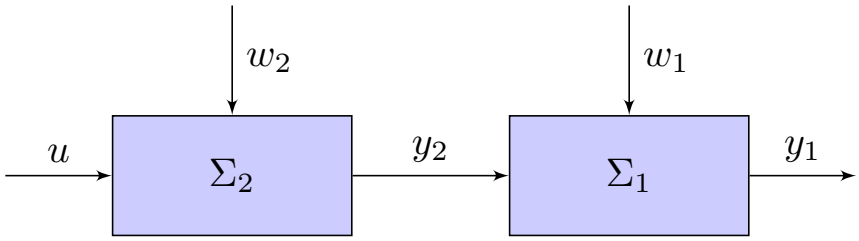
UPPSALA UNIVERSITET

Department of signals and systems

Distributed Kalman Filtering and Optimal control with packet-loss



By

P. Wijnbergen

Supervisor: Dr. S. Knorn
Second supervisor: Dr. ir. B. Besselink

June 2018

Abstract

In this report the optimal control problem with packet drop-out is investigated. First the Kalman filter is analyzed and simulations are done on different types of state estimation for a cascaded system. We made a distinction between local and global estimation, where local refers to using multiple outputs for the Kalman filter process and global to using only one output. In a similar fashion the construction of the optimal controller for stochastic systems is analyzed and simulations are done on an optimal controller for a cascaded system. We simulated that a part of the state is arrived at the controller with different arrival probabilities and relate the dependence of the controller performance to this probability.

Acknowledgements

First of all I would like to thank my supervisor Steffi Knorn for hosting this project at the Uppsala University. It has been a great experience and without you it would not have been possible. I would also like to thank you for your advice and supervision. I learned a lot during my stay in Uppsala.

Secondly I would like to thank Bart Besselink for introducing me to Steffi. Without you this internship would not have been possible.

Thank you Steffi and Bart.

Paul

Contents

1 Introduction

Wireless sensor technology is of growing interest for process and automation industry. The driving force behind using wireless technology in monitoring and control applications is its lower deployment and reconfiguration cost. Furthermore, wireless devices can be positioned where wires cannot go, or where there is no steady electricity supply, for transmitters can use energy from a possibly rechargeable battery or a local source like a solar cell.

In classical wired communication systems the probability of information getting lost is very low and there are many ways to minimize the influence of external noise sources. This is in contrast to wireless communication technology where information loss is much more probable due to a lack of energy for transmittance, data corruption or external electric fields. Besides this, the external noise is more prominent and very difficult if not impossible to reduce.

Earlier research on wireless communication systems was done in [1] [2]. This research was followed up by a stability analysis in [3] and and extension to packet loss under energy harvesting constraints in [4] [5]. The research done so far concentrates on the communication of the state estimate and control of a single system. With the increase in computational power and the renewed interest in complex systems, the concept of wireless communication might be extended to a network of systems. Computer networks or multi-agent power grids are only two out of many applications of networks of systems. Throughout this report we will be mainly interested in a network consisting of two systems. In particular we consider cascaded systems, which are systems where the output of one system acts as input to the second system. These systems occur regularly in practice and examples are often systems representing physical phenomena. One can think of a water tank, whose level is controlled by a pump, or even two vehicles following each other with a specified distance.

The main difference with respect to the optimal control problem as it is defined for a single system, is that we have access to information from different sensors in a cascaded setting. This means that we are able to estimate states from different sensors and that gives rise to the question of how to get the optimal estimates. This is also known as the distributed Kalman filtering problem. The main goal of this research is to get some insights in this problem with respect to cascaded systems and show that it is not always obvious from which sensors one should estimate the states. Furthermore, we aim to establish some results on the performance requirements of the wireless communication system with packet dropout.

In order to do so, we will start by giving an introduction to the Kalman filter problem for a single system in the next section. Once this is fully understood, we will compare two cases to estimate the states of a cascaded system in Section 3. By means of a simulation study we will show that it is not straightforward which method of estimation is optimal. In the section thereafter, Section 4, we will investigate the optimal control problem for stochastic systems. As it will turn out, the separation principle with respect to state estimation and actuation also holds for stochastic systems. Finally, in Section 5, we will perform a simulation study where we will investigate the influence of packet dropout on the controller performance.

2 Kalman Filtering

2.1 Filter construction

As mentioned above, we will start by introducing the Kalman filter for state estimation of a linear system. Let us first define our system Σ, from which we desire to estimate the state, as follows

$$\Sigma = \begin{cases} x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + v_k, \end{cases} \tag{2.1}$$

where $x_k \in \mathbb{R}^n$ is the state vector at a discrete time step k, $u_k \in \mathbb{R}^m$ an input function, $y_k \in \mathbb{R}^p$ an output function and $A : \mathcal{X} \to \mathcal{X}$, $B : \mathcal{U} \to \mathcal{X}$ and $C : \mathcal{X} \to \mathcal{Y}$ linear maps of appropriate dimension. Here, w_k and v_k are the process and measurements noise vectors respectively, which are both assumed to be i.i.d. Gaussian with zero mean and covariances $W = \mathbb{E}\{w_k w_k^T\} \geq 0$ and $R = \mathbb{E}\{v_k v_k^T\} > 0$. $\mathbb{E}\{\cdot\}$ denotes the expected value. The initial state x_0 is also Gaussian with mean $\bar{x}_0$ and covariance P_0.

Due to the process and measurement noise, the state of the system becomes a stochastic variable. This implies that it is impossible to have an observe that generates the state with full certainty. This means that we desire to have an observer that generates the expected value of the state $\hat{x}_k = \mathbb{E}\{x_k\}$. The Kalman filter problem deals with finding such an observer, such that the error of the estimate is minimized. To be more specific, if we define the error of the estimate to be $e_k = x_k - \hat{x}_k$, we would like to minimize the expected value of the square of the norm of e_k, i.e. $\mathbb{E}\{||e_k||^2\}$. As it turns out, minimizing this norm is equivalent to minimizing the trace of the error covariance matrix. To see this, consider the following equation

$$\begin{aligned}\mathbb{E}\{||e_k||^2\} &= \mathbb{E}\{e_k^T e_k\},\\ &= \mathbb{E}\{\operatorname{tr} e_k e_k^T\},\\ &= \operatorname{tr}\mathbb{E}\{e_k e_k^T\}.\end{aligned} \tag{2.2}$$

The key concepts in the Kalman filtering process are prediction and correction. The main idea is to first use the knowledge of the system dynamics to predict the next state based on the previous estimate. This will be influenced by the process noise. Therefore, secondly, we use the output y to correct the predicted state. In order to differentiate between the predicted state and the corrected state, the predicted state is denoted $\hat{x}_{k|k-1}$ and the corrected state $\hat{x}_{k|k}$. So in the Kalman filtering process we first make a prediction on the state given the system dynamics and a previous estimate $x_{k-1|k-1}$:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1}, \tag{2.3}$$

The prediction is then corrected using the difference between the measurement and the expected measurement $C\hat{x}_{k|k-1}$. This yields

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}). \tag{2.4}$$

Combining equations (2.3) and (2.4), we can recognize the structure of a state observer. The remaining question is, however, how we should choose our matrix K_k such that the estimation error is minimized. To this extend we consider the error $e_k = x_k - \hat{x}_{k|k}$. Denote the covariance of the error

$$\mathbb{E}\{e_k e_k^T\} = \mathbb{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T\} := P_{k|k}. \tag{2.5}$$

Given this structure an expression for the update equation of $P_{k|k}$ in terms of K_k and $P_{k-1|k-1}$ can be constructed. Based on this we can calculate how to choose K_k. If we substitute equation (2.4) in equation (2.5) we see that

$$\begin{aligned}P_{k|k} &= \mathbb{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T\},\\ &= \mathbb{E}\{(x_k - \hat{x}_{k|k-1} - K_k(y_k - C\hat{x}_{k|k-1}))(x_k - \hat{x}_{k|k-1} - K_k(y_k - C\hat{x}_{k|k-1}))^T\},\\ &= \mathbb{E}\{(x_k - \hat{x}_{k|k-1} - K_k(Cx_k + v_k - C\hat{x}_{k|k-1}))(x_k - \hat{x}_{k|k-1} - K_k(Cx_k + v_k - C\hat{x}_{k|k-1}))^T\},\\ &= \mathbb{E}\{((I - K_kC)(x_k - \hat{x}_{k|k-1}) + K_kv_k)((I - K_kC)(x_k - \hat{x}_{k|k-1}) + K_kv_k)^T\}.\end{aligned} \tag{2.6}$$

Note that $(x_k - \hat{x}_{k|k-1})$ is the error of the prior estimate before the correction has been applied. This term is clearly uncorrelated to the measurement noise and hence we can rewrite equation (2.6) as

$$\begin{aligned}P_{k|k} &= (I - K_kC)P_{k|k-1}(I - K_kC)^T + K_kRK_k^T\\ &= P_{k|k-1} - K_kCP_{k|k-1} - P_{k|k-1}C^TK_k^T + K_k(CP_{k|k-1}C^T + R)K_k^T\end{aligned} \tag{2.7}$$

As mentioned earlier, minimizing the error of the estimate, is equivalent with minimizing the trace of the error covariance matrix. Since we have derived a full expression for the error covariance matrix, we can minimize its trace. To do so, we need to calculate the gradient of the trace of $P_{k+1|k+1}$ from equation (2.7) with respect to the coefficients of K_k and set it equal to zero, i.e.

$$\frac{\partial(\operatorname{tr} P_{k|k})}{\partial K_k} = -2CP_{k|k-1} + 2(CP_{k|k-1}C^T + R)K_k^T = 0, \tag{2.8}$$

which leads to the solution for K_k

$$K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}. \tag{2.9}$$

Substituting K_k back into equation (2.7) and rewriting some terms leads to the Riccati difference equation as an update equation for $P_{k+1|k+1}$:

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}. \tag{2.10}$$

If we then consider the prior estimation error of the next step somewhat closer and take into account that the prior estimation error is also not correlated to the process noise we can calculate

$$\begin{aligned} P_{k+1|k} &= \mathbb{E}\{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^T\} \\ &= \mathbb{E}\{(A(x_k - \hat{x}_{k|k}) + w_k)(A(x_k - \hat{x}_{k|k}) + w_k)^T\}, \\ &= AP_{k|k}A^T + W. \end{aligned} \tag{2.11}$$

By plugging in (2.10) we see that

$$P_{k+1|k} = AP_{k|k-1}A^T - AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T + W. \tag{2.12}$$

We can rewrite this equation in as what will turn out, to be a very useful formulation

$$\begin{aligned} P_{k+1|k} &= AP_{k|k-1}A^T - AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T + W \\ &= AP_{k|k-1}A^T - AK_kCP_{k|k-1}A^T + W \\ &= AP_{k|k-1}A^T - AK_kCP_{k|k-1}A^T - AP_{k|k-1}C^TK_k^TA^T + AP_{k|k-1}C^TK_k^TA^T + W, \\ &= AP_{k|k-1}A^T - AK_kCP_{k|k-1}A^T - AP_{k|k-1}C^TK_k^TA^T + K_k(CP_{k|k-1}C^T + R)K_k^TA^T + W, \\ &= AP_{k|k-1}A^T - AK_kCP_{k|k-1}A^T - AP_{k|k-1}C^TK_k^TA^T + A^TK_kCP_{k|k-1}C^TK_kA^T \\ &\quad + AK_kRK_k^TA^T + W, \\ &= (A - AK_kC)P_{k|k-1}(A - AK_kC)^T + AK_kRK_k^TA^T + W, \end{aligned} \tag{2.13}$$

Hence we we have constructed to following update equation for $P_{k+1|k}$ as an alternative for (2.10)

$$P_{k+1|k} = (A - AK_kC)P_{k|k-1}(A - AK_kC)^T + AK_kRK_k^TA^T + W. \tag{2.14}$$

2.2 Filter convergence

Given this update equation the question arises what happens if $k \to \infty$. If the error covariance grows unbounded, the state estimate becomes rather useless. In order to have the Kalman filter work properly, that is, to generate an estimate with a bounded covariance, we need to make some assumptions on the system. The two assumptions that we need to do, is that the system is (C, A) detectable and $(A, W^{\frac{1}{2}})$ stabilizable. Intuitively this makes sense. The detectability assumption is also a necessary condition for the existence of an observer for a deterministic system. The stabilizability condition can be interpreted as the condition that all states are excited by the noise. With these two assumptions we can guarantee the error covariance to converge to a limit $P^* \geq 0$, even if the state of the system grows unbounded.

The idea of the proof to this statement is captured in several steps. First we show that given the observability condition, the sequence generated by equation (2.12), i.e. $\{P_{k|k-1}\}$, is monotonic and bounded for zero initial condition, i.e. $P_0 = 0$. This implies that the sequence converges and hence it follows from equation (2.10) that $\{P_{k|k}\}$ converges as well for zero initial condition. This means that K_k also converges to some K^*. It will follow from the stabilizability condition that $A - AK^*C$ is a stable matrix. With this proven, we will be able to prove the final step, which says that given the conditions, the sequence will converge for any initial condition. This method of filtering and the proofs, that we are about to see, originate from [6] and can be found in many papers and book on Kalman filtering such as [7].

In the next lemma we will prove that if the system is detectable, the error covariance will remain bounded in every step.

Lemma 1. *For all $P(0) = P_0 \geq 0$ and $P_0 < \infty$, the sequence $\{P_{k+1|k}\}$ is bounded by some $\overline{P} \geq 0$, if the system is (C, A) detectable.*

Proof. Since the system is (C, A) detectable, there exists a K such that $A - KC$ has its eigenvalues strictly within the complex unit circle. Consider a regular observer, which is a suboptimal filter

$$\hat{x}_{k+1} = A\hat{x}_k - K(C\hat{x}_k - y_k) + Bu_k. \tag{2.15}$$

The equation for the error is then given by

$$\begin{aligned} e_{k+1} &= x_{k+1} - \hat{x}_{k+1}, \\ &= (A - KC)e_k + w_k + Kv_k, \end{aligned} \tag{2.16}$$

which results in an update equation for the error covariance matrix

$$P_{k+1} = (A - KC)P_k(A - KC)^T + KRK^T + W. \tag{2.17}$$

We can rewrite this in terms of P_0 as follows:

$$P_{k+1} = (A - KC)^k P_0((A - KC)^T)^k + \sum_{n=0}^{k}(A - KC)^n(KRK^T + W)((A - KC)^T)^n. \tag{2.18}$$

By the singular value decomposition we have that, since $A - KC$ has its eigenvalues strictly within the unit circle that $(A - KC) \leq \lambda Z$ for some Z and $|\lambda| \in [0, 1)$. To see this, consider the singular value decomposition of $A - KC$, where Σ is a diagonal matrix containing the singular values σ of $A - KC$.

$$\begin{aligned} A - KC &= U\Sigma V^T, \\ &\leq U\sigma_{max} I V^T, \\ &\leq \sigma_{max} U V^T, \\ &= \lambda Z. \end{aligned} \tag{2.19}$$

Therefore the $P_{k+1|k}$ is bounded by

$$\begin{aligned} P_{k+1} &= (A - KC)^k P_0((A - KC)^T)^k + \sum_{n=0}^{k}(A - KC)^n(KRK^T + W)((A - KC)^T)^n, \\ &\leq \lambda^{2k} Z P_0 Z^T + \sum_{n=0}^{k} \lambda^{2n}(Z(KRK^T + W)Z^T). \end{aligned} \tag{2.20}$$

Since this filter is suboptimal, it follows that the sequence is also bounded for the optimal filter. □

Next we show that given an initial condition P_0, the sequence is either increasing or decreasing, i.e. monotonic.

Lemma 2. *If $P_{N+1|N} \leq P_{N|N-1}$, for some N then $P_{k+1|k} \leq P_{k|k-1}$ for all $k > N$. On the other hand if $P_{N+1|N} \geq P_{N|N-1}$, for some N then $P_{k+1|k} \geq P_{k|k-1}$ for all $k > N$*

Proof. Define the function

$$g(P_{k|k-1}, K) = (A - AKC)P_{k|k-1}(A - AKC)^T + AKRK^T A^T + W. \tag{2.21}$$

Note that g is a positive monotonic function in $P_{k|k-1}$. Also $P_{k+1|k} = \min_K g(P_{k|k-1}, K)$. Hence if $P_{k+1|k} \leq P_{k|k-1}$ we see that

$$\begin{aligned} P_{k+1|k} &= \min_K g(P_{k|k-1}, K), \\ &= g(P_{k|k-1}, K_k^*), \\ &\geq g(P_{k+1|k}, K_k^*), \\ &\geq \min_K g(P_{k+1|k}, K), \\ &= g(P_{k+1|k}, K_{k+1}^*), \\ &= P_{k+2|k+1}. \end{aligned} \tag{2.22}$$

Conversely we see that if $P_{k|k-1} \leq P_{k+1|k}$ then

$$\begin{aligned} P_{k+2|k+1} &= \min_K g(P_{k+1|k}, K), \\ &= g(P_{k+1|k}, K^*_{k+1}), \\ &\geq g(P_{k|k-1}, K^*_{k+1}), \\ &\geq \min_K g(P_{k|k-1}, K), \\ &= g(P_{k|k-1}, K^*_k), \\ &= P_{k+1|k}. \end{aligned} \tag{2.23}$$

□

With the proof that the sequence is monotonic, the next lemma is in fact a mere consequence. However, it is worth stating and proving it.

Lemma 3. *If $P_0 = 0$, then $P_{k|k}$ converges to a steady state error covariance matrix P^*.*

Proof. Since $P_0 = 0$, we have that $P_{1,0} = W$, and $P_{2,1} = (A - AK_1C)W(A - AK_1C)^T + AK_1RK_1^TA^T + W$ and hence $P_{1,0} \leq P_{2,1}$. By the previous lemma we have that $P_{k|k-1} \leq P_{k+1|k}$ for all k. We are using an optimal filter here, and hence the error covariance will be less then when a regular observer is used. Hence by Lemma 1 we have for all k that $\{P_{k|k-1}\}$ is bounded. Therefore $\{P_{k|k-1}\}$ converges and hence according to equation (2.10) we have that $P_{k|k} \to P^*$ for some P^*. □

With the previous results we have already proven that if we have an exact state estimate at a certain time step k, the uncertainty will only grow. If the detectability condition is met, we have that the error covariance will converge to a steady state value. The next lemma shows that for the steady state matrix K^* we have that $A - AK^*C$ is a stable matrix, i.e. has its eigenvalues within the complex unit circle.

Lemma 4. *Let the system be (C, A) detectable and $(A, W^{\frac{1}{2}})$ stabilizable. Denote $P^* = \lim_{k\to\infty} P_{k|k}$ and K^* as the corresponding filter gain, then $A - AK^*C$ has its eigenvalues strictly within the unit circle.*

Proof. With this stationary filter gain K^* we have that P^* is given by the Ricatti equation

$$P^* = (A - AK^*C)P^*(A - AK^*C)^T + AK^*RK^{*T}A^T + W. \tag{2.24}$$

Let x be an eigenvector of $(A - AK^*C)$ then we have that

$$\begin{aligned} x^TP^*x &= x^T((A - AK^*C)P^*(A - AK^*C)^T + AK^*RK^{*T}A^T + W)x, \\ &= |\lambda|^2x^TP^*x + x^T(AK^*RK^{T*}A^T + W)x. \end{aligned} \tag{2.25}$$

From this it follows that

$$(1 - |\lambda|^2)x^TP^*x = x^T(AK^*RK^{*T}A^T + W)x. \tag{2.26}$$

Since P^*, R and W are positive (semi)-definite, λ cannot be greater than 1. If $\lambda = 1$ we have that the following equations must hold:

a) $x^TW^{\frac{1}{2}} = 0$,

b) $x^TAK^* = 0$,

c) $x^T(A - AK^*C) = \lambda x^T$.

But a) and b) together imply that $x^TA = \lambda x^T$, i.e. $x^T(A - \lambda I) = 0$. Together with c) this means that $x^T\begin{pmatrix} A - \lambda I & W^{\frac{1}{2}} \end{pmatrix} = 0$. However, we assumed that the system was $(A, W^{\frac{1}{2}})$ stabilizable. This means that for all unstable eigenvalues of A, we assumed that $(A - \lambda I \quad W^{\frac{1}{2}})$ has rank n. Hence $x^T\begin{pmatrix} A - \lambda I & W^{\frac{1}{2}} \end{pmatrix} = 0$ if and only if $x = 0$. This means that λ cannot equal one. □

If we combine what we have proven so far, we come to the main result of Kalman filtering.

Theorem 1. *Consider the system as in* (2.1)*. If the system is* (C, A) *detectable and* $(A, W^{\frac{1}{2}})$ *stabilizable, then for any* $P_0 \geq 0$*, it holds that* $P_{k|k} \to P^*$*.*

Proof. From equation (2.14) we have that the update equation for the prior error covariance is given by $P_{k+1|k} = (A - AK_kC)P_{k|k-1}(A - AK_kC)^T + AK_kRK_k^TA^T + W$. By Lemma 3 we have that $\{P_{k,k-1}\}$ converges to some limit, which we denote Φ, for zero initial condition. From this it follows that if $P_0 = 0$:

$$\begin{aligned}\lim_{k\to\infty} P_{k|k-1} &= \lim_{k\to\infty} \sum_{n=0}^{k} (A - AK^*C)^n (AK^*R(K^*)^TA^T + W)((A - AK^*C)^T)^n, \\ &:= \Phi.\end{aligned} \tag{2.27}$$

By Lemma 4 $A - AK^*C$ has its eigenvalues strictly within the complex unit circle and by the singular value decomposition we have that that $A - AK^*C \leq \lambda Z$ for some Z and $|\lambda| \in [0, 1)$. Then for all $P_0 \geq 0$ we have that

$$\lim_{k\to\infty} (A - AK^*C)^k P_0((A - AK^*C)^T)^k \leq \lim_{k\to\infty} \lambda^{2k} Z P_0 Z^T = 0. \tag{2.28}$$

Suppose we have an arbitrary positive semi-definite initial condition and use the suboptimal steady state Kalman gain $K_k = K^*$ for all steps. Then it holds for any initial condition $P_0 \geq 0$ that

$$\begin{aligned}\lim_{k\to\infty} P_{k|k-1} &= \lim_{k\to\infty} \left((A - AK^*C)^k P_0((A - AK^*C)^T)^k + \sum_{n=0}^{k} (A - K^*C)^n (AK^*R(K^*)^TA^T + W)((A - AK^*C)^T)^n \right), \\ &= \lim_{k\to\infty} \sum_{n=0}^{k} (A - AK^*C)^n (AK^*R(K^*)^TA^T + W)((A - AK^*C)^T)^n, \\ &= \Phi.\end{aligned} \tag{2.29}$$

This shows that, if K^* is used in every step, $\{P_{k|k-1}\}$ converges to Φ, for all $P_0 \geq 0$. Since K^* is suboptimal in every step, we have that, if the optimal gain matrix K_k is used in every step, $\{P_{k|k-1}\}$ is bounded for all $P_0 \geq 0$. By Lemma 2 the sequence $\{P_{k|k-1}\}$ is also monotonic and thus it converges for any initial condition $P_0 \geq 0$. It follows from (2.10) that $\lim_{k\to\infty} P_{k|k} = P^*$ for some P^* for all $P_0 \geq 0$. □

2.3 Filter performance

Now that we have conditions on the system for the convergence of the error covariance with a Kalman filter, we will investigate the performance of a Kalman filter. The main question we would like to answer, is how we can minimize the error covariance given the system. Would it be useful to reduce the process noise if the measurement noise is really small? If one has access to two different measurements, which one is optimal to use for a Kalman filter? Whereas the first question is rather straightforward to answer, the second one is not straightforward to answer as we will show.

The answer to the question on noise reduction follows from Theorem 1. We state it as a corollary and it says in fact that any reduction on the noise, both the process and measurement noise, will result in a lower error covariance matrix.

Corollary 1. *Consider a system* $x_{k+1} = Ax_k + Bu_k + w_k$*,* $y_k = Cx_k + v_k$*, where the the covariance of the process and measurement noise is given by* W *and* R *respectively. If* W *and* R *are changed to some* $\overline{W} \leq W$ *and* $\overline{R} \leq R$*, then the estimates of the state resulting from the Kalman filter have an error covariance* $\overline{P}^* \leq P^*$*.*

Proof. Let K^* be the steady state Kalman gain resulting from R and W and $\overline{K}^*$ with respect to $\overline{R}$ and $\overline{W}$. We have that

$$\begin{aligned}
\lim_{k\to\infty} \overline{P_{k|k-1}} = \overline{\Phi} &:= \lim_{k\to\infty} \left(\min_K \left[\sum_{n=0}^{k} (A - AKC)^n (AK\overline{R}K^T A^T + \overline{W})((A - AKC)^T)^n \right] \right), \\
&= \lim_{k\to\infty} \sum_{n=0}^{k} (A - A\overline{K}^* C)^n (A\overline{K}^* \overline{R} (\overline{K}^*)^T A^T + \overline{W})((A - A\overline{K}^* C)^T)^n, \\
&\leq \lim_{k\to\infty} \sum_{n=0}^{k} (A - AK^* C)^n (AK^* \overline{R} K^{*T} A^T + \overline{W})(A - AK^* C)^{nT}, \\
&\leq \lim_{k\to\infty} \sum_{n=0}^{k} (A - AK^* C)^n (AK^* R K^{*T} A^T + W)(A - AK^* C)^{nT}, \\
&= \Phi := \lim_{k\to\infty} P_{k|k-1}.
\end{aligned} \tag{2.30}$$

Then in a similar fashion we see

$$\begin{aligned}
\overline{P^*} &= \min_K \left[(I - KC)\overline{\Phi}(I - KC)^T + KRK^T \right], \\
&\leq \min_K \left[(I - KC)\Phi(I - KC)^T + KRK^T \right], \\
&= P^*.
\end{aligned} \tag{2.31}$$

□

The second question is more difficult two answer. The filtering theory and current literature on it focuses mainly on the optimization given a certain measurement. In a multi-agent network and also in the cascaded setting as we will see later on, one might have access to multiple measurements. Therefore it is useful to see how we can how the Kalman filter performs with respect to the measurements.

One might assume, that given two measurements y_1 and y_2 with the same noise and the system is observable from both y_1 and y_2, that this might lead to the same error covariance, once a Kalman filter is applied. This is however not true, as the following example will show. Consider the following system

$$\Sigma = \begin{cases} x_{k+1} &= \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} x_k + Bu_k + w_k, \\ y_{1k} &= \begin{pmatrix} 0 & 1 \end{pmatrix} x_k + v_k, \\ y_{2k} &= \begin{pmatrix} 1 & 0 \end{pmatrix} x_k + v_k, \end{cases} \tag{2.32}$$

with covariance matrices $R = W = I$. If we take the initial error covariance matrix $P(0) = 0$ and run the Kalman filter, we find the result in Figure 1.

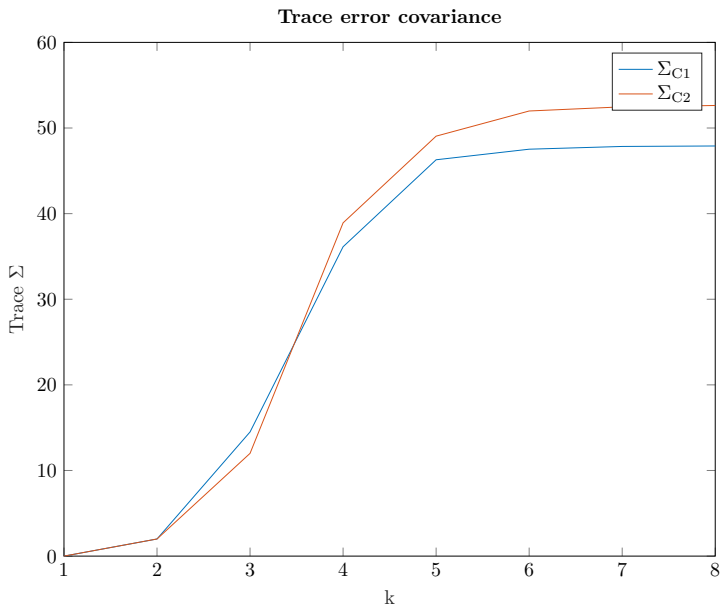


Figure 1: The comparison of the error covariance using y_1 and y_2.

As a respons to this example, one would like to formulate necessary and sufficient conditions on C_1 and C_2, such that we can tell by the system matrices how to estimate the states optimally. An intuition tells us that it might be worth while investigating what the relative influence of the noise it compared to the measurement part $C_i x_k$. Another suggestion might be to see how K_k evolves as a function of C_i. However, these subjects are not trivial. Hence we will give some more straightforward results. In order to do so we first need the next two lemma's.

Lemma 5. *Consider two positive definite matrices A and B, such that $A \leq B$. Then we have that $B^{-1} \leq A^{-1}$.*

Proof. First note that since $0 < A$ we have that $0 < AA^{-1}A$ and hence $0 < A^{-1}$. Since A and B are positive definite and $B - A$ is positive semidefinite, by the Schur complement we have that:

$$\begin{pmatrix} B & I \\ I & A^{-1} \end{pmatrix} \geq 0. \tag{2.33}$$

Since B is positive definite and hence invertible, we can take the Schur complement again to find

$$A^{-1} - B^{-1} \geq 0. \tag{2.34}$$

□

In order to prove the results we will use the next lemma.

Lemma 6. *Consider a system $x_{k+1} = Ax_k + Bu_k + w_k$ with two outputs $y_{1,k} = C_1 x_k + v_{1,k}$ and $y_{2,k} = C_2 x_k + v_{2,k}$ with $\mathbb{E}\{v_{1,k} v_{1,k}^T\} = R_1$ and $\mathbb{E}\{v_{2,k} v_{2,k}^T\} = R_2$. Let $P_{1k|k-1}$ be the error covariance of the prior estimate due to a Kalman filter using $y_{1,k}$ and let $P_{2,k|k-1}$ be the error covariance if $y_{2,k}$ is used. Denote $P^*_{C_1} = \lim_{k\to\infty} P_{1,k|k-1}$ and $P^*_{C_2} = \lim_{k\to\infty} P_{2,k|k-1}$. If $P^*_{C_1} \leq P^*_{C_2}$, then we have $\lim_{k\to\infty} P_{1,k|k} \leq \lim_{k\to\infty} P_{2,k|k}$.*

Proof. By equation (2.11) we have

$$P^*_{C_1} = \lim_{k\to\infty} AP_{1,k|k}A^T + W, \tag{2.35}$$

and

$$P^*_{C_2} = \lim_{k\to\infty} AP_{2,k|k}A^T + W. \tag{2.36}$$

By assumption we have that $P^*_{C_1} \leq P^*_{C_2}$ and thus $P^*_{C_1} - P^*_{C_2} \leq 0$. Hence we see that

$$\begin{aligned} P^*_{C_1} - P^*_{C_2} &= \lim_{k\to\infty} \left(AP_{1,k|k}A^T - AP_{2,k|k}A^T\right), \\ &= \lim_{k\to\infty} A(P_{1,k|k} - P_{2,k|k})A^T, \leq 0. \end{aligned} \tag{2.37}$$

From this it follows that

$$\lim_{k\to\infty} \left(P_{1,k|k} - P_{2,k|k}\right) = \lim_{k\to\infty} P_{1,k|k} - \lim_{k\to\infty} P_{2,k|k} \leq 0. \tag{2.38}$$

□

This last lemma means that optimizing our prior estimate, will result in a better posterior estimate. Hence to increase the performance of a Kalman filter, we can do this by optimizing the prior or the posterior estimate. Equipped with these lemma's, we can prove the next theorem.

Theorem 2. *Consider a system $x_{k+1} = Ax_k + Bu_k + w_k$ with two outputs $y_{1,k} = C_1x_k + v_{1,k}$ and $y_{2,k} = C_2x_k + v_{2,k}$ with $\mathbb{E}\{v_{1,k}v^T_{1,k}\} = \mathbb{E}\{v_{2,k}v^T_{2,k}\} = R$. Let $P_{1,k|k}$ be the error covariance of the estimate due to a Kalman filter using $y_{1,k}$ and let $P_{2,k|k}$ be the error covariance if $y_{2,k}$ is used. Denote $P^*_{C_1} = \lim_{k\to\infty} P_{1,k|k-1}$ and $P^*_{C_2} = \lim_{k\to\infty} P_{2,k|k-1}$. If*

$$C_2^T(C_2P^*_{C_1}C_2^T + R)^{-1}C_2 \leq C_1^T(C_1P^*_{C_1}C_1^T + R)^{-1}C_1, \tag{2.39}$$

then $P^*_{C_1} \leq P^*_{C_2}$.

Proof. Recall that by equation (2.12)

$$P^*_{C1} = AP^*_{C_1}A^T - AP^*_{C_1}C_1^T(C_1P^*_{C_1}C_1^T + R)^{-1}CP^*_{C_1}A^T + W \tag{2.40}$$

Hence we see

$$\begin{aligned} P^*_{C1} &= AP^*_{C_1}A^T - AP^*_{C_1}C_1^T(C_1P^*_{C_1}C_1^T + R)^{-1}C_1P^*_{C_1}A^T + W \\ &\leq AP^*_{C_1}A^T - AP^*_{C_1}C_2^T(C_2P^*_{C_1}C_2^T + R)^{-1}C_2P^*_{C_1}A^T + W, \\ &= P_{2,k+1|k}. \end{aligned} \tag{2.41}$$

By Lemma 2 $P_{2,k+1|k}$ will be increasing for all k, and hence $P^*_{C_1} \leq P^*_{C_2}$. □

The next result shows that if the noise becomes relatively smaller compared to the measurement, or differently stated, that Cx_k is amplified, the error covariance of the estimate is reduced.

Theorem 3. *If $C_2 = \alpha C_1$ for some $1 < \alpha$ then $P^*_{C_2} \leq P^*_{C_1}$.*

Proof. Recall that by equation (2.12)

$$P_{k+1|k} = AP_{k|k-1}A^T - AP_{k|k-1}C_1^T(C_1P_{k|k-1}C_1^T + R)^{-1}C_1P_{k|k-1}A^T + W. \tag{2.42}$$

From this we see that in the limit that $k \to \infty$:

$$\begin{aligned} P^*_{C_1} &= AP^*_{C_1}A^T - AP^*_{C_1}C_1^T(C_1P^*_{C_1}C_1^T + R)^{-1}C_1P^*_{C_1}A^T + W \\ &= AP^*_{C_1}A^T - AP^*_{C_1}C_2^T\frac{1}{\alpha^2}\left(\frac{1}{\alpha^2}C_2P^*_{C_1}C_2^T + R\right)^{-1}C_2P^*_{C_1}A^T + W, \\ &= AP^*_{C_1}A^T - AP^*_{C_1}C_2^T\frac{1}{\alpha^2}\left(\frac{1}{\alpha^2}(C_2P^*_{C_1}C_2^T + \alpha^2R)\right)^{-1}C_2P^*_{C_1}A^T + W, \\ &= AP^*_{C_1}A^T - AP^*_{C_1}C_2^T(C_2P^*_{C_1}C_2^T + \alpha^2R))^{-1}C_2P^*_{C_1}A^T + W, \\ &= AP^*_{C_1}A^T - AP^*_{C_1}C_2^T(C_2P^*_{C_1}C_2^T + \overline{R}))^{-1}C_2P^*_{C_1}A^T + W. \end{aligned} \tag{2.43}$$

From this we see that $P^*_{C_1}$ is the same error covariance as we would get by estimating the error covariance using $\overline{y_2} = C_2x + \overline{v_k}$, where $\overline{v_k}$ has covariance $\overline{R}$. However, since $R \leq \overline{R}$ we have by using Corollary 1, that using $y_2 = C_2x_k + v_k$ would result in a lower error covariance. □

In the case that C_1 and C_2 are invertible, we can state the following result.

Theorem 4. *Consider a system $x_{k+1} = Ax_k + Bu_k + w_k$ with two outputs $y_{1,k} = C_1x_k + v_{1,k}$ and $y_{2,k} = C_1x_k + v_{2,k}$. Assume both C_1 and C_2 are invertible and that $R = R_1 = R_2$. Then the state estimate can be estimated optimally from $y_{1,k}$ if and only if*

$$C_2R^{-1}C_2^T \leq C_1R^{-1}C_1^T \tag{2.44}$$

Proof. Consider the Riccati update equation

$$P_{k+1|k} = AP_{k|k-1}A^T - AP_{k|k-1}C_1^T(C_1P_{k|k-1}C_1^T + R)^{-1}C_1P_{k|k-1}A^T + W. \tag{2.45}$$

The second term in this equation, omitting A and A^T and denoting $(C^{-1})^T = C^{-T}$, can be rewritten as

$$\begin{aligned}
P_{k|k-1}C_1^T(C_1P_{k|k-1}C_1^T + R)^{-1}C_1P_{k|k-1} &= P_{k|k-1}C_1^T(C_1C_2^{-1}C_2P_{k|k-1}C_2^TC_2^{-T}C_1^T + R)^{-1}C_1P_{k|k-1}, \\
&= P_{k|k-1}C_1^T(C_1(C_2^{-1}C_2P_{k|k-1}C_2^TC_2^{-T} + C_1^{-1}RC_1^{-T})C_1^T)^{-1}C_1P_{k|k-1}, \\
&= P_{k|k-1}C_1^TC_1^{-T}((C_2^{-1}C_2P_{k|k-1}C_2^TC_2^{-T} + C_1^{-1}RC_1^{-T}))^{-1}C_1^{-1}C_1P_{k|k-1}, \\
&= P_{k|k-1}(C_2^{-1}(C_2P_{k|k-1}C_2^T + C_2C_1^{-1}RC_1^{-T}C_2^T)C_2^{-T})^{-1}P_{k|k-1}, \\
&= P_{k|k-1}C_2^T(C_2P_{k|k-1}C_2^T + C_2C_1^{-1}RC_1^{-T}C_2^T)^{-1}C_2P_{k|k-1}.
\end{aligned} \tag{2.46}$$

Then by Corollary 1 we have that if $C_2C_1^{-1}RC_1^{-1T}C_2^T \leq R$ then y_1 will result in a better estimate of the state. This is equivalent with

$$C_2R^{-1}C_2^T \leq C_1R^{-1}C_1^T. \tag{2.47}$$

□

3 Extension to Cascaded systems

Now that we have a solid understanding of how the Kalman filter works for a single system, we will extend the filtering problem to cascaded systems. With the extension to cascaded systems several questions arise, namely how to estimates the states. First we will define what we mean by a cascaded system more explicitly. Consider two systems of the form

$$\Sigma_i = \begin{cases} x_{i,k+1} &= A_ix_{i,k} + Bu_{i,k} + w_{i,k}, \\ y_{i,k} &= C_ix_{i,k} + v_{i,k}, \end{cases} \quad , \qquad i \in \{1, 2\}, \tag{3.1}$$

where $x_{i,k} \in \mathbb{R}^n$ is the state vector for a discrete time step k, $u_{i,k} \in \mathbb{R}^m$ an input function, $y_{i,k} \in \mathbb{R}^p$ an output function and $A_i : \mathcal{X} \to \mathcal{X}$, $B_i : \mathcal{U} \to \mathcal{X}$ and $C_i : \mathcal{X} \to \mathcal{Y}$ linear maps of appropriate dimension. The process and measurement noise is assumed to be i.i.d. Gaussian noise with both vectors with zero mean and covariance $W_i = \mathbb{E}\{w_{i,k}w_{i,k}^T\} \geq 0$ and $R_i = \mathbb{E}\{v_{i,k}v_{i,k}^T\} > 0$, respectively. The initial state $x_{i,0}$ is also Gaussian with mean $\bar{x}_{i,0}$ and covariance $P_{i,0}$. Furthermore, it is assumed that (A_i, B_i) and $(A_i, W_i^{\frac{1}{2}})$ are stabilizable and (C_i, A_i) is detectable for $i \in \{1, 2\}$.

A cascaded system is defined as the interconnection of two of these systems, where the output of the first system serves as an input of the second system, such that $y_2 = u_1$. This is a cascaded system and a block diagram is shown in Figure 2.

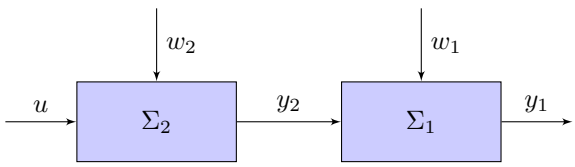


Figure 2: A cascaded system

The interconnection of these two systems can be modelled as follows

$$\Sigma_1 \times \Sigma_2 = \begin{cases} x_{k+1} = & \begin{pmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{pmatrix} x_k + \begin{pmatrix} 0 \\ B_2 \end{pmatrix} u_k + w_k, \\ & = Ax_k + Bu_k + w_k, \\ y_{1,k} & = \begin{pmatrix} C_1 & 0 \end{pmatrix} x_k + v_{1k}, \\ y_{2,k} & = \begin{pmatrix} 0 & C_2 \end{pmatrix} x_k + v_{2k}, \end{cases} \tag{3.2}$$

where $x_k = \begin{pmatrix} x_{1,k} \\ x_{2,k} \end{pmatrix}$ and $w_k = \begin{pmatrix} w_{1,k} \\ w_{2,k} \end{pmatrix}$ and the corresponding (error) covariance matrices are given by

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{pmatrix}, \qquad \text{and} \qquad W = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}, \tag{3.3}$$

respectively, where we omitted the dependence on k for clarity reasons. By P we mean however $P_{k|k-1}$.

3.1 Detectability and Stabilizability

As we have seen that detectability is an important notion for the existence of an optimal filter, we start by investigating the detectability of a cascaded system, consisting of two detectable subsystems. We will prove a stronger result, namely on observability of a cascaded system with observable subsystems. As it will be shown, since both system Σ_1 and Σ_2 are observable from y_1 and y_2, respectively, and $B_1C_2 \neq 0$ it follows that $\Sigma_1 \times \Sigma_2$ is also observable from y_2. By proving this result, the case where the systems are detectable instead of observable is guaranteed as well.

Theorem 5. *Consider two systems of the form of* (3.1). $\Sigma_1 \times \Sigma_2$ *as in* (3.2) *is observable from* y_1 *if and only if* Σ_1 *and* Σ_2 *are observable from* y_1 *and* y_2 *respectively and* $B_1 \neq 0$.

Proof. ($\Rightarrow$) Assume that the cascaded system $\Sigma_1 \times \Sigma_2$ is observable. Then we have for all complex λ

$$rank \begin{pmatrix} A_1 - \lambda I & B_1C_2 \\ 0 & A_2 - \lambda I \\ C_1 & 0 \end{pmatrix} = n_1 + n_2. \tag{3.4}$$

Then by Fact 2.11.8 in [8] it holds that

$$rank \begin{pmatrix} A_1 - \lambda I & B_1C_2 \\ 0 & A_2 - \lambda I \\ C_1 & 0 \end{pmatrix} \leq rank \begin{pmatrix} A_1 - \lambda I \\ C_1 \end{pmatrix} + rank \begin{pmatrix} B_1C_2 \\ A_2 - \lambda I \end{pmatrix}. \tag{3.5}$$

Since the rank of $\begin{pmatrix} A_1 - \lambda I \\ C_1 \end{pmatrix}$ and $\begin{pmatrix} B_1C_2 \\ A_2 - \lambda I \end{pmatrix}$ are at most n_1 and n_2 respectively, it holds that if $\Sigma_1 \times \Sigma_2$ is observable, these matrices have maximum rank for all $\lambda \in \mathbb{C}$. So we can conclude that (C_1, A_1) is observable.

To see that (C_2, A_2) is observable as well, note that by Corollary 2.5.10 in [8]

$$\begin{aligned} rank \begin{pmatrix} B_1C_2 \\ A_2 - \lambda I \end{pmatrix} &= rank \left(\begin{pmatrix} B_1 & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} C_2 \\ A_2 - \lambda I \end{pmatrix} \right), \\ &\leq \min \left(rank \begin{pmatrix} B_1 & 0 \\ 0 & I \end{pmatrix}, rank \begin{pmatrix} C_2 \\ A_2 - \lambda I \end{pmatrix} \right). \end{aligned} \tag{3.6}$$

Hence the rank of $\begin{pmatrix} C_2 \\ A_2 - \lambda I \end{pmatrix}$ is at least n_2 and (C_2, A_2) is thus observable as well.

($\Leftarrow$) Now assume that (C_1, A_1) and (C_2, A_2) are observable. Since Σ_1 and Σ_2 are observable from y_1 and y_2, we have for all $\lambda \in \mathbb{C}$ that

$$rank \begin{pmatrix} C_1 \\ A_1 - \lambda I \end{pmatrix} = n_1, \qquad rank \begin{pmatrix} C_2 \\ A_2 - \lambda I \end{pmatrix} = n_2. \tag{3.7}$$

If we consider the following we see that

$$rank\begin{pmatrix} A_1-\lambda I & B_1C_2 \\ 0 & A_2-\lambda I \\ C_1 & 0 \end{pmatrix} = rank\left(\begin{pmatrix} A_1-\lambda I & B_1 & 0 \\ 0 & 0 & I \\ C_1 & 0 & 0 \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & C_2 \\ 0 & A_2-\lambda I \end{pmatrix}\right), \tag{3.8}$$

Since we have that (C_2, A_2) is observable, we have that for all λ

$$rank\begin{pmatrix} I & 0 \\ 0 & C_2 \\ 0 & A_2-\lambda I \end{pmatrix} = n_1+n_2. \tag{3.9}$$

Since this matrix is of full rank we have

$$\begin{aligned} rank\begin{pmatrix} A_1-\lambda I & B_1C_2 \\ 0 & A_2-\lambda I \\ C_1 & 0 \end{pmatrix} &= rank\left(\begin{pmatrix} A_1-\lambda I & B_1 & 0 \\ 0 & 0 & I \\ C_1 & 0 & 0 \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & C_2 \\ 0 & A_2-\lambda I \end{pmatrix}\right), \\ &= rank\begin{pmatrix} A_1-\lambda I & B_1 & 0 \\ 0 & 0 & I \\ C_1 & 0 & 0 \end{pmatrix}, \\ &= n_1+n_2. \end{aligned} \tag{3.10}$$

□

A second condition for the convergence of the Kalman filter is the stabilizability question with respect to the covariance noise of the process noise. To ascertain ourselves that this won't pose a problem, we proof the following theorem.

Theorem 6. *Consider two systems of the form of* (3.1). $\Sigma_1 \times \Sigma_2$ *is* $(A, W^{\frac{1}{2}})$ *stabilizable if and only if* Σ_1 *and* Σ_2 *are* $(A_1, W_1^{\frac{1}{2}})$ *and* $(A_2, W_2^{\frac{1}{2}})$ *respectively.*

Proof. We have that Σ is $(A, W^{\frac{1}{2}})$ stabilizable if and only if for all unstable eigenvalues λ

$$rank\begin{pmatrix} A_1-\lambda I & B_1C_2 & W_1^{\frac{1}{2}} & 0 \\ 0 & A_2-\lambda I & 0 & W_2^{\frac{1}{2}} \end{pmatrix} = n_1+n_2. \tag{3.11}$$

Premultiplying this matrix with some vector $\begin{pmatrix} x^T & y^T \end{pmatrix}$ yields

$$\begin{pmatrix} x^T & y^T \end{pmatrix}\begin{pmatrix} A_1-\lambda I & B_1C_2 & W_1^{\frac{1}{2}} & 0 \\ 0 & A_2-\lambda I & 0 & W_2^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} x^T(A_1-\lambda I) & x^TB_1C_2+y^T(A_2-\lambda I) & x^TW_1^{\frac{1}{2}} & y^TW_2^{\frac{1}{2}} \end{pmatrix} \tag{3.12}$$

If $(A_1, W_1^{\frac{1}{2}})$ is stabilizable, then for all unstable eigenvalues of A_1 we have that $x^T\begin{pmatrix} A_1-\lambda I & W^{\frac{1}{2}} \end{pmatrix} = 0$ if and only if $x^T = 0$, see [9]. If $(A_2, W_2^{\frac{1}{2}})$ is stabilizable as well and $x^T = 0$, then (3.12) is zero if and only if $y^T = 0$ too, and thus $(A, W^{\frac{1}{2}})$ is stabilizable. Conversely, assume $(A, W^{\frac{1}{2}})$ is stabilizable. Then we have

$$n_1+n_2 \leq rank\begin{pmatrix} A_1-\lambda I & W_1^{\frac{1}{2}} \end{pmatrix} + rank\begin{pmatrix} A_2-\lambda I & W_2^{\frac{1}{2}} \end{pmatrix}, \tag{3.13}$$

which implies the stabilizability of $(A_1, W_1^{\frac{1}{2}})$ and $(A_2, W_2^{\frac{1}{2}})$. □

3.2 Local and Global estimation

The previous theorems prove that if we are dealing with a cascaded system as in (3.2) and there exists an optimal filter based on the measurement y_1, there also exist optimal estimators of the states of the two subsystems based on y_1 and y_2. This poses the problem of how to estimate the states optimally. We can distinguish two cases

Case 1: Local estimation The Kalman filter is applied after the output y_1 and y_2. The update equations for the error covariance matrices would be given by

$$P_{1,k+1|k} = A_1 P_{1,k|k-1} A_1^T - A_1 P_{1,k|k-1} C_1 (C_1 P_{1,k|k-1} C_1^T + R_1)^{-1} C_1^T P_{1,k|k-1} A_1^T + W_1, \qquad P_{1,k=0} = P_{1,0}. \tag{3.14}$$

$$P_{2,k+1|k} = A_2 P_{2,k|k-1} A_2^T - A_2 P_{2,k|k-1} C_2 (C_2 P_{2,k|k-1} C_2^T + R_2)^{-1} C_2^T P_{2,k|k-1} A_{22}^T + W_2, \qquad P_{2,k=0} = P_{2,0}. \tag{3.15}$$

In this way the states of the subsystems are estimated locally. For the estimation of the states of Σ_1, one could however argue, that it is not reasonable to have perfect information about the input signal $C_2 x_{2k}$ for Σ_1. It is for a Kalman filter however necessary to have information on the input signal to the system. Therefore it is suggested to use the measurement y_2 as an input for the Kalman filter. This results in the following Kalman prior estimate:

$$\begin{aligned} \hat{x}_{1,k|k-1} &= A_1 \hat{x}_{k-1|k-1} + B_1 (C_2 x_{2,k-1} + v_{2,k-1}), \\ &= A_1 \hat{x}_{k-1|k-1} + B_1 C_2 x_{2,k-1} + B_1 v_{2,k-1}, \end{aligned} \tag{3.16}$$

and posterior update

$$\hat{x}_{1,k|k} = \hat{x}_{1,k|k-1} - K_k (C_1 \hat{x}_{1,k|k-1} - y_{1,k}). \tag{3.17}$$

From the prior and posterior estimate we can see that the Kalman filter gain K_k will be influenced by this noise in the input. The exact influence of this noise on the filter gain is, however, beyond the scope of this thesis. We modeled a penalty for this noise by increasing the process noise. This yields that we model system 1 as

$$\Sigma_s = \begin{cases} x_{1,k+1} &= A x_{1,k} + B_1 u_k + \bar{w}_{1,k}, \\ y_{1,k} &= C_1 x_{1,k} + v_{1,k}, \end{cases} \tag{3.18}$$

where $\bar{w}_{1,k}$ has covariance $W_1 + B_1 R_2 B_1^T$. The error covariance of the prior estimate will thus be updated according to

$$\begin{aligned} P_{1,k+1|k} &= A_1 P_{1,k|k-1} A_1^T - A_1 P_{1,k|k-1} C_1 (C_1 P_{1,k|k-1} C_1^T + R_1)^{-1} C_1^T P_{1,k|k-1} A_1^T + W_1 + B_1 R_2 B_1^T, \\ P_{1,k=0} &= P_{1,0}. \end{aligned} \tag{3.19}$$

Case 2: Global estimation The cascaded system is treated as one system and only the output y_1 is used and the filter is applied to jointly optimize the estimates of the states. In this case we are estimating the state of the system as given in (3.2). The update equation for the error covariance due to a Kalman filter is given by

$$P_{k+1|k} = A P_{k|k-1} A^T - A P_{k|k-1} C^T (C P_{k|k-1} C^T + R_1)^{-1} C P_{k|k-1} A^T + W, \qquad P_{k=0} = P_0, \tag{3.20}$$

were $P_0 = \begin{pmatrix} P_{11,0} & P_{12,0} \\ P_{12,0}^T & P_{22,0} \end{pmatrix}$.

From now on we will omit the dependence on k of the error covariance matrix P unless it is necessary for clarity reasons. The trace of P_{11} and P_{22} correspond to the error in the estimate of Σ_1 and Σ_2 respectively. Since eventually we are interested in minimizing these traces we are mainly interested in P_{11} and P_{22}. To make a comparison with Case 2, we will derive an explicit expression for these update equations from equation (3.20). Denote $A_{12} = B_1 C_2$ and note that we are dealing with the following matrices:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \qquad C = \begin{pmatrix} C_1 & 0 \end{pmatrix}, \qquad W = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}, \tag{3.21}$$

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{pmatrix}, \qquad R = R_1, \tag{3.22}$$

First we compute APA^T and APC^T, such that we can alos compute $APC^T(CPC^T + R)CPA^T$:

$$\begin{aligned} APA^T = &\begin{pmatrix} A_{11}(P_{11}A_{11}^T + P_{12}A_{12}^T) + A_{12}(P_{12}A_{11}^T + P_{22}A_{12}^T) & 0 \\ 0 & A_{22}P_{22}A_{22}^T \end{pmatrix} \\ &+ \begin{pmatrix} 0 & A_{11}P_{12}A_{22}^T + A_{12}P_{22}A_{22}^T \\ A_{11}P_{12}A_{22}^T + A_{12}P_{22}A_{22}^T & 0 \end{pmatrix}, \end{aligned} \tag{3.23}$$

$$APC^T = \begin{pmatrix} A_{11}P_{11}C_1^T + A_{12}P_{21}C_1^T \\ A_{22}P_{21}C_1^T \end{pmatrix}, \tag{3.24}$$

$$APC^T(CPC^T + R)^{-1}CPA^T = \begin{pmatrix} A_{11}P_{11}C_1^T + A_{12}P_{21}C_1^T \\ A_{22}P_{21}C_1^T \end{pmatrix} (C_1P_1C_1^T + R)^{-1} \begin{pmatrix} A_{11}P_{11}C_1^T + A_{12}P_{21}C_1^T \\ A_{22}P_{21}C_1^T \end{pmatrix}^T. \tag{3.25}$$

Multiplying these results in a block matrix P, where the blocks are given by

$$P_{22}(k+1) = A_{22}P_{22}A_{22}^T - A_{22}P_{21}C_1^T(C_1P_{11}C_1^T + R)^{-1}C_1P_{12}A_{22} + W_2, \tag{3.26}$$

$$\begin{aligned} P_{11}(k+1) = &A_{11}P_{11}A_{11}^T - A_{11}P_{11}C_1^T(C_1P_{11}C_1^T + R)^{-1}C_1P_{11}A_{11}^T \\ &+ A_{11}P_{12}A_{12}^T - A_{11}P_{11}C_1^T(C_1P_{11}C_1^T + R)^{-1}C_1P_{12}A_{12}^T \\ &+ A_{12}P_{12}A_{11}^T - A_{12}P_{21}C_1^T(C_1P_{11}C_1^T + R)^{-1}C_1P_{11}A_{11}^T \\ &+ A_{12}P_{22}A_{12}^T - A_{12}P_{21}C_1^T(C_1P_{11}C_1^T + R)^{-1}C_1P_{12}A_{12}^T + W_1, \end{aligned} \tag{3.27}$$

$$P_{12}(k+1) = (A_{11}P_{11} + A_{12}P_{21})A_{22}^T - (A_{11}P_{11}C_1^T + A_{12}P_{21}C_1^T)(C_1P_{11}C_1^T + R)^{-1}C_1P_{12}A_{22}^T. \tag{3.28}$$

Whereas there is not much rewriting of $P_{22}(k+1)$ or $P_{12}(k+1)$, we can rewrite $P_{11}(k+1)$ in the following ways

$$\begin{aligned} P_{11}(k+1) = &\begin{pmatrix} A_{11} & A_{12} \end{pmatrix} P \begin{pmatrix} A_{11} & A_{12} \end{pmatrix}^T \\ &- \begin{pmatrix} A_{11} & A_{12} \end{pmatrix} \begin{pmatrix} P_{11} & 0 \\ 0 & P_{12} \end{pmatrix} C_1(C_1P_{11}C_1^T + R_1)^{-1}C_1^T \begin{pmatrix} P_{11} & 0 \\ 0 & P_{12} \end{pmatrix} \begin{pmatrix} A_{11}^T \\ A_{12}^T \end{pmatrix} + W_1, \end{aligned} \tag{3.29}$$

3.3 Performance comparison

The estimation error covariances resulting from estimating the states in Case 1 and Case 2 will in general not be the same. This has bee proven by [10]. This means that one of the two cases will in general be optimal, given the parameters. It is, however, difficult to predict on forehand from the parameters which case will be the optimal case.

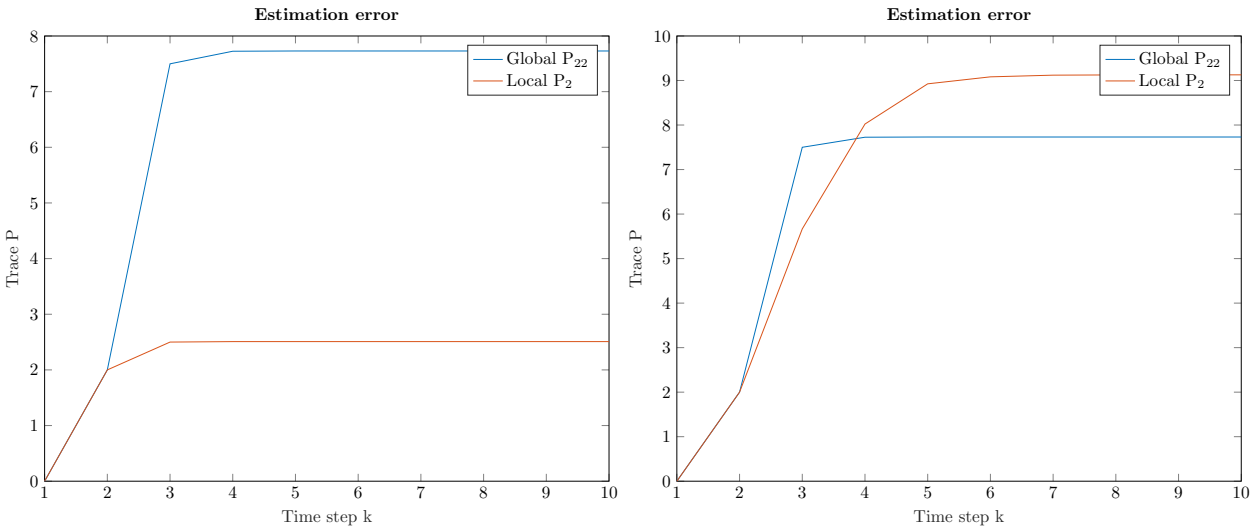
If one considers the error covariance matrices from the Σ_2 in both cases one notices that the global estimation of the state of Σ_2 depends not only on the measurement y_1. It also depends on the cross covariance P_{12}. The local estimation depends mainly on y_2 and hence one has to compare two very different filters. As an example, consider the systems

$$\Sigma_2 = \begin{cases} x_{2,k+1} &= \begin{pmatrix} \frac{1}{2} & 1 \\ 2 & \frac{1}{2} \end{pmatrix} x_{2,k} + B_2u_k + w_{1,k}, \\ y_{2,k} &= x_{2,k} + v_{2,k}, \end{cases} \tag{3.30}$$

and

$$\Sigma_1 = \begin{cases} x_{1,k+1} & = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} x_{1,k} + B_1 x_{2k} + w_{1,k}, \\ y_{1,k} & = x_{1,k} + v_{1,k}. \end{cases} \tag{3.31}$$

If we have that $W_1 = W_2 = I$ and take $B_1 = 10I$, then we have that with two different covariance matrices R_2 we get two relatively different error covariance matrices for the state. The result of a simulation with $R_1 = R_2 = 0.1$ and a simulation with $R_1 = 0.1I$ and $R_2 = 2I$ is shown in Figure 3.



(a) The trace of the error covariance with $R_2 = 0.1I$ and $B_1 = 10I$.

(b) The trace of the error covariance with $R_2 = 2I$ and $B_1 = 10I$.

Figure 3: A comparison of Kalman filter performance with different noise covariance.

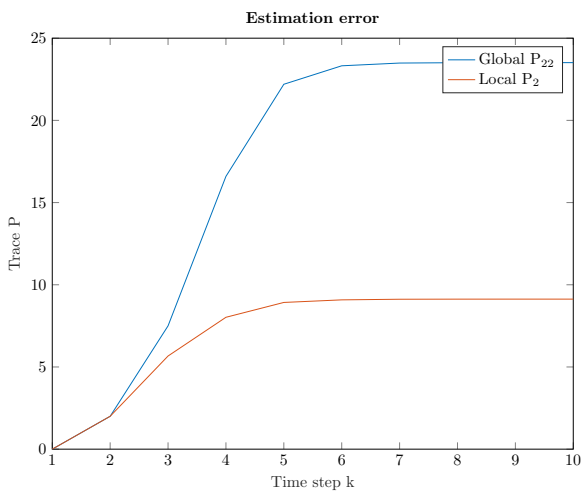


Figure 4: The trace of the error covariance with $R_2 = 2I$ and $B_1 = I$.

The results of this simulation come with a rather intuitive explanation. As the measurement y_2 becomes worse, i.e. has noise with a large covariance, the measurement y_1 will become favourable to use for estimating

x_2. Similarly, if there is a weak coupling between Σ_1 and Σ_1, that is, the influence of x_2 on x_1 can be considered small, then the estimate of x_2 from y_1 will become worse. This is shown in Figure 4. Here we modelled the weak coupling by having $B_1 = I$ instead of $B_1 = 10I$.

Measurements with a low covariance matrix R_2 will not only result in a good estimation of x_2, but also in a good estimation of x_1. As one can see in equation (3.19), the error covariance of x_1 also depends on R_2. This is due to the fact that having a very accurate measurement y_2, is equivalent to knowing the input of Σ_1 very accurately. The optimal Kalman filter assumes that the input of a system is known, and therefore one obtains the best linear estimate if $R_2 = 0$. This means that the error covariance of the local estimate of x_1 tends to the optimal estimate of x_1 as R_2 tends to zero.

As an example, consider the system

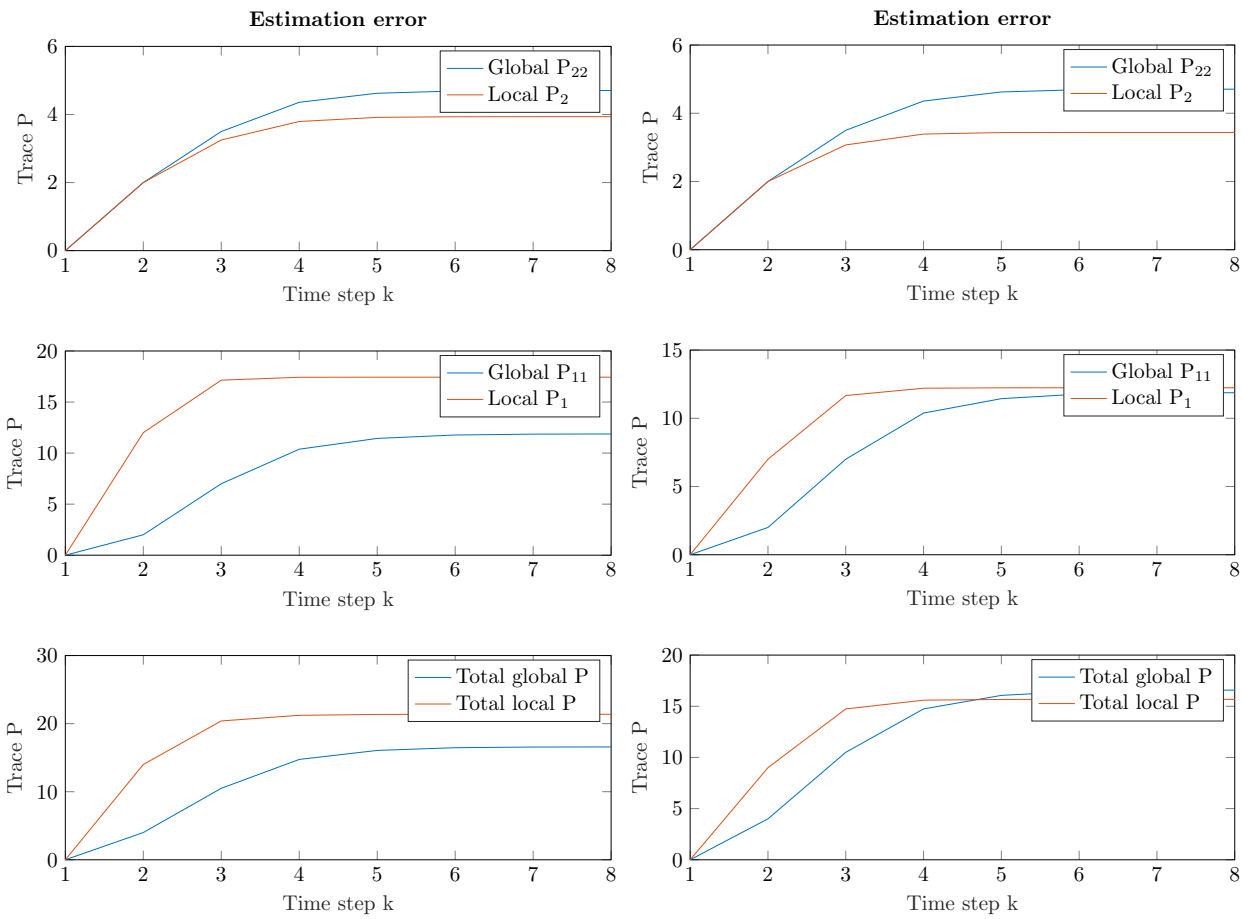
$$\Sigma_2 = \begin{cases} x_{2,k+1} &= \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & \frac{1}{2} \end{pmatrix} x_{2,k} + B_2 u_k + w_{1,k}, \\ y_{2,k} &= x_{2,k} + v_{2,k}, \end{cases} \tag{3.32}$$

and

$$\Sigma_1 = \begin{cases} x_{1,k+1} &= \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} x_{1,k} + x_{2,k} + w_{1,k}, \\ y_{1,k} &= x_{1,k} + v_{1,k}, \end{cases} \tag{3.33}$$

with covariance matrices $R_2 = 5I$, $R_1 = I$ and $W_1 = W_2 = I$. The result of a simulation is shown in Figure 5a. The fact that the local estimation of x_1 becomes worse than the global estimation does not necessarily mean that the overall global estimation of x_1 and x_2 is better than the sum of the local ones. This is shown in Figure 5b, where we modelled $R_2 = 2.5I$ instead of $5I$.

This shows that in general, the better the measurement y_2, the better both local estimates become. In general one can conclude that if one has a very good sensor to measure y_2, it is very likely that the local Kalman filters will result in a general lower error covariance of x_1 and x_2. If it is, however, the case, that the sensor measuring y_1 is much better than the sensor of y_2, it will be more probable that using global estimation will result in a better state estimate.



(a) The trace of the error covariance with $R_2 = 5I$.

(b) The trace of the error covariance with $R_2 = 2.5I$.

Figure 5: A comparison of Kalman filter performance with different noise covariance.

4 Optimal control

Now that we have investigated the Kalman filtering problem, we can start investigating the optimal control problem for our cascaded system. Since we can model our cascaded system as a system of the form $x_{k+1} = Ax_k + Bu_k + w_k$, we can design an optimal controller for this system. To do so properly, we will first investigate the optimal control problem for stochastic systems. This is followed by a simulation, where we investigate the influence of packet dropout on the performance of the controller.

4.1 Optimal control for stochastic systems

It is well known that for deterministic systems the optimal controller consists of an optimal state estimator and an optimal linear actuator. This is due to the fact that there is perfect communication assumed between the controller and the estimator. This would yield a standard LQG controller that aims to minimize the control cost function and it is assumed that the controller has access to the exact value of the state. When dealing with stochastic systems this is not a feasible assumption, and one has to deal with imperfect state information. Therefore, if one is dealing with a stochastic system such as

$$\Sigma = \begin{cases} x_{k+1} & = Ax_k + Bu_k + w_k, \\ y_k & = Cx_k + v_k, \end{cases} \tag{4.1}$$

with covariance matrices W and R for the process and measurement noise respectively, the main goal in the optimal control problem, is to find an input sequence $\{u(k)\}$, such that the expected value of the cost function is minimized. This means that we desire to minimize

$$J_N = \mathbb{E}\left\{x_N^T Q_N x_N + \sum_{k=0}^{N-1}(x_k^T Q_k x_k + u_k^T V_k u_k)\right\}. \tag{4.2}$$

As it will turn out, we can reduce this problem, to a problem with perfect state information. We will show that for the system in equation (4.1) the optimal controller is also separable into a state estimator and an actuator. Once this problem is understood, we can investigate it's application to our cascaded control system as we have done in the previous section. The investigation of this problem follows very similar lines as [11].

To reformulate this problem such that it is a perfect state information problem, define as a new state of the system the information vector

$$\mathcal{I}_k = (y_0, y_1, ..., y_k, u_0, u_1, ..., u_{k-1}). \tag{4.3}$$

Assume that this information vector is available at the controller. This means that the controller has access to the observations and inputs up to time step k and $k-1$, respectively, and thus we have perfect knowledge of this new state. For this new state, $\mathcal{I}_k$, the input is u_k and y_{k+1} can be viewed as a random disturbance. Furthermore we have

$$\mathbb{P}(y_{k+1} \mid \mathcal{I}_k, u_k) = \mathbb{P}(y_{k+1} \mid \mathcal{I}_k, u_k, y_0, ..., y_k), \tag{4.4}$$

since $y_0, ..., y_k$ are part of the information vector $\mathcal{I}_k$. Here $\mathbb{P}(x \mid y)$ denotes the probability of event x occurring given the occurrence of event y. This means that the probability distribution of y_{k+1} depends explicitly only on the state $\mathcal{I}_k$ and the input u_k. Denote $\mathbb{E}_{a,b}\{\cdot\}$ for calculating the expected value for $\cdot$ with respect to a and b and setting all other variables to constant. By writing

$$\mathbb{E}\{Ax_k + Bu_k + w_k\} = \mathbb{E}\{\mathbb{E}_{x_k,w_k}\{Ax_k + Bu_k + w_k \mid \mathcal{I}_k, u_k\}\}, \tag{4.5}$$

we can similarly reformulate the cost per stage as a function of the new state $\mathcal{I}_k$ and the control input u_k:

$$\begin{aligned} J_{N-1}(\mathcal{I}_{N-1}) = \min_{u_{N-1}} [&\mathbb{E}_{u_{N-1},w_{N-1}}\{x_{N-1}^T Q_{N-1} x_{N-1} + u_{N-1}^T V_{N-1} u_{N-1} \\ &+ (Ax_{N-1} + Bu_{N-1} + w_{N-1})^T \\ &\cdot Q_N((Ax_{N-1} + Bu_{N-1} + w_{N-1}) \mid \mathcal{I}_{N-1}\}]. \end{aligned} \tag{4.6}$$

Since $\mathbb{E}\{w_{N-1} \mid \mathcal{I}_{N-1}\} = \mathbb{E}\{w_{N-1}\} = 0$, this expression can be written as

$$\begin{aligned} J_{N-1}(\mathcal{I}_{N-1}) =& \mathbb{E}_{x_{N-1}}\{x_{N-1}^T(A^T Q_N A + Q_{N-1})x_{N-1} \mid \mathcal{I}_{N-1}\}, \\ &+ \mathbb{E}_{w_{N-1}}\{w_{N-1}^T Q_N w_{N-1}\} \\ &+ \min_{u_{N-1}}[u_{N-1}^T(B^T Q_N B + V_{N-1})u_{N-1} + 2\mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\}^T A^T Q_N B u_{N-1}]. \end{aligned} \tag{4.7}$$

Equation (4.7) is minimized when the last term is zero. To find the value for u_{N-1} we take the derivative with respect to u_{N-1} and set it equal to zero. This gives an optimal input u_{N-1}^* as follows

$$u_{N-1}^* = -(B^T Q_N B + V_{N-1})^{-1} B Q_N A \mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\}, \tag{4.8}$$

and upon substitution in (4.7) we obtain

$$\begin{aligned} J_{N-1}(\mathcal{I}_{N-1}) =& \mathbb{E}_{x_{N-1}}\{x_{N-1}^T M_{k-1} x_{N-1} \mid \mathcal{I}_{N-1}, \\ &+ \mathbb{E}_{x_{N-1}}\left\{(x_{N-1} - \mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\})^T S_{N-1}(x_{N-1} - \mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\}\right) \\ &+ \mathbb{E}_{w_{N-1}}\{w_{N-1}^T Q_N w_{N-1}\}, \end{aligned} \tag{4.9}$$

where the matrices M_{N-1} and S_{N-1} are given by

$$\begin{aligned} M_{N-1} &= A^T Q_N B(BQ_N B + V_{N-1})^{-1} BQ_N A, \\ S_{N-1} &= A^T Q_N A - P_{N-1} + Q_{N-1}. \end{aligned} \tag{4.10}$$

Note that $\mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\}$ is the expected value of the state at the time instance $N-1$ given the inputs and outputs up to time step $N-2$ and N. This is exactly equal to the estimate of x_{N-1} generated by a Kalman filter. If we continue the dynamic programming algorithm for period $N-2$, we have that

$$\begin{aligned} J_{N-2}(\mathcal{I}_{N-2}) = &\min_{u_{N-2}} \left[\mathbb{E}_{x_{N-2},w_{N-2},y_{N-2}}\{x_{N-2}^T Q_{N-2} x_{N-2} + u_{N-2}^T V_{N-2} u_{N-2} + J_{N-2}(\mathcal{I}_{N-2} \mid \mathcal{I}_{N-2}, u_{N-2})\}\right] \\ =&\mathbb{E}\{x_{N-2}^T Q_{N-2} x_{N-2} \mid \mathcal{I}_{N-2}\} \\ &+ \min_{u_{N-2}} \left[u_{N-2}^T V_{N-2} u_{N-2} + \mathbb{E}\{x_{N-1}^T S_{N-1} x_{N-1} \mid \mathcal{I}_{N-2}, u_{N-2}\}\right] \\ &+ \mathbb{E}\{w_{N-1}^T Q_N w_{N-1}\} \\ &+ \mathbb{E}\left\{(x_{N-1} - \mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\})^T\right\} M_{N-1} \mathbb{E}\{(x_{N-1} - \mathbb{E}\{x_{N-1} \mid \mathcal{I}_{N-1}\}) \end{aligned} \tag{4.11}$$

The last line in equation (4.11) is left out of the minimization, for it does not depend on u_{N-2}. This is proven by the next lemma. The lemma essentially says that the quality of the estimate $\mathbb{E}\{x_k \mid \mathcal{I}_k\}$ is not influenced by the choice of control input.

Lemma 7. *For every k, there is a function M_k such that we have*

$$x_k - \mathbb{E}\{x_k \mid \mathcal{I}_k\} = M_k(x_0, w_0, ..., w_{k-1}, v_0, ..., v_{k-1}), \tag{4.12}$$

independently of the input being used.

Proof. Fix an input and consider the following two systems. In the first system, the control input is used,

$$x_{k+1} = Ax_k + Bu_k + w_k, \qquad y_k = Cx_k + v_k, \tag{4.13}$$

whereas it is not considered in the second system,

$$\bar{x}_{k+1} = A\bar{x}_k + \bar{w}(k), \qquad \bar{y}(k) = C\bar{x}(k) + \bar{v}(k). \tag{4.14}$$

Consider the evolution of these two systems when their initial conditions are identical, i.e. $x_0 = \bar{x}_0$ as well as the noise vectors $w(k) = \bar{w}(k)$, $v(k) = \bar{v}(k)$, for all k. Consider the vectors

$$\begin{aligned} Y_k &= (y_0, ..., y_k)^T, \qquad \bar{Y}_k = (\bar{y}_0, ..., \bar{y}_k)^T \\ W_k &= (w_0, ..., w_k)^T, \qquad V_k = (v_0, ..., v_k)^T, \qquad U_k = (u_0, ..., u_k)^T. \end{aligned} \tag{4.15}$$

Linearity implies the existence of matrices F_k, G_k and H_k such that

$$\begin{aligned} x_k &= F_k x_0 + G_k U_{k-1} + H_k W_{k-1} k, \\ \bar{x}_k &= F_k \bar{x}_0 + H_k W_{k-1}. \end{aligned} \tag{4.16}$$

Since the vector U_{k-1} is part of the information vector I_k, we have $U_{k-1} = \mathbb{E}\{U_{k-1} \mid \mathcal{I}_k\}$, so

$$\begin{aligned} \mathbb{E}\{x_k \mid \mathcal{I}_k\} &= F_k \mathbb{E}\{x_0 \mid \mathcal{I}_k\} + G_k U_{k-1} + H_k \mathbb{E}\{W_k \mid \mathcal{I}_k\} \\ \mathbb{E}\{\bar{x}_k \mid \mathcal{I}_k\} &= F_k \mathbb{E}\{\bar{x}_0 \mid \mathcal{I}_k\} + H_k \mathbb{E}\{\bar{W}_k \mid \mathcal{I}_k\} \end{aligned} \tag{4.17}$$

From this we can see that

$$x_k - \mathbb{E}\{x_k \mid \mathcal{I}_k\} = \bar{x}_k - \mathbb{E}\{\bar{x}_k \mid \mathcal{I}_k\}. \tag{4.18}$$

From the equations for y_k and $\bar{y}_k$ in equations (4.13) and (4.14), we see that

$$\bar{Y}_k = Y_k - R_k U_{k-1} = S_k W_{k-1} + T_k V_k, \tag{4.19}$$

where R_k, S_k and T_k are some matrices of appropriate dimension. Thus, the information provided by $\mathcal{I}_k = (Y_k, U_{k-1}\})$ regarding $\bar{x}_k$ is summarized in $\bar{Y}_k$, and we have $\mathbb{E}\{\bar{x}_k \mid \mathcal{I}_k\} = \mathbb{E}\{\bar{x}_k \mid \bar{Y}_k\}$, such that

$$x_k - \mathbb{E}\{x_k \mid \mathcal{I}_k) = \bar{x}_k - \mathbb{E}\{\bar{x}_k \mid \bar{Y}_k\} = M_k(x_0, w_0, ..., w_{k-1}, v_0, ..., v_k). \tag{4.20}$$

□

If we now return to (4.11), we see that the dynamic programming algorithm for period $N-2$ is optimized by

$$u^*_{N-2} = -(B^T S_{N-1} B + V_{N-2})^{-1} B^T S_{N-1} A \mathbb{E}\{x_{N-2} \mid \mathcal{I}_{N-2}\}. \tag{4.21}$$

This proves that we can proceed similarly to obtain the optimal input sequence $\{u(k)\}$ for every stage where $u(k)$ is given by

$$u^*(k) = L_k \mathbb{E}\{x(k) \mid \mathcal{I}_k\}, \tag{4.22}$$

where the matrix L_k is given by

$$L_k = -(B^T S_{k+1} B + V_k)^{-1} B^T S_{k+1} A, \tag{4.23}$$

with the matrices K_k given recursively by the Riccati equation

$$\begin{aligned} S_N &= Q_N, \\ E_k &= A^T S_{k+1} B (B^T S_{k+1} B + V_k)^{-1} B^T S_{k+1} A, \\ S_k &= A^T S_{k+1} A - E_k + Q_k. \end{aligned} \tag{4.24}$$

An interesting observation is that the optimal controller for the imperfect state information case is again a linear combination of an optimal state estimator and an actuator. This make the Kalman filtering problem in the previous section an even more important topic. For the better the state estimate will be, i.e. the better the Kalman filter performs, the lower the control cost will be. With this understanding of how the the optimal controller works, we can continue applying this theory the cascaded setting.

4.2 The cascaded setting and packet drop-out

As has been shown in the previous section, the expected value of the cost function is minimized by applying a linear actuator and an optimal state estimator. In the first section we have investigated how to estimate the state of a cascaded system optimally. As it turned out, in some cases, estimating the states of the subsystems locally was optimal compared to estimating them globally. With respect to wireless communication systems, this means that both estimates also have to be send and received by the controller. This might pose a problem, in case one of the two state estimates does not arrive for some reason. To investigate this, we will consider such a setting. Recall that we consider the cascaded setting of the form

$$\Sigma_1 \times \Sigma_2 = \begin{cases} x_{k+1} = & \begin{pmatrix} A_1 & B_1 C_2 \\ 0 & A_2 \end{pmatrix} x_k + \begin{pmatrix} 0 \\ B_2 \end{pmatrix} u_k + w_k, \\ & = A x_k + B u_k + w_k, \\ y_{1,k} & = \begin{pmatrix} C_1 & 0 \end{pmatrix} x_k + v_{1,k}, \\ y_{2,k} & = C_2 x_{2,k} + v_{2,k}. \end{cases} \tag{4.25}$$

Let us define a control cost as follows:

$$J_T = \mathbb{E}\left\{ x_T^T W x_T + \sum_{k=0}^{T-1} x_{1,k}^T W_1 x_{1,k} + x_{2,k}^T W_2 x_{2,k} + u_k^T U u_k \right\}. \tag{4.26}$$

It follows from the theory above that the input sequence $\{u_k\}$, that minimized the expected value of the cost function, can be computed backwards by the following formula

$$\begin{aligned} u_k &= -(B^T Q_{k+1} B + U)^{-1} B^T Q_{k+1} A \hat{x}_{2k|k} \\ &= -G_k^{-1} L_k \hat{x}_{k|k}, \end{aligned} \tag{4.27}$$

with

$$Q_k = A^T Q_{k+1} A + W_k - A^T Q_{k+1} B (B^T Q_{k+1} B + U)^{-1} B^T Q_{k+1} A, \qquad Q_N = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}. \tag{4.28}$$

If we consider a cascaded system with an optimal controller, where the states are estimated locally by a Kalman filter and sent to the controller, we are dealing with a setting as is depicted in Figure 6. We can assign a probability to the event of a packet arriving. Some properties of the performance of the controller in relation to the packet arrival probability will be investigated in a simulation study.

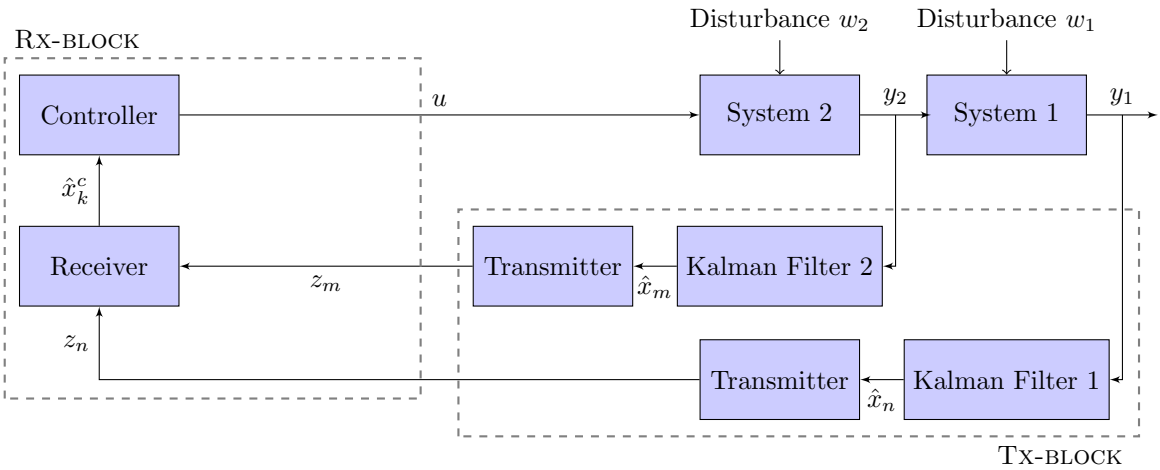


Figure 6: Block diagram of the cascaded system with control block

As an example, consider the system

$$\Sigma_2 = \begin{cases} x_{2,k+1} &= \begin{pmatrix} \frac{1}{4} & 0.9 \\ 0 & \frac{1}{2} \end{pmatrix} x_{2,k} + u_k + w_{2,k}, \\ y_{2,k} &= 2x_{2k} + v_{2,k}, \end{cases} \tag{4.29}$$

and

$$\Sigma_1 = \begin{cases} x_{1,k+1} &= \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} x_{1,k} + x_{1,k} + w_{1,k}, \\ y_{1,k} &= x_{1,k} + v_{1,k}, \end{cases} \tag{4.30}$$

with covariance matrices $W_1 = W_2 = I$ and $R_1 = R_2 = 0.1I$. These matrices guarantee that the local estimation is indeed optimal. As a cost function consider

$$J_T = \mathbb{E}\left\{ x_T^T I x_T + \sum_{k=0}^{T-1} x_{1,k}^T I x_{1,k} + 4x_{2,k}^T I x_{2,k} + u_k^T I u_k \right\}. \tag{4.31}$$

In most engineering applications Σ_1 is an unstable system and hence loosing information on the state of Σ_1 can pose a big problem. Since Σ_2 is often a stable system, information loss on the state of Σ_2 is often less of a problem. Investigating packet drop-out between the state estimator of Σ_1 is, however, beyond the scope of this research. Therefore we will focus on the packet drop out between the estimator of Σ_2 and the controller. It is thus assumed that the state of Σ_1 is always delivered. In this case one can think of the state of Σ_2 being communicated wireless, whereas the state of Σ_1 is communicated via a wire. We denote the Boolean variable $\gamma = 1$ if a packet arrives and hence the probability of the packet arrival is denoted $\mathbb{P}(\gamma = 1)$. The results of three simulations are shown in Figures 7, 8 and 9. The probability of a packet dropout for the wireless link is given by $\mathbb{P}(\gamma = 1) = 0.85$, $\mathbb{P}(\gamma = 1) = 0.90$ and $\mathbb{P}(\gamma = 1) = 0.95$ respectively. We plotted the mean cost, which is the sum over every cost per step, devided by the amount of steps.

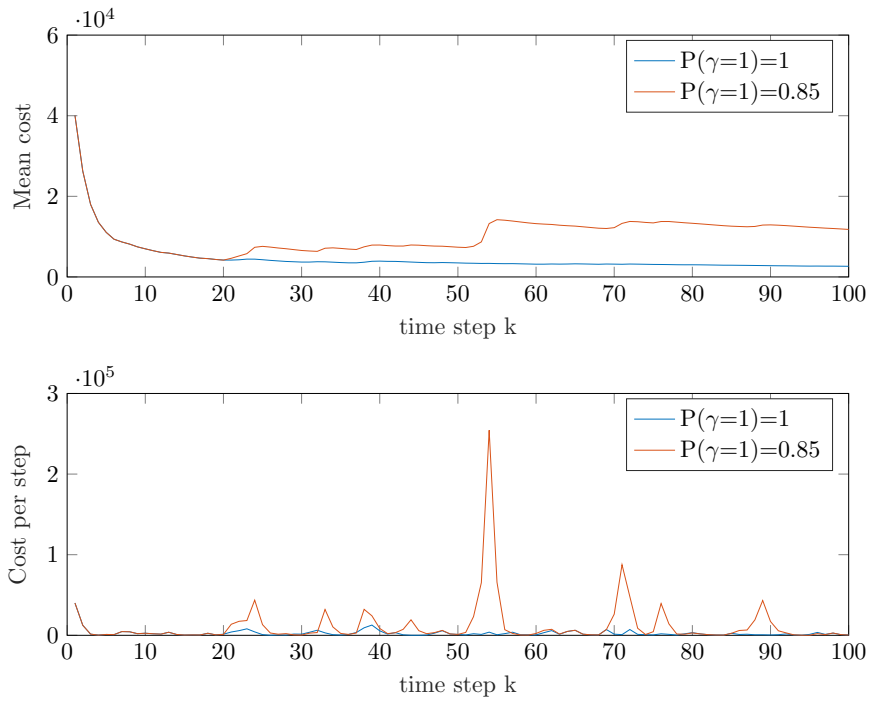


Figure 7: The mean cost and cost per step of when 85% of the packages arrive.

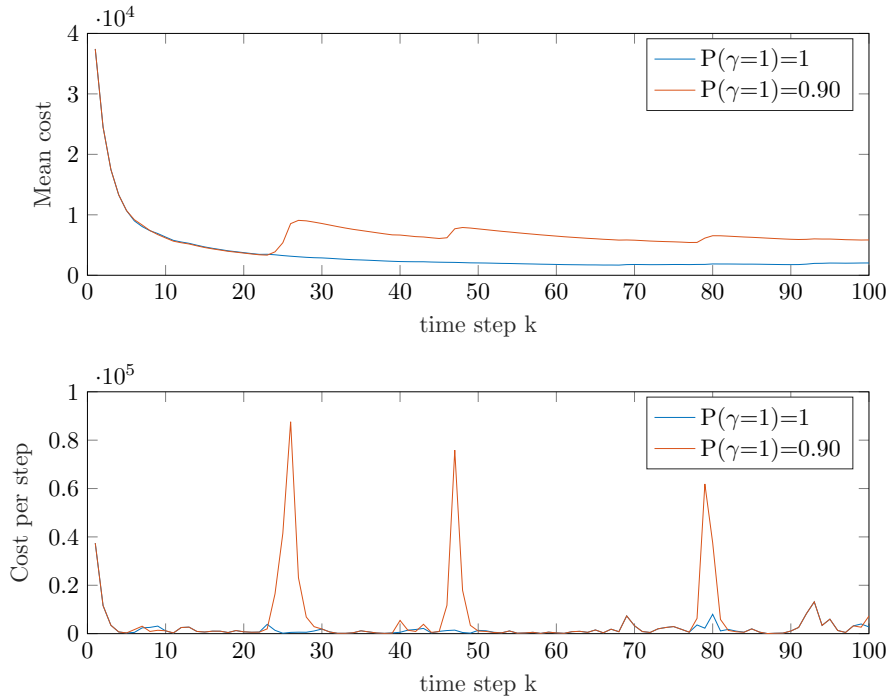


Figure 8: The mean cost and cost per step of when 90% of the packages arrive.

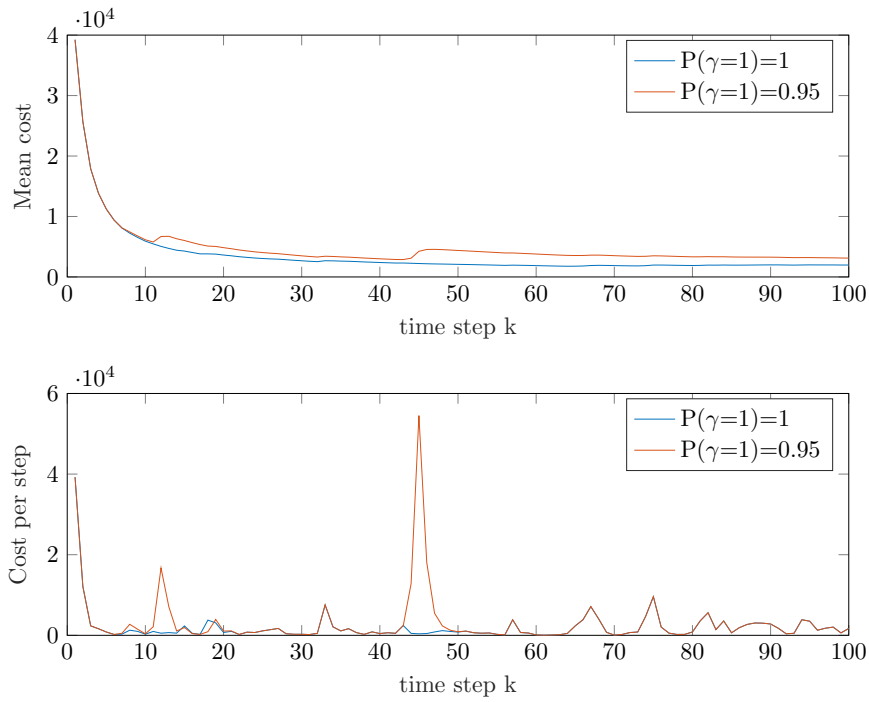


Figure 9: The mean cost and cost per step of when 95% of the packages arrive.

As one can see, the average control cost per step decreases for growing packet arrival probability. One sees also that at a certain pack-out probability the stability is not guaranteed anymore. This is in accordance with the results in [3]. In order to find the expected value of the control cost,the simulations were repeated 500 times for different packet arrival probabilities. The results are shown in Figure 10.

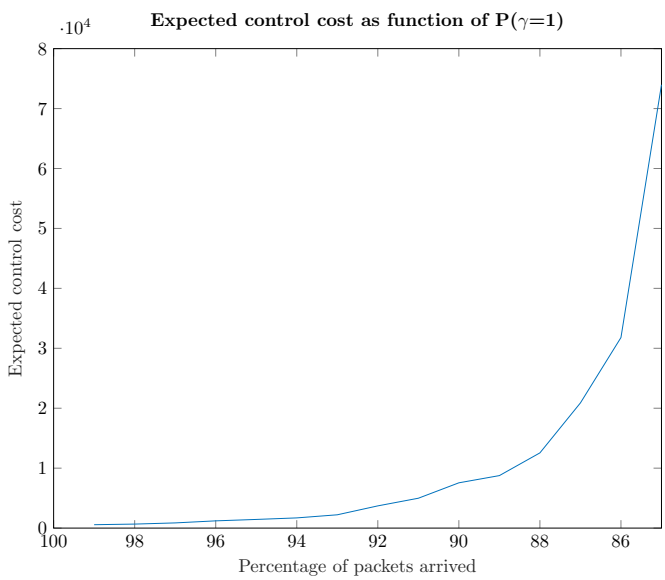


Figure 10: Expected mean control cost per step.

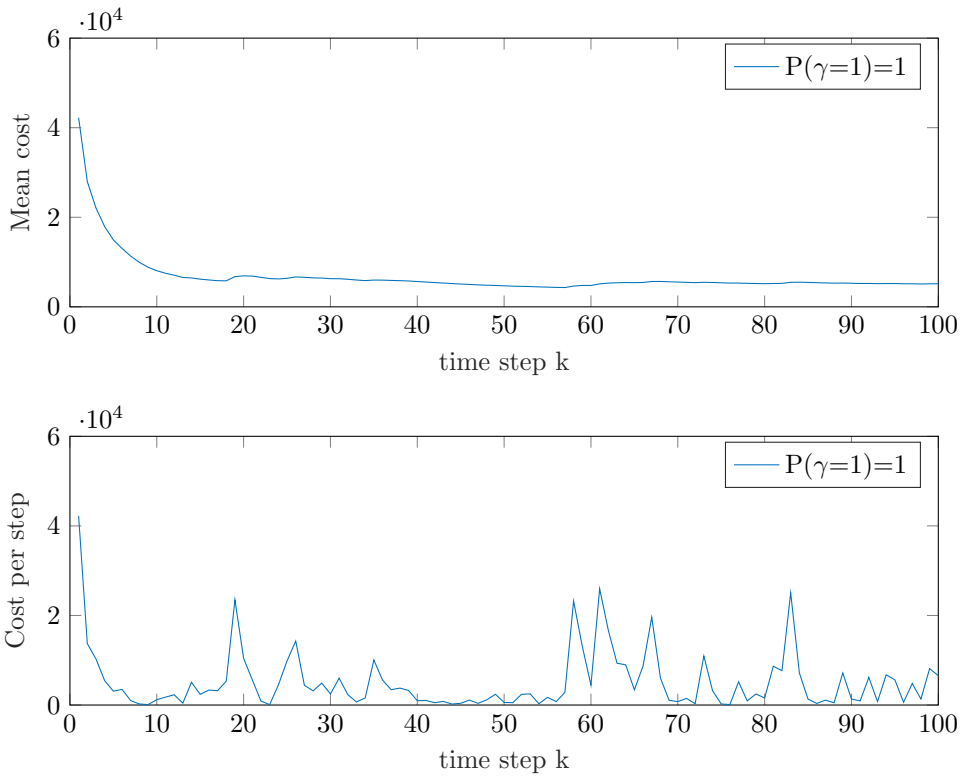


Figure 11: The mean cost and cost per step for estimating the state globally.

A simulation where we estimate the state globally is shown in Figure 11. Running these simulations also 500 times and calculating the average mean control cost, result in an expected average mean control cost of $5.7278 * 10^3$. This also shows, that for a sufficiently low packet arrival probability of the second Kalman filter, estimating the state globally becomes more favorable, if the packet arrival probability of the first Kalman filter is $\mathbb{P}\{\gamma = 1\} = 1$ in both cases.

5 Conclusion and recommendation

In retrospect we have investigated the optimal control problem for cascaded stochastic systems and analyzed the Kalman filtering problem for this type of systems. We analyzed the Kalman filter and proved under which conditions the resulting error covariance matrix converges to a steady state value. As it turned out, having less process and measurement noise, results in a lower error in the state estimate. We have also seen that if there are multiple outputs available for the Kalman filter, they do not necessarily result in the same error covariance. We achieved some results on the dependence of the steady state error covariance on the choice of measurement, but there are still questions unanswered and lead to our first recommendation for further research.

In our analysis of the dependence on the choice of output, we have results depending on the steady state value of the error covariance. Furthermore we have results comparing outputs with an invertible C matrices or C matrices which where scalar multiples. It remains to investigate how one can predict what output will result in the optimal error covariance matrix without computing the steady state values for at least one output. Similarly, one would like to obtain results for comparing outputs of different dimensions. This could also help in the analysis of the performance of a distributed Kalman filter.

So far we have compared two ways of state estimation for a cascaded system, which we called local and global estimation. We did this by means of a simulation. These simulations show that the performance of the two cases mainly depended on the covariance of the measurement noise and the coupling strength between the two systems. The fact, that there are no firm theorems on the conditions for a specific case to

result in an optimal state estimate, gives rise to more research.

As it seems so far, the current paradigm in doing research on (distributed) Kalman filtering is to investigate a given system with a given output. Given a specific set up, there are many papers on optimizing algorithms and combining information from the given sensors. The question of which general output equation is more favorable compared to another one is a relatively new question resulting from the increase of system complexity and networks. Therefore, a lot of results are yet to be obtained in this area.

After the investigation of the Kalman filter problem for the cascaded setting, we gave an introduction to optimal control for stochastic systems. It turned out, that the optimal controller was separable into an optimal state estimator and a linear actuator. With this theory, we designed an optimal controller for a cascaded system and investigated the influence of packet drop-out on the performance of the controller. By means of a simulation study we have shown that the expected control cost increases as the packet arrival probability decreases. In the simulation it was assumed that only a part of the state was communicated wireless, whereas one could also investigate the case where the arrival of the whole state is uncertain. An interesting question would be if the results of [3] could be extended to the cascaded setting. The simulations suggest, that the stabilizing property of the controller can be guaranteed if the arrival probability is above a certain threshold. However no firm claims can be made based on these simulations.

Furthermore, the simulation study shows, that estimating the states globally becomes favorable compared to local estimation if the packet arrival probability of the second Kalman filter drops below a certain threshold value. It remains to investigate the case where the communication link between the first Kalman filter and the controller also becomes uncertain. Perhaps one can find a trade off between the Kalman filter performance and the packet arrival probabilities. It remains to investigate how the behaviour of the controller would be influenced by the packet drop-out between the estimator of Σ_1 and the controller.

In general, one can conclude that extending the optimal control problem with packet drop out to cascaded systems is not straight forward. There are a lot of questions yet to be answered.

References

[1] Bruno Sinopoli, Luca Schenato, Massimo Franceschetti, Kameshwar Poolla, Michael I Jordan, and Shankar S Sastry. Kalman filtering with intermittent observations. *IEEE transactions on Automatic Control*, 49(9):1453–1464, 2004.

[2] Luca Schenato, Bruno Sinopoli, Massimo Franceschetti, Kameshwar Poolla, and S Shankar Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163–187, 2007.

[3] Daniel E Quevedo, Anders Ahlen, and Karl H Johansson. State estimation over sensor networks with correlated wireless fading channels. *IEEE Transactions on Automatic Control*, 58(3):581–593, 2013.

[4] Steffi Knorn and Subhrakanti Dey. Optimal sensor transmission energy allocation for linear control over a packet dropping link with energy harvesting. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, pages 1199–1204. IEEE, 2015.

[5] Steffi Knorn and Subhrakanti Dey. Optimal energy allocation for linear control with packet loss under energy harvesting constraints. *Automatica*, 77:259–267, 2017.

[6] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[7] Graham C Goodwin and Kwai Sang Sin. *Adaptive filtering prediction and control*. Courier Corporation, 2014.

[8] Dennis S Bernstein. Matrix mathematics: Theory, facts, and formulas . princeton reference, 2009.

[9] Harry L Trentelman, Anton A Stoorvogel, and Malo Hautus. *Control theory for linear systems*. Springer Science & Business Media, 2012.

[10] Zsófia Lendek, R Babuška, and Bart De Schutter. Distributed kalman filtering for cascaded systems. *Engineering applications of artificial intelligence*, 21(3):457–469, 2008.

[11] Dimitri P. Bertsekas et al. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.