairs the optimality of the Kalman filter. Note that in this situation this is not a problem, because we aim to reformulate covariance intersection. Covariance intersection also impairs the optimality of the Kalman filter. In Theorem 4.3.3 we will show that the Kalman filter with the adjustment of Lemma 4.3.1 is the same as covariance intersection, as given in Definition 4.1.1 for a fixed ω.

Theorem 4.3.3. *Given the dynamic system of Equation* (4.1)*, covariance intersection described in Definition 4.1.1 and the method described in Lemma 4.3.1 are the same algorithm for a fixed $\omega \in [0,1]$.*

The proof of this theorem is given at the end of this section.

Remark 4.3.4. *The reformulation given in Lemma 4.3.1 removes the need to calculate two partial estimates and fuse them together, because it fuses the covariance matrices of the noise before applying the Kalman filter. This reduces the number of times that the Kalman filter needs to be applied compared with the method described in Definition 4.1.1 while obtaining the same result for a fixed $\omega \in [0,1]$, see Theorem 4.3.3. Furthermore, it means that it is possible to apply covariance intersection to the continuous time Kalman filter by changing the input in a similar manner as in the discrete case. Thus, if it is possible to fuse the intensities of the measurement noise from source A and B, then covariance intersection can be applied to the continuous Kalman filter.*

Along with the calculation of the a priori and a posteriori steps of covariance intersection, the calculation of ω can also we rewritten. This is done in Theorem 4.3.5.

Theorem 4.3.5. *The minimisation problem of Equation* (4.6) *with $\chi(\omega, P_{k|k-1}^{CI}, \hat{x}_{k|k-1}, y_k, H_k, \tilde{R}_k) = P_{k|k}^{CI}$ is equivalent to:*

$$\omega = \arg\min_{\omega} \phi \left(\left(\left(P_{k|k-1}^{CI}\right)^{-1} + \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}^T \begin{bmatrix} \frac{1}{\omega} R_k^A & 0 \\ 0 & \frac{1}{(1-\omega)} R_k^B \end{bmatrix}^{-1} \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix} \right)^{-1} \right).$$

This theorem is proven below.

Remark 4.3.6. *Applying covariance intersection to the continuous time Kalman filter requires continuous optimisation of ω. The optimisation of ω is needed to ensure the convergence of covariance intersection. Thus ω must influence the convergence rate. Since there is no consensus in the literature on which measure for ϕ leads to the fastest convergence rate, it might be possible to change the map χ without impairing the convergence rate. It is possible that there exists a map χ that is a linear function of ω, $(H_k^A)^T(R_k^A)^{-1}H_k^A$ and $(H_k^B)^T(R_k^B)^{-1}H_k^B$, which makes it possible to translate the optimisation of ω to continuous time and to apply covariance intersection to the continuous time Kalman filter. Note that this would lead to a different algorithm than covariance intersection described in Definition 4.1.1.*

Below the lemma used in Remark 4.3.2 is given.

Lemma 4.3.7. *Let X_1, X_2 be two $\mathbb{R}^d$ valued random variables. Let (x_1, Σ_1) and (x_2, Σ_2) be conservative for X_1 and X_2 respectively. Let $\omega \in (0,1)$ and let*

$$U \succeq \frac{1}{\omega}\Sigma_1 + \frac{1}{1-\omega}\Sigma_2$$

Then $(x_1 + x_2, U)$ is conservative for $X_1 + X_2$.

This lemma is proven later in this section.

We start by proving the results of this section.

Proof of Lemma 4.3.1. Changing the covariance matrix of the measurement noise does not change the expressions for the a priori steps, because the covariance matrix of the measurement noise is not involved in those steps. For the a posteriori steps, we will start with the covariance matrix of the estimate $P^{RCI}_{k|k}$. By Equation (2.6) the expression of $P^{RCI}_{k|k}$ is:

$$\begin{aligned}
P^{RCI}_{k|k} &= \left((P^{RCI}_{k|k-1})^{-1} + H_k^T (R_k^{RCI})^{-1} H_k\right)^{-1} \\
&= \left((P^{RCI}_{k|k-1})^{-1} + \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}^T \begin{bmatrix} \frac{1}{\omega} R_k^A & 0 \\ 0 & \frac{1}{1-\omega} R_k^B \end{bmatrix}^{-1} \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}\right)^{-1} \\
&= \left((P^{RCI}_{k|k})^{-1} + \omega (H_k^A)^T (R_k^A)^{-1} H_k^A + (1-\omega)(H_k^B)^T (R_k^B)^{-1} H_k^B\right)^{-1}.
\end{aligned}$$

We continue with the Kalman gain, again starting from the expression in Equation (2.6):

$$\begin{aligned}
K_k^{RCI} &= P^{RCI}_{k|k} H_k^T (R_k^{RCI})^{-1} \\
&= P^{RCI}_{k|k} \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}^T \begin{bmatrix} \frac{1}{\omega} R_k^A & 0 \\ 0 & \frac{1}{1-\omega} R_k^B \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \omega P^{RCI}_{k|k} (H_k^A)^T (R_k^A)^{-1} \\ (1-\omega) P^{RCI}_{k|k} (H_k^B)^T (R_k^B)^{-1} \end{bmatrix}^T.
\end{aligned}$$

Lastly, we derive the expression of $\hat{x}^{RCI}_{k|k}$ starting from the expression in Equation (2.6):

$$\begin{aligned}
\hat{x}^{RCI}_{k|k} &= \hat{x}^{RCI}_{k|k-1} + K_k^{RCI}(y_k - H_k \hat{x}^{RCI}_{k|k-1}) \\
&= \hat{x}^{RCI}_{k|k-1} + \begin{bmatrix} \omega P^{RCI}_{k|k} (H_k^A)^T (R_k^A)^{-1} \\ (1-\omega) P^{RCI}_{k|k} (H_k^B)^T (R_k^B)^{-1} \end{bmatrix}^T \left(\begin{bmatrix} y_k^A \\ y_k^B \end{bmatrix} - \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix} \hat{x}^{RCI}_{k|k-1}\right) \\
&= \hat{x}^{RCI}_{k|k-1} + \begin{bmatrix} \omega P^{RCI}_{k|k} (H_k^A)^T (R_k^A)^{-1} \\ (1-\omega) P^{RCI}_{k|k} (H_k^B)^T (R_k^B)^{-1} \end{bmatrix}^T \begin{bmatrix} y_k^A - H_k^A \hat{x}^{RCI}_{k|k-1} \\ y_k^B - H_k^B \hat{x}^{RCI}_{k|k-1} \end{bmatrix} \\
&= \hat{x}^{RCI}_{k|k-1} + \omega P^{RCI}_{k|k} (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \hat{x}^{RCI}_{k|k-1}) + (1-\omega) P^{RCI}_{k|k} (H_k^B)^T (R_k^B)^{-1} (y_k^B - H_k^B \hat{x}^{RCI}_{k|k-1}).
\end{aligned}$$

□

Proof of Theorem 4.3.3. We denote the estimates calculated by covariance intersection, as given in Definition 4.1.1, by $(\hat{x}^{CI}_{k|k}, P^{CI}_{k|k})$. The estimates calculated by the method described in Lemma 4.3.1 are denoted by $(\hat{x}^{RCI}_{k|k}, P^{RCI}_{k|k})$. Fix $\omega \in [0,1]$. We prove that these two estimates are equal using induction. We start with the base step and then continue to the induction step.

Base step: Both algorithms have the same starting position, thus:

- Covariance of state estimate: $P^{CI}_{0|0} = P^{RCI}_{0|0}$,

- Estimate of the state: $\hat{x}_{0|0}^{CI} = \hat{x}_{0|0}^{RCI}$.

This concludes the base step.

Induction step: Assume $\hat{x}_{k|k}^{CI} = \hat{x}_{k|k}^{RCI}$ and $P_{k|k}^{CI} = P_{k|k}^{RCI}$. We derive the induction step for the a priori and a posteriori estimates separately.

- A priori covariance of state estimate:

$$\begin{aligned} P_{k+1|k}^{CI} &= F_k \hat{x}_{k|k}^{CI} F_k + Q_k \\ &= F_k \hat{x}_{k|k}^{RCI} F_k + Q_k \\ &= P_{K+1|k}^{RCI}. \end{aligned}$$

- A priori estimate of the state:

$$\begin{aligned} \hat{x}_{k+1|k}^{CI} &= F_k \hat{x}_{k|k}^{CI} \\ &= F_k \hat{x}_{k|k}^{RCI} \\ &= \hat{x}_{k+1|k}^{RCI}. \end{aligned}$$

- A posteriori covariance of state estimate:

$$P_{k+1|k+1}^{CI} = (\omega (P_{k+1|k+1}^{A})^{-1} + (1-\omega)(P_{k+1|k+1}^{B})^{-1})^{-1}.$$

We use the expression of $P_{k+1|k+1}^{A}$ and $P_{k+1|k+1}^{B}$ as in Equation (4.4):

$$\begin{aligned} P_{k+1|k+1}^{CI} &= (\omega((P_{k+1|k}^{CI})^{-1} + (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} H_{k+1}^{A}) + (1-\omega)((P_{k+1|k}^{CI})^{-1} + (H_{k+1}^{B})^T (R_{k+1}^{B})^{-1} H_{k+1}^{B})^{-1} \\ &= ((P_{k+1|k}^{CI})^{-1} + \omega (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} H_{k+1}^{A}) + (1-\omega)(H_{k+1}^{B})^T (R_{k+1}^{B})^{-1} H_{k+1}^{B})^{-1}. \end{aligned}$$

We use $P_{k+1|k}^{CI} = P_{k+1|k}^{RCI}$:

$$\begin{aligned} P_{k+1|k+1}^{CI} &= ((P_{k+1|k}^{RCI})^{-1} + \omega (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} H_{k+1}^{A}) + (1-\omega)(H_{k+1}^{B})^T (R_{k+1}^{B})^{-1} H_{k+1}^{B})^{-1} \\ &= P_{k+1|k+1}^{RCI}. \end{aligned}$$

- A posteriori estimate of the state:

$$\begin{aligned} \hat{x}_{k+1|k+1}^{CI} &= P_{k+1|k+1}^{CI} (\omega (P_{k+1|k+1}^{A})^{-1} \hat{x}_{k+1|k+1}^{A} + (1-\omega)(P_{k+1|k+1}^{B})^{-1} \hat{x}_{k+1|k+1}^{B}) \\ &= \omega P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{A})^{-1} \hat{x}_{k+1|k+1}^{A} + (1-\omega) P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{B})^{-1} \hat{x}_{k+1|k+1}^{B}). \end{aligned}$$

We use the expression of $\hat{x}_{k+1|k+1}^{A}$ and $\hat{x}_{k+1|k+1}^{B}$ of Equation (4.4):

$$\begin{aligned} \hat{x}_{k+1|k+1}^{CI} &= \omega P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{A})^{-1} (\hat{x}_{k+1|k}^{CI} + P_{k+1|k+1}^{A} (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} (y_{k+1}^{A} - H_{k+1}^{A} \hat{x}_{k+1|k}^{CI}) \\ &\quad + (1-\omega) P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{B})^{-1} (\hat{x}_{k+1|k}^{CI} + P_{k+1|k+1}^{B} (H_{k+1}^{B})^T (R_{k+1}^{B})^{-1} (y_{k+1}^{B} - H_{k+1}^{B} \hat{x}_{k+1|k}^{CI}) \\ &= (\omega P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{A})^{-1} + (1-\omega) P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{B})^{-1}) \hat{x}_{k+1|k}^{CI} \\ &\quad + \omega P_{k+1|k+1}^{CI} (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} (y_{k+1}^{A} - H_{k+1}^{A} \hat{x}_{k+1|k}^{CI}) \\ &\quad + (1-\omega) P_{k+1|k+1}^{CI} (H_k + 1^B)^T (R_k + 1^B)^{-1} (y_{k+1}^{B} - H_{k+1}^{B} \hat{x}_{k+1|k}^{CI}). \end{aligned}$$

Using Lemma 4.2.8 with $L_k^{CI} = \omega P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{A})^{-1}$ and $M_k^{CI} = (1-\omega) P_{k+1|k+1}^{CI} (P_{k+1|k+1}^{B})^{-1}$ such that $L_k^{CI} + M_k^{CI} = I$, we find:

$$\begin{aligned} \hat{x}_{k+1|k+1}^{CI} &= \hat{x}_{k+1|k}^{CI} + \omega P_{k+1|k+1}^{CI} (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} (y_{k+1}^{A} - H_{k+1}^{A} \hat{x}_{k+1|k}^{CI}) \\ &\quad + (1-\omega) P_{k+1|k+1}^{CI} (H_{k+1}^{B})^T (R_{k+1}^{B})^{-1} (y_{k+1}^{B} - H_{k+1}^{B} \hat{x}_{k+1|k}^{CI}). \end{aligned}$$

We use $\hat{x}_{k+1|k}^{CI} = \hat{x}_{k+1|k}^{RCI}$:

$$\begin{aligned} \hat{x}_{k+1|k+1}^{CI} &= \hat{x}_{k+1|k}^{RCI} + \omega P_{k+1|k+1}^{RCI} (H_{k+1}^{A})^T (R_{k+1}^{A})^{-1} (y_{k+1}^{A} - H_{k+1}^{A} \hat{x}_{k+1|k}^{RCI}) \\ &\quad + (1-\omega) P_{k+1|k+1}^{RCI} (H_{k+1}^{B})^T (R_{k+1}^{B})^{-1} (y_{k+1}^{B} - H_{k+1}^{B} \hat{x}_{k+1|k}^{RCI}) \\ &= \hat{x}_{k+1|k+1}^{RCI}. \end{aligned}$$

In conclusion, for a fixed omega, covariance intersection can be rewritten as in Lemma 4.3.1. $\square$

Proof of Theorem 4.3.5. We will start with Equation (4.6) and rewrite it using the expressions of $P_{k|k}^{CI}$, $P_{k|k}^{A}$ and $P_{k|k}^{B}$, given in Equations (4.5a) and Equation (4.4) respectively:

$$
\begin{aligned}
\omega &= \arg\min_{\omega} \phi\left(P_{k|k}^{CI}\right) \\
&= \arg\min_{\omega} \phi\left(\left(\omega\left(P_{k|k}^{A}\right)^{-1} + (1-\omega)\left(P_{k|k}^{B}\right)^{-1}\right)^{-1}\right) \\
&= \arg\min_{\omega} \phi\left(\left(\omega\left(\left(P_{k|k-1}^{CI}\right)^{-1} + \left(H_k^A\right)^T\left(R_k^A\right)^{-1}H_k^A\right) + (1-\omega)\left(\left(P_{k|k-1}^{CI}\right)^{-1} + \left(H_k^B\right)^T\left(R_k^B\right)^{-1}H_k^B\right)\right)^{-1}\right) \\
&= \arg\min_{\omega} \phi\left(\left(\left(P_{k|k-1}^{CI}\right)^{-1} + \omega\left(H_k^A\right)^T\left(R_k^A\right)^{-1}H_k^A + (1-\omega)\left(H_k^B\right)^T\left(R_k^B\right)^{-1}H_k^B\right)^{-1}\right) \\
&= \arg\min_{\omega} \phi\left(\left(\left(P_{k|k-1}^{CI}\right)^{-1} + \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}^T \begin{bmatrix} \frac{1}{\omega}R_k^A & 0 \\ 0 & \frac{1}{(1-\omega)}R_k^B \end{bmatrix}^{-1} \begin{bmatrix} H_k^A \\ H_k^B \end{bmatrix}\right)^{-1}\right).
\end{aligned}
$$

This establishes the claim. $\square$

Proof of Lemma 4.3.7. We aim to establish:

$$
U - \mathbb{E}\left[\left(X_1 + X_2 - (x_1 + x_2)\right)\left(X_1 + X_2 - (x_1 + x_2)\right)^T\right] \succeq 0.
$$

Working out the expectation, we have:

$$
\begin{aligned}
&\mathbb{E}\left[\left(X_1 + X_2 - (x_1 + x_2)\right)\left(X_1 + X_2 - (x_1 + x_2)\right)^T\right] \\
&\quad = \mathbb{E}\left[\left(X_1 - x_1\right)\left(X_1 - x_1\right)^T\right] + \mathbb{E}\left[\left(X_2 - x_2\right)\left(X_2 - x_2\right)^T\right] \\
&\qquad + \mathbb{E}\left[\left(X_1 - x_1\right)\left(X_2 - x_2\right)^T + \left(X_2 - x_2\right)\left(X_1 - x_1\right)^T\right].
\end{aligned}
$$

Set $\gamma = \sqrt{\frac{1-\omega}{\omega}}$ to simplify the computation below. Using the definition of U and the properties of (x_1, Σ_1) and (x_2, Σ_2) we obtain:

$$
\begin{aligned}
&U - \mathbb{E}\left[\left(X_1 + X_2 - (x_1 + x_2)\right)\left(X_1 + X_2 - (x_1 + x_2)\right)^T\right] \\
&\quad \succeq \frac{1-\omega}{\omega}\Sigma_1 + \frac{\omega}{1-\omega}\Sigma_2 - \mathbb{E}\left[\left(X_1 - x_1\right)\left(X_2 - x_2\right)^T + \left(X_2 - x_2\right)\left(X_1 - x_1\right)^T\right].
\end{aligned}
$$

Using that if (x, Σ) is consistent for X, then $\Sigma \succeq \mathbb{E}[(X-x)(X-x)^T]$ gives:

$$
\begin{aligned}
&U - \mathbb{E}\left[\left(X_1 + X_2 - (x_1 + x_2)\right)\left(X_1 + X_2 - (x_1 + x_2)\right)^T\right] \\
&\quad \succeq \gamma^2\mathbb{E}\left[\left(X_1 - x_1\right)\left(X_1 - x_1\right)^T\right] + \gamma^{-2}\mathbb{E}\left[\left(X_2 - x_2\right)\left(X_2 - x_2\right)^T\right] \\
&\qquad - \mathbb{E}\left[\left(X_1 - x_1\right)\left(X_2 - x_2\right)^T + \left(X_2 - x_2\right)\left(X_1 - x_1\right)^T\right] \\
&\quad = \mathbb{E}\left[\left(\gamma\left(X_1 - x_1\right) - \gamma^{-1}\left(X_2 - x_2\right)\right)^2\right] \\
&\quad \succeq 0.
\end{aligned}
$$

This concludes the claim. $\square$

5

Covariance intersection in H_∞ filter

In Section 5.1 we modify the cost function of the H_∞ filter such that it is a generalisation of covariance intersection applied to the Kalman filter. Since a continuous version of the H_∞ filter exists, this approach offers a different method to extend the fusion step of covariance intersection to continuous time.

5.1. Covariance intersection in H_∞ filter

We consider the following dynamic system with two measurement sources A and B:

$$\begin{aligned} x_{k+1} &= F_k x_k + w_k, \\ y_k^A &= H_k^A x_k + v_k^A, \\ y_k^B &= H_k^B x_k + v_k^B, \end{aligned} \tag{5.1}$$

where $x_k \in \mathbb{R}^n$ is the state vector; $F_k \in \mathbb{R}^n \times \mathbb{R}^n$ is the state transition matrix, which can describe the physics underlying the process that is modelled; $w_k \in \mathbb{R}^n$ is the noise term. $y_k^A \in \mathbb{R}^m$ represents the measurements from source A at time k; $H_k^A \in \mathbb{R}^m \times \mathbb{R}^n$ provides a linear connection between the state vector and the measurement vector for measurement source A; $v_k^A \in \mathbb{R}^m$ is the measurement noise. Measurement source B is allowed to have different dimensions, so $y_k^B \in \mathbb{R}^q$, $H_k^B \in \mathbb{R}^q \times \mathbb{R}^n$, and $v_k^B \in \mathbb{R}^q$. The noise terms w_k, v_k^A and v_k^B do not have to satisfy any assumptions and may even be deterministic. Note that it is possible to add as many sources of measurements as desired, but for the sake of simplicity we use two measurement sources.

Remember the cost function of the H_∞ filter, given in Equation (3.2):

$$J = -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1}\left[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\|w_k\|^2_{Q_k^{-1}} - \frac{1}{\theta}\|v_k\|^2_{R_k^{-1}}\right],$$

with $S_k \in \mathbb{R}^n \times \mathbb{R}^n$; $P_0 \in \mathbb{R}^n \times \mathbb{R}^n$; $Q_k \in \mathbb{R}^n \times \mathbb{R}^n$ and $R_k \in \mathbb{R}^m \times \mathbb{R}^m$ positive semi-definite matrices. The vector norm is defined as $\|x\|_A^2 = x^T A x$. To introduce the structure of covariance intersection to the H_∞ filter, the term containing v_k is split as follows:

$$\|v_k\|^2_{R_k^{-1}} = \|v_k^A\|^2_{(R_k^A)^{-1}} + \|v_k^B\|^2_{(R_k^B)^{-1}},$$

where $R_k^A \in \mathbb{R}^m \times \mathbb{R}^m$ and $R_k^B \in \mathbb{R}^q \times \mathbb{R}^q$ are positive semi-definite matrices. The two terms on the right hand side get their own different θ, which will take on the role that ω has in covariance intersection, see Definition 4.1.1. Note that this split essentially constrains the allowed structure of the R_k matrix in the original H_∞ filter. If we write $R_k = \begin{bmatrix} R_k^A & R_k^{AB} \\ R_k^{BA} & R_k^B \end{bmatrix}$, then this split forces $R_k^{AB} = R_k^{BA} = 0$. This reflects that covariance intersection treats the measurement noise as independent until fusing the two estimates. Forcing this structure on the matrix R_k negatively impacts the performance of the H_∞ filter, because it reduces the opportunity for tuning. The split of $\|v_k\|^2_{R_k^{-1}}$ gives rise to the following cost function, for $\theta, \theta_{R_k^A}, \theta_{R_k^B} > 0$:

$$\begin{aligned} J &= -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1}\left[\|z_k - \hat{z}_k\|^2_{S_k} - \frac{1}{\theta}\|w_k\|^2_{Q_k^{-1}} - \frac{1}{\theta_{R^A}}\|v_k^A\|^2_{(R_k^A)^{-1}} - \frac{1}{\theta_{R^B}}\|v_k^B\|^2_{(R_k^B)^{-1}}\right] \\ &= -\frac{1}{\theta}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N-1}\left[\|x_k - \hat{x}_k\|^2_{\bar{S}_k} - \frac{1}{\theta}\|w_k\|^2_{Q_k^{-1}} - \frac{1}{\theta_{R^A}}\|y_k^A - H_k^A x_k\|^2_{(R_k^A)^{-1}} \right. \\ &\qquad \left. - \frac{1}{\theta_{R^B}}\|y_k^B - H_k^B x_k\|^2_{(R_k^B)^{-1}}\right] \\ &= \varphi(x_0) + \sum_{k=0}^{N-1}\mathcal{L}_k(x_k, w_k, y_k^A, y_k^B), \end{aligned} \tag{5.2}$$

where $\bar{S}_k = L_k^T S_k L_k$. Theorem 5.1.1 derives the H_∞ filter belonging to this cost function.

Theorem 5.1.1. *Given the dynamic system of Equation* (5.1) *and the cost function defined in Equation* (5.2)*. As well as $\theta, \theta_{R_k^A}, \theta_{R_k^B} > 0$. Using the following notation:*

$$H_k = \begin{bmatrix} H_k^A \\ H_k^k \end{bmatrix}, \qquad R_k = \begin{bmatrix} \frac{\theta_{R^A}}{\theta} R_k^A & 0 \\ 0 & \frac{\theta_{R^B}}{\theta} R_k^B \end{bmatrix}, \qquad y_k = \begin{bmatrix} y_k^A \\ y_k^k \end{bmatrix},$$

the solution to the following optimisation problem:

$$\min_{x_k} \max_{\substack{x_0, w_k, \\ y_k^A, y_k^B}} J, \tag{5.3}$$

is given by:

Prediction step:

$$\begin{aligned} \bar{S}_k &= L_k^T S_k L_k, \\ \hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1|k-1}, \\ P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1} + Q_{k-1}. \end{aligned}$$

Update step:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}) \\ &= \hat{x}_{k|k-1} + P_{k|k}\left(\frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\hat{x}_{k|k-1}) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\hat{x}_{k|k-1})\right), \\ P_{k|k} &= (P_{k|k-1}^{-1} - \theta\bar{S}_k + H_k^T R_k^{-1} H_k)^{-1} \\ &= \left(P_{k|k-1}^{-1} - \theta\bar{S}_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^B\right)^{-1}, \\ K_k &= P_{k|k}H_k^T R_k^{-1} \\ &= P_{k|k}\begin{bmatrix} \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1} \\ \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1} \end{bmatrix}^T, \end{aligned} \tag{5.4}$$

provided that:

$$\begin{aligned}
P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k &\succ 0, \\
R_k^A - \frac{\theta}{\theta_{R^A}} H_k^A P_{k|k} (H_k^A)^T &\succ 0, \\
R_k^B - \frac{\theta}{\theta_{R^B}} H_k^B P_{k|k} (H_k^B)^T &\succ 0.
\end{aligned}$$

The proof of this theorem can be found in Appendix A.

In the solution given in Theorem 5.1.1 three different θ's arise. The following proposition shows that it is possible to choose these such that we find the Kalman filter with covariance intersection applied to the measurement covariance matrices.

Proposition 5.1.2. *Given the dynamic system in Equation* (5.1) *with* v_k, w_k^A *and* w_k^B *Gaussian white noise such that* $v_k \sim N(0, Q_k)$, $w_k^A \sim N(0, R_k^A)$ *and* $w_k^B \sim N(0, R_k^B)$ *and* v_k *independent of* w_k^A *and* w_k^B. *The algorithm given in Lemma 4.3.1 can be obtained with two additional constraints from the algorithm given in Theorem 5.1.1 by choosing:*

$$\begin{aligned}
\theta_{R^A} &= \frac{\theta}{\omega}, \\
\theta_{R^B} &= \frac{\theta}{1-\omega},
\end{aligned}$$

and sending $\theta \to 0$. *The additional constraints to the algorithm in Lemma 4.3.1 are:*

$$\begin{aligned}
R_k^A - \omega H_k^A P_{k|k} (H_k^A)^T &\succ 0, \\
R_k^B - (1-\omega) H_k^B P_{k|k} (H_k^B)^T &\succ 0.
\end{aligned}$$

Remark 5.1.3. *It is important that* v_k, w_k^A *and* w_k^B *are Gaussian white noise with the correct covariance matrices. If these noise terms are not white noise, then the Kalman filter will not be obtained when sending* $\theta \to 0$. *The choice of* S_k *is unrestricted, because as* $\theta \to 0$, *the role of* S_k *disappears. This can also be seen from Lemma 3.3.4, where it is shown that the Kalman filter minimises the mean square error of the* S_k *weighted sum for every symmetric* S_k. *Note that the additional constraints arise from the structure of the* H_∞ *filter.*

Proof of Proposition 5.1.2. For $P_{k|k}$ in Equation (5.4) to be equal to $P_{k|k}$ in Equation (4.16), we need $\frac{\theta}{\theta_{R^A}} = \omega$. This is obtained when $\theta_{R^A} = \frac{\theta}{\omega}$. The reasoning is similar for the choice of θ_{R^B}.

With these choices, we are left with the following algorithm:

Prediction step:

$$\begin{aligned}
\bar{S}_k &= L_k^T S_k L_k, \\
\hat{x}_{k|k-1} &= F_{k-1} \hat{x}_{k-1|k-1}, \\
P_{k|k-1} &= F_{k-1} P_{k-1|k-1} F_{k-1} + Q_{k-1}.
\end{aligned}$$

Update step:

$$\begin{aligned}
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k \begin{bmatrix} y_k^A - H_k^A \hat{x}_{k|k-1} \\ y_k^B - H_k^B \hat{x}_{k|k-1} \end{bmatrix}, \\
P_{k|k} &= \left(P_{k|k-1}^{-1} - \theta \bar{S}_k + \omega (H_k^A)^T (R_k^A)^{-1} H_k^A + (1-\omega) (H_k^B)^T (R_k^B)^{-1} H_k^B \right)^{-1}, \\
K_k &= P_{k|k} \begin{bmatrix} \omega (H_k^A)^T (R_k^A)^{-1} \\ (1-\omega)(H_k^B)^T (R_k^B)^{-1} \end{bmatrix}^T.
\end{aligned} \tag{5.5}$$

The only difference between Equation (5.5) and (4.16) is the term $\theta \bar{S}_k$ in the expression for $P_{k|k}$. When sending $\theta \to 0$, this term vanishes and the two expressions are the equal.

To ensure that both the minimum and the maximum of the optimisation problem in Equation (5.3) are attained, we need:

$$
\begin{aligned}
P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k &\succ 0, \\
R_k^A - \frac{\theta}{\theta_{R^A}} H_k^A P_{k|k} (H_k^A)^T &\succ 0, \\
R_k^B - \frac{\theta}{\theta_{R^B}} H_k^B P_{k|k} (H_k^B)^T &\succ 0.
\end{aligned}
$$

The first equation is automatically satisfied when sending $\theta \to 0$. The last two equations turn into the following when $\frac{\theta}{\theta_{R^A}} = \omega$ and $\frac{\theta}{\theta_{R^B}} = 1 - \omega$:

$$
\begin{aligned}
R_k^A - \omega H_k^A P_{k|k} (H_k^A)^T &\succ 0, \\
R_k^B - (1-\omega) H_k^B P_{k|k} (H_k^B)^T &\succ 0.
\end{aligned}
$$

This concludes the claim. $\square$

6 Conclusion

This thesis studied the Kalman filter with observations from one and multiple measurement sources. A problem arises when the correlation between multiple measurement sources is not known when using the Kalman filter. There exist multiple solutions to this problem. The solutions discussed in this work are the H_∞ filter, see Chapter 3, and covariance intersection, see Chapter 4. The overall aim of this thesis was to investigate whether it is possible to apply covariance intersection to the continuous time Kalman filter. To this end three topics were discussed:

- Section 4.2 highlights the assumption made in the proof of the consistency of covariance intersection given by S. J. Julier and J. K. Uhlmann. It is found that this assumption which is treated as trivial in the literature is found to be non-trivial. For this assumption to hold, the a posteriori Kalman filter steps must be consistent. This is likely not true in general. However, the assumption made in the proof probably holds in expectation over many time steps. So in practice covariance intersection behaves as a consistent algorithm.
- In Section 4.3, the linear combination in covariance intersection is applied to the covariance matrices of the measurement noise rather than to the covariance matrices of the estimates. This reformulation essentially changes the input of the Kalman filter. This means that the Kalman filter only needs to be applied once per time step rather than once per dependent measurement source per time step. It also means, that if this change in input can be translated to continuous time, then the continuous Kalman filter can be used with the altered input, making it possible to apply covariance intersection to the continuous time Kalman filter.
- In Section 5.1, the fact that the H_∞ filter is a generalisation of the Kalman filter is used to include the limit case of covariance intersection described in Section 4.3 in the H_∞ filter. While this has no practical application, the ability to formulate the H_∞ filter in continuous time provides a different path to formulate covariance intersection such that it can be applied to the continuous time Kalman filter.

6.1. Recommendations for future work

In Section 4.2 we saw that the assumption made in the proof of the consistency of covariance intersection given by S. J. Julier and J. K. Uhlmann, is neither trivial nor proven. Does assumption does not hold in general, but it probably holds in expectation. A possible avenue for further research is to show that the terms preventing consistency for every time step approach zero on average. This would explain why there are no problems with consistency when applying covariance intersection in practice.

The literature on covariance intersection states that covariance intersection is consistent for every $\omega \in [0, 1]$, but that the optimisation of ω to an indifferent (convex) measure is needed to ensure convergence. It would be interesting to investigate why the algorithm converges for every (convex) measure. Does the measure influence the convergence rate of covariance intersection?

6.1.1. Possible extensions to continuous time

To apply covariance intersection to the continuous time Kalman filter, the fusion of the covariance matrices of the measurement noise should be formulated in continuous time. A possible way of doing this, is by changing the map χ in the optimisation problem for ω. Perhaps χ can be chosen as linear map of $\omega, (H_k^A)^T(R_k^A)^{-1}H_k^A$ and $(H_k^B)^T(R_k^B)^{-1}H_k^B$ such that it is possible to find a differential equation for ω. It is however, not clear what the consequences of this change would be. Covariance intersection will remain consistent (in expectation), because this property is not dependent on the value of ω. However, ω controls the convergence of covariance intersection, as convergence is ensured only by optimizing ω at each time step. Therefore, the map χ should be chosen such that convergence is still ensured. Investigating possible choices for the map χ such that covariance intersection converges will likely make it possible to formulate covariance intersection such that it can be applied to the continuous time Kalman filter.

References

[1] A. P. Andrews and M. S. Grewal. *Kalman Filtering: Theory and Practice with MATLAB, 3rd Edition*. John Wiley & Sons Inc., 2008.

[2] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons Inc., 2001.

[3] T. M. Cover, J. A. Thomas, et al. “Entropy, relative entropy and mutual information”. In: *Elements of information theory* 2.1 (1991), pp. 12–13.

[4] I. Ekeland and R. Téman. *Convex Analysis and Variational Problemns*. Elsevier, 1976.

[5] Robin Forsling, Benjamin Noack, and Gustaf Hendeby. “A Quarter Century of Covariance Intersection: Correlations Still Unknown?” In: *IEEE Control Systems Magazine* 44.2 (2024), pp. 81–105.

[6] G. R. Grimmett and D. J. A. Welsh. *Probability: An Introduction, 2nd Edition*. Oxford University Press, 2014.

[7] D. Hall and J. Llinas. *Handbook of Multisensor Data Fusion*. CRC Press, 2001.

[8] S. J. Julier and J. K. Uhlmann. “A non-divergent estimation algorithm in the presence of unknown correlations”. In: *Proceedings of the 1997 American Control Conference (Cat. No. 97CH36041)*. Vol. 4. IEEE. 1997, pp. 2369–2373.

[9] S. J. Julier and J. K. Uhlmann. “Using covariance intersection for SLAM”. In: *Robotics and Autonomous Systems* 55.1 (2007), pp. 3–20.

[10] F.L. Lewis, L. Xie, and D. Popa. *Optimal and robust estimation; with an introduction to stochastic control theory, 2nd Edition*. CRC Press, 2008.

[11] J. I. Marden. *Multivariate Statistics*. University of Illinois at Urbana-Champaign, 2015.

[12] H. Masnadi-Shirazi, A. Masnadi-Shirazi, and M.-A. Dastgheib. “A Step by Step Mathematical Derivation and Tutorial on Kalman Filters”. In: *arXiv preprint arXiv:1910.03558* (2019).

[13] S. F. Schmidt. “The Kalman filter-Its recognition and development for aerospace applications”. In: *Journal of Guidance and Control* 4.1 (1981), pp. 4–7.

[14] D. Simon. *Optimal state estimation*. John Wiley & Sons Inc., 2006.

[15] J. K. Uhlmann. “Dynamic Map Building and Localization: New Theoretical Foundations”. PhD thesis. University of Oxford, 1995.

[16] D. Wang and X. Cai. “Robust data assimilation in hydrological modeling–A comparison of Kalman and H-infinity filters”. In: *Advances in water resources* 31.3 (2008), pp. 455–472.

A

Proof of Theorem 5.1.1

This appendix proves Theorem 5.1.1. The proof in this appendix is based on the derivation of the H_∞ filter given in [14]. Before proving Theorem 5.1.1 a preliminary result is needed. Lagrange multipliers are used to solve constrained optimisation problems. The constraints are assumed to be constant over time. In the context of the H_∞ filter, the constraints have the following form: $x_{k+1} = F_k x_k + w_k$, which means that the constraint might change for every time step. We use a sequence of Lagrange multipliers, instead of a constant Lagrange multiplier to account for the time varying constraints. Lemma A.0.1 shows that the method of Lagrange multipliers can be extended in this way. This extended method of Lagrange multipliers is then used to prove Theorem 5.1.1.

Lemma A.0.1. *Given the following process dynamics:*

$$x_{k+1} = F_k x_k + w_k, \tag{A.1}$$

with $x_k \in \mathbb{R}^n$, $w_k \in \mathbb{R}^n$ and $F_k \in \mathbb{R}^n \times \mathbb{R}^n$. For some differentiable function $\phi : \mathbb{R}^n \to \mathbb{R}$ and some differentiable function $\mathcal{L}_k : \mathbb{R}^{2n} \to \mathbb{R}$ define the following cost function:

$$J(x_k, w_k) = \phi(x_0) + \sum_{k=0}^{N-1} \mathcal{L}_k(x_k, w_k).$$

The cost function can be augmented with a sequence of Lagrange multipliers to obtain:

$$J_\lambda = \phi(x_0) + \sum_{k=0}^{N-1} \left(\mathcal{H}_k(x_k, w_k) - \lambda_k^T x_k\right) - \lambda_N^T x_N + \lambda_0^T x_0,$$

with $\mathcal{H}_k(x_k, w_k) = \mathcal{L}_k(x_k, w_k) + \lambda_{k+1}^T(F_k x_k + w_k)$, which is called the Hamiltonian. The Hamiltonian is analogous to the Lagrangian except that λ is a sequence instead of a constant. The necessary conditions of the stationary points for the following constrained optimisation problem:

$$\begin{aligned} &\max_{x_0, w_k} J(x_k, w_k) \\ &\text{s.t. } x_{k+1} = F_k x_k + w_k, \end{aligned} \tag{A.2}$$

are:

$$\lambda_0^T + \frac{\partial \phi}{\partial x_0}(x_0) = 0, \tag{A.3a}$$

$$-\lambda_N^T = 0, \tag{A.3b}$$

$$\frac{\partial \mathcal{H}_k}{\partial w_k}(x_k, w_k) = 0, \qquad (k = 0, \ldots, N-1), \tag{A.3c}$$

$$\lambda_k = \frac{\partial \mathcal{H}_k}{\partial x_k}(x_k, w_k), \qquad (k = 0, \ldots, N-1). \tag{A.3d}$$

Proof. We intent to find the necessary constraints for the stationary points for the constraint optimisation problem defined in Equation (A.2). We add the constraint posed by the dynamic system via the method of Lagrange multipliers. This gives:

$$\begin{aligned} J_\lambda(x_k, w_k) &= \phi(x_0) + \sum_{k=0}^{N-1} [\mathcal{L}_k(x_k, w_k) + \lambda_{k+1}^T (F_k x_k + w_k - x_{k+1})] \\ &= \phi(x_0) + \sum_{k=0}^{N-1} [\mathcal{L}_k(x_k, w_k) + \lambda_{k+1}^T (F_k x_k + w_k)] - \sum_{k=0}^{N-1} \lambda_{k+1}^T x_{k+1}. \end{aligned}$$

Adding and subtracting $\lambda_0^T x_0$ gives:

$$J_\lambda(x_k, w_k) = \phi(x_0) + \sum_{k=0}^{N-1} [\mathcal{L}_k(x_k, w_k) + \lambda_{k+1}^T (F_k x_k + w_k)] - \sum_{k=0}^{N} \lambda_k^T x_k + \lambda_0^T x_0.$$

We define the Hamiltonian:

$$\mathcal{H}_k(x_k, w_k) = \mathcal{L}_k(x_k, w_k) + \lambda_{k+1}^T (F_k x_k + w_k).$$

We use the Hamiltonian to rewrite the cost function:

$$\begin{aligned} J_\lambda(x_k, w_k) &= \phi(x_0) + \sum_{k=0}^{N-1} \mathcal{H}_k(x_k, w_k) - \sum_{k=0}^{N} \lambda_k^T x_k + \lambda_0^T x_0 \\ &= \phi(x_0) + \sum_{k=0}^{N-1} \mathcal{H}_k(x_k, w_k) - \sum_{k=0}^{N-1} \lambda_k^T x_k - \lambda_N^T x_N + \lambda_0^T x_0 \\ &= \phi(x_0) + \sum_{k=0}^{N-1} (\mathcal{H}_k(x_k, w_k) - \lambda_k^T x_k) - \lambda_N^T x_N + \lambda_0^T x_0. \end{aligned} \tag{A.4}$$

Just like when using a constant λ the following conditions need to be satisfied for the constrained stationary point:

$$\frac{\partial J_\lambda}{\partial x_k}(x_k, w_k) = 0, \qquad (k = 0, \ldots, N), \tag{A.5a}$$

$$\frac{\partial J_\lambda}{\partial w_k}(x_k, w_k) = 0, \qquad (k = 0, \ldots, N-1), \tag{A.5b}$$

$$\frac{\partial J_\lambda}{\partial \lambda_k}(x_k, w_k) = 0, \qquad (k = 0, \ldots, N). \tag{A.5c}$$

Equation (A.5c) turns out to be superfluous, because in this context the first two equations give enough information to solve the constrained optimisation problem. We start with rewriting Equation (A.5a) using the definition of the cost function defined in Equation (A.4). For $k = 1, \ldots, N-1$, we find:

$$\begin{aligned} \frac{\partial J_\lambda}{\partial x_k}(x_k, w_k) &= \frac{\partial \mathcal{H}_k}{\partial x_k}(x_k, w_k) - \lambda_k = 0, \\ \Leftrightarrow \quad \lambda_k &= \frac{\partial \mathcal{H}_k}{\partial x_k}(x_k, w_k). \end{aligned} \tag{A.6}$$

For $k = N$ Equation (A.5a) can be rewritten as:

$$\frac{\partial J_\lambda}{\partial x_N}(x_N, w_N) = -\lambda_N^T = 0.$$

For $k = 0$ Equation (A.5a) becomes:

$$\begin{aligned} \frac{\partial J_\lambda}{\partial x_0}(x_0, w_0) &= \frac{\partial \phi}{\partial x_0}(x_0) + \frac{\partial \mathcal{H}_0}{\partial x_0}(x_0, w_0) - \lambda_0 + \lambda_0 = 0, \\ \Leftrightarrow \quad \lambda_0 + \frac{\partial \phi}{\partial x_0}(x_0) &= -\frac{\partial \mathcal{H}_0}{\partial x_0}(x_0, w_0) - \lambda_0. \end{aligned}$$

Note that this last equality holds when both sides are 0. So we add $k = 0$ to the constraint in Equation (A.6). And we add the constraint:

$$\lambda_0 + \frac{\partial \phi}{\partial x_0}(x_0) = 0.$$

Using the definition of $J_\lambda(x_k, w_k)$ of Equation (A.4), Equation (A.5b) can be rewritten for $k = 0, \ldots N-1$:

$$\frac{\partial J_\lambda}{\partial w_k}(x_k, w_k) = \frac{\partial \mathcal{H}_k}{\partial w_k}(x_k, w_k) = 0.$$

So in conclusion we find the following constraints:

$$\begin{aligned} \lambda_0^T + \frac{\phi}{\partial x_0}(x_0) &= 0, & \\ -\lambda_N^T &= 0, & \\ \lambda_k &= \frac{\partial \mathcal{H}_k}{\partial x_k}(x_k, w_k), & (k = 0, \ldots, N-1), \\ \frac{\partial \mathcal{H}_k}{\partial w_k}(x_k, w_k) &= 0, & (k = 0, \ldots, N-1). \end{aligned}$$

□

We move on to the proof of Theorem 5.1.1.

A.1. Proof of Theorem 5.1.1

Proof of Theorem 5.1.1. This proof will consist of the following steps:

1. Finding x_0 and w_k that satisfy the constraints of Equations (A.3a) through (A.3d), see Subsection A.1.1.
2. Incorporating the equations found from the constraints of Equations (A.3a) through (A.3d) into the cost function defined in Equation (5.2), see Subsection A.1.2.
3. Calculating the partial derivatives of the cost function defined in Equation (5.2) with respect to x_k, y_k^A and y_k^B to find the stationary points with respect to x_k, y_k^A and y_k^B, see Subsection A.1.3.
4. Finding the conditions such that the stationary points found are a minimum for x_k and a maximum for y_k^A and y_k^B, see Subsection A.1.4.

A.1.1. Satisfying necessary conditions

Before solving the conditions of Equation (A.3) we change the Lagrange multiplier slightly. This will not change the solution, but makes the calculations easier. We use $\frac{2\lambda_{k+1}}{\theta}$ instead of λ_{k+1}. Note that θ is different from θ_{R^A} and θ_{R^B}. The Hamiltonian then becomes:

$$\mathcal{H}_k(x_k, w_k) = \mathcal{L}_k(x_k, w_k, y_k^A, y_k^B) + \frac{2\lambda_{k+1}^T}{\theta}(F_k x_k + w_k).$$

From the condition of Equation (A.3a) we obtain:

$$\frac{2\lambda_0}{\theta} - \frac{2}{\theta}P_0^{-1}(x_0 - \hat{x}_0) = 0.$$

Multiplying from the left with P_0 and dividing by $\frac{2}{\theta}$ gives:

$$\begin{aligned} & P_0\lambda_0 - (x_0 - \hat{x}_0) = 0, \\ \Leftrightarrow \quad & x_0 = \hat{x}_0 + P_0\lambda_0. \end{aligned} \tag{A.7}$$

Equation (A.3b) gives:

$$\lambda_N = 0. \tag{A.8}$$

Satisfying Equation (A.3c) requires:

$$-\frac{2}{\theta}Q_k^{-1}w_k + \frac{2\lambda_{k+1}}{\theta} = 0.$$

Multiplying from the left by Q_k and dividing by $\frac{2}{\theta}$ gives:

$$w_k = Q_k\lambda_{k+1}. \tag{A.9}$$

Equation (A.9) can be substituted into the process dynamics of Equation (A.1) to obtain:

$$x_{k+1} = F_k x_k + Q_k\lambda_{k+1}. \tag{A.10}$$

The fourth condition given in Equation (A.3d) gives:

$$\begin{aligned}\frac{2\lambda_k}{\theta} &= 2\bar{S}_k(x_k - \hat{x}_k) + \frac{2}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A x_k) + \frac{2}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B x_k) + \frac{2}{\theta}F_k^T\lambda_{k+1},\\ \lambda_k &= \theta\bar{S}_k(x_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A x_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B x_k) + F_k^T\lambda_{k+1}.\end{aligned} \tag{A.11}$$

To move forward, we have to make an assumption. From Equation (A.7) we know that x_0 is a linear function of λ_0, so we will assume that x_k is a linear function of λ_k. Thus:

$$x_k = \mu_k + P_k\lambda_k, \tag{A.12}$$

for all $k = 0, \dots, N$, where μ_k and P_k are some vector and matrix to be determined, with P_0 given, and the initial condition:

$$\mu_0 = \hat{x}_0. \tag{A.13}$$

This assumption may or may not turn out to be valid. Substituting Equation (A.12) into Equation (A.10) gives:

$$\mu_{k+1} + P_{k+1}\lambda_{k+1} = F_k\mu_k + F_kP_k\lambda_k + Q_k\lambda_{k+1}. \tag{A.14}$$

Substituting Equation (A.12) into Equation (A.11) gives:

$$\begin{aligned}\lambda_k &= F_k^T\lambda_{k+1} + \theta\bar{S}_k(\mu_k + P_k\lambda_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A(\mu_k + P_k\lambda_k))\\ &\quad + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B(\mu_k + P_k\lambda_k)),\end{aligned} \tag{A.15a}$$

$$\begin{aligned}\Leftrightarrow\quad & \lambda_k - \theta\bar{S}_kP_k\lambda_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k\lambda_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\lambda_k\\ &\quad = F_k^T\lambda_{k+1} + \theta\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k),\\ \Leftrightarrow\quad & \lambda_k = [I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k]^{-1}[F_k^T\lambda_{k+1} + \theta\bar{S}_k(\mu_k - \hat{x}_k)\\ &\quad + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)].\end{aligned} \tag{A.15b}$$

Substituting Equation (A.15b) into Equation (A.14) gives:

$$\begin{aligned}\mu_{k+1} + P_{k+1}\lambda_{k+1} &= F_k\mu_k\\ &\quad + F_kP_k\left[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\right]^{-1}\times\\ &\quad \left[F_k^T\lambda_{k+1} + \theta\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\right]\\ &\quad + Q_k\lambda_{k+1}.\end{aligned}$$

Bringing all terms with μ_k to the left hand side gives:

$$\begin{aligned}
&\mu_{k+1} - F_k\mu_k \\
&\quad - F_kP_k\Big[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\Big]^{-1} \times \\
&\quad \Big[\theta\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\Big] \\
&\quad = \Big[-P_{k+1} + F_kP_k\Big[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\Big]^{-1}F_k^T \\
&\quad + Q_k\Big]\lambda_{k+1}.
\end{aligned} \tag{A.16}$$

This equation is satisfied if both sides are zero. Setting left hand side to zero gives:

$$\begin{aligned}
\mu_{k+1} = F_k\mu_k + F_kP_k&\Big[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\Big]^{-1} \times \\
&\Big[\theta\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\Big],
\end{aligned}$$

with the initial condition given in Equation (A.13). Setting the right hand side of Equation (A.16) to zero gives:

$$P_{k+1} = F_kP_k[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k]^{-1}F_k^T + Q_k. \tag{A.17}$$

We define:

$$\begin{aligned}
\tilde{P}_k &= P_k[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k]^{-1} \\
&= [P_k^{-1} - \theta\bar{S}_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^B]^{-1}.
\end{aligned} \tag{A.18}$$

And we can thus write:

$$P_{k+1} = F_k\tilde{P}_kF_k^T + Q_k. \tag{A.19}$$

From Equation (A.18) we see that if $P_k, \bar{S}_k$, and R_k are symmetric, then $\tilde{P}_k$ will be symmetric. We see from Equation (A.17) that if Q_k is also symmetric, then P_{k+1} will be symmetric. So if P_0, Q_k, R_k, and $\bar{S}_k$ are symmetric for all k, then $\tilde{P}_k$ and P_k will be symmetric for all k. Note that P_0, Q_k, R_k are not necessarily covariance matrices thus the symmetry of these matrices is not automatically guaranteed.

The necessary conditions to ensure that the constraints of the optimisation problem are satisfied can be summarised as follows:

$$x_0 = \hat{x}_0 + P_0\lambda_0, \tag{A.20a}$$
$$w_k = Q_k\lambda_{k+1}, \tag{A.20b}$$
$$\lambda_N = 0, \tag{A.20c}$$
$$\lambda_k = \left[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\right]^{-1} \times \left[F_k^T\lambda_{k+1} + \theta\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\right], \tag{A.20d}$$
$$P_{k+1} = F_kP_k\left[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\right]^{-1}F_k^T + Q_k, \tag{A.20e}$$
$$\mu_0 = \hat{x}_0, \tag{A.20f}$$
$$\mu_{k+1} = F_k\mu_k + F_kP_k\left[I - \theta\bar{S}_kP_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^AP_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\right]^{-1} \times \left[\theta\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\right]. \tag{A.20g}$$

The next section will incorporate the relevant conditions into the cost function of Equation (5.2).

A.1.2. Incorporating necessary conditions into cost function

In this subsection the conditions of Equation (A.20a) through Equation (A.20f) will be incorporated into the cost function of Equation (5.2). The condition in Equation (A.20g) is used to create the H_∞ algorithm.

By incorporating these conditions the condition in Equation (A.20g) is also satisfied.

Using the condition of Equation (A.20a) and (A.20f) in Equation (A.12), we see that:

$$\begin{aligned} \lambda_k &= P_k^{-1}(x_k - \mu_k), \\ \lambda_0 &= P_0^{-1}(x_0 - \hat{x}_0). \end{aligned} \tag{A.21}$$

We therefore obtain:

$$\begin{aligned} \|x_0 - \hat{x}_0\|^2_{P_0^{-1}} &= (x_0 - \hat{x}_0)^TP_0^{-1}(x_0 - \hat{x}_0) \\ &= (x_0 - \hat{x}_0)^TP_0^{-T}P_0P_0^{-1}(x_0 - \hat{x}_0) \\ &= \lambda_0^TP_0\lambda_0 \\ &= \|\lambda_0\|^2_{P_0}. \end{aligned} \tag{A.22}$$

Therefore, Equation (5.2) can be written as:

$$J = -\frac{1}{\theta}\|\lambda_0\|^2_{P_0} + \sum_{k=0}^{N-1}\left[\|x_k - \hat{x}_k\|^2_{\bar{S}_k} - \left(\frac{1}{\theta}\|w_k\|^2_{Q_k^{-1}} + \frac{1}{\theta_{R^A}}\|y_k^A - H_k^Ax_k\|^2_{(R_k^A)^{-1}} + \frac{1}{\theta_{R^B}}\|y_k^B - H_k^Bx_k\|^2_{(R_k^B)^{-1}}\right)\right]. \tag{A.23}$$

Substituting the expression for x_k from Equation (A.12) into Equation (A.23) gives:

$$J = -\frac{1}{\theta}\|\lambda_0\|^2_{P_0} + \sum_{k=0}^{N-1}\left[\|\mu_k + P_k\lambda_k - \hat{x}_k\|^2_{\bar{s}_k} - \left(\frac{1}{\theta}\|w_k\|^2_{Q_k^{-1}} + \frac{1}{\theta_{R^A}}\|y_k^A - H_k^A(\mu_k + P_k\lambda_k)\|^2_{(R_k^A)^{-1}} + \frac{1}{\theta_{R^B}}\|y_k^B - H_k^B(\mu_k + P_k\lambda_k)\|^2_{(R_k^B)^{-1}}\right)\right]. \tag{A.24}$$

We rewrite the term $\|w_k\|^2_{Q_k^{-1}}$ as follows:

$$\|w_k\|^2_{Q_k^{-1}} = w_k^T Q_k^{-1} w_k.$$

Using the condition in Equation (A.20b) for w_k gives:

$$\begin{aligned}\|w_k\|^2_{Q_k^{-1}} &= \lambda_{k+1}^T Q_k^T Q_k^{-1} Q_k \lambda_{k+1} \\ &= \lambda_{k+1}^T Q_k \lambda_{k+1} \\ &= \|\lambda_{k+1}\|^2_{Q_k}.\end{aligned}$$

Equation (A.24) can therefore be written as:

$$\begin{aligned}J = -\frac{1}{\theta}\|\lambda_0\|^2_{P_0} + \sum_{k=0}^{N-1}\Big[&\|\mu_k + P_k\lambda_k - \hat{x}_k\|^2_{\bar{s}_k} - \frac{1}{\theta_{R^A}}\|y_k^A - H_k^A(\mu_k + P_k\lambda_k)\|^2_{(R_k^A)^{-1}} \\ &- \frac{1}{\theta_{R^B}}\|y_k^B - H_k^B(\mu_k + P_k\lambda_k)\|^2_{(R_k^B)^{-1}}\Big] - \frac{1}{\theta}\sum_{k=0}^{N-1}\|\lambda_{k+1}\|^2_{Q_k}.\end{aligned} \tag{A.25}$$

Using the condition in Equation (A.20c), we can write:

$$\begin{aligned}0 &= \sum_{k=0}^{N}\lambda_k^T P_k \lambda_k - \sum_{k=0}^{N-1}\lambda_k^T P_k \lambda_k \\ &= \lambda_0^T P_0 \lambda_0 + \sum_{k=1}^{N}\lambda_k^T P_k \lambda_k - \sum_{k=0}^{N-1}\lambda_k^T P_k \lambda_k \\ &= \lambda_0^T P_0 \lambda_0 + \sum_{k=0}^{N-1}\lambda_{k+1}^T P_{k+1} \lambda_{k+1} - \sum_{k=0}^{N-1}\lambda_k^T P_k \lambda_k \\ &= \|\lambda_0\|^2_{P_0} + \sum_{k=0}^{N-1}\left(\|\lambda_{k+1}\|^2_{P_{k+1}} - \|\lambda_k\|^2_{P_k}\right).\end{aligned}$$

Dividing both sides by θ gives:

$$0 = \frac{1}{\theta}\|\lambda_0\|^2_{P_0} + \frac{1}{\theta}\sum_{k=0}^{N-1}\left(\|\lambda_{k+1}\|^2_{P_{k+1}} - \|\lambda_k\|^2_{P_k}\right).$$

We can add this zero term from the cost function of Equation (A.25) to obtain:

$$\begin{aligned}J = -\frac{1}{\theta}\|\lambda_0\|^2_{P_0} + \sum_{k=0}^{N-1}\Big[&\|\mu_k + P_k\lambda_k - \hat{x}_k\|^2_{\bar{S}_k} - \frac{1}{\theta_{R^A}}\|y_k^A - H_k^A(\mu_k + P_k\lambda_k)\|^2_{(R_k^A)^{-1}} \\ &- \frac{1}{\theta_{R^B}}\|y_k^B - H_k^B(\mu_k + P_k\lambda_k)\|^2_{(R_k^B)^{-1}}\Big] - \frac{1}{\theta}\sum_{k=0}^{N-1}\|\lambda_{k+1}\|^2_{Q_k} \\ &+ \frac{1}{\theta}\|\lambda_0\|^2_{P_0} + \frac{1}{\theta}\sum_{k=0}^{N-1}\left(\|\lambda_{k+1}\|^2_{P_{k+1}} - \|\lambda_k\|^2_{P_k}\right).\end{aligned} \tag{A.26}$$

Note that we can rewrite the matrix norm $\|\mu_k + P_k\lambda_k - \hat{x}_k\|^2_{\bar{S}_k}$ as follows:

$$\begin{aligned}\|\mu_k + P_k\lambda_k - \hat{x}_k\|^2_{\bar{S}_k} &= (\mu_k + P_k\lambda_k - \hat{x}_k)^T \bar{S}_k(\mu_k + P_k\lambda_k - \hat{x}_k) \\ &= (\mu_k - \hat{x}_k)^T \bar{S}_k(\mu_k - \hat{x}_k) + (\mu_k - \hat{x}_k)^T \bar{S}_k P_k\lambda_k + (P_k\lambda_k)^T \bar{S}_k(\mu_k - \hat{x}_k) \\ &\quad + (P_k\lambda_k)^T \bar{S}_k P_k\lambda_k.\end{aligned} \tag{A.27}$$

Since $(P_k\lambda_k)^T\bar{S}_k(\mu_k-\hat{x}_k)\in\mathbb{R}$ it is equal to its transpose. Thus we find:

$$\|\mu_k+P_k\lambda_k-\hat{x}_k\|^2_{\bar{S}_k}=(\mu_k-\hat{x}_k)^T\bar{S}_k(\mu_k-\hat{x}_k)+2(\mu_k-\hat{x}_k)^T\bar{S}_kP_k\lambda_k+(P_k\lambda_k)^T\bar{S}_kP_k\lambda_k. \tag{A.28}$$

Using similar reasoning we find:

$$\begin{aligned}\frac{1}{\theta_{R^A}}\|y_k^A-H_k^A(\mu_k+P_k\lambda_k)\|^2_{(R_k^A)^{-1}}&=\frac{1}{\theta_{R^A}}(y_k^A-H_k^A\mu_k)^T(R_k^A)^{-1}(y_k^A-H_k^A\mu_k)\\&\quad-\frac{2}{\theta_{R^A}}(y_k^A-H_k^A\mu_k)^T(R_k^A)^{-1}H_k^AP_k\lambda_k+\frac{1}{\theta_{R^A}}\lambda_k^TP_k(H_k^A)^T(R_k^A)^{-1}H_k^AP_k\lambda_k,\end{aligned} \tag{A.29a}$$

$$\begin{aligned}\frac{1}{\theta_{R^B}}\|y_k^B-H_k^B(\mu_k+P_k\lambda_k)\|^2_{(R_k^B)^{-1}}&=\frac{1}{\theta_{R^B}}(y_k^B-H_k^B\mu_k)^T(R_k^B)^{-1}(y_k^B-H_k^B\mu_k)\\&\quad-\frac{2}{\theta_{R^B}}(y_k^B-H_k^B\mu_k)^T(R_k^B)^{-1}H_k^BP_k\lambda_k+\frac{1}{\theta_{R^B}}\lambda_k^TP_k(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\lambda_k.\end{aligned} \tag{A.29b}$$

Substituting Equations (A.27), Equation (A.29a) and Equation (A.29b) into Equation (A.26) and using that P_k is symmetric gives:

$$\begin{aligned}J=\sum_{k=0}^{k-1}\Big[&(\mu_k-\hat{x}_k)^T\bar{S}_k(\mu_k-\hat{x}_k)+2(\mu_k-\hat{x}_k)^T\bar{S}_kP_k\lambda_k+\lambda_k^TP_k\bar{S}_kP_k\lambda_k\\&+\frac{1}{\theta}\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1}-\frac{1}{\theta}\lambda_k^TP_k\lambda_k-\frac{1}{\theta_{R^A}}(y_k^A-H_k^A\mu_k)^T(R_k^A)^{-1}(y_k^A-H_k^A\mu_k)\\&+\frac{2}{\theta_{R^A}}(y_k^A-H_k^A\mu_k)^T(R_k^A)^{-1}H_k^AP_k\lambda_k-\frac{1}{\theta_{R^A}}\lambda_k^TP_k(H_k^A)^T(R_k^A)^{-1}H_k^AP_k\lambda_k\\&-\frac{1}{\theta_{R^B}}(y_k^B-H_k^B\mu_k)^T(R_k^B)^{-1}(y_k^B-H_k^B\mu_k)+\frac{2}{\theta_{R^B}}(y_k^B-H_k^B\mu_k)^T(R_k^B)^{-1}H_k^BP_k\lambda_k\\&-\frac{1}{\theta_{R^B}}\lambda_k^TP_k(H_k^B)^T(R_k^B)^{-1}H_k^BP_k\lambda_k\Big].\end{aligned} \tag{A.30}$$

We consider the term $\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1}$ in the above expression. Substituting P_{k+1} from the condition in Equation (A.20e) in this term gives:

$$\begin{aligned}\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1}&=\lambda_{k+1}^T(F_k\tilde{P}_kF_k^T+Q_k-Q_k)\lambda_{k+1}\\&=\lambda_{k+1}^TF_k\tilde{P}_kF_k^T\lambda_{k+1}.\end{aligned} \tag{A.31}$$

And from Equation (A.15a) (which is equivalent to the necessary condition in Equation (A.20d)) we see that:

$$\begin{aligned}F_k^T\lambda_{k+1}=\lambda_k-\theta\bar{S}_k(\mu_k+P_k\lambda_k-\hat{x}_k)&-\frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A-H_k^A(\mu_k+P_k\lambda_k))\\&-\frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B-H_k^B(\mu_k+P_k\lambda_k)).\end{aligned} \tag{A.32}$$

Substituting this expression for $F_k^T\lambda_{k+1}$ into Equation (A.31) gives:

$$\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1} = \Big[\lambda_k - \theta\bar{S}_k(\mu_k + P_k\lambda_k - \hat{x}_k) - \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A(\mu_k + P_k\lambda_k)) - \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B(\mu_k + P_k\lambda_k))\Big]^T \tilde{P}_k\Big[\lambda_k - \theta\bar{S}_k(\mu_k + P_k\lambda_k - \hat{x}_k) - \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A(\mu_k + P_k\lambda_k)) - \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B(\mu_k + P_k\lambda_k))\Big].$$

Rewriting the brackets and working out the transposes gives:

$$\begin{aligned}\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1} = &\Big[\lambda_k^T(I - \theta P_k\bar{S}_k + \frac{\theta}{\theta_{R^A}}P_k(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}P_k(H_k^B)^T(R_k^B)^{-1}H_k^B)\\ &- \theta(\mu_k - \hat{x}_k^T)^T\bar{S}_k - \frac{\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A - \frac{\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\Big]\tilde{P}_k\times\\ &\Big[\lambda_k^T(I - \theta P_k\bar{S}_k + \frac{\theta}{\theta_{R^A}}P_k(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}P_k(H_k^B)^T(R_k^B)^{-1}H_k^B) - \theta(\mu_k - \hat{x}_k)^T\bar{S}_k\\ &- \frac{\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A - \frac{\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\Big]^T.\end{aligned} \tag{A.33}$$

Note from Equation (A.18) that $I - \theta P_k\bar{S}_k + \frac{\theta}{\theta_{R^A}}P_k(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}P_k(H_k^B)^T(R_k^B)^{-1}H_k^B = P_k\tilde{P}_k^{-1}$. Substituting this in Equation (A.33) gives the following:

$$\begin{aligned}&\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1}\\ &= \Big[\lambda_k^TP_k\tilde{P}_k^{-1} - \theta(\mu_k - \hat{x}_k)^T\bar{S}_k - \frac{\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A - \frac{\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\Big]\times\\ &\quad\tilde{P}_k\Big[\lambda_k^TP_k\tilde{P}_k^{-1} - \theta(\mu_k - \hat{x}_k)^T\bar{S}_k - \frac{\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A - \frac{\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\Big]^T.\end{aligned}$$

Working out the brackets gives:

$$\begin{aligned}\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1} = &\lambda_k^TP_k\tilde{P}_k^{-1}P_k\lambda_k - \theta(\mu_k - \hat{x}_k)^T\bar{S}_kP_k\lambda_k - \frac{\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^AP_k\lambda_k\\ &- \frac{\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^BP_k\lambda_k - \theta\lambda_k^TP_k\bar{S}_k(\mu_k - \hat{x}_k) + \theta^2(\mu_k - \hat{x}_k)^T\bar{S}_k\tilde{P}_k\bar{S}_k(\mu_k - \hat{x}_k)\\ &+ \frac{\theta^2}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k\bar{S}_k(\mu_k - \hat{x}_k) + \frac{\theta^2}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k\bar{S}_k(\mu_k - \hat{x}_k)\\ &- \frac{\theta}{\theta_{R^A}}\lambda_k^TP_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) + \frac{\theta^2}{\theta_{R^A}}(\mu_k - \hat{x}_k)^T\bar{S}_k\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k)\\ &+ \frac{\theta^2}{\theta_{R^A}^2}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k)\\ &+ \frac{\theta^2}{\theta_{R^A}\theta_{R^B}}(y^BH_k - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k)\\ &- \frac{\theta}{\theta_{R^B}}\lambda_k^TP_k(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k) + \frac{\theta^2}{\theta_{R^B}}(\mu_k - \hat{x}_k)^T\bar{S}_k\tilde{P}_k(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\\ &+ \frac{\theta^2}{\theta_{R^A}\theta_{R^B}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k)\\ &+ \frac{\theta^2}{\theta_{R^B}^2}(y^BH_k - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k).\end{aligned} \tag{A.34}$$

Notice that the left hand side of the above expression is a scalar. Therefore, the right hand side must also be a scalar and thus every term on the right hand side must also be a scaler. The transpose of a scalar is the scalar itself, so each term on the right hand side side is equal to its transpose. For example, consider the second term on the right side. Since it is a scalar, we see that $(\mu_k - \hat{x}_k)^T \bar{S}_k P_k \lambda_k = \lambda_k^T P_k \bar{S}_k(\mu_k - \hat{x}_k)$. Where we use that P_k and $\bar{S}_k$ are symmetric. Equation (A.34) can therefore be written as:

$$
\begin{aligned}
\lambda_{k+1}^T(P_{k+1} - Q_k)\lambda_{k+1} = {} & \lambda_k^T P_k \tilde{P}_k^{-1} P_k \lambda_k - 2\theta(\mu_k - \hat{x}_k)^T \bar{S}_k P_k \lambda_k - \frac{2\theta}{\theta_{R^A}}(y_k^A - H_k^A \mu_k)^T (R_k^A)^{-1} H_k^A P_k \lambda_k \\
& - \frac{2\theta}{\theta_{R^B}}(y_k^B - H_k^B \mu_k)^T (R_k^B)^{-1} H_k^B P_k \lambda_k + \theta^2 (\mu_k - \hat{x}_k)^T \bar{S}_k \tilde{P}_k \bar{S}_k (\mu_k - \hat{x}_k) \\
& + \frac{2\theta^2}{\theta_{R^A}}(y_k^A - H_k^A \mu_k)^T (R_k^A)^{-1} H_k^A \tilde{P}_k \bar{S}_k (\mu_k - \hat{x}_k) + \frac{2\theta^2}{\theta_{R^B}}(y_k^B - H_k^B \mu_k)^T (R_k^B)^{-1} H_k^B \tilde{P}_k \bar{S}_k (\mu_k - \hat{x}_k) \\
& + \frac{\theta^2}{\theta_{R^A}^2}(y_k^A - H_k^A \mu_k)^T (R_k^A)^{-1} H_k^A \tilde{P}_k (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \mu_k) \\
& + \frac{2\theta^2}{\theta_{R^A}\theta_{R^B}}(y^B H_k - H_k^B \mu_k)^T (R_k^B)^{-1} H_k^B \tilde{P}_k (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \mu_k) \\
& + \frac{\theta^2}{\theta_{R^B}^2}(y^B H_k - H_k^B \mu_k)^T (R_k^B)^{-1} H_k^B \tilde{P}_k (H_k^B)^T (R_k^B)^{-1} (y_k^B - H_k^B \mu_k).
\end{aligned}
\tag{A.35}
$$

Equation (A.18) can be rewritten as follows:

$$
\begin{aligned}
\tilde{P}_k^{-1} &= (P_k[I - \theta \bar{S}_k P_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T (R_k^A)^{-1} H_k^A P_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T (R_k^B)^{-1} H_k^B P_k]^{-1})^{-1} \\
&= [I - \theta \bar{S}_k P_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T (R_k^A)^{-1} H_k^A P_k + \frac{\theta}{\theta_{R^B}}(H_k^B)^T (R_k^B)^{-1} H_k^B P_k] P_k^{-1} \\
&= [P_k^{-1} - \theta \bar{S}_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T (R_k^A)^{-1} H_k^A + \frac{\theta}{\theta_{R^B}}(H_k^B)^T (R_k^B)^{-1} H_k^B] P_k P_k^{-1} \\
&= P_k^{-1}[I - \theta P_k \bar{S}_k + \frac{\theta}{\theta_{R^A}} P_k (H_k^A)^T (R_k^A)^{-1} H_k^A + \frac{\theta}{\theta_{R^B}} P_k (H_k^B)^T (R_k^B)^{-1} H_k^B].
\end{aligned}
$$

We therefore see that:

$$
\begin{aligned}
\lambda_k^T P_k \tilde{P}_k^{-1} P_k \lambda_k &= \lambda_k^T [I - \theta P_k \bar{S}_k + \frac{\theta}{\theta_{R^A}} P_k (H_k^A)^T (R_k^A)^{-1} H_k^A + \frac{\theta}{\theta_{R^B}} P_k (H_k^B)^T (R_k^B)^{-1} H_k^B] P_k \lambda_k \\
&= \lambda_k^T P_k \lambda_k - \theta \lambda_k^T P_k \bar{S}_k P_k \lambda_k + \frac{\theta}{\theta_{R^A}} \lambda_k^T P_k (H_k^A)^T (R_k^A)^{-1} H_k^A P_k \lambda_k \\
&\quad + \frac{\theta}{\theta_{R^B}} \lambda_k^T P_k (H_k^B)^T (R_k^B)^{-1} H_k^B P_k \lambda_k.
\end{aligned}
\tag{A.36}
$$

Substituting Equation (A.36) into Equation (A.35) gives:

$$
\begin{aligned}
&\lambda_{k+1}^T(P_{k+1}-Q_k)\lambda_{k+1} \\
&\quad= \lambda_k^T P_k\lambda_k - \theta\lambda_k^T P_k\bar{S}_k P_k\lambda_k + \frac{\theta}{\theta_{R^A}}\lambda_k^T P_k(H_k^A)^T(R_k^A)^{-1}H_k^A P_k\lambda_k + \frac{\theta}{\theta_{R^B}}\lambda_k^T P_k(H_k^B)^T(R_k^B)^{-1}H_k^B P_k\lambda_k \\
&\qquad - 2\theta(\mu_k-\hat{x}_k)^T\bar{S}_k P_k\lambda_k - \frac{2\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A P_k\lambda_k - \frac{2\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B P_k\lambda_k \\
&\qquad + \theta^2(\mu_k-\hat{x}_k)^T\bar{S}_k\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) + \frac{2\theta^2}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) \\
&\qquad + \frac{2\theta^2}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) \\
&\qquad + \frac{\theta^2}{\theta_{R^A}^2}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) \\
&\qquad + \frac{2\theta^2}{\theta_{R^A}\theta_{R^B}}(y^B H_k - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) \\
&\qquad + \frac{\theta^2}{\theta_{R^B}^2}(y^B H_k - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^B)^T(R_k^B)^{-1}(y_k^B - H_k^B\mu_k).
\end{aligned}
\tag{A.37}
$$

We substitute Equation (A.37) into Equation (A.30). In the interest of brevity, we will immediately collect the terms and leave out the terms that are cancelled out:

$$
\begin{aligned}
J = \sum_{k=0}^{k-1}&[(\mu_k-\hat{x}_k)^T(\bar{S}_k + \theta\bar{S}_k\tilde{P}_k\bar{S}_k)(\mu_k-\hat{x}_k) \\
&+ \frac{2\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) + \frac{2\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) \\
&+ \frac{1}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T\left(\frac{\theta}{\theta_{R^A}}(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1} - (R_k^A)^{-1}\right)(y_k^A - H_k^A\mu_k) \\
&+ \frac{1}{\theta_{R^B}}(y^B H_k - H_k^B\mu_k)^T\left(\frac{\theta}{\theta_{R^B}}(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^B)^T(R_k^B)^{-1} - (R_k^B)^{-1}\right)(y_k^B - H_k^B\mu_k) \\
&+ \frac{2\theta}{\theta_{R^A}\theta_{R^B}}(y^B H_k - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k)].
\end{aligned}
$$

A.1.3. Stationary points with respect to $\hat{x}_k$, y_k^A and y_k^B

We intent to find a stationary point with respect to $\hat{x}_k$, y_k^A and y_k^B. So we will take the partial derivative of J with respect to $\hat{x}_k$, y_k^A and y_k^B and set them to 0. This gives us the following:

$$
\begin{aligned}
\frac{\partial J}{\partial \hat{x}_k} &= -2(\bar{S}_k + \theta\bar{S}_k\tilde{P}_k\bar{S}_k)(\mu_k-\hat{x}_k) - \frac{2\theta}{\theta_{R^A}}(y_k^A - H_k^A\mu_k)^T(R_k^A)^{-1}H_k^A\tilde{P}_k\bar{S}_k \\
&\quad - \frac{2\theta}{\theta_{R^B}}(y_k^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k\bar{S}_k = 0, \\
\frac{\partial J}{\partial y_k^A} &= \frac{2\theta}{\theta_{R^A}}(R_k^A)^{-1}H_k^A\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) + \frac{2}{\theta_{R^A}}\left(\frac{\theta}{\theta_{R^A}}(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1} - (R_k^A)^{-1}\right)(y_k^A - H_k^A\mu_k) \\
&\quad + \frac{2\theta}{\theta_{R^A}\theta_{R^B}}(y^B - H_k^B\mu_k)^T(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^A)^T(R_k^A)^{-1} = 0, \\
\frac{\partial J}{\partial y_k^B} &= \frac{2\theta}{\theta_{R^B}}(R_k^B)^{-1}H_k^B\tilde{P}_k\bar{S}_k(\mu_k-\hat{x}_k) + \frac{2}{\theta_{R^B}}\left(\frac{\theta}{\theta_{R^B}}(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^B)^T(R_k^B)^{-1} - (R_k^B)^{-1}\right)(y_k^B - H_k^B\mu_k) \\
&\quad + \frac{2\theta}{\theta_{R^A}\theta_{R^B}}(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^A)^T(R_k^A)^{-1}(y_k^A - H_k^A\mu_k) = 0.
\end{aligned}
$$

These equations are satisfied when:

$$\begin{aligned}\hat{x}_k &= \mu_k,\\ y_k^A &= H_k^A\mu_k,\\ y_k^B &= H_k^B\mu_k.\end{aligned} \tag{A.38}$$

A.1.4. Type of stationary points

We of course want that $\hat{x}_k$ minimises J and that y_k^A and y_k^B maximise J. So we want the second partial derivative of J with respect to $\hat{x}_k$ to be positive and the second partial derivative of J with respect to y_k^A and y_k^B to be negative. We calculate the second partial derivatives:

$$\begin{aligned}\frac{\partial^2 J}{\partial \hat{x}_k^2} &= 2(\bar{S}_k + \theta\bar{S}_k\tilde{P}_k\bar{S}_k) \succ 0,\\ \frac{\partial^2 J}{\partial (y_k^A)^2} &= \frac{2}{\theta_{R^A}}\left(\frac{\theta}{\theta_{R^A}}(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1} - (R_k^A)^{-1}\right) \prec 0,\\ \frac{\partial^2 J}{\partial (y_k^B)^2} &= \frac{2}{\theta_{R^B}}\left(\frac{\theta}{\theta_{R^B}}(R_k^B)^{-1}H_k^B\tilde{P}_k(H_k^B)^T(R_k^B)^{-1} - (R_k^B)^{-1}\right) \prec 0.\end{aligned}$$

Let us first analyse the second partial derivative with respect to $\hat{x}_k$. Note that $\bar{S}_k$ is chosen such that it is positive definite. So we conclude that $\frac{\partial^2 J}{\partial \hat{x}_k^2}$ is positive definite if and only if $\tilde{P}_k$ is positive definite. Thus we find:

$$\begin{aligned}&[P_k^{-1} - \theta\bar{S}_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^B]^{-1} \succ 0,\\ \Leftrightarrow\quad &P_k^{-1} - \theta\bar{S}_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^B \succ 0.\end{aligned}$$

We continue with analysing the second partial derivative with respect to y_k^A. Thus:

$$\begin{aligned}&\frac{2}{\theta_{R^A}}\left(\frac{\theta}{\theta_{R^A}}(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1} - (R_k^A)^{-1}\right) \prec 0,\\ \Leftrightarrow\quad &(R_k^A)^{-1} - \frac{\theta}{\theta_{R^A}}(R_k^A)^{-1}H_k^A\tilde{P}_k(H_k^A)^T(R_k^A)^{-1} \succ 0,\\ \Leftrightarrow\quad &(R_k^A)^{-1}\left(R_k^A - \frac{\theta}{\theta_{R^A}}H_k^A\tilde{P}_k(H_k^A)^T\right)(R_k^A)^{-1} \succ 0,\\ \Leftrightarrow\quad &R_k^A - \frac{\theta}{\theta_{R^A}}H_k^A\tilde{P}_k(H_k^A)^T \succ 0.\end{aligned}$$

Of course we find a similar expression as a result of setting the second partial derivative with respect to y_k^B to be negative definite. So in conclusion the following need to be satisfied for every time step:

$$\begin{aligned}P_k^{-1} - \theta\bar{S}_k + \frac{\theta}{\theta_{R^A}}(H_k^A)^T(R_k^A)^{-1}H_k^A + \frac{\theta}{\theta_{R^B}}(H_k^B)^T(R_k^B)^{-1}H_k^B &\succ 0,\\ R_k^A - \frac{\theta}{\theta_{R^A}}H_k^A\tilde{P}_k(H_k^A)^T &\succ 0,\\ R_k^B - \frac{\theta}{\theta_{R^B}}H_k^B\tilde{P}_k(H_k^B)^T &\succ 0.\end{aligned}$$

In conclusion, using Equation (A.38) in the conditions in Equation (A.20e) and Equation (A.20g), we find:

$$P_{k+1} = F_k P_k \Big[I - \theta \bar{S}_k P_k + \frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} H_k^A P_k + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} H_k^B P_k \Big]^{-1} F_k^T + Q_k,$$

$$\hat{x}_{k+1} = F_k \hat{x}_k + F_k P_k \Big[I - \theta \bar{S}_k P_k + \frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} H_k^A P_k + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} H_k^B P_k \Big]^{-1} \times \Big[\frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \hat{x}_k) + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} (y_k^B - H_k^B \hat{x}_k) \Big],$$

with $\bar{S}_k = L_k^T S_k L_k$. This can be rewritten as:

$$\begin{aligned} P_{k+1} &= F_k \tilde{P}_k F_k^T + Q_k, \\ \tilde{P}_k &= P_k \Big[I - \theta \bar{S}_k P_k + \frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} H_k^A P_k + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} H_k^B P_k \Big]^{-1} \\ \hat{x}_{k+1} &= F_k \hat{x}_k + F_k \tilde{P}_k \Big[\frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \hat{x}_k) + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} (y_k^B - H_k^B \hat{x}_k) \Big]. \end{aligned}$$

Choosing $P_{k+1} = P_{k+1|k}$; $\tilde{P}_k = P_{k|k}$; $\hat{x}_{k+1} = \hat{x}_{k|k}$ gives us the H_∞ filter:

Prediction step:

$$\begin{aligned} \bar{S}_k &= L_k^T S_k L_k, \\ \hat{x}_{k|k-1} &= F_{k-1} \hat{x}_{k-1|k-1}, \\ P_{k|k-1} &= F_{k-1} P_{k-1|k-1} F_{k-1} + Q_{k-1}. \end{aligned}$$

Update step:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (y_k - H_k \hat{x}_{k|k-1}) \\ &= \hat{x}_{k|k-1} + P_{k|k} \Big(\frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} (y_k^A - H_k^A \hat{x}_{k|k-1}) + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} (y_k^B - H_k^B \hat{x}_{k|k-1}) \Big), \\ P_{k|k} &= (P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k)^{-1} \\ &= \Big(P_{k|k-1}^{-1} - \theta \bar{S}_k + \frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} H_k^A + \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} H_k^B \Big)^{-1}, \\ K_k &= P_{k|k} H_k^T R_k^{-1} \\ &= P_{k|k} \begin{bmatrix} \frac{\theta}{\theta_{R^A}} (H_k^A)^T (R_k^A)^{-1} \\ \frac{\theta}{\theta_{R^B}} (H_k^B)^T (R_k^B)^{-1} \end{bmatrix}^T, \end{aligned}$$

provided that:

$$\begin{aligned} P_{k|k-1}^{-1} - \theta \bar{S}_k + H_k^T R_k^{-1} H_k &\succ 0, \\ R_k^A - \frac{\theta}{\theta_{R^A}} H_k^A P_{k|k} (H_k^A)^T &\succ 0, \\ R_k^B - \frac{\theta}{\theta_{R^B}} H_k^B P_{k|k} (H_k^B)^T &\succ 0. \end{aligned}$$

□