Nonlinear estimation and control in the iron ore pelletizing process

An application and analysis of the Extended Kalman Filter

Thesis by

Knut Rapp

Submitted in partial fulfillment of the requirements of the Doktor ingeniør degree

Abstract

This thesis deals with estimation and control of the balling drums used in the iron ore pelletizing process.

First a conceptually new control scheme for balling drums is proposed. In this control scheme a cluster of drums is considered rather than considering each drum to be an independent unit, and the output of each drum is assumed to be oscillating during normal operation. Three states are essential in this scheme, and these are the amplitude of the oscillation, frequency of the oscillation, and its phase angle. These three states are estimated by use of an extended Kalman filter (EKF). The new control scheme thus depends heavily on the state estimation. A simulation of this scheme controlling a cluster of three drums is carried out. In this simulation the drums are modelled as van der Pol oscillators with varying amplitude and frequency of the output.

Next the problem of automatically tuning and optimizing the extended Kalman filter is addressed. In spite of its importance, there has been published surprisingly few results on this subject, however, two different methods for EKF tuning has been proposed recently in the literature. One of these is tuning by use of a genetic algorithm (GA), which is also applied in this thesis. It is shown that by use of a simple GA, the extended Kalman filter is well tuned in a reasonable amount of time, provided that the tuning criterion is well defined. Furthermore, the same simple GA is applied for optimizing the EKF with respect to its performance in high noise environment.

Finally the stability properties of the EKF is analyzed. It is not possible to guarantee that the states in the signal model used for the state estimation in the new control scheme will be bounded from above, as required in the stability results presented so far. By applying a different bound on the Kalman gain matrix, it is shown that this requirement can be relaxed, and thus convergence of the state estimator applied in the new control scheme can be guaranteed under some given conditions.

Preface and Acknowledgements

This thesis is submitted in partial fulfillment of the requirements for the degree of Doktor Ingeniør at Norwegian University of Science and Technology. The work has been carried out at Narvik University College in co-operation with the research department at LKAB in Kiruna, Sweden, during the period from August 2000 to June 2004. The funds for this work has been granted by LKAB.

I am indeed grateful to my supervisor Professor Dr. Per-Ole Nyman at Narvik University College for being an excellent source of inspiration, knowledge and guidance through all stages of this work. I also want to thank Professor Dr.ing Rolf Henriksen at Norwegian University of Science and Technology for interesting discussions and helpful comments, and my colleagues at Narvik University College for all help, all the interesting and nice discussions and for making an excellent working environment.

Our former M.Sc. student Dongni Sun deserves a thank for implementing the control algorithms in SIMULINK.

The research department at LKAB in Kiruna, headed by Dr. Per-Olof Samskog, deserves a great thank for both the financial support and for the co-operation during these years. Magnus Ruthfors deserves a special thank for always being willing to help and for providing necessary data from the process.

Finally, I want to thank my dear wife Tanja for her patience, support and love. Her ability to motivate me and organize our days has been of great importance for this work.

Narvik, July 1, 2004
Knut Rapp

Contents

List of Figures

Part I

Introductory chapters

Chapter 1

Introduction and motivation

These introductory chapters are intended to give an overview of the research field in which this dissertation gives its contribution. The new results obtained during this research are mainly presented in the papers in Part II. However, in section 3.4 and 4.3 some new material additional to what is found in the papers are included. Most of the material presented about iron ore pelletizing can be found in the references given, however, some is information delivered orally. The sources for this are Mr. Roland Drügge, Mr. Ola Erikson and Mr. Magnus Ruthfors at LKAB, Sweden. As this information is not necessarily easily accessible, it has been used with care, and only when it is supported by other written sources. This does not apply to historical information, which has been used to illustrate the developments in the iron ore pelletizing, and is as such not considered to be essential for our research.

The thesis is divided into two parts, one introductory part and one part which consists of papers. The contents of this thesis is also divided into two parts; a theoretical part and one part in which some of the new results are applied to a specific estimation and control problem. The area to which the research results presented in this thesis are intended to be applied, can very briefly be described as the cold process segment in iron ore pelletizing. The cold process segment is the part where the material (iron ore fines) is tumbled together in rotating drums into raw pellets, or green pellets as they often are referred to. To be more specific, it is the classical problem of stabilizing the pelletizing process to give a constant, or nearly constant rate of production of green pellets which is our main concern. However, this has been a research field for about 50 year, so any attempt to completely solve this problem in less than 4 years may be too ambitious. Therefore the area has been limited to some specific problems. Even if this area has been a research field for a long time, there has been relatively little work on balling drums since the 1970's. This is in fact rather surprising, as the importance of iron ore pelletizing has increased tremendously during the last thirty years. It is outside the scope of this work to analyze and explain why scientists have not found this area attractive for the last three decades, so we only state this fact without any further discussion. However, as the scientists today are equipped with more modern techniques, there should be a big potential for bringing this field a step further.

Iron ore pelletizing is quite complex, and thus includs several fields like metallurgy, chemistry, measuring and instrumentation techniques and estimation and control theory. This dissertation only deals with the estimation and control theory part, which is a considerable part in almost all industrial processes, the iron ore industry being no exception in that respect. In such a complex plant there is of course lot of different interesting control problems to solve. The one in question here is to stabilize the balling drum's output. To achieve this we propose a new control scheme which is highly based upon estimated states from an Extended Kalman Filter. The contribution in this dissertation is thus within several areas, which are:

1) A new Extended Kalman Filter based control scheme for balling drums

2) Tuning of the Extended Kalman Filter used as state estimator[1]

3) Stability analysis of the Extended Kalman Filter

Part 1 is treated in paper 1, part two is treated in paper 2 and 3, and part 3 is treated in paper 4 and 5. All papers are either published or accepted for publication. This is indicated separately for each paper.

The first item is a conceptually new way of thinking when it comes to balling drum control. The traditional approach has been to consider one drum as the total system, thus handling a multi drum plant with n drums as n parallel working systems. Our approach is to consider the balling drum cluster as the total system. This may give the plant designers some new challenges, but the amount of additional equipment required is quite low for most of the existing modern pelletizing plants. The problem of controlling the balling drum is not solved properly with the traditional approach as no well functioning control scheme has been able to replace manual control, which still is the normal control method. With the new concept it may be easier to automatically control the drums, but a successful control system depends heavily on the state estimation of the states to be controlled. This system is nonlinear, so the well known extended Kalman filter (EKF) has been used for this purpose. The level of noise will in general affect the quality of the estimated states, and in this application the level of noise is rather challenging. However, as in the linear case, a well tuned filter yields better results. This introduce the next item,

[1] Tuning may be a too restrictive term. Also optimization with respect to different filter properties is included under this item.

which is tuning and optimization of the (EKF).

Since the optimality of the Kalman filter requires that the signal model is linear, optimality can no longer be guaranteed for the extended Kalman filter. By clever tuning it is, however, possible to get a well working filter, but EKF tuning is by no means a trivial task. Since high quality estimation is crucial for the control part, this is an important second step in our work. Included in this, we have also carried out some work in order to design an EKF with high noise rejection and fast response. This is reported in section 3.4, and is based on some result from stability analysis of the EKF, which bring us to the third and final item.

Unlike the linear case, stability, or more precise bounded estimation error, can not be guaranteed for the EKF independently of the size of the noise processes and the initial error. The conditions for which stability can be guaranteed is of course of great interest and importance. Earlier results have been very conservative and as such not very informative. In this thesis these results are extended. In this context, these three parts, as listed above, are equally important parts of this research project.

Even though it is estimation and control theory which is the main part in this work, the pelletizing process will be briefly described in the following chapter in order to give a very short introduction into this field, but also to establish the notion. This introduction is followed by a chapter about estimation and the Extended Kalman filter in particular. The next introductory chapter, gives an introduction to balling drum control. This chapter gives a overview of previous research in this area, and gives a detailed introduction to the new control scheme proposed in paper 1. As a final introductory chapter, some conclusions and suggestions for further work are given.

Two appendixes are included at the end of this thesis. Appendix 1 gives an short introduction to Genetic Algorithms (GA's). GA is the tool used for tuning the state estimator in this work. In the second appendix the definition of local input-to-stable discrete time systems is given.

Chapter 2

Iron ore pelletizing, a short process description

2.1 Background

In earlier days iron ore was sold and transported mainly as bulk material. The patent describing how to make pellets from iron ore fines is Swedish, and dated back to 1911. Even if this invention is quite old, iron ore pellets did not start to be common until the seventies. However, after the second world war the technique started to spread, even though in a slow tempo. Today the majority of iron ore sold on the world market is in pellets form. Pellets offers some great advantages compared to bulk material, and among these are more easy handling and a reduction of dust losses. The latter is of course desirable also from an environment point of view, and in particular for the local environment.

The market demand is of course the most important aspect to be taken into account when evaluating in which direction product development should be conducted. Important parameters for the price of the pellets includes, among others, purity and size of the pellets. The size is specified by the buyer in terms of a nominal diameter ± some small variance. The nominal diameter may vary from one iron mill to another, but a quite normal value is between $10 - 12$ mm.

Crushed pellets, which always will be present in some extent, is highly undesirable for the iron mill as it complicates their process. The amount of such material is thus required to be at a minimum. It also represents an unacceptable economic loss for the pellets manufacturer, as it is not possible to recycle this material into an earlier stage of the process, and therefore has to be delivered as a low quality product. Crushed pellets is normally a result of too high moisture content in the green pellets, which is caused by a badly controlled cold process. However, a small amount of crushed pellets due to the transport and handling will always be present. This illustrates that the quality of the final product and the productivity depends highly upon both the production of green pellets and the firing process. In the following the firing process will be referred to as the warm process. In Figure 2.1 (printed with permission from LKAB, Sweden) a complete pelletizing plant is shown.

It is customary to divide the pelletizing process into two main process segments, which are:

- The cold process segment
- The warm process segment

With reference to Figure 2.1, the main parts of each these segments are listed below.

The *cold process* consists of the following main parts:

1) Slurry tank, in which water and iron ore is mixed
2) Filters, in which most of the water is removed from the slurry
3) A mixer were binder is added to the fines in order to obtain sufficient mechanical strength of the green pellets
4) Balling drums, in which the fines are tumbled together to pellets
5) Screen, where the different fractions are separated. Undersized pellets is recycled back to the drum, onsize pellets is the output from the cold process and oversized pellets are crushed and fed to the fines tank

The *warm process* consists of the following main parts:

1) Drying, in which most of the remaining moisture is removed before the green pellets enter the firing process
2) Firing, in which the green pellets are turned into the final product
3) Cooling

In the next subsection some more detail are given about the different stages of the pelletizing process.

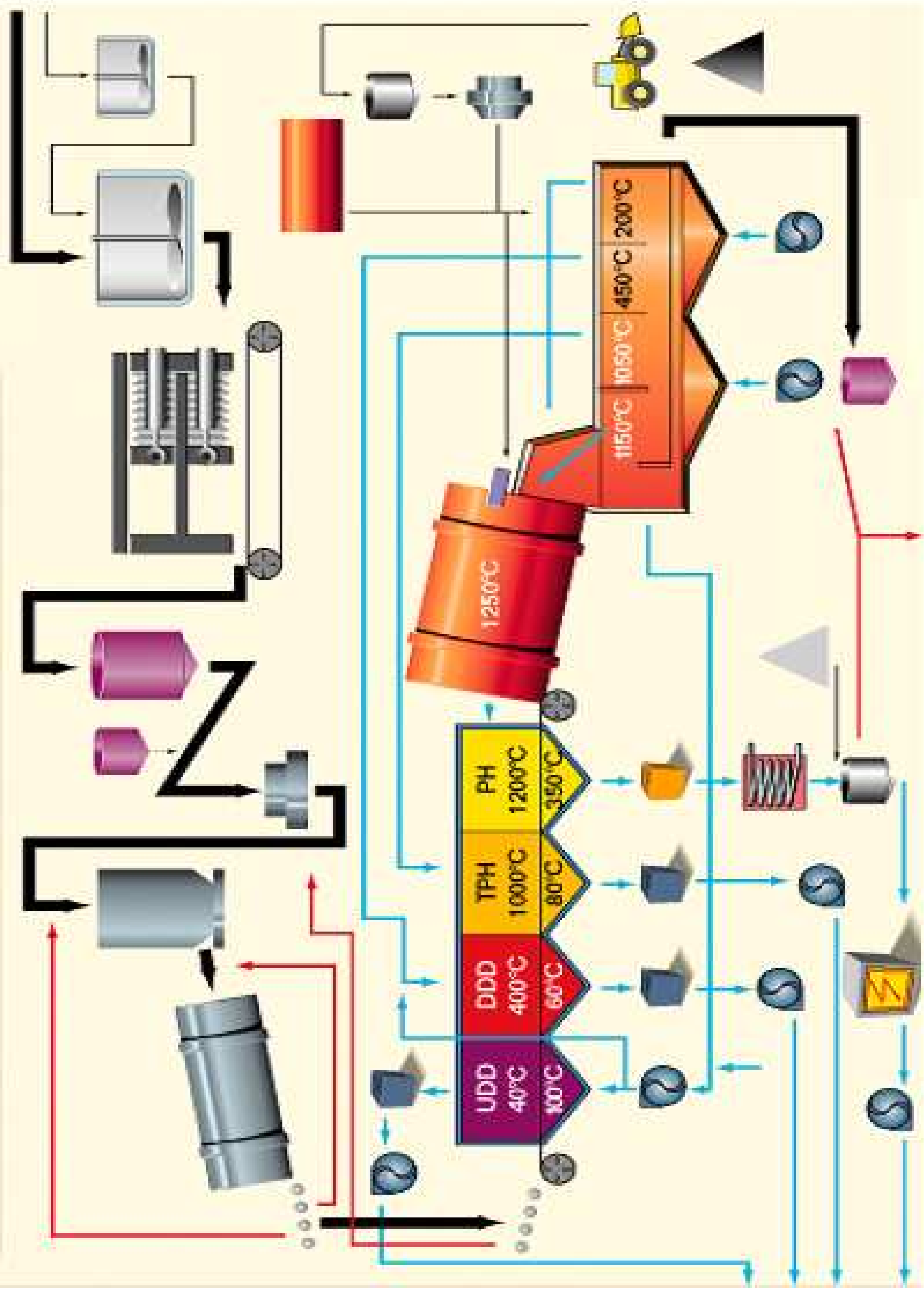

Figure 2.1: Overview of the iron ore pelletizing process

2.2 Sub-processes

The pelletizing process is a process which contains numerous sub-processes, or process segments. On the way from the mine to a finalized product, the iron ore goes through the following main process segments:

i) The iron ore is crushed and the waste rock is removed. About 85 % of the particles should be less than $44\,\mu m$ (in length, width, or height). It is then possible to extract the valuable mineral, which is magnetite

ii) Water is added to the magnetite to make magnetite slurry

iii) Additive material (dolomite or olivine, depending on the product) is added to the slurry

iv) Most of the water is then removed from the slurry by use of press filters. The water content after this filtering is about 9 % of the weight.

v) After the filtering, binder (bentonite or organic binder) is added

vi) Green pellets are made by use of balling drums. When leaving the drum the pellets are screened, and pellets with too small diameter is fed back to go through the drum once more. Oversized pellets are crushed and recycled. The rest is onsized pellets which forms the drum's output.

vii) The onsize pellets are transported on a conveyor to the drying process, where they are dried by hot air flowing through the bed.

viii) The pellets are fired (1250-1300 degree Celsius) and then cooled down to about 200 degree Celsius.

Several of the above listed items describe process segments which may be operated with classical control techniques, and some of them like item ii), iii), and v) are already automatic in most pelletizing plants today. Clearly some process segments depend highly on a well functioning preceding segment. If the particle size is too large, the fines are too dry, or the drying is not working, no pellets can be produced. A less dramatic situation is when some segments are working suboptimally. If for instance the process in which binder is added gives too varying output, then the balling drum operation will suffer. Therefore, the iron ore pelletizing process may be described as a chain of several sub-processes which may depend highly upon each others performance.

Chapter 3

Nonlinear State Estimation

3.1 State estimation and the Kalman filter

In general, estimation may be defined as a process in which information is extracted from data (see e.g. [13]). In this work the term estimation always refers to the more specific term state estimation, and we work with signal models, linear or nonlinear, described by the state space model

$$
\begin{aligned}
x(k+1) &= f(x(k), k) + w(k) && (3.1)\\
y(k) &= h(x(k)) + v(k) && (3.2)
\end{aligned}
$$

where $f : R^n \times Z_+ \rightarrow R^n$ is the state map, $h : R^n \rightarrow R^m$ is the output map, $x(k) \in R^n$ the state vector, $y(k) \in R^m$ the output vector and $w(k)$ and $v(k)$ are the process and measurement noise respectively. Z_+ denotes the set of all positive integers.

Estimation includes a number of different techniques developed for a broad class of problems. In this work only the well known Kalman filter will be considered , see [9], [24] or [2] for a thorough treatment of this subject. As the signal model considered in the balling drum application has a nonlinear output map, we will mostly refer to the Extended Kalman Filter (EKF).

The Kalman filter has some nice properties which make it very suitable for real-time applications. First of all, it does not require much space for storing data as it is a recursive algorithm. Secondly, it is designed for operating in the time domain rather than in the frequency domain. Furthermore, the Kalman filter is optimal in the sense that it yields a minimum variance estimate of the states if the signal model is linear. This optimality is lost when turning to the nonlinear signal model. However, the transition from the Kalman filter to the extended Kalman filter is simple, and therefore nonlinear estimation problems may be solved in an easy way. Even though optimality, and until recently stability, no longer can be guaranteed, the extended Kalman filter has gained vast popularity since it was introduced by the NASA scientist Dr. Stanley F. Schmidt in 1960.

For later reference, the extended Kalman filter equations are given below (see e.g. [9] or [13]):

Measurement update:

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k \left[y_k - h(\hat{x}_{k,k-1}) \right] \tag{3.3}$$

$$P_{k,k} = \left[I - K_k H_k \right] P_{k,k-1} \tag{3.4}$$

where

$$H_k = \left[\frac{\partial h}{\partial x} \right]_{\hat{x}=\hat{x}_{k,k-1}} \tag{3.5}$$

Time update:

$$\hat{x}_{k,k-1} = f(\hat{x}_{k-1,k-1}) \tag{3.6}$$

$$P_{k,k-1} = F_{k-1} \cdot P_{k-1,k-1} \cdot F_{k-1}^T + Q_k \tag{3.7}$$

where

$$F_k = \left[\frac{\partial f}{\partial x} \right]_{\hat{x}=\hat{x}_{k,k}} \tag{3.8}$$

The filter gain matrix is given by

$$K_k = P_{k,k-1} H_k^T \left[H_k P_{k,k-1} H_k^T + R_k \right]^{-1} \tag{3.9}$$

3.2 Stability of the Extended Kalman Filters

When referring to the term stability, it should be mentioned which kind of stability one has in mind. In this thesis, we mostly refer to stability of an equilibrium point. It is common to assume that a nonlinear system has its equilibrium in the origin, i.e. $f(0, k) = 0$. Some fundamental stability concepts of the equilibrium point $x = 0$ are defined as follows (see e.g. [20] or [37]):

Definition 3.1. *The equilibrium point $x = 0$ is*

1) stable if, for each $\epsilon > 0$ and each $k_0 \in Z_+$, there exists a $\delta = \delta(\epsilon, k_0)$ such that

$$\|x_0\| < \delta(\epsilon, k_0) \Rightarrow \|x_k\| < \epsilon, \quad \forall\, k \geq k_0 \tag{3.10}$$

2) uniformly stable if, for each $\epsilon > 0$*, there exists a* $\delta = \delta(\epsilon)$ *independent of* k_0*, such that*

$$\|x_0\| < \delta(\epsilon), \;\; k_0 \geq 0 \quad \Rightarrow \|x_k\| < \epsilon, \qquad \forall\, k \geq k_0 \tag{3.11}$$

3) uniformly asymptotically stable if it is uniformly stable and there exists a positive constant η*, independent of* k_0*, such that*

$$\|x_0\| < \eta, \;\; k_0 \geq 0 \quad \Rightarrow \|x_k\| \to 0 \text{ as } k \to \infty \tag{3.12}$$

4) exponentially stable if there exists constants $a, b, c > 0$ *such that*

$$\|x_k\| \leq a\|x_0\|e^{-bk}, \forall\; \|x_0\| < c \tag{3.13}$$

In the noise free case ($w = v = 0$) of the EKF we refer to local (exponential) stability of the equilibrium point. When noise is present, this definition is no longer valid. In this case, it is more correct to use the term bounded rather than stable. In this thesis the term EKF stability refers to both stability of the equilibrium point of the EKF's error dynamic and, when noise is present, boundedness of the estimation error.

3.2.1 Stability by Lyapunov analysis and the total stability theorem

In the linear case, it can be shown that the Kalman filter is stable regardless of the initial error and the size of the noise processes, provided that the signal model is both observable and controllable, see e.g. [2]. When turning to the extended Kalman filter, this picture changes dramatically. Theoretical results presented so far are very conservative and are therefore of limited practical interest. Stability has been possible to prove only if the initial error and noise processes are so small that they in practice are absent, see e.g. [28]. However, it is of course of great importance to actually know that stability can be guaranteed also for the extended Kalman filter, even if the results are conservative.

In Paper 4 and 5 we present results which extends the results published so far on EKF stability. In the special case when the state equation is linear, we show that stability can be guaranteed for initial errors and noise processes much larger than proved earlier. Furthermore, in the general case we show that stability can be proved without requiring the matrix

$$H_k = \left[\frac{\partial h}{\partial x}\right]_{\hat{x}=\hat{x}_{k,k-1}} \tag{3.14}$$

to be bounded in norm as required earlier, provided that the Hessian matrix

$$\mathrm{Hess}(h_k) = \left[\frac{\partial^2 h}{\partial x^2}\right]_{\hat{x}=\hat{x}_{k,k-1}} \tag{3.15}$$

is bounded for all $x \in R^n$. This also allows the signal model to be unstable in certain cases. These cases are not handled in proofs presented earlier.

Before we turn to the general filter case where noise is present in both the state and the measurement equation, a short overview is given of results presented so far on EKF stability when the EKF is used as a state observer for a deterministic system.

When used as a state observer for the deterministic system

$$\begin{aligned} x(k+1) &= f(x(k), k) && (3.16) \\ y(k) &= h(x(k)) && (3.17) \end{aligned}$$

it can be shown that the EKF converges very quickly when the filter algorithm is slightly modified, or when the filter tuning matrices R_k and Q_k are chosen in a certain way. See [29] and [6]. These two papers handles the problem in a different manner, although the effect on the filtering algorithm is quite similar. Consider the covariance time-update equation

$$P_{k,k-1} = F_{k-1}P_{k-1,k-1}F_{k-1}^T + Q_k \tag{3.18}$$

In [29] this equation is replaced by

$$P_{k,k-1} = \alpha^2 F_{k-1}P_{k-1,k-1}F_{k-1}^T + Q_k \tag{3.19}$$

By a standard Lyapunov argument, using the Lyapunov function

$$V(e_k) = e_k^T P_{k,k}^{-1} e_k^T \tag{3.20}$$

where $e_k = x_k - \hat{x}_{k,k}$ is the estimation error at time k, it is shown that the observer under certain conditions is exponentially stable and that the rate of convergence depends upon the choice of the constant α such that a large α results in fast convergence. That is, an observer with a prescribed degree of stability.

In [6] the matrices R_k and Q_k are shown to be the key for obtaining fast convergence. Consider again equation (3.18). Now the matrix Q_k is varied as a function of the filter innovation, such that

$$P_{k,k-1} = F_{k-1}P_{k-1,k-1}F_{k-1}^T + Q(e_{k-1}) \tag{3.21}$$

One particular choice of Q_k mentioned in [6] is

$$Q_k = \alpha e_k^T e_k I + \beta I \tag{3.22}$$

where $e_k = y_k - h_k(\hat{x}_{k,k-1})$ and the identity matrix is of appropriate dimension $(n \times n)$.

Roughly speaking, the effect on the filter algorithm is the same for this two methods. When the error is large the covariance is large which in turn results in a large Kalman gain matrix. Therefore the filter responds quickly on any step in the state to be observed. This is in fact a very desirable property of any observer or filter, however, it seems to be difficult to transfer this approach to the filter case. For some reason the filter tends to diverge even for quite small modifications of the covariance matrix $P_{k,k-1}$, when it is done as shown in equation (3.19). The reason for this divergence is somewhat unclear, however, increasing the convergence speed, may lead to a more noise sensitive filter. One solution to avoid divergence is to apply the modification only temporarily, as in the method described in [6]. However, in the filtering case, the residual will be corrupted by noise and may therefore not be used without some sort of filtering or de-noising.

In Paper 4 we show that the following two items are crucial for obtaining useful theoretical stability results of the EKF.

- The lower and upper bounds of the Kalman gain matrix K_k and the matrix $[I - K_k H_k]$ should be as tight as possible.
- When choosing the filter tuning matrix Q_k, the stability properties must be taken into consideration.

It should be noted that Paper 4 only treats the case of linear state map and nonlinear measurement equation. The first item listed above will, however, be valid also for the general case.

The results obtained when letting Q_k be constant (and small) are tremendously conservative, even for almost linear signal models. With the term ”almost linear signal model” we refer to a nonlinear signal model with such a small nonlinearity that it would have been impossible to observe it for a physical system. When simulating such systems no differences compared to a corresponding linear signal model can be observed. An example of a ”almost linear signal model” is the following scalar system

$$x_{k+1} = a x_k + w_k \tag{3.23}$$

$$y_k = x_k + \eta x_k^2 + v_k \tag{3.24}$$

where $0 < \eta \ll 1$.

For an unfortunate choice of Q_k, stability of the EKF associated with this signal model can only be guaranteed theoretically for an η so small the it would have been virtually impossible to observe the term ηx^2 in a practical application (unless of course, the state is allowed to become extremely large). Alternatively, the initial error must be very small. More details about this example is found in Paper 4.

The stability proof in this work, as well as in [28] and [6], is based on the Lyapunov stability theorem (see e.g. [37]) using the Lyapunov function candidate

$$V(e_k) = e_k^T P_k^{-1} e_k \tag{3.25}$$

which has the following lower and upper bounds

$$\frac{1}{p_2}\|e_{k,k}\| \leq V(e_{k,k}) \leq \frac{1}{p_1}\|e_{k,k}\| \tag{3.26}$$

when we assume that

$$p_1 I \leq P_{k,k} \leq p_2 I \tag{3.27}$$

The error dynamics in the general case is given by (see Paper 4 or 5 for supplementary details)

$$e_{k,k} = \tilde{F}_k e_{k-1,k-1} + n_k + l_k \tag{3.28}$$

where:

$$\tilde{F}_k = \big[I - K_k H_k\big] F_{k-1} \tag{3.29}$$
$$n_k = \big[I - K_k H_k\big] w_{k-1} - K_k v_k \tag{3.30}$$
$$l_k = \big[I - K_k H_k\big] \theta_f(x_k, \hat{x}_{k,k}) + K_k \phi_h(x_k, \hat{x}_{k,k-1}) \tag{3.31}$$

Using 3.25 we obtain

$$\begin{aligned}\Delta V &:= e_k^T P_k^{-1} e_k - e_{k-1}^T P_{k-1}^{-1} e_{k-1}\\ &= \big(\tilde{F}_k e_{k-1} + n_k + l_k\big)^T P_k^{-1} \big(\tilde{F}_k e_{k-1} + n_k + l_k\big) - e_{k-1}^T P_{k-1}^{-1} e_{k-1}\\ &= e_{k-1}^T \Big[\tilde{F}_k^T P_k^{-1} \tilde{F}_k - P_{k-1}^{-1}\Big] e_{k-1} + n_k^T P_k^{-1} n_k + l_k^T P_k^{-1} \big(2\tilde{F}_k e_{k-1} + l_k\big)\\ &\quad + 2 n_k^T P_k^{-1} \big(\tilde{F}_k e_{k-1} + l_k\big)\end{aligned} \tag{3.32}$$

Requiring the term

$$\tilde{F}_k^T P_k^{-1} \tilde{F}_k - P_{k-1}^{-1} \tag{3.33}$$

to be negative definite for all k, yields exponential stability of the EKF in the absence of noise. Under the assumption that the covariance matrices $P_{k,k}$ and $P_{k,k-1}$ are bounded from below and above, and that the matrix F_k is non-singular and bounded from above, it can be showed that

$$\tilde{F}_k^T P_k^{-1} \tilde{F}_k - P_{k-1}^{-1} \leq (1-\gamma) P_{k-1}^{-1} \tag{3.34}$$

where $0 < \gamma < 1$. The proof is given in Paper 4 and 5. See also [28], Lemma 3.1.

Using (3.34) and taking the norm of the remaining terms, the following inequality is obtained (see Paper 4 and 5)

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi} V(e_{k-1}) + \rho(\bar{w}, \bar{v}) \tag{3.35}$$

where $\bar{w}$ and $\bar{v}$ are upper bounds of the process noise and measurement noise respectively and $\psi > 1$ and $\gamma < 1$ are positive real numbers.

The function ρ will be zero for $\bar{w} = \bar{v} = 0$, and thus exponential stability can be guaranteed for the noise free case, i.e.

$$\|e_{k,k}\| \leq \Gamma \|e_{0,0}\| \xi^{-k} \tag{3.36}$$

where $\Gamma > 0$ and $\xi > 1$, holds for $\|e_{0,0}\| \leq \epsilon$.

When noise is present, it follows that if $\bar{w}$ and $\bar{v}$ are sufficiently small, for some $\tilde{\epsilon} > 0$ ΔV will be negative definite for $0 < \tilde{\epsilon} \leq \|e_{0,0}\| \leq \epsilon$. That is, the rate of increase of V along the trajectory is negative definite. The error will therefore be bounded for all k. See Paper 4 and 5 for more details and an explicit expression of the error bound.

When trying to apply these results on examples, it becomes clear that the results will be too conservative in most cases. Therefore, the results obtained so far are mainly of theoretical interest. This is also reported in [28]. There may be several reasons for this conservatism, and in the following this will be discuss into more details.

The approximation in the EKF is only valid locally. It should therefore not be surprising that the initial error must bounded. What is surprising, however, is that this bound turns out to be very small.

If $\gamma \approx 0$ then the results will be very conservative, and this is exactly what happens if the matrix Q_k is not carefully chosen. In Paper 4 it is showed that the following choice of Q_k is favorable

$$Q_k = \delta \cdot F_{k-1} P_{k-1} F_{k-1}^T \tag{3.37}$$

This choice of Q_k yields

$$\gamma = 1 - \frac{1}{1+\delta} \tag{3.38}$$

so γ can be chosen arbitrarily in the interval $(0, 1)$. However, inserting (3.37) into (3.18) yields

$$P_{k,k-1} = \big(1+\delta\big) \cdot F_{k-1} P_{k-1,k-1} F_{k-1}^T \tag{3.39}$$

which is almost equal to equation (3.19), and thus may give a noise sensitive filter as the gain matrix is increased, even if the conservatism regarding the initial error is taken care of. This leads to the following paradox: it is reasonable to expect that a well tuned filter (as close as possible to optimality) will ensure that upper bounds on initial error and noise processes are close to their maximum values, however, there are no evidence in the analysis that this is a reasonable assumption. It is observed for some cases that the largest upper bounds for the initial error and noise processes are obtained when the filter is badly tuned.

An alternative proof of EKF convergence is reported in [21]. This paper is restricted to the case where the signal model has a nonlinear state equation and a linear output map. The proof is based on the *total stability theorem* (see [1]). The total stability theorem requires the equation in question to have an exponential stable linear part, i.e.

$$\|\Phi_A(k,j)\| \leq \Gamma \lambda^{k-j} \tag{3.40}$$

where $\Gamma \geq 1$ and $0 \leq \lambda < 1$ are finite positive constants and $\Phi_A(k,j)$ is the transition matrix for the linear state equation

$$x(k+1) = A(k)x(k) \tag{3.41}$$

Theorem 3.1 (Total stability theorem). *Assume that the linear state equation*

$$x(k+1) = A(k)x(k) \tag{3.42}$$

is exponential stable and $A(k)$, $f(k,x)$ and $g(k,x)$ are all fixed in $\|x\| \leq \epsilon_r$ for each k. Let $f(k,x)$ and $g(k,x)$ be $n \times 1$ vector functions which satisfy the following

$$f(k,0) = 0 \tag{3.43}$$

$$\|f(k,x_1) - f(k,x_2)\| \leq \alpha \|x_1 - x_2\| \tag{3.44}$$

$$\|g(k,x_1)\| \leq \beta \epsilon_r \tag{3.45}$$

$$\|g(k,x_1) - g(k,x_2)\| \leq \beta \|x_1 - x_2\| \tag{3.46}$$

Then $\|x_0\| \leq \epsilon_r/\alpha$ *and* $\Gamma(\alpha+\beta)+\lambda < 1$, *where* Γ *and* λ *are given by (3.40), implies that the solution of*

$$x(k+1) = A(k)x(k) + f(k,x) + g(k,x) \tag{3.47}$$

will be bounded by

$$\|x(k)\| \leq \Gamma(\lambda+\alpha\Gamma)^k\|x_0\| + \frac{\Gamma\beta\epsilon_r}{1-(\lambda+\alpha\Gamma)} \leq \epsilon_r \tag{3.48}$$

Proof: See [1]. □

Comparing this result with the result obtained by Lyapunov analysis may be of interest. Consider the rate of convergence in the deterministic case. From inequality (3.35) we obtain, when following the same track as in [29]:

$$V(e_k) - V(e_{k-1}) \leq \frac{\gamma(1-\psi)}{\psi}V(e_{k-1}) \tag{3.49}$$

Thus

$$V(e_k) \leq \frac{\gamma(1-\psi)+\psi}{\psi}V(e_{k-1}) \tag{3.50}$$

Therefore

$$V(e_k) \leq V(e_0)\left(\frac{\gamma(1-\psi)+\psi}{\psi}\right)^k = V(e_0)\left(\frac{\psi}{\gamma(1-\psi)+\psi}\right)^{-k} \tag{3.51}$$

Using (3.26) we obtain

$$\|e_{k,k}\| \leq \Gamma_1\|e_{0,0}\|\xi_1^{-k} \tag{3.52}$$

where $\Gamma_1 = \sqrt{\frac{p_2}{p_1}} \geq 1$ and $\xi_1^2 = \frac{\psi}{\gamma(1-\psi)+\psi} > 1$, since $\gamma < 1$ and $\psi > 1$.

Using the result from the total stability theorem we obtain

$$\|e_{k,k}\| \leq \Gamma\|e_{0,0}\|(\lambda+\alpha\Gamma)^k = \Gamma\|e_{0,0}\|\left(\frac{1}{\lambda+\alpha\Gamma}\right)^{-k} = \Gamma\|e_0\|\xi^{-k} \tag{3.53}$$

where $\Gamma \geq 1$ by assumption. Obviously, this theorem collapse if $\alpha \geq 1$, which is a serious limitation as a large α is a result of a large nonlinearity. For modest and small nonlinearities the theorem will work. This is consistent with the result obtained from the Lyapunov analysis. It should be mentioned, as a final remark on the total stability theorem, that since this theorem requires the linear part of the error dynamics to be exponential stable, i.e.

$$e_{k,k} = [I - K_k H_k] F_{k-1} e_{k-1,k-1} \tag{3.54}$$

is required to be exponential stable, Lyapunov analysis must be used anyway. It is only the part of the stability proof which defines the upper bounds on the initial error and noise processes that can be derived by the total stability theorem. Therefore, any conservatism in the results from the Lyapunov analysis, may transfer to the results obtained by use of the total stability theorem.

3.2.2 Stochastic stability of the EKF

In [28] stochastic stability of the EKF is considered. Then the noise processes are assumed to be zero-mean white noise rather than bounded in ∞-norm. The analysis is based on standard results for convergence of a positive supermartingale (see e.g. [14]). A significant difference between the work in [28] compared with the work in e.g. [21] and the present thesis, is the formulation of the EKF. Two common formulation of the EKF exists in the literature: a one-step formulation in terms of the *a-priori* variables and a two-step formulation consisting of time-update and measurement-update with a re-linearization between these two steps. In [28] the one-step formulation is used. This is indeed a great advantage when considering stochastic stability since all variables are defined at only one time instant k, and not at time k, k and $k, k-1$, as is the case in the two-step formulation. When using the formulation in terms of the a-priori variables, the inequalities developed during the Lyapunov analysis simplifies considerably by use of the white-noise property. This fails when using the two-step formulation because higher moments will be included in the inequalities.

Consider the EKF associated with the signal model (3.1)-(3.2) formulated in terms of the a-priori variables (see e.g. [14])

$$\hat{x}_{k+1} = f(x_k) + K_k \left(y_k - h(\hat{x}_k) \right) \tag{3.55}$$

$$P_{k+1} = F_k P_k F_k^T + Q_k - K_k \left(H_k P_K H_k^T + R_k \right) K_k^T \tag{3.56}$$

where

$$F_k = \frac{\partial f}{\partial x}(\hat{x}_k) \qquad \text{and} \qquad H_k = \frac{\partial h}{\partial x}(\hat{x}_k) \tag{3.57}$$

and the Kalman gain matrix is given by

$$K_k = F_k P_k H_k^T \left(H_k P_k H_k^T + R_k \right)^{-1} \tag{3.58}$$

The error dynamics is given by (see [28]):

$$e_{k+1} = (F_k - K_k H_k)\, e_k + n_k + l_k \tag{3.59}$$

where

$$n_k = w_k - K_k v_k \tag{3.60}$$

$$l_k = \theta_f(x, \hat{x}) - K_k \phi_h(x, \hat{x}) \tag{3.61}$$

(see [28] for more detail about the functions $\theta_f(x, \hat{x})$ and $\phi_h(x, \hat{x})$).

Using the Lyapunov function (3.25) it can be shown that

$$\begin{aligned} V(e_{k+1}) - V(e_k) \leq (1-\gamma)V(e_k) + l_k^T P_{k+1}^{-1} \left[2(F_k - K_k H_k)e_k + l_k\right] + n_k^T P_{k+1}^{-1} n_k \\ + 2n_k^T P_{k+1}^{-1} \left[(F_k - K_k H_k)e_k + l_k\right] \end{aligned} \tag{3.62}$$

When taking the conditional expectation $E\{V(e_{k+1})|e_k\}$ the term $E\{2n_k^T P_{k+1}^{-1} \left[(F_k - K_k H_k)e_k + l_k\right]|e_e\}$ will vanish since none of the terms depends on the noise processes.

For the two-step formulation this is quite different. Consider the error dynamics given by (see Paper 4 and 5 for details)

$$e_{k,k} = \left[F_{k-1} - K_k H_K F_{k-1}\right] e_{k-1,k-1} + n_k + l_k \tag{3.63}$$

where n_k and l_k is given by (3.30) and (3.31) respectively.

Using the same Lyapunov function we obtain, as before

$$\begin{aligned} V(e_k) - V(e_{k-1}) \leq (1-\gamma)V(e_{k-1}) + l_k^T P_k^{-1} \left[2(F_{k-1} - K_k H_k)e_k + l_k\right] + n_k^T P_k^{-1} n_k \\ + 2n_k^T P_k^{-1} \left[(F_{k-1} - K_k H_k)e_k + l_k\right] \end{aligned} \tag{3.64}$$

Now the functions $\theta_f(x_{k-1}, \hat{x}_{k-1})$ and $\phi_h(x_k, \hat{x}_{k,k-1})$ (remainder terms from Taylor expansion) are taken at different time, so before taking the conditional expectation this must be corrected. The correction is easily done by using the following inequality

$$\|e_{k,k-1}\| \leq \vartheta \|e_{k-1,k-1}\|^2 + f\|e_{k-1,k-1}\| + w^2 \tag{3.65}$$

The term $2n_k^T P_{k+1}^{-1} \left[(F_k - K_k H_k)e_k + l_k\right]$ contains the variance of the process noise, which is not zero. Higher moments, which meaning may be unclear, will also appear in the equations. Therefore, taking the conditional expectation $E\{2n_k^T P_{k+1}^{-1} \left[(F_k - K_k H_k)e_k + l_k\right]|e_e\}$ and assuming the noise processes to be zero-mean and white, will not yield any simplification as in the case of the EKF formulated by the a-priori variables. Furthermore, the higher moments $E\{w\,\mathrm{cov}(w)\}$ and $E\{\mathrm{cov}(w)\mathrm{cov}(w)\}$ which will be included in the analysis, make the picture more blurred than when assuming the noise processes to be bounded in ∞-norm.

3.3 Extended Kalman filter tuning

The filter tuning process is normally a time consuming and cumbersome task. This process is also complicated by the fact that it is difficult to give a physical interpretation of the different parameters in the covariance matrix. For a practical application, especially for a higher order signal model, these two problems may be a significant obstruction for applying an extended Kalman filter, and for some application this has been a motivation for developing other techniques. One example of importance is observers for velocity and wave frequency motions for ships, see [12].

Recently, two different techniques for automatic tuning of the EKF have been proposed, see [26] and [25]. In [26] the use of a simplex downhill method is described. This method will not be used in this work, so no further description is given. In [25] the use of Genetic Algorithms (GA) for EKF tuning is discussed. This method turns out to be very flexible and easy to implement. In this work the free MATLAB toolbox GAOT (see [17]) is used. Genetic algorithms in general is shortly described in Appendix 1.

During the tuning process the desired properties of the filter must be taken into account. Doing this manually is certainly not a trivial task, especially for filters of high order. Some rules of thumb on how the different parameters should be chosen may be given, but this is far from sufficient for obtaining a well tuned filter. In addition, the design has to be checked by simulations for each choice of tuning parameters. When tuning the EKF by use of an genetic algorithm, this is rather easy to implement. In this respect the GA is an advanced numerical optimization tool, and as in most numerical optimization procedures, different solutions are obtained by putting weights on the different parts of the function in question. In this application the function to be minimized (or maximized) is a performance function based on the estimation error for the different states, see Paper 2 and 3 for details. Then different properties are obtained by putting different weights on each component of the performance function. It is concluded in Paper 2 and 3 that genetic algorithms is a tool well suited for tuning the EKF.

3.4 EKF with high noise rejection and quick convergence

A reasonable tuning objective of the EKF in a high noise environment is to maximize the filters noise rejection. This leads to a filter with low variance in the estimated states. The drawback is that the filter will react slowly on changes in the states to be estimated, such that the estimation error will become large for quite a long time if a quick change in the state should occur. In this section an ad-hoc method which will solve this problem, provided that a reliable method for detecting the changes is available, is described. Methods for detecting changes in the state are described in e.g. [16].

3.4.1 Modification of the EKF algorithm

The method is based on the idea presented in [29], with the important difference that the parameter α is no longer a constant. The covariance time update equation is now given by

$$P_{k,k-1} = \alpha(k) F_{k-1} P_{k-1,k-1} F_{k-1}^T + Q \tag{3.66}$$

where $\alpha(k)$ is a real valued function defined for all $k > k_0$ and:

$$\alpha(k) \geq 0 \tag{3.67}$$

$$\alpha(k) \geq \alpha(k+1) \tag{3.68}$$

$$\lim_{k \to \infty} \alpha(k) \longrightarrow 0 \tag{3.69}$$

The set of functions candidates $\alpha(k)$ which are considered in this section is:

$$\alpha_1(k) = \alpha_0 \frac{1}{k - k_0} \tag{3.70}$$

$$\alpha_2(k) = \alpha_0 \frac{1}{(k - k_0)^2} \tag{3.71}$$

$$\alpha_3(k) = \alpha_0 \frac{1}{\sqrt{k - k_0}} \tag{3.72}$$

There is of course a great possibility that none of these functions are very well suited for this purpose, so better functions may be found by searching more carefully. However, the functions listed are simple and as such desirable.

The filter tuning matrix Q is kept constant and is chosen in such a way that the variance in the filtered state is kept as low as possible. The filters slow transient response is compensated for by letting $\alpha(k)$ be greater then zero when a step, or a major deviation, in the real state occurs, and then decrease to zero when k increases. With this method the gain matrix will be increased for a short period after a step, and then converge to the original gain matrix without destabilizing the filter.

3.4.2 Results from simulations

To illustrate the idea described in the previous subsection, a set of simulations has been carried out. In this section four different cases are considered. The first case is the unmodified filter. The second case is the filter modified with the function $\alpha_1(k)$. The third case is the filter modified with the function $\alpha_2(k)$, and in the fourth it is modified with $\alpha_3(k)$. The filter used in this cases, is a filter design for tracking a low frequency oscillating signal. The signal model is given by

$$x_{k+1} = Ax_k + Bw_k \tag{3.73}$$

$$y_k = h(x_k) + v_k \tag{3.74}$$

where $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$

and

$$h(x_k) = x_k^{(3)} \sin(x_k^{(2)}) \tag{3.75}$$

where x_1 is the phase increment (or frequency), x_2 is the phase and x_3 is the amplitude of the oscillation.

For each case a step in the amplitude occur at $k = 1500$. In all cases considered the filter parameters are: $R = 1$ and $Q = \text{diag}\{1 \cdot 10^{-10}, 1 \cdot 10^{-12}\}$.

The search for α_0 is one-dimensional, so no advanced search method is required. In the simple cases considered in this section a manual search may even be sufficient. This filter is tuned for working at low SNR values (down to $SNR = -10$ dB) and it is also assume that every change in the states will be slow. In all simulations, the signal-to-noise ratio (SNR) is $SNR = -7.8$ dB.

Unmodified filter algorithm

In Figure 3.1 and 3.2 the estimation error and real and estimated amplitude is shown.

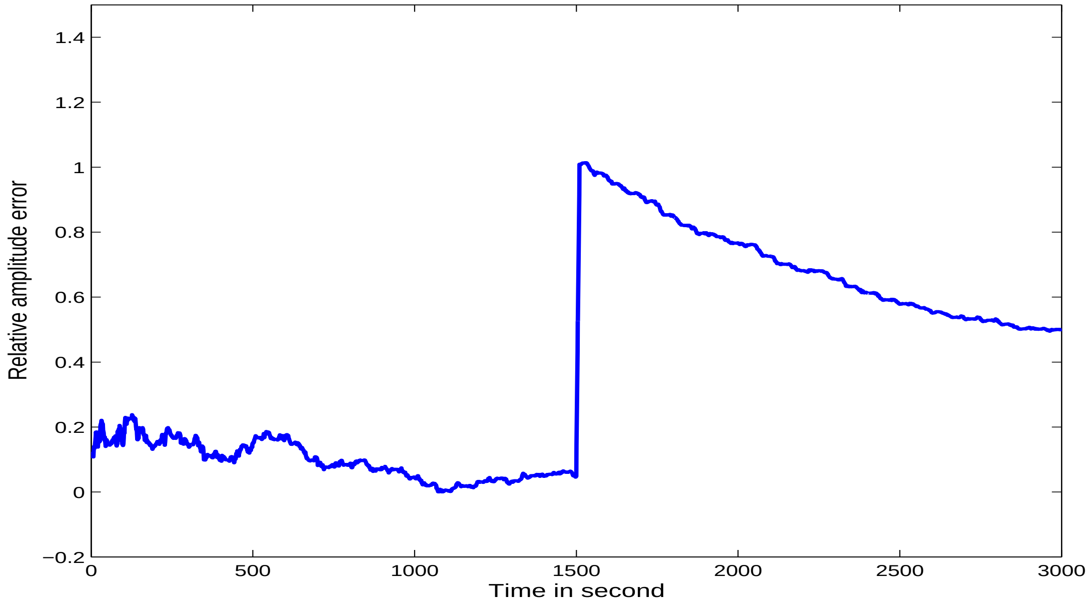

Figure 3.1: Estimation error

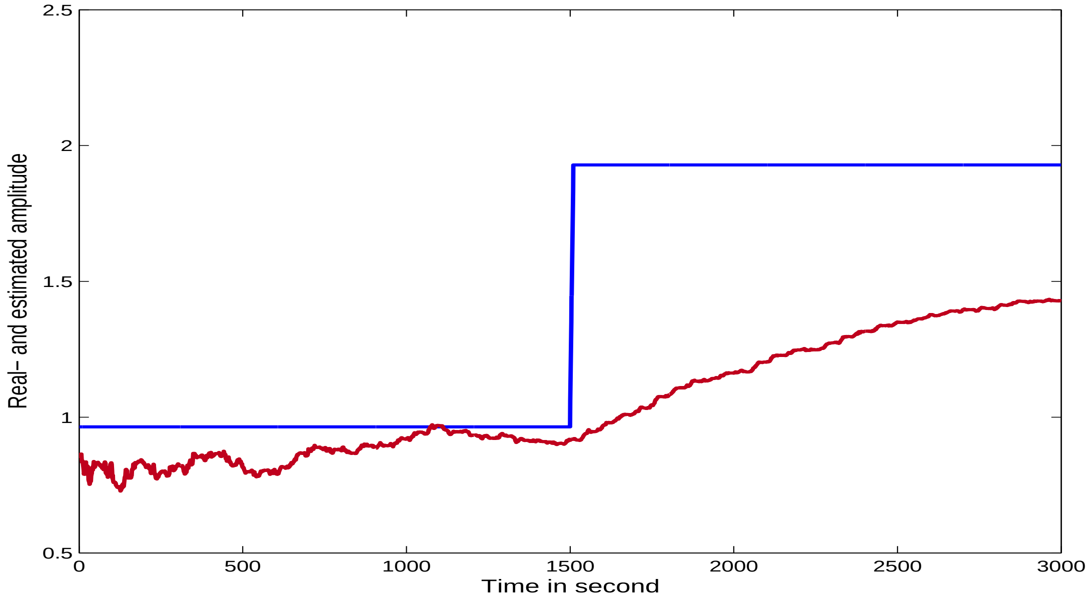

Figure 3.2: Real and estimated amplitude

Taking into account the high level of noise, the estimate is rather noise free. On the other hand it is obvious that this filter will not perform satisfactory unless the states to be estimated are almost constant. The response time after initialization and the step at $k = 1500$ is tremendously long. However, this filter was not tuned for quick response, so this is as expected.

Filter algorithm modified with $\alpha_1(k)$

In Figure 3.3 and 3.4 the estimation error and real and estimated amplitude is shown for a filter which is modified with the function $\alpha_1(k)$, as shown in equation (3.70).

The function $\alpha_1(k)$ is given by:

$$\alpha_1(k) = \alpha_0 \frac{1}{k - k_0} = 2.2 \cdot \frac{1}{k - 1499} \tag{3.76}$$

Compared to the previous case, the difference in error convergence time is quite dramatic; from still having more than half the value of the error after 1500 samples to compensating for most of the step in less than 100 samples. The filter is still tuned for working at the same level of noise, while at the same time the transient response is very satisfactory.

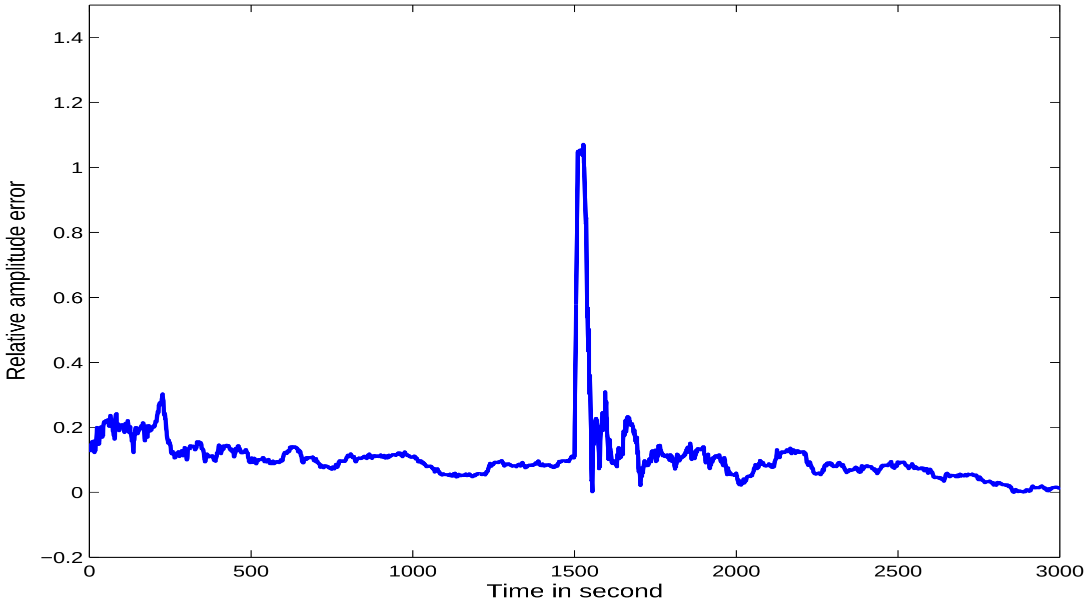

Figure 3.3: Estimation error

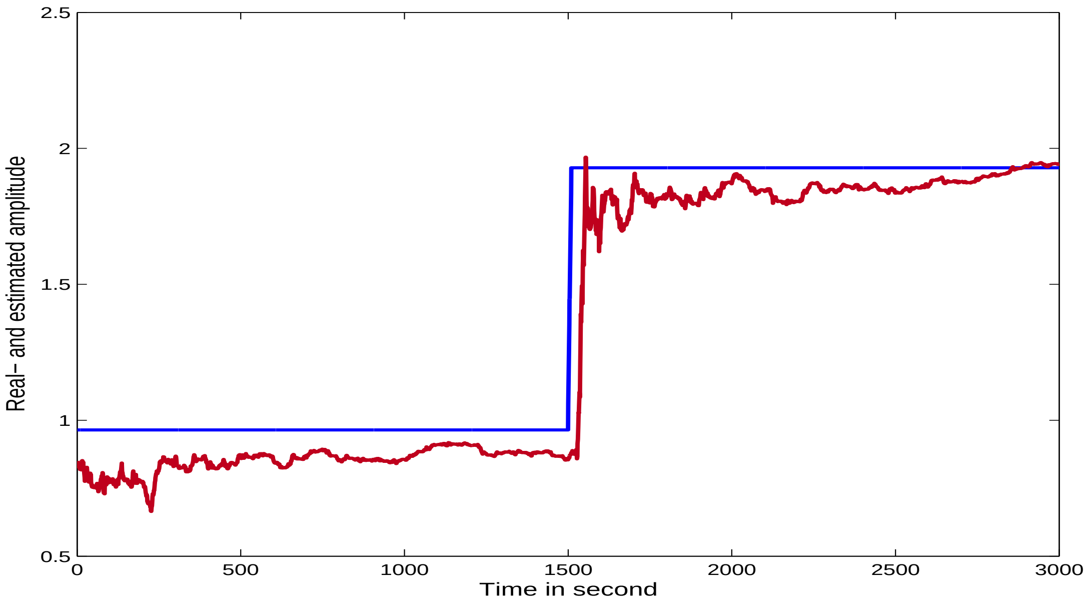

Figure 3.4: Real and estimated amplitude

The drawback is a more noisy estimate close after any step or quick change in the states. For steady state, this filters performance is equal to the previous one. This is also clearly illustrated by Fig. 3.4. Until the step occur the noise components in the estimate is on the same level as seen in Figure 3.2, immediately after the step it is considerably higher, but settles in about 500 samples.

Filter algorithm modified with $\alpha_2(k)$

In Figure 3.5 and 3.6 the estimation error and real and estimated amplitude is shown for a filter which is modified with the function $\alpha_2(k)$, as shown in equation (3.71).

The function $\alpha_2(k)$ is given by:

$$\alpha_2(k) = \alpha_0 \frac{1}{(k-k_0)^2} = 25 \cdot \frac{1}{(k-1499)^2} \tag{3.77}$$

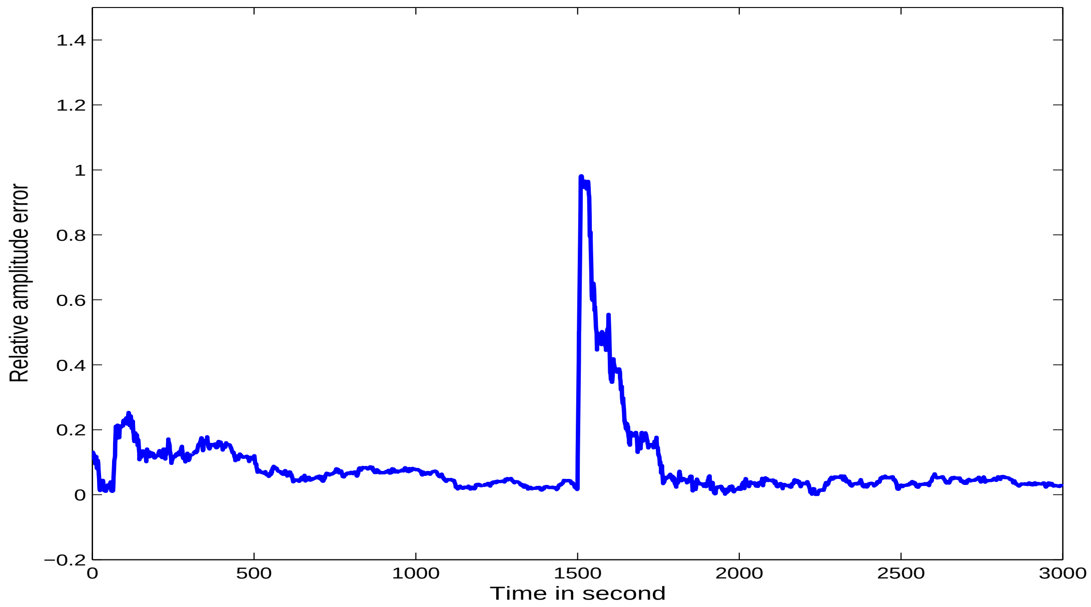

Figure 3.5: Estimation error

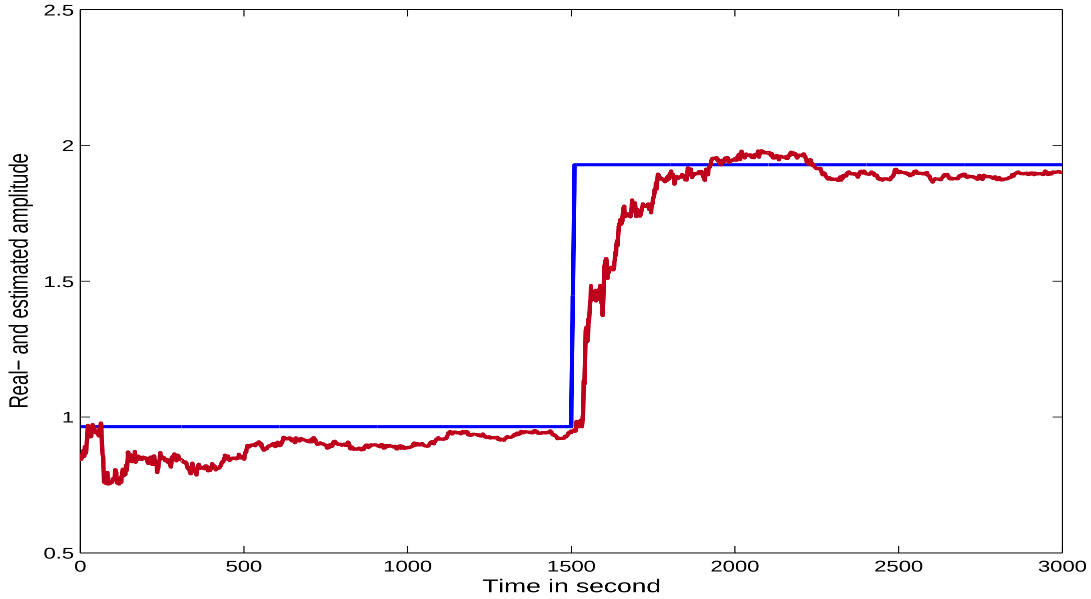

Figure 3.6: Real and estimated amplitude

The effect from $\alpha_2(k)$ is almost the same as for $\alpha_1(k)$. The speed of convergence after a step is dramatically increased. However, the value of $\alpha_2(k)$ is decreasing much faster than $\alpha_1(k)$. Moreover, if the initial value is too large the estimate will contain some large ”spikes.” It seems to be difficult to obtain the same convergence speed with $\alpha_2(k)$ as with $\alpha_1(k)$. This may indicate that the value of $\alpha_2(k)$ diminishes too quickly.

Filter algorithm modified with $\alpha_3(k)$

In Figure 3.7 and 3.8 the estimation error and real and estimated amplitude is shown for a filter which is modified with the function $\alpha_3(k)$, as shown in equation (3.72).

The function $\alpha_3(k)$ is given by:

$$\alpha_3(k) = \alpha_0 \frac{1}{\sqrt{k-k_0}} = 0.30 \cdot \frac{1}{\sqrt{k-k_0}} \tag{3.78}$$

As in the two previous cases, the error converges fast down to its bound, even if α_0 in this case has a surprisingly low value.

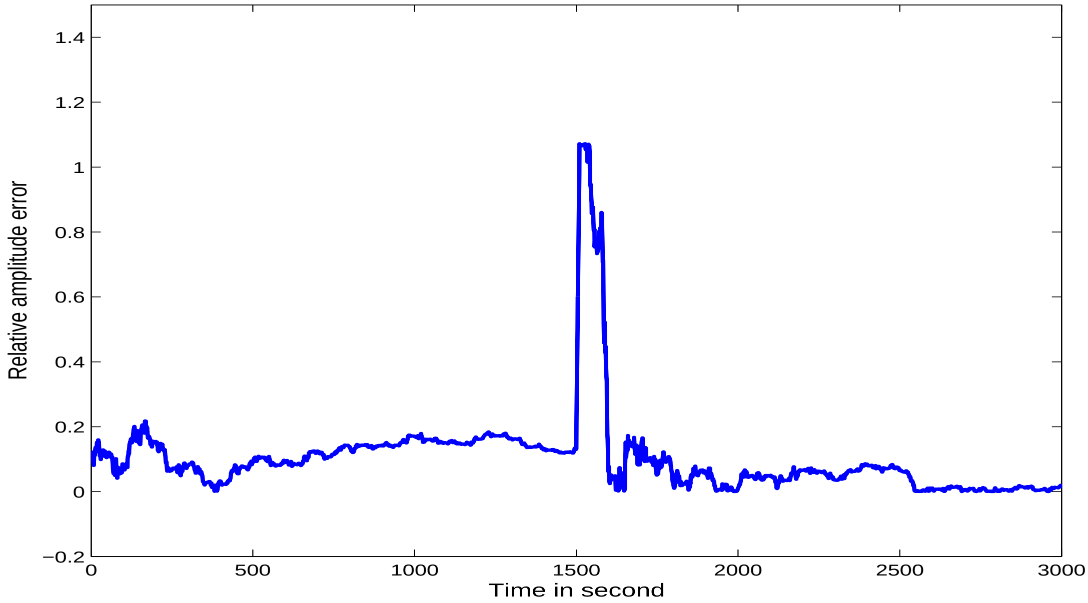

Figure 3.7: Estimation error

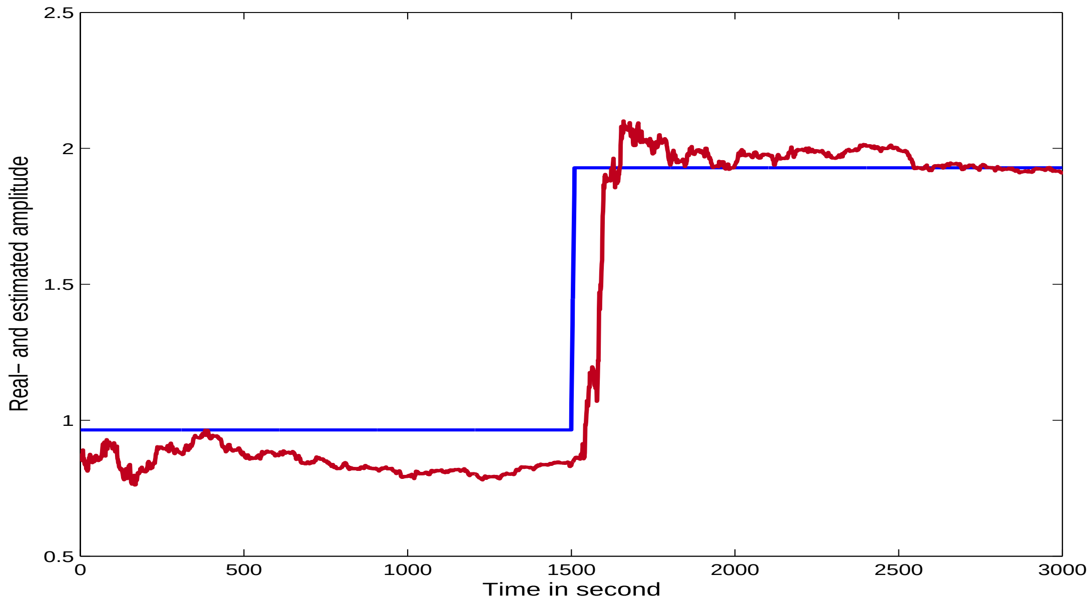

Figure 3.8: Real and estimated amplitude

Discussion

In this subsection it is shown how a specific EKF may be modified, by temporarily increasing the Kalman gain matrix, to obtain fast transient response even if it is tuned for a high noise environment. It should be emphasized that it is important that the gain matrix is modified for a short period only. A permanent modification, as in the observer case, will destroy the filters performance for low signal-to-noise ratios (SNR) and may lead to divergence of the estimation error. This is demonstrated in Fig. 3.9 and 3.10, which shows the estimation error for the filter when it is modified as shown in equation (3.19) with a constant $\alpha = 1.2$. Obviously this results in a filter useless for such low SNR values.

When comparing the different cases, it is not clear which function α_1, α_2 or α_3 gives the best result. However, some differences are evident. When using a fast decreasing function $\alpha(k)$ its initial value should be chosen large. This will initially increase the gain matrix quite a lot, but only for a very limited number of samples. A slowly decreasing function will change the gain matrix less, but for a longer time. This time frame is of course an important parameter, and for the particular filter used as example in this section, a function $\alpha(k)$ converging faster than $\alpha_2(k)$ will result in too slow convergence.

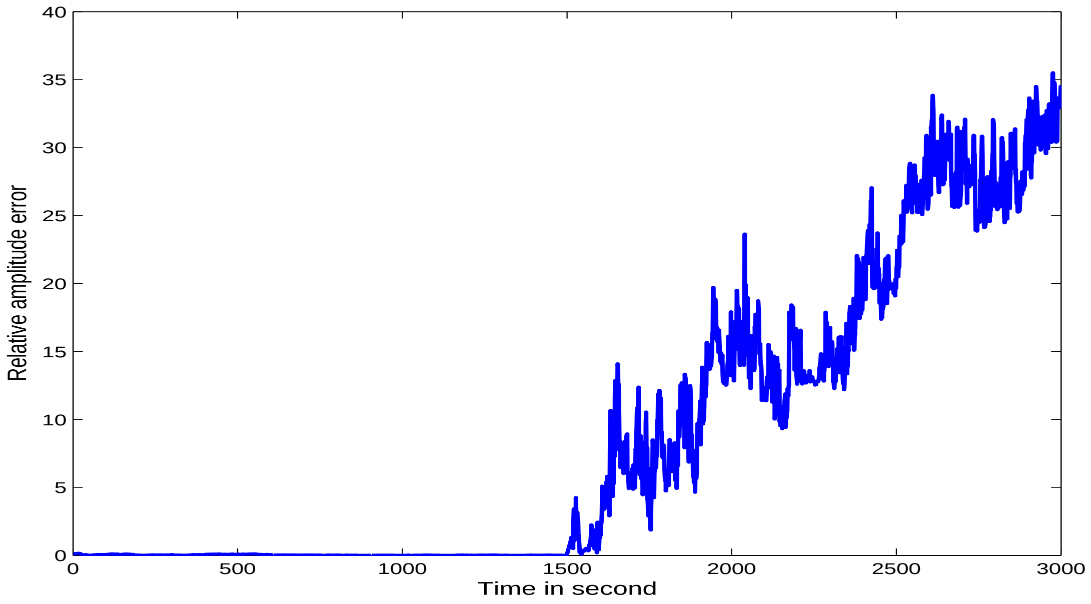

Figure 3.9: Estimation error

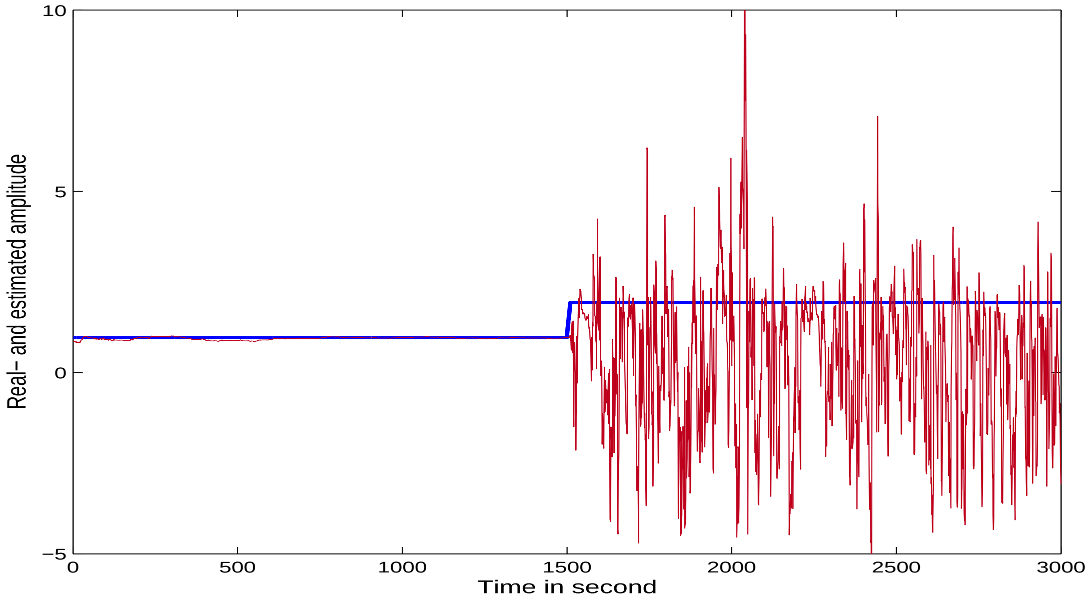

Figure 3.10: Estimation error

One might wonder if some measure of $\alpha(k)$, like the l_1-norm over the time interval as given by

$$\sum_{i=k_0}^{K} \alpha(k) \tag{3.79}$$

where $\alpha > 0$ for $0 \leq k \leq K$ and $\alpha = 0$ for $k < k_0, \quad k > K$, should be the same for all function candidates if the result should be comparable in some sense, i.e.

$$\sum_{i=k_0}^{K} \alpha_1(k) = \sum_{i=k_0}^{K} \alpha_2(k) = ... = \sum_{i=k_0}^{K} \alpha_n(k) \tag{3.80}$$

where $\alpha_1 ... \alpha_n$ are different function candidates.

Using such a measure, or a similar one, would make it more easy to apply different functions $\alpha(k)$ if some special forms of these functions should be more desirable than others. Then the only needed value is the measurement value, and a desirable function fulfilling this requirement could be designed. From the experimental results presented in this thesis, there is not possible to conclude that such a relation exists.

3.4.3 Some implementation issues

In the previous sections the subject of speeding up the convergence rate while the filter is still able to work in the same level of noise, has been treated somewhat too simply, as some items, which may become very evident during an implementation phase, are not mentioned. In the following, some important ones of these are listed.

1) The filter tuning may e.g. be carried out by use of an genetic algorithm (see Paper 2 and 3 for details). The filter performance index should include some penalty for very quick changes of the states to avoid big ”spikes” in the amplitude estimate. The following type of function has been tested and found to work:

$$J(q_1, q_2, \alpha) = \left[\frac{1}{T+1} \sum_{t=0}^{T} \left(\hat{e}_t^T W \hat{e}_t \right) \right] \tag{3.81}$$

where $\hat{e} = [\hat{e}_1, \hat{e}_2]$, with components: $\hat{e}_1 = [(\hat{x}_1 - x_1), (\hat{x}_2 - x_2), (\hat{x}_3 - x_3)]$ (which is the estimation error vector), $\hat{e}_2 = \sum_{t=0}^{T} |\hat{x}(t) - \hat{x}(t-1)|$ for all $|\hat{x}(t) - \hat{x}(t-1)| > a$ (which punishes too large ”derivatives” of the states), T is the final time, which equals the number of samples and $W = \text{diag}\,(w_1 \;\, w_2 \;\, w_3 \;\, w_4)$ is the 4x4 weighting matrix.

2) During the filter tuning it is assumed that only one step in the states will occur within the given time frame. In practice, it is of course not possible to guarantee that several steps will not occur within any given sufficiently large time interval. This situation will require some additional criterion in the filtering algorithm to avoid that the increase in gain does not destabilize the filter.

3) The property given by (3.69) may very well be tightened. After some time the gain will be almost equal to the one without a modifying function $\alpha(k)$ so there is only waste of capacity to continue the calculation of the function $\alpha(k)$. For $\alpha_1(k)$ and $\alpha_3(k)$ α_0 can be set equal to zero after 400-600 samples, and for $\alpha_2(k)$ even earlier.

4) How to discover that a step actually has occurred? This is not a trivial question. An overview and description of possible techniques is outside the scope of this thesis, however, a simple method like the CUSUM test may be effective. For a description of this and other methods for change detection, the reader may consult e.g. [16].

5) Even if there may be difficult to discover a step in the state, the difference between the actual state and the estimated state after initialization can be regarded as caused by a step. This method may therefore be applied a short time after the starting time to faster obtain a reliable estimate.

Chapter 4

Balling drum operation and control

4.1 Balling drum operation

When the balling drums were introduced in the iron ore industry, each drum was manually controlled by one dedicated operator. They gain valuable experience in how to operate the drum in a efficient and safe manner, and some rule of thumbs can be stated based on this experience (Ref. R. Drügge, LKAB).

1) If the drum starts to surge, which means that the onsize and undersize fractions starts to fluctuate, the surging is damped out by adding some water into the drum.

2) When the drum is operated close to the point when surging may be expected to occur, the quality of the final pellets is good

This two rules of thumb is confirmed by later research, see e.g. [30] and [10]. Operating the drums in accordance with Rule 2) may be an tiring job for the operator, so during the day the quality would normally vary (Ref. R. Drügge, LKAB).

Today the picture is quite different. By use of monitoring systems, one operator is now able to control a whole plant (both the cold and warm part) from a centralized control room. Unfortunately, the level of automation when it comes to operating the drum, has not changed dramatically. The ”finger tip” feeling the earlier operators trusted, is no longer a part of the operators tool box due to the long distance from the operator to the drum. The drum itself is highly nonlinear, and it has been claimed from the operators that each drum has its own characteristic and may be rather unpredictable. A quite normal situation today is therefore that the amount of moisture in the fines is kept above the necessary level in order to avoid surging. This may reduces the risk that something undesirable happens, like the stop of a drum, but leads to reduced product quality, and is suboptimal also from the productivity point of view.

It is also seen that the moisture level is kept far to high, and that this is compensated for by adding more binder. This situation is very undesirable for at least four reasons (see also [30] and [10]):

- The extra water must be removed before the pellets enter the firing process to avoid that they crack. This is done by evaporation, which is very expensive compared to removing water in the press filters
- Binder contaminate the iron ore, and thus the level should be kept low. Too high level of water and binder gives poor product quality
- Binder is expensive so it is uneconomical to add more than necessary for obtaining suitable mechanical strength of the green pellets
- The drums total through-put is reduced

An automatic control scheme which takes care of the drum operation is therefore highly demanded, however, no such which works satisfactory is available today.

4.2 Previously research on balling drums

In the whole post war time, research in the field of pelletizing has been carried out, and some old references on balling drums goes back to the late forties, see e.g. [10] and the references therein. However, the intensity has been highly varying over the decades. The seventies seems to have been a very productive decade, and most of the fundamental work on balling drum modelling and control were carried out during this decade. After this time only a few scientist has found interest in this field, which is illustrated by the low number of publications dated after 1980.

The following subsection is dedicated to some earlier obtained essential results in modelling and control of the balling drums.

4.2.1 Modelling of balling drums

Model 1

During the seventies two different models for describing the drums dynamics were developed. The first one is due to K. V. S. Sastry (see [30] or [31]) and is a population model based on the number of pellets in the drum. During the sixties and the seventies the different growth mechanisms which takes place in a balling drum and disk were investigated and described, see e.g. [30] and [33], and this forms the basis for this model. Unfortunately it is of infinite order, and as such not suitable for control purposes.

The model is given by the following set of integral equations:

$$\begin{aligned}\frac{\partial n\left(m;\underline{x},t\right)}{\partial t} &+ \nabla\ \left[\underline{v}n\left(m;\underline{x},t\right)\right] = -\overset{o}{N}\left(m;\underline{x},t\right) - \frac{\partial}{\partial m}\left[k_l(m)s(m)n\left(m;\underline{x},t\right)\right] \\ &- \frac{1}{N\left(\underline{x},t\right)}\int_0^\infty \lambda\left(m,m';\underline{x},t\right)n\left(m;\underline{x},t\right)n\left(m';\underline{x},t\right)dm' \\ &+ \frac{1}{2N\left(\underline{x},t\right)}\int_0^m \lambda\left(m',m-m';\underline{x},t\right)n\left(m';\underline{x},t\right)n\left(m-m';\underline{x},t\right)dm'\end{aligned} \tag{4.1}$$

$$\frac{\partial F\left(\underline{x},t\right)}{\partial t} + \nabla\ \left[\underline{v}F\left(\underline{x},t\right)\right] = -\int_0^\infty m\overset{o}{N}\left(m;\underline{x},t\right)dm - \int_0^\infty k_l(m)s(m)n\left(m;\underline{x},t\right)dm \tag{4.2}$$

where:

- s is the surface area of a pellets of mass m
- $\underline{v}$ is the convective velocity vector
- $\overset{o}{N}$ is the rate of generation of nuclei in [number per time unit]
- k_l is the layering growth parameter in [mass per unit surface per time unit]
- λ is the coalescence parameter in [mass per unit surface per (invers) time unit]
- $N\left(\underline{x},t\right)$ is the total number of pellets given by the following equation:

$$N\left(\underline{x},t\right) = \int_0^\infty n\left(m;\underline{x},t\right)dm \tag{4.3}$$

It is further assumed that the pelletization takes place by the mechanism of nucleation, layering (snowball effect), and coalescence only. See [30] for a thoroughly description of the pellets growth mechanisms which may take place in such drums.

Based on this model, a simulation program for balling drums called CCBDrum, was developed by K. V. S. Sastry and marketed through the company SPEX Inc. For a description of this program see [32].

Model 2

The second model is due to M. Cross (see [10]). This model describe the growth of the pellets by considering the path it follows through the drum. It is not given explicitly by use of equations, but may be illustrated by a flow chart, see [10] for the original publication. This model has been used to investigate how the drums output changes when different drum and material parameters are changes by use of the simulation program BALSIM, see [11] and [18]. The results of these investigations are fairly consistent with observations done on a real plant (Ref. R. Drügge, LKAB)

4.2.2 Automatic control of balling drums

Over the years different control schemes have been proposed. All of them have one goal in common, namely stabilizing the drums. Some included other goals like controlling the moisture content in the pellets, which may be regarded as one of the most important output parameter, and controlling the amount of onsize pellets.

The task of stabilizing a surging drum has turned out not to be trivial. Two different methods have been proposed:

I) Automatic water spray which turns on when the output oscillates with an amplitude larger than a predefined value and turns off when the amplitude small enough or zero (see [18])

II) Remove some fraction (10-15 %) of the recycled undersized pellets. This is basically the same as decreasing the loop gain in a control loop, which by the small gain theorem (see e.g. [34]) results in a stable controlled system (see [38]).

However, both of these methods include major disadvantages which are discussed in more detail in Paper 1.

Stabilizing the drums is in fact only a part of the control problem. Another very important task is to get good product quality. The main parameter in this respect is the moisture content in the pellets. Another parameter is the amount of binder added to the fines, but this parameter will not be considered here.

Maybe the first automatic control scheme for balling drums was proposed by R. Dügge at LKAB, Sweden. This control scheme was only intended to stabilize the drums, and is identical to the one described as Method I above.

A multivariable control scheme proposed by P. E. Wellstead et. al. (see [39, 38]) controls both the amount of pellets and its moisture content. Stabilizing of the drum is not considered in this control scheme and this must therefore be taken care of by other means. One method used in both [39] and [18] is Method II described above. Even if the moisture content is a key parameter for stable operation, it can be varied between quite wide limits when the drum is stabilized by Method II, and thus make the proposed control scheme possible. In Figure 4.1 the block diagram for this controller structure is shown, where u_1 is the iron ore feed rate, u_2 is the water spray rate, y_1 is the pellets production rate and y_2 is the pellets moisture content. The controller $K(s)$ is given by

$$K(s) = K_p(s)K_d(s) \tag{4.4}$$

where $K_d(s) = \text{diag}\{k_i(s)\}$ is a diagonal matrix of single loop controllers for the pellets production rate and the moisture content respectively. $G(s)$ is the stabilized balling drum.

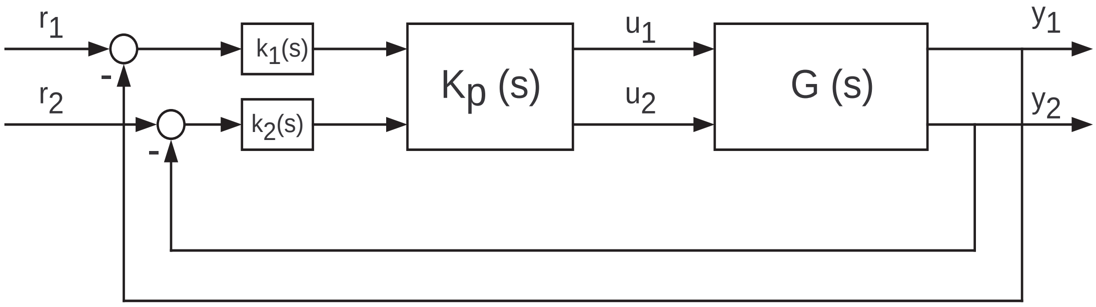

Figure 4.1: Block diagram, Multivariable control scheme

4.3 The new control scheme

4.3.1 Description of the control scheme

A new control scheme for stabilizing the output of a cluster of balling drums rather than controlling each drum as one unit, is proposed in Paper 1. A disadvantage of this method is of course that it can not be applied for a single drum plant, however, most plant are multi drum plants. In the following a description will be given. At some points this extends the description given in Paper 1.

The basic idea is to permit the drums to oscillate with a modest amplitude, and then adjust each drum's phase angle, relative to the other drums, so that the total output flow from all drums is constant, or nearly constant. The phase angles between the drums are calculated based on estimates of the amplitude and frequency of each drum's output. The estimator used in this work is the well known Extended Kalman Filter (see e.g. [9], chapter 8).

It is concluded earlier that a modest oscillation of a drum does not degrade the quality of pellets (see [30, 10]). In fact, due to a favorable moisture content the quality may be on its best when the drum is oscillating with a modest amplitude.

If it is assumed that the output from each drum is sine shaped, then it is possible to compensate exactly and get a constant output if the cluster consists of three or more drums. A prerequisite for this is that the largest amplitude is less or equal to the sum of the other. A rephrase of this in mathematical language is

Proposition 4.1. *The equation*

$$a_1 \sin(\varphi) + a_2 \sin(\varphi + k_1) + a_3 \sin(\varphi + k_2) + ... + a_n \sin(\varphi + k_{n-1}) = 0 \tag{4.5}$$

were $a_1 \geq a_2 \geq ... \geq a_n > 0$ *are fixed constants, has a solution if* $n \geq 3$ *and* $a_1 \leq \sum_{i=2}^{n} a_i$.

Here $a_1 \dots a_n$ are the amplitudes of the drums and k_1 is the difference in phase between drum 1 and 2, k_2 is the difference in phase between drum 2 and 3, and so forth.

Proof: The left hand side of equation (4.5) can be expressed as a linear combination of two trigonometric functions as follows

$$\begin{aligned} &a_1 \sin(\varphi) + a_2 \sin(\varphi + k_1) + a_3 \sin(\varphi + k_2) + ... + a_n \sin(\varphi + k_{n-1}) \\ &= \left[a_1 + \sum_{m=2}^{n} a_m \cos(k_{m-1})\right] \sin(\varphi) + \left[\sum_{m=2}^{n} a_k \sin(k_{m-1})\right] \cos(\varphi) \end{aligned} \tag{4.6}$$

Therefore, a solution of (4.5) exists if

$$a_1 + \sum_{m=2}^{n} a_m \cos(k_{m-1}) = 0 \tag{4.7}$$

$$\sum_{m=2}^{n} a_k \sin(k_{m-1}) = 0 \tag{4.8}$$

By the substitution $q = [\pi/2 - k_1, \pi/2 - k_2, ..., \pi/2 - k_{n-1}]$ we obtain

$$a_1 + \sum_{m=2}^{n} a_m \sin(q_{m-1}) = 0 \tag{4.9}$$

$$\sum_{m=2}^{n} a_k \cos(q_{m-1}) = 0 \tag{4.10}$$

Since $\sin(\varphi) \leq 1$, a solution of (4.9) exists if and only if $a_1 \leq \sum_{m=2}^{n} a_m$. Assume that $Q_1 = [q_{1,1}, q_{1,2}, ..., q_{1,n-1}]$ is a solution to (4.9) such that the left side of (4.10) is negative. Let Q_- be the set of all solutions of (4.9) such that the left side of (4.10) is negative, that is

$$S_- = \left\{ Q_1 \;\middle|\; a_1 + \sum_{m=2}^{n} a_m \cos(q_{m-1}) = 0 \text{ and } \sum_{m=2}^{n} a_k \sin(k_{m-1}) < 0 \right\} \tag{4.11}$$

Now we define the set

$$S_+ = \left\{ Q_2 \;\middle|\; a_1 + \sum_{m=2}^{n} a_m \cos(q_{m-1}) = 0 \text{ and } \sum_{m=2}^{n} a_k \sin(k_{m-1}) > 0 \right\} \tag{4.12}$$

which is the set of all solutions $Q_2 = [q_{2,1}, q_{2,2}, ..., q_{2,n-1}]$ of (4.9) such that the left side of (4.10) is positive. A valid solution of (4.9) may also be of the form $\tilde{Q} = [\pi/2 - q_{i,1}, q_{i,2}, ..., q_{i,n-1}]$, $i \in (1, 2)$, so the elements can pass from one set to the other through π. The two sets S_- and S_+ are therefore connected, and a solution to the equation system (4.9)-(4.10), and hence (4.5), exists. □

In the real plant the output is of course not exactly sine shaped, and in addition, noise will always be present. The level of measurement noise may at times become rather high. In Figure 4.2 the power spectrum of the measured drum output is shown. This measurement is carried out at 3.5 RPM. In Figure 4.3 the power spectrum is shown for 4.0 RPM. The three large spikes, which occur at 0.5 rad/sec, 1 rad/sec and 1.5 rad/sec, are due to slits through which the pellets leave the drum. The frequency at which these spikes occur, depends therefore upon the drum's rotational speed. This is also seen from Figures 4.2 and 4.3. These disturbances are not taken into account in the signal model, as this should require a more complex model and filter algorithm. They are simply considered to be noise. It is possible that including these in the model may yield a less noise sensitive filter, however, as shown in Paper 1, the filter works quite satisfactory even for a noise level considerably higher than what is expected in the real plant. The filtering part should therefore not be the limiting factor.

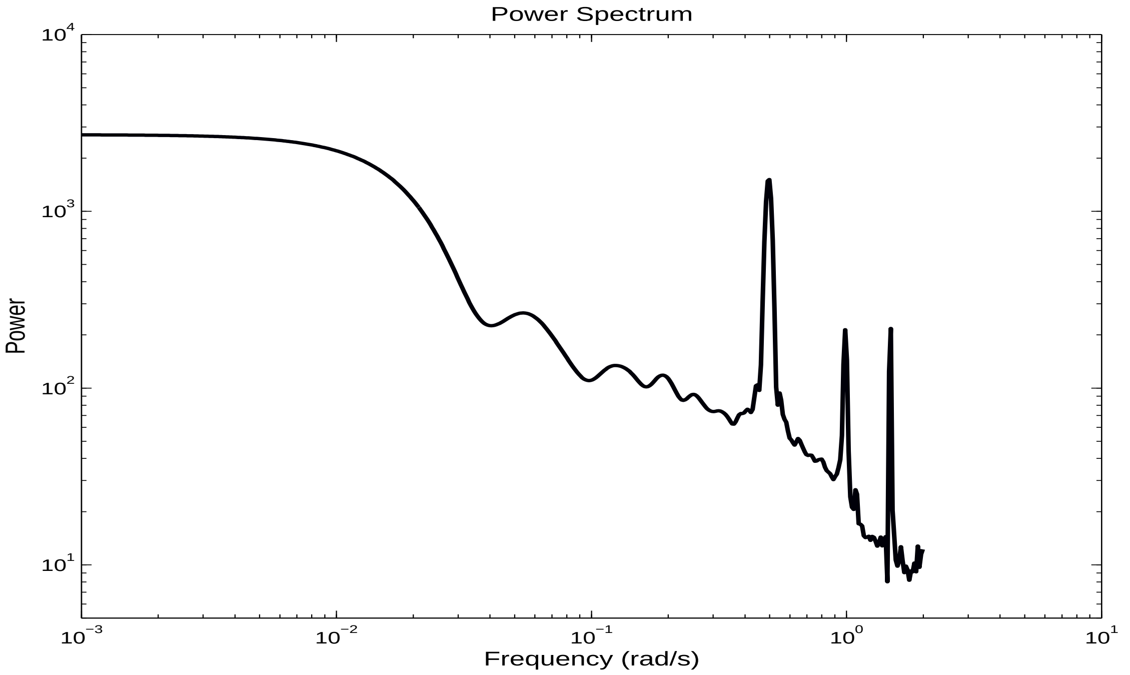

Figure 4.2: Power spectrum of the measured drum output at 3.5 RPM

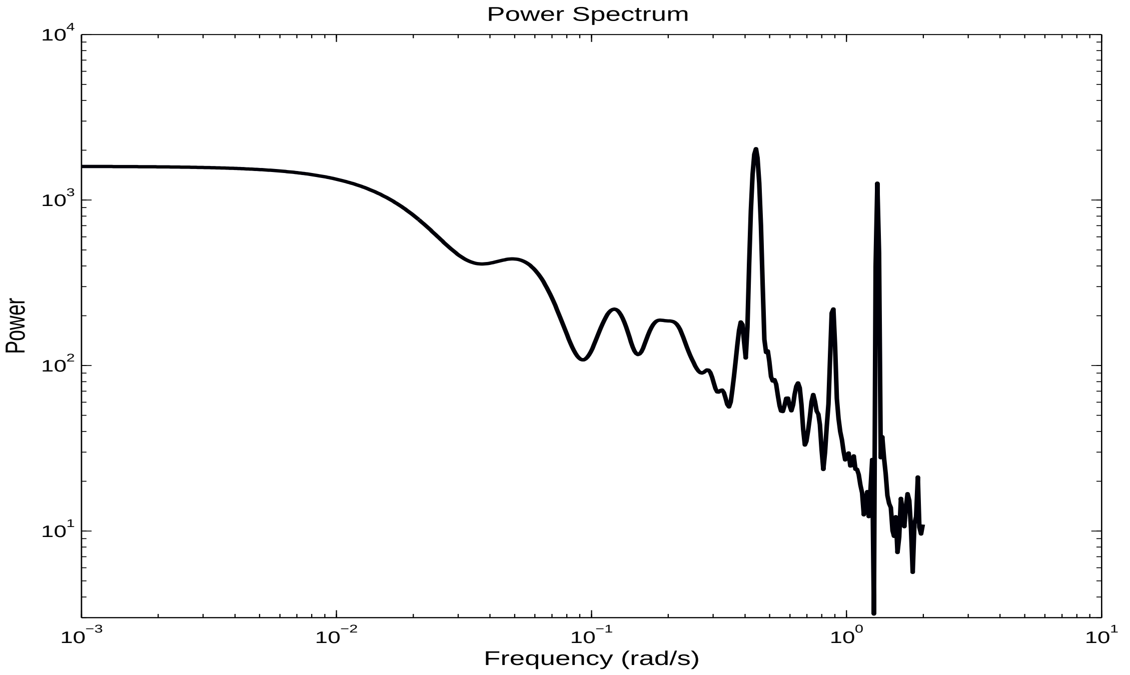

Figure 4.3: Power spectrum of the measured drum output at 4.0 RPM

Even if the output from the drums is not exactly sine shaped, it is still possible to obtain an almost constant mass flow from the drums. Some harmonic components will then be present, but these does not cause any major problems. In the next subsection it is shown by simulations that this principle works well when assuming that the oscillation obeys the van der Pol equation, which introduces over harmonic components. Furthermore, noise is added to the oscillators output in order to make the simulations more realistic.

4.3.2 Results from simulations

In this section two cases are considered. The first case is with two balling circuits and the second case is with three balling circuits. As the control objective is to obtain a small amplitude, preferably zero, in the fluctuations of the total mass flow, the mean value of the output from each drum is not of any interest in these simulations. We therefore only consider the oscillation, which has zero mean value. In Figure 4.4 the structure of the control system with two drums is shown.

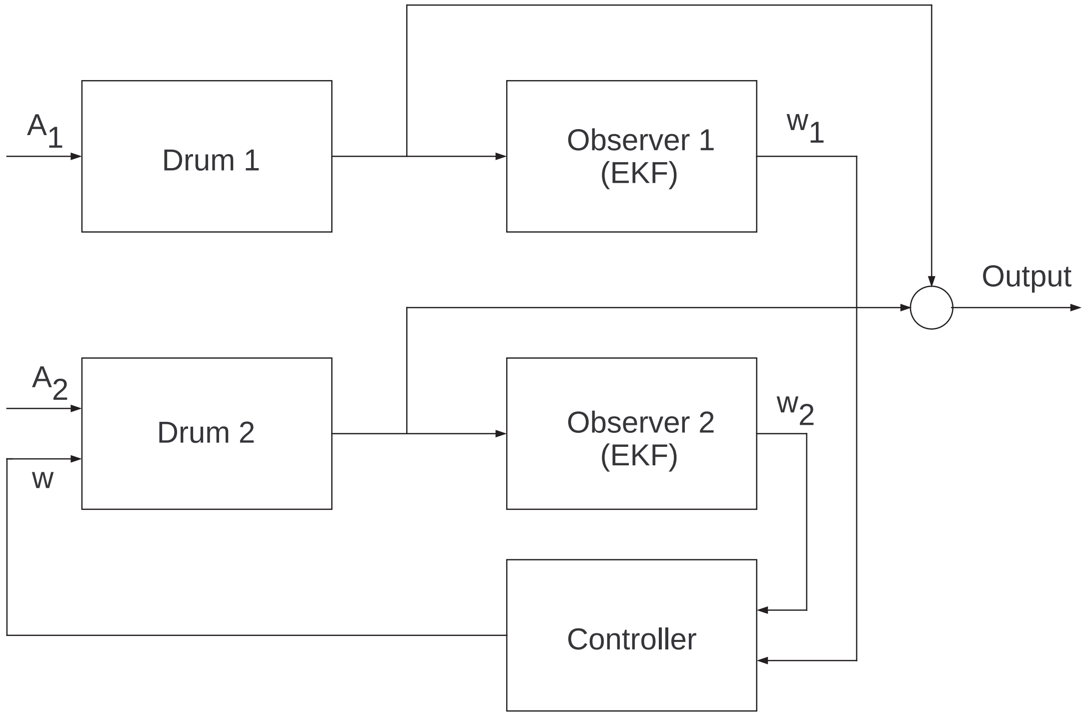

Figure 4.4: Block Diagram, Control of two drums

A_1 and A_2 denote the amplitudes, which are fixed in these simulations, and w_1 and w_2 denote the phase angel for drum 1 and 2 respectively. The desired phase angel for drum two is calculated (it is always $w_2 = w_1 + \pi$ in the case of two drums) and this value is used as set point for the controller. In Figure 4.4 the controller block also contains a measurement of the difference between w_1 and w_2. The controller used in these simulations is a PID controller.

In the case of three drums one more controller is required to take care of drum 3. The phase angel of drum 3 is kept fixed until drum 2 has correct phase angel. This is not an optimal solution as the synchronization will take very long time if the number of drums is high. Furthermore, if one drum should fail it will be difficult to deal with the transient which then will occur.

In case one the two drums have equal amplitude ($6\ \frac{\text{tons}}{\text{hour}}$) and level of noise. The result is shown in Figure 4.5, where the upper figure shows the output from each circuit, and the lower figure shows the sum. The principle works well, but the settling time for the amplitude of the total mass flow is quite long. The amplitude is reasonable low after approximately 3000 s, and after 4500 s the deviation from zero is mostly caused by noise. However, due to over harmonic components from the van der Pol oscillator a small periodic component remains, although hard to observe due to the level of noise.

In case two the three drums have equal level of noise, but different amplitude. The amplitudes are given in Table 4.1. The colors given for each drum refer to Figure 4.6, which shows the result from the simulation. See also Paper 1 for the noiseless case.

Table 4.1: Amplitudes for each drum

Drum No.	1 (blue line)	2 (green line)	3 (red line)
Amplitude in $\frac{\text{tons}}{\text{hour}}$	10	9	8.5

Compared to the previous case with two drums, the most obvious difference is that the settling time is much longer (approximately twice the time for two drums).

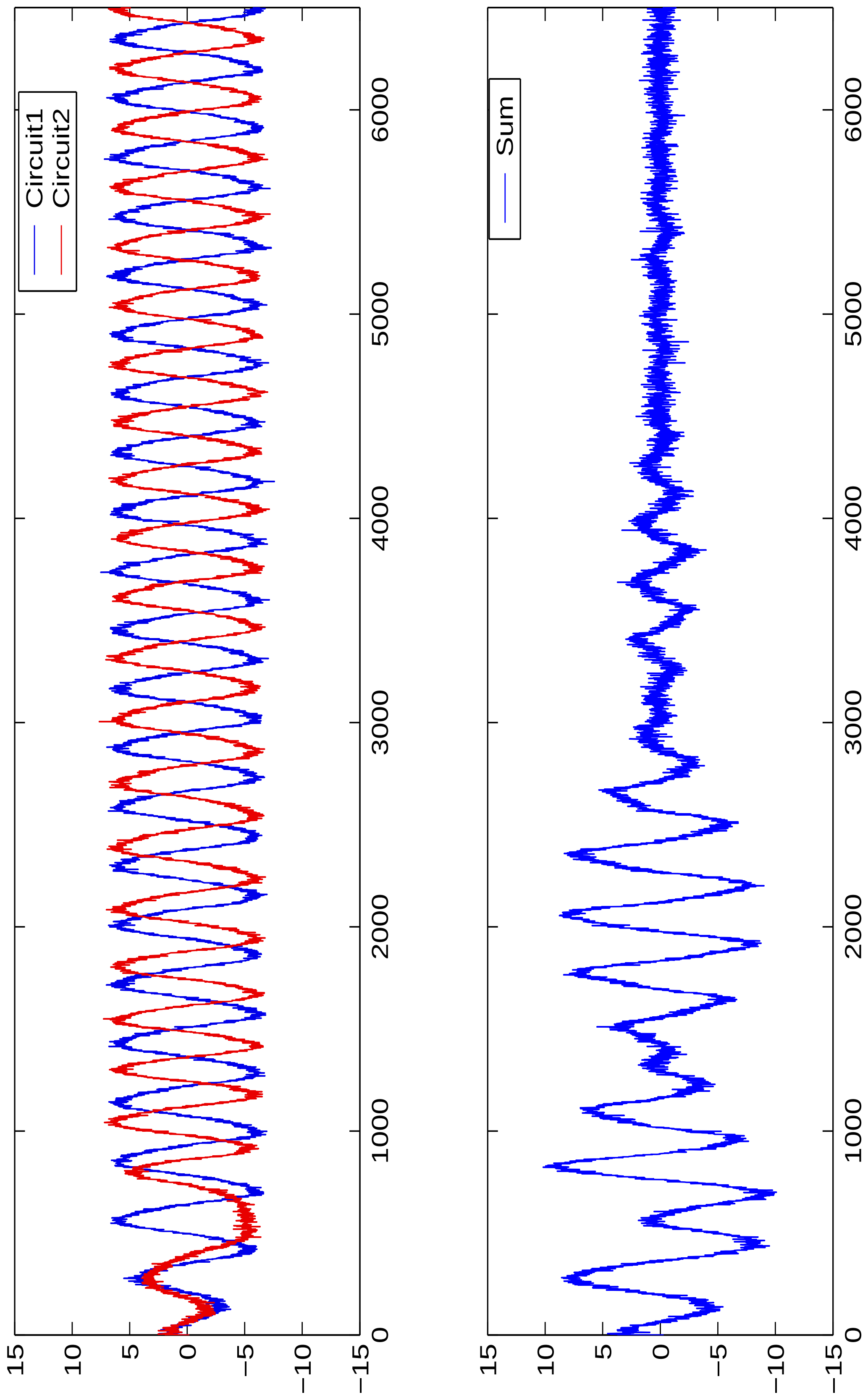

Figure 4.5: Simulation with two circuits

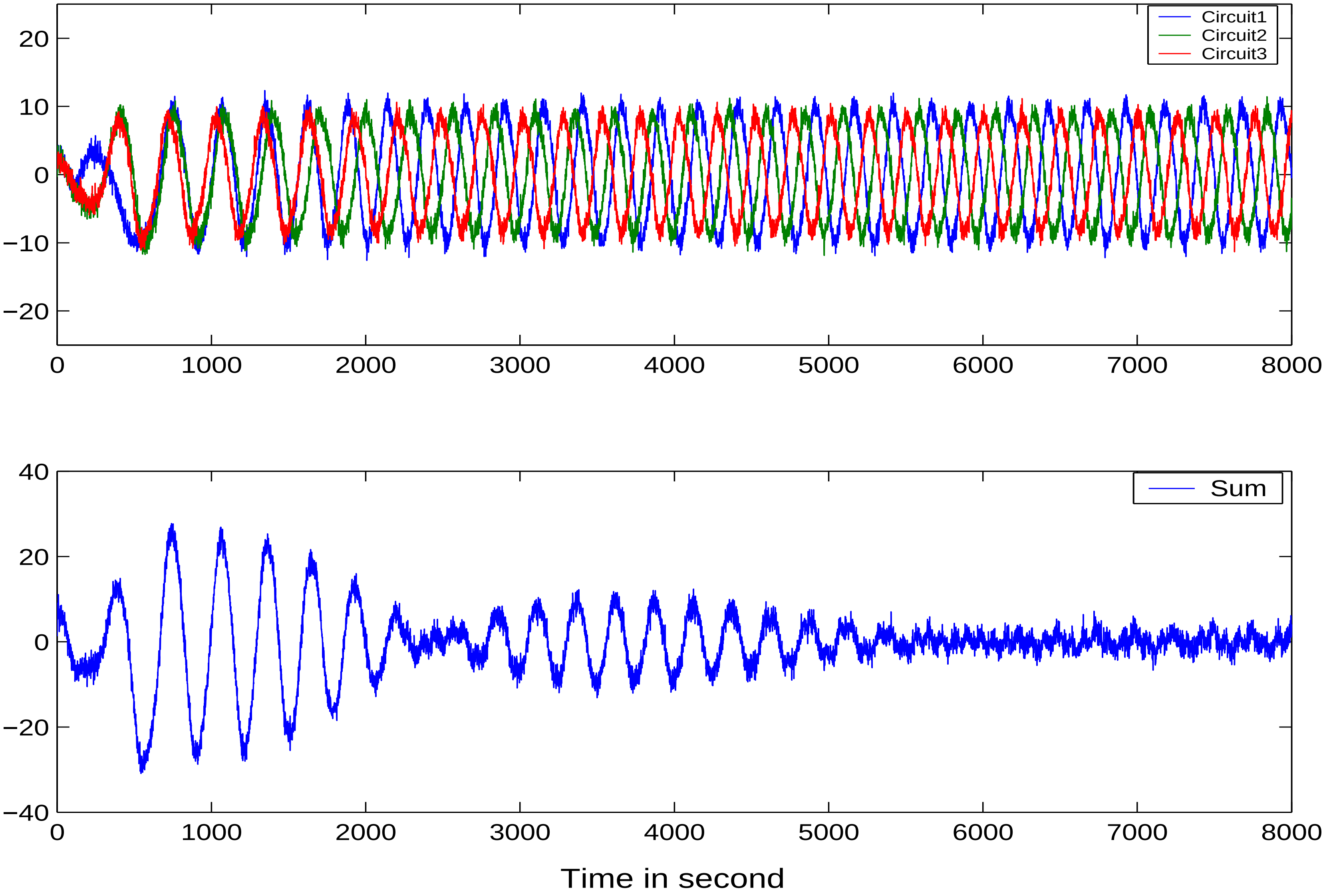

Figure 4.6: Simulation with three circuits

Chapter 5

Concluding remarks and suggestions for further work

5.1 Concluding remarks

In this thesis some estimation and control problems connected to the iron ore pelletizing process are described and discussed. A new concept for controlling and stabilizing the balling drums used in the iron ore pelletizing process is suggested. In this new control scheme a cluster of drums are controlled collectively rather than controlling each drum individually. This allow each drum to deliver a fluctuating output while the total mass flow is still nearly constant, as required by the subsequent process segment. A fluctuating output from the drums has been a major problem for the iron ore industry for several decades, and the normal solution has been to use fines with higher moisture content to damp out the oscillations. By use of the new concept the moisture content of the fines can be reduced, which reduces the total power consumption.

In order to obtain a constant output from the drum cluster the oscillation must be sine shaped. It is, however, confirmed by simulations that this concept also works in the case of non-sinusoidal oscillation if the total mass flow is allowed to oscillate with a small amplitude. In these simulations each drum is modelled as a van der Pol oscillator, and PID controllers are applied for controlling the amplitude of the total mass flow. The results presented from simulations indicates that PID controllers may be to simple for this control problem. It is commonly known that PID controllers works well for simple systems of first and second order, but may be insufficient for more complex systems, see e.g. [3]. The settling time is too long, but for steady state operation the principle works well with PID control. The EKF used as state estimator is tuned by use of an genetic algorithm, as described in Paper 2, and the result is a well performing estimator. Re-tuning of the filter for working in higher measurement noise is easily done by a slight modification of the performance function used by the genetic algorithm.

One major disadvantage from using a filter with constant tuning parameters, is that its performance will be poor if the level of noise should change. In chapter 3.4 it is shown how the EKF algorithm can be modified to yield a filter with both fast transient response, which is suitable in a low noise environment, and high noise rejection, provided that a method for detecting a step in the state is available (see [16]). At start-up this modification can be the default in order to ensure reliable estimates in shorter time.

In Papers 4 and 5, the stability properties of the Extended Kalman filter are addressed. Previously published results require that the state $x \in \mathcal{M}$, where $\mathcal{M}$ is a compact subset of $\mathbb{R}^n$. This cannot be guaranteed for some signal models, e.g. the ramp function and the signal model used in the balling drum control scheme. It is proved in Papers 4 and 5 that this requirement can be relaxed to only require the state to belong to an open convex subset $\mathcal{M}$ of $\mathbb{R}^n$, i.e. $x \in \mathcal{M} \subseteq \mathbb{R}^n$, provided that the Hessian matrix of the output map is bounded in $\mathbb{R}^n$ and that the Jacobian of the output map has a finite ratio between its largest and smallest singular value. When assuming that the covariance matrices are bounded from above and below, and that the matrix

$$F_k = \left[\frac{\partial f}{\partial x}\right]_{\hat{x}=\hat{x}_{k,k}} \tag{5.1}$$

where $f(x)$ is the signal state map, is both nonsingular and bounded from above, it is proved that the estimation error will be bounded by

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 - \frac{p_2}{\xi}\rho(\bar{w}, \bar{v}, \epsilon)$$

if the initial error satisfies $\|e_{0,0}\| \leq \epsilon$, and $\bar{w}$, $\bar{v}$ are sufficiently small. Here $\xi \in (-1, 0)$ is a constant and $\rho(\bar{w}, \bar{v}, \epsilon) > 0 \ \forall \ k \geq 0$ is a function of the maximum process noise, measurement noise and maximum initial error (see Paper 4 or Paper 5 for supplementary details). In the noiseless case, i.e. $\bar{w} = \bar{v} = 0$, the bound on the estimation error is

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2$$

which implies that it is an exponentially observer.

Earlier published results indicated that the stability results are very conservative, that is, the maximum allowed initial error and noise processes are very small. The same is concluded here for the general case. In Paper 4 it is shown that when the state map is linear, the results can be considerably improved by a proper choice of the filter tuning matrix Q. These results can be further improved by applying tighter upper bounds on the Kalman gain matrix K_k and the matrix $(I - K_k H_k)$. This also applies for the general case. One disadvantage with choosing Q such that the stability results are improved is that the filter will become more noise sensitive. In the context of filter tuning, this is the traditional trade-off between fast transient response (speed of convergence) and low variance in the estimated state. In the noiseless case the transient response can be chosen arbitrary fast.

5.2 Suggestions for further work

When a work is based on a mathematical model there is always a possibility to make this model more precise, and thus improve the accuracy of the obtained results. This applies for all parts of this work, as mathematical models are used when

i) Simulating the behavior of a cluster of oscillating balling drums when applying PID controllers to make the total output from the cluster constant

ii) Designing a state observer for the balling drum

The model used for simulating the control scheme were sufficient for investigating the principle. However, for controller design intended for implementation in the real plant this model may not be adequate. Future work could be based on the model due to Sastry (see chapter 4.2.1). This model is of infinite order, but an reasonable discretization of this model yields a more accurate description of the balling drums behavior than the simplified model used in this thesis.

It is also reasonable to assume that the performance of the control system can be considerably improved by applying a more sophisticated controller than a PID controller. One possible choice, which should be investigation in future work, is to use a model predictive controller (MPC), in which any necessary constraint is easily included. During the simulations it is observed that a maximum allowed value for the frequency of the oscillation from each drum should be included in the controller algorithm, and this is easily done in a model predictive controller.

The signal model used for the state observer seems to be well suited for the purpose. Still there may be convenient to have an observer which yield an estimate with lower variance in the estimated states. Future work could include different filter algorithms, e.g. an adaptive filter or a wavelet based filter.

The stability results presented in this thesis, and also in previous work, are very conservative for the general case. In the case of a signal model with linear state map it is shown that the results can be considerably improved, but they are still conservative compared to the results obtained by simulations. In addition, some conditions under which stability is proved, are rather strong. For instance, the state map is required to be invertible. Further work on this topic could focus more on the reasons for the conservative results. It may also be of interest to search for a different proof which does not include the Lyapunov function used in this thesis. Finally, relaxing the conditions under which EKF stability can be proved, will always be of interest.

References

[1] B. D. O. Anderson, R. R. Bitmead, C. R. Johnson Jr, P. V. Kokotovic, R. L. Kosnut, I. M. Y. Mareels, L. Pray, and B. D. Riedle. *Stability of Adaptive Systems, Passivity and Averaging Analysis.* MIT Press, Cambridge, Massachusetts, 1986.

[2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentic-Hall, New Jersey, 1979.

[3] Karl J. Åstrøm and Tore Hägglund. *PID Controllers: Theory, Design, and Tuning, 2nd Edition.* Instrument Society of America, USA, 1995.

[4] Karl J. Åstrøm and Björn Wittenmark. *Computer Controlled Systems, Theory And Design.* Prentice Hall, New Jersey, 1997.

[5] Thomas Bäck. *Evolutionary Algorithms in Theory and Practice.* Oxford University Press, Oxford, 1996.

[6] M. Boutayeb and D. Aubry. A strong tracking extended Kalman observer for nonlinear discrete-time systems. *IEEE Transaction on automatic control*, 44: 1550–1556, 1999.

[7] Karl Brammer and Gerhard Siffling. *Kalman-Bucy Filters.* Artech House, Norwood, 1989.

[8] Jeffrey B. Burl. *Linear Optimal Control, $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Methods.* Addison Wesley, California, 1999.

[9] C. K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications.* Springer, Berlin, 1999.

[10] M. Cross. Mathematical model of balling-drum circuit of a pelletizing plant. *Ironmaking and steelmaking*, pages 159–169, 1977.

[11] M. Cross, R. W. Young, P. E. Wellstead, and R. D. Gibson. The mathematical modelling and control aspects of the pelletizing of iron ores. *Agglomeration 77, AIME New York*, pages 403–424, 1977.

[12] T. I. Fossen. *Marine Control Systems, Guidance, Navigation, and Control of Ships, Rigs and Underwater Vehicles.* Marine Cybernetics, Trondheim, Norway, 2002.

[13] Arthur Gelb. *Applied Optimal Filtering.* MIT Press, Cambridge, Massachusetts, 1974.

[14] Graham C. Goodwin and Kwai Sang Sin. *Adaptive Filtering Prediction and Control*. Rentice Hall, New Jersey, 1984.

[15] Michael J. Grimble and Michael A. Johnson. *Optimal Control and Stochastic Estimation, Theory and Applications*. John Wiley & Sons, Chichester New York Brisbane Toronto Singapore, 1988.

[16] F. Gustafsson. *Adaptive Filtering and Change Detection*. John Wiley, West Sussex, England, 2000.

[17] C. R. Houck, J. A. Joines, and M. G. Kay. The genetic algorithm optimization toolbox (gaot) for matlab 5. *http://www.ie.ncsu.edu/mirage/GAToolBox /gaot/*, 1996.

[18] D. Ibrahim. Control of the balling drum circuit of an iron ore pelletising plant. *M.Sc. thesis, University of Manchester*, 1977.

[19] Mo Jamshidi, Leandros dos Santos Coelho, Renato A. Krohling, and Peter J. Fleming. *Robust Control Systems with Genetic Algorithms*. CRC Press, Boca Raton London New York Washington DC, 2003.

[20] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, New Jersey, 2002.

[21] B. F. La Scala, R.R. Bitmead, and M.R. James. Conditions for stability of the extended Kalman filter and their applications to the frequency tracking problem. *Mathematics of Control, Signals, and Systems*, 8:1–26, 1995.

[22] L. Ljung. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transaction on automatic control*, AC-24: 36–50, 1979.

[23] Peter S. Maybeck. *Stochastic Models, Estimation and Control*. Academic Press, New York, 1979.

[24] G. Minkler and J. Minkler. *Theory and Application of Kalman Filtering*. Magellan Book Company, Palm Bay, Florida, 1993.

[25] Y. Oshman and Ilan G. Shaviv. Optimal tuning of a Kalman filter using genetic algorithms. *AIAA Paper 2000-4558*, 2000.

[26] T. D. Powell. Automated tuning of an extended Kalman filter using the downhill simplex algorithm. *Journal of guidance, control and dynamics*, 25:901–908, 2002.

[27] B. G. Quinn and E. J. Hannan. *The Estimation and Tracking of Frequency*. Cambridge University Press, Cambridge, 2001.

[28] K. Reif, Stefan Günter, Engin Yaz, and Rolf Unbehauen. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transaction on automatic control*, 44:714–728, 1999.

[29] K. Reif and R. Unbehauen. The extended Kalman filter as an exponential observer for nonlinear systems. *IEEE Transaction on Signal Processing*, 47: 2324–2328, 1999.

[30] K. V. S. Sastry. The agglomeration of particulate materials by green pelletization. *Ph.D thesis, University of California, Berkeley*, 1970.

[31] K. V. S. Sastry. Similarity size distribution of agglomerates during their growth by coalescence in granulation or green pelletization. *International Journal of Mineral Processing*, pages 187–203, 1975.

[32] K. V. S. Sastry. Process engineering of agglomeration systems. *The Sixth International Symposium on Agglomeration, Nagoya, JAPAN*, 1993.

[33] K. V. S. Sastry and D. W. Fuerstenau. Kinetic and process analysis of the agglomeration of particulate materials by green pelletization. *Agglomeration 77, AIME New York*, pages 381–402, 1977.

[34] S. Skogestad and I. Postletwaite. *Mulitivariable Feedback Control, Analysis and Design.* John Wiley, New York, 1996.

[35] Robert F. Stengel. *Optimal Control And Estimation.* Dover Publications, INC, New York, 1994.

[36] Tzyh-Jong Tarn and Yona Rasis. Observers for nonlinear stochastic systems. *IEEE Transaction on automatic control*, AC-21:441–448, 1976.

[37] M. Vidyasagar. *Nonlinear Systems Analysis, 2nd ed.* Prentice Hall, New Jersey, 1993.

[38] P. E. Wellstead, M. Cross, N. Munro, and D. Ibrahim. On the design and assessment of control schemes for balling-drum circuits used in pelletizing. *International Journal of Mining Processing*, pages 45–67, 1978.

[39] P. E. Wellstead and N. Munro. Multivariable control of cold iron ore agglomeration plant. 1977.

Part II

The papers in the thesis

PAPER 1

Control of the Amplitude in a Surging Balling Drum Circuit, a New Approach to an Old Problem[1]

Knut Rapp*, **Per-Ole Nyman***

* Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, kr@hin.no, fax: +47 76 96 68 10

† Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, pon@hin.no, fax: +47 76 96 68 10

Abstract: In this paper we suggest a new method for controlling the balling drums used in the iron ore industry. We suggest that a cluster of drums are controlled collectively rather than individually. Further, we investigate the possibility of using an extended Kalman filter for estimating the amplitude and frequency of the oscillations in such drums. The filters thresholding point is identified, and the area for which the filter is usable is given.

[1] In MODELLING, IDENTIFICATION AND CONTROL, 2003, vol. 24, No. 4, 193-203, and presented at European Control Conference 2003, 01-04 September, Cambridge, UK

1.1 Introduction

Use of balling drums has become common in many parts of the industry. In the iron ore industry, balling drums used in pellets production has a long tradition. The main problem areas associated with such drums are therefore well described and to some extent also analysed. In Figure 1.1, a typical single drum circuit used for pelletizing iron ore is shown.

The orange arrows below the drum represents the undersize flow of pellets with too low diameter. These pellets are transported back and fed into the drum together with the fines. The orange arrows above the drum are the oversize flow of pellets with too big diameter. These pellets are crushed and transported bach to the fines tank. The pellets on the lower right conveyor belt are the onsize pellets, which form the process output.

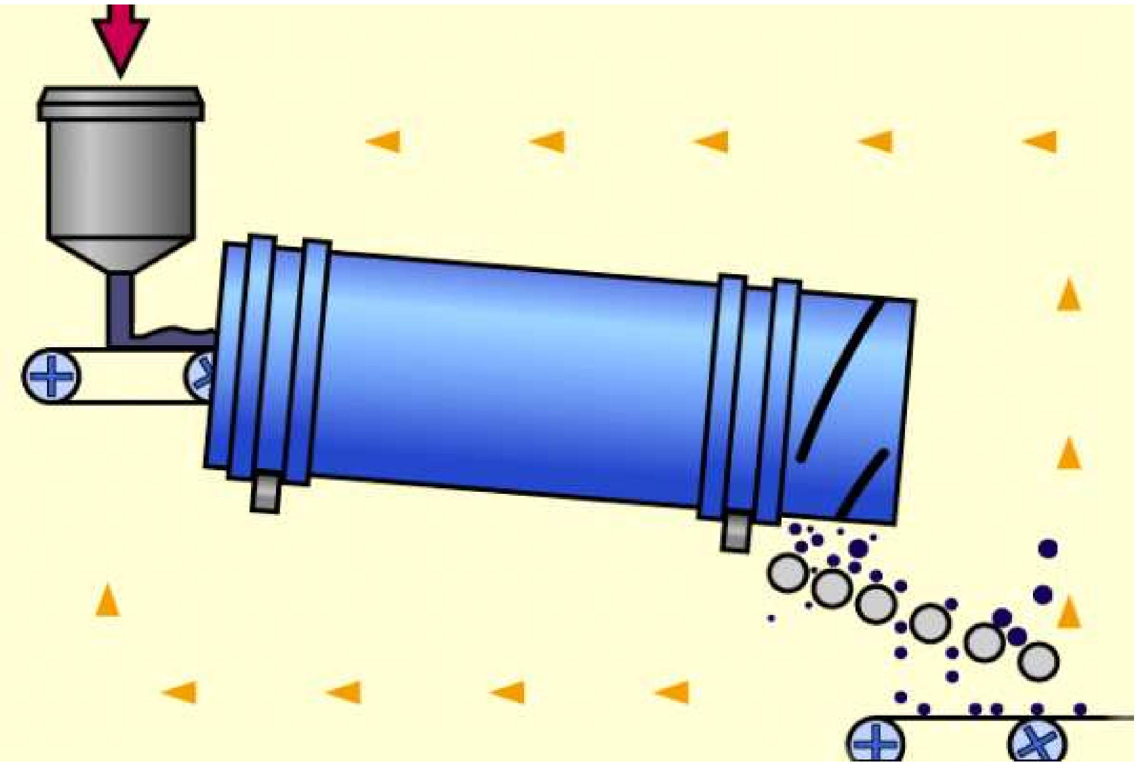

Figure 1.1: Single balling drum circuit

One problem the iron ore industry has been dealing with for as long as balling drums have been used, is that they tend to give a surging output under some operational conditions. This surging poses a problem for the balling drum circuit only if the amplitudes get too high. A process shutdown will then normally be the result, but danger for some equipment, such as the conveyor belt for recirculation of undersized pellets, is also evident. The major problem with surging lies however, in the subsequent process segment, which is the induration process, also denoted as the ”warm process”.

If the input to the warm process is fluctuating, then the efficiency of this segment is reduced, accompanied by a considerably increase in energy consumption and poor product quality. Unfortunately, the conditions which cause the drums output to oscillate, coincidence with those required for good product quality (see e.g. [2] and [7]). Several attempts have therefore been made to design an automatic control scheme for regulating the amplitude to zero. See e.g. [9].

So far the problem has only been considered for a single drum and tandem drums even if a pelletizing plant normally consists of several drums, see [3],[8] and [6]. In this paper we point out why such a strategy may not be optimal and suggest an alternative way to solve the problem.

1.2 Preliminary results

In this section we give some preliminary results from previous research within this area.

Over the last four decades the problem with surging drums has been investigated from several points of view. A great deal of work has been carried out in developing mathematical models of the balling drum. The first model was established about thirty years ago, see e.g. [5] or [9]. This model has infinity dimension, and is as such not useful for control purposes. A lot of knowledge about the system is, however, gained from this model. Later, simplified models have been presented, see e.g. [2]. Simulations based on these models reveals that the moisture content in the pellets is the most important process parameter to be controlled if one wishes to stabilize and keep the process stable. The binder, which is added in order to obtain sufficient mechanical strength, tends to have an contrary effect. This suggests that a simple multivariable control could be applied to regulate the moisture content, mechanical strength and the amount of onsized pellets. Unfortunately, this does not turn out to be possible. One of reasons for this is that some plant parameters are hard to measure on-line, and consequently difficult to control. An other reason is that small changes in the operating conditions, e.g the moisture content, may cause large changes in the operating point, see [2]. Together with noisy measurements, this offers a big challenge for this approach.

1.3 Stabilizing and controlling the drums

As mentioned in the previous section, a quite simple multivariable control system could be designed if some practical problems were solved. One of the most important unsolved problem is to develop a method for on-line measurement of the moisture content in the fines. As far as we are aware of, the best equipment today offer an accuracy no better than ± 0.5 % in absolute error, ref [4], and this is not sufficient, as a change less than this may cause the drum to surge. Alternative ways of stabilizing the drums have been tried, and in the following we describe two different methods described in the literature, see [9] and [8], and we introduce a new concept as a third method.

1.3.1 Method 1

The simplest method suggested to stabilize the drum circuit is to add moisture in the recycle circuit when the surging occur. If a proper amount of moisture is added then the surging will decrease and finally stop. This method has been implemented on a real drum (ref [4]), but some serious problems were revealed. First of all, a method for detecting that surging really was present was not sufficiently developed. Secondly, as the surging disappear, the system will no longer be observable in the sense that all information needed to decide the amount of moisture to be added in order to keep the process stable without overcompensating, is lost. For this reasons the method was rejected, ref [4].

Another serious drawback with this method is that it is suboptimal from an energy point of view, as moisture is added without any knowledge about other important parameters, such as the amount of binder in the fines. This may result in overcompensating, i.e. too high water content in the onsize pellets, which, in addition to increased energy consumption, will result in poor quality and reduced drum throughput. In other words, the product quality and productivity are not considered in this control scheme.

1.3.2 Method 2

A second method was presented in 1976 by P. E. Wellstead and N. Munro, ref [9]. This method, which is the first based on control theoretical analysis, concludes that the surging is a limit cycle caused by to high loop gain. Then the well known method of reducing the gain, which is exactly the same as reducing the amount of recycled undersized pellets, is applied to stabilize the drum. Results from simulations based on the model described in [2], shows that a reduction of the recycled pellets of about 10-12 % will be sufficient to bring the surging down to an acceptable level. When the drum is stabilized, a multivariable control system for controlling the amount of onsize pellets and the moisture content in the pellets, can be designed. Since the moisture content is an important quality parameter, this model represent therefore a great achievement compared to the previous method.

From an economical or a productivity point of view the method is not optimal, since a mass balance shows that by taking some fraction out of the recycled mass, then the amount of onsize pellets will be reduced accordingly. This method of stabilizing the drum will therefore lower the output of onsize pellets when the input is constant.

As mentioned in the previous section, one of the problems with the first method is that the observability is lost at the same moment as surging disappears. A similar problem appears also in this method. If we assume that the amount of undersized pellets which is removed from the system should be as small as possible, then we should remove just enough to get the system stabilized. If the conditions are then changed in such a direction that the system becomes more stable, i.e. the moisture content is increased, then it may be quite difficult to discover that less material could be removed. The problem with unobservability is therefore still present.

1.3.3 Method 3, a new approach

A basic assumption in this method is that the surging does no harm to the pellets quality, see [2]. From a quality point of view there should therefore not be any problem to let the surging be present. The remaining problem is then to make sure that the amount of green pellets transferred to the warm part is constant. In order to obtain this goal, a cluster of drums is controlled collectively, rather than controlling each drum individually. The control scheme may be described in the following way:

1. All drums are operated with surging output, and the amplitude is kept on a desirable value by controlling the water and binder content in the fines.

2. One drum, preferably the one with largest amplitude, is chosen as reference (fixed RPM and return conveyor speed).

3. By adjusting the return conveyors speed the drums phase angel, relative to the reference drum, is controlled in such a way that the amplitude in the total output from all drums is kept at its minimum.

Controlling the amplitude to the desired value will in this setting mean that the quality parameters (the content of moisture and binder) will decide the amplitude. As the amplitudes in practice are not equal, the drum with largest amplitude should preferably be chosen as reference, as this allow us to more easily derive criteria for when the total amplitude is zero. Furthermore, as this drum normally will have the lowest signal to noise ratio (SNR), it is likely to have the best estimate of amplitude and phase.
The advantages of this method can be summarize as follows:

- The process parameters can be kept at a level which gives high product quality
- The total pelletizing process will consume considerable less energy compared to the situation where the moisture content is kept at a higher lever to avoid surging
- The total capacity (throughput) of the drum is not reduced
- There is no need for an advanced process model as the drums are treated as oscillators with controllable amplitude and frequency

A prerequisite for this method is that the following assumptions hold:

1. it is possible to detect the oscillation and estimate its amplitude, phase and frequency with sufficient accuracy
2. the amplitude is controllable in some range
3. the frequency is controllable in some range

The two latter items are well documented in the literature, see e.g. [2]. We will therefore concentrate on the first item in the following section.

1.4 Estimation of amplitude, phase and frequency

1.4.1 Signal model

An extended Kalman filter (EKF) is used for estimation of the frequency, amplitude and phase of the oscillations. The signal model used by the EKF is

$$x_1(k+1) = x_1(k) + v_1 \tag{1.1}$$

$$x_2(k+1) = x_1(k) + x_2(k) \tag{1.2}$$

$$x_3(k+1) = x_3(k) + v_2 \tag{1.3}$$

$$y(k) = x_3(k) \sin x_2(k) + z(k) \tag{1.4}$$

where x_1 is the phase increment (or frequency), x_2 is the phase and x_3 is the amplitude of the oscillation. $v = [v_1, v_2]^T$ and z are white, zero mean processes with covariance matrices Q =diag(q_1, q_2) and R. Locally the state is uniquely determined by the output y, but owing to the factor $\sin x_2$ in output equation, this does not hold globally. In fact, a simultaneous change of sign in x_1 and x_2, or a shift of x_2 by any number of periods, does not change the output. However, with a reasonable initialization of the EKF this mild nonuniqueness does in general not cause problems. The choice of the matrix Q is a compromise between accuracy in steady state and capability to track a changing amplitude or frequency. R is set equal to the covariance of assumed measurement noise of the true, measured output.

1.4.2 EKF equations

Written in a more compact form, equation (1)-(4) are given by

$$x_{k+1} = Ax_k + Bv_k \tag{1.5}$$

$$y_k = g_k(x_k) + z_k \tag{1.6}$$

In the literature two different formulations of the discrete-time extended Kalman filter are widely used. The first one is a one-step formulation in terms of the a-priori variables, and the second one is a two-step recursion consisting of a time update and a measurement update with a re-linearization between the two steps. In this paper the latter formulation is used.
The EKF equations derived from equation (1)-(4) are given by (see e.g. [1]):

Time update

$$\hat{x}_{k,k-1} = A\hat{x}_{k-1} \tag{1.7}$$

$$P_{k,k-1} = A \cdot [P_{k-1,k-1}] \cdot A^T + BQB^T \tag{1.8}$$

Measurement update

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k(y_k - g_k(\hat{x}_{k,k-1})) \tag{1.9}$$

$$P_{k,k} = P_{k,k-1} - K_k \left[\frac{\partial g_k}{\partial x_k}(\hat{x}_{k,k-1})\right] P_{k,k-1} \tag{1.10}$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$G_k = \left[\frac{\partial g_k}{\partial x_k}\right]_{x=x_{k,k-1}} = \begin{bmatrix} 0 & x_{k,k-1}^{(3)} \cos(x_{k,k-1}^{(2)}) & \sin(x_{k,k-1}^{(2)}) \end{bmatrix} \tag{1.11}$$

The filter gain matrix is given by

$$K_k = P_{k,k-1} \left[G_k(\hat{x}_{k,k-1})\right]^T \left[G_k(\hat{x}_{k,k-1}) P_{k,k-1} G_k(\hat{x}_{k,k-1})^T + R\right]^{-1} \tag{1.12}$$

1.4.3 Estimation of the amplitude

In this work we assume that control of the first harmonic component of the oscillations will give sufficient accuracy for this application. Figure 1.2 shows a typical situation from a physical plant in LKAB's works in Kiruna, Sweden. The measurement is done with a sampling period of 1 second. As we see, the signal has considerable noise components. The variance is calculated to be approximately 135 $[\frac{tons}{h}]^2$.

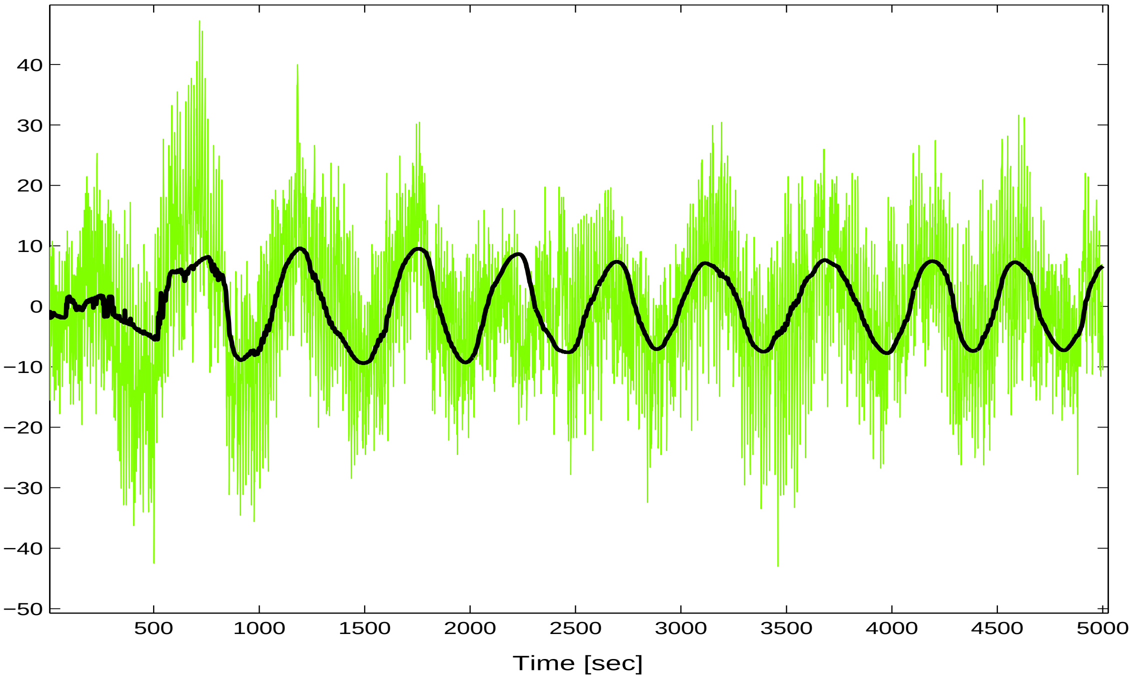

Figure 1.2: Variations around the working point

The dark solid-drawn line is the estimate from the Kalman filter, and the green solid-drawn line is the real, noisy signal. The filter model is based on a sinusoid, which in this application corresponds to the first harmonic of the oscillation. The solid-drawn line is the filters estimate of the amplitude. An important question at this stage is as follows: what is the lowest possible amplitude the filter is capable of detecting, given a specified level of noise. If we permit the amplitude to get below this limit we will no longer be able to control the phase angel to the desired value. The signal to noise ratio is given by the following equation

$$SNR = 10 \log \frac{A^2}{2\mathrm{Var}(v)} \tag{1.13}$$

where A is the amplitude of the oscillation and v is the signal noise, which is assumed to be white.

In figure 1.3 the filters signal to noise ratio curve with respect to relative amplitude error is shown. The solid line is for $A = 20\frac{ton}{h}$, the dashed line in the middle represents $A = 10\frac{ton}{h}$, and the dotted line represents $A = 5\frac{ton}{h}$. Clearly the area where the thresholding phenomenon occurs depends on the amplitude.

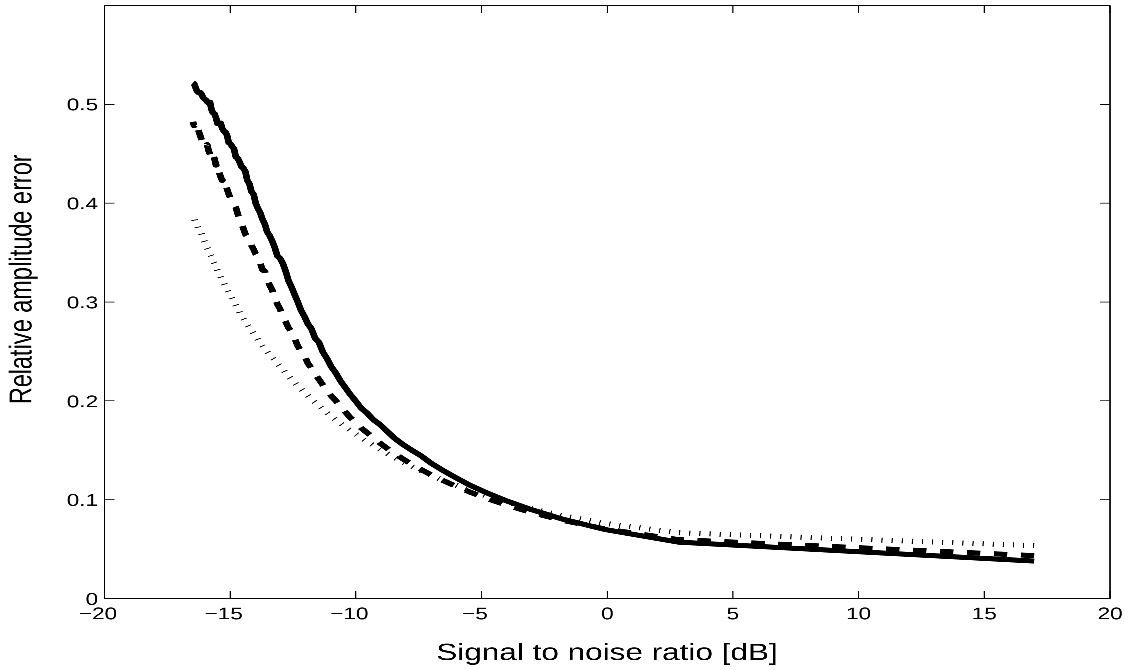

Figure 1.3: Relative amplitude error vs. signal-to-noise ratio

1.4.4 Estimation of the phase

In figure 1.4 the filters signal to noise ratio curve with respect to phase error is shown. As in the previous case, the solid line represents $A = 20\frac{ton}{h}$, the dashed line in the middle represents $A = 10\frac{ton}{h}$, and the dotted line represents $A = 5\frac{ton}{h}$.

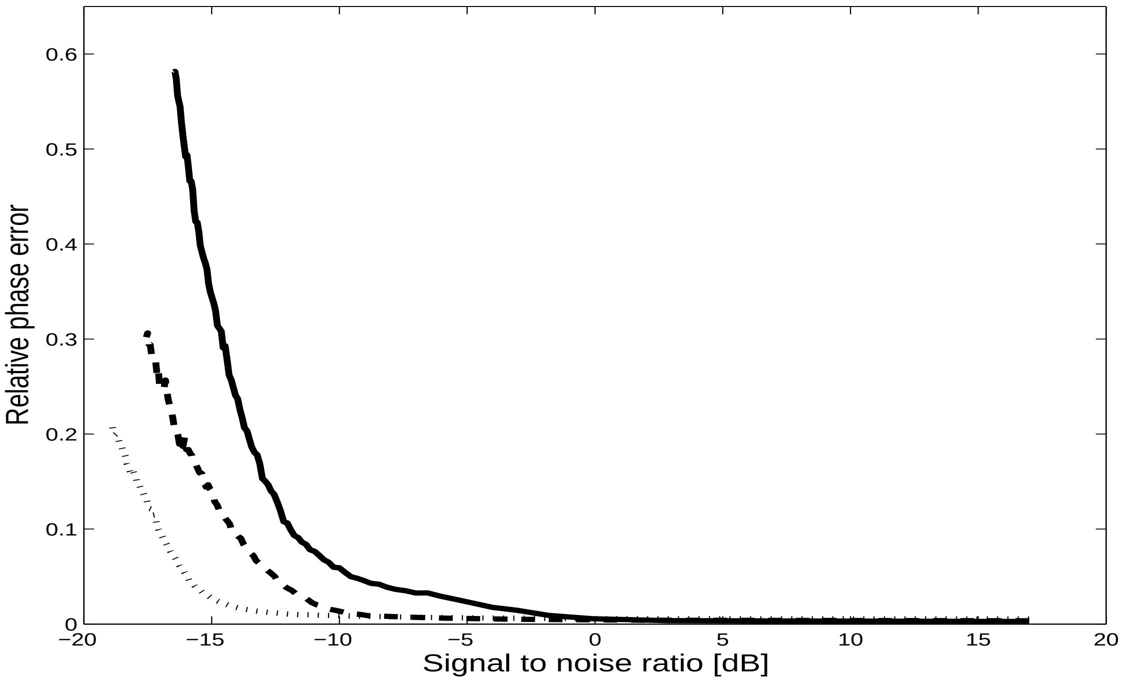

Figure 1.4: Phase error vs. signal-to-noise ratio

As in subsection 4.2, the thresholding phenomenon occurs at different SNR levels for different amplitudes. It is also clear from figure 1.4 that the filters phase estimate is more robust to noise than the amplitude estimate.

1.4.5 Estimation of the frequency

In figure 1.5 the filters signal to noise ratio curve with respect to frequency error is shown.

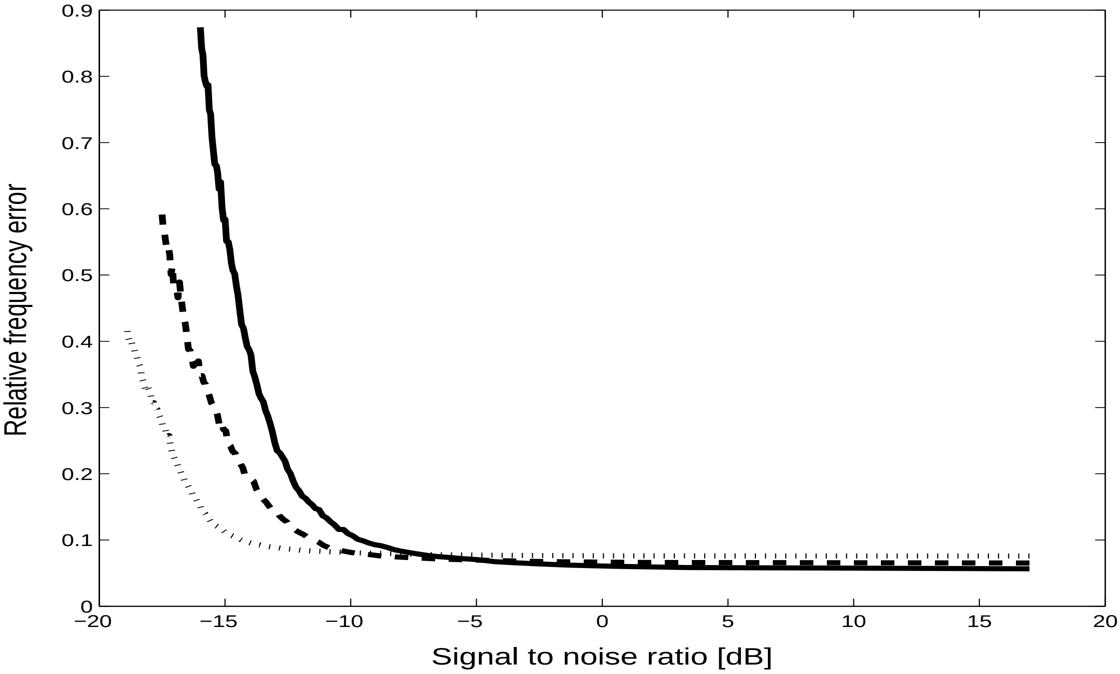

Figure 1.5: Frequency error vs. signal-to-noise ratio

Also in this case we see the same pattern as in the previous two subsections. It is also clearly seen from the figure that the frequency estimate is more robust to noise than the amplitude estimate.

1.4.6 Discussion

From the previous subsections, we see that it is the filters ability to estimate the amplitude which is the most critical part. Using the case with A=10 $[\frac{tons}{h}]$ as example, we see that thresholding in the amplitude estimate will occure for SNR below approximately -10 dB. For the frequency and phase estimate thresholding will occure below -13 dB and -12 dB respectively.

In a typical plant the noise level will normally be between 50 $[\frac{tons}{h}]^2$ and 250 $[\frac{tons}{h}]^2$. Within this limits we may expect that the filter will be able to track a signal with a amplitude down to 7.5 $[\frac{tons}{h}]$, i.e SNR = -10 dB. Repeated simulations shows that the filter in fact does work in this situation. We should, however, not expect the filter to have good performance close to the thresholding point. In figure 1.6 a representative result is shown. In the interval $0 \leq t \leq 5000$ the signal to noise ratio is -9.8 dB and in the interval $5000 \leq t \leq 10000$ it is increased to 3 dB by decreasing the signal measurement noise.

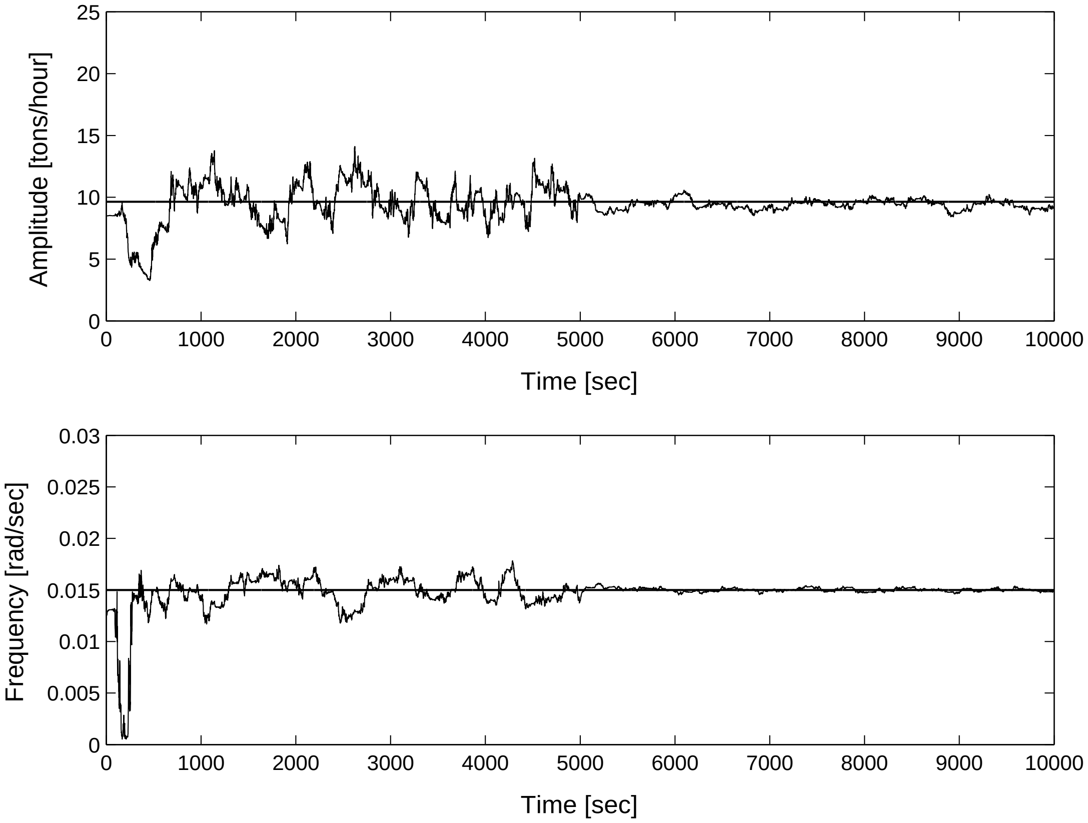

Figure 1.6: Amplitude and frequency estimate

1.5 Results from simulations

In this section results from simulations are shown. A simple model based on Van der Pol's equation is used to describe the drums oscillating behavior. This model may only be used for a quite narrow range of the process parameters, as it does not adequately describe the transition between the drums oscillation mode and steady mode. It should be pointed out that this model gives no information about the plants efficiency, it is only intended to illustrate the proposed control scheme.

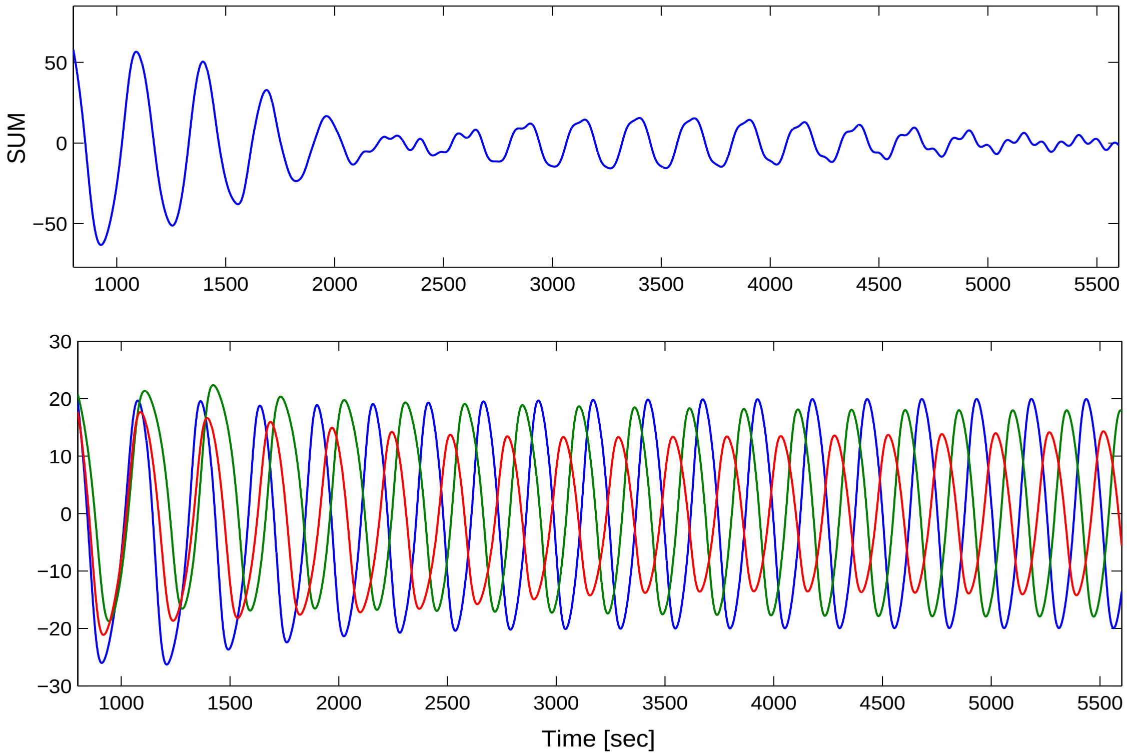

Figure 1.7: Result from simulation with a cluster of three balling drums

A standing assumption in this section is that the moisture content in the pellets is sufficiently low, so that the drums will oscillate. It is also assumed that the system is deterministic.

The simulated system is a cluster of three drums. The systems outputs are each drums onsize flow and the total flow. One drum is chosen as phase reference drum, and PID controllers are applied for controlling each of the slave-drums phase angel relative to the reference drum. In figure 1.7, the upper figure shows the total flow and the lower figure shows each drums onsize flow. As in figure 1.2, only the variation around the working point is shown.

In the first 800 sec. the drums are started and the oscillations are established. After this time delay the PID controllers starts to adjust the phase angels, and the amplitude in the total flow starts to decrease. From figure 1.7 we clearly see that the control is quite slow. Experiments with this model shows that it is quite difficult to speed up the control. The reason for this is somewhat unclear, however we expect that one of the following modifications may allow faster control:

a) Use of a multivariable controller structure rather than single loop controllers

b) The basis for the signal model used in the EKF has no local support, so that any changes will affect the output for all future. A change to a basis with local support e.g. a wavelet basis may therefore be desirable.

1.6 Concluding remarks

In this paper we have suggested a new approach to the old problem of controlling the balling drums used for pelletizing iron ore. A prerequisite for this method is that is has to be possible to estimate the oscillations amplitude, phase and frequency with sufficient accuracy. We have investigated the possibility for using an extended Kalman filter for this purpose. In section 4 we show that this is indeed possible.

The proposed control scheme has been simulated for a plant consisting of three drums. One major problem which remains to be solved is that it takes too long time for the output to settle down. One of the reasons for this may be that the basis for the signal model is a trigonometric basis, which has no local support. A possible solution of this problem may therefore be to use a signal model with local support, e.g. a wavelet basis. Another possible solution may be to use a multivariable controller structure rather than single loop controllers, which are used in the simulation presented in section 5.

ACKNOWLEDGEMENT

The authors are grateful to LKAB, Sweden, for support both financially and practical during this research. In particular Mr. Magnus Rutfors deserves a great thanks for providing us with measurement data from the real process.

References

[1] C. K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications*. Springer, Berlin, 1999.

[2] M. Cross. Mathematical model of balling-drum circuit of a pelletizing plant. *Ironmaking and steelmaking*, pages 159–169, 1977.

[3] M. Cross, R. W. Young, P. E. Wellstead, and R. D. Gibson. The mathematical modelling and control aspects of the pelletizing of iron ores. *Agglomeration 77, AIME New York*, pages 403–424, 1977.

[4] R. Dügge, O. Erikson, and Magnus Rutfors. Unpublished work. 2000.

[5] K. V. S. Sastry. The agglomeration of particulate materials by green pelletization. *Ph.D thesis, University of California, Berkeley*, 1970.

[6] K. V. S. Sastry. Process engineering of agglomeration systems. *The Sixth International Symposium on Agglomeration, Nagoya, JAPAN*, 1993.

[7] K. V. S. Sastry and D. W. Fuerstenau. Kinetic and process analysis of the agglomeration of particulate materials by green pelletization. *Agglomeration 77, AIME New York*, pages 381–402, 1977.

[8] P. E. Wellstead, M. Cross, N. Munro, and D. Ibrahim. On the design and assessment of control schemes for balling-drum circuits used in pelletizing. *International Journal of Mining Processing*, pages 45–67, 1978.

[9] P. E. Wellstead and N. Munro. Multivariable control of cold iron ore agglomeration plant. 1977.

PAPER 2

Genetic algorithm based tuning of an Extended Kalman Filter[1]

Knut Rapp[*], **Per-Ole Nyman**[*]

[*] Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, kr@hin.no, fax: +47 76 96 68 10
[†] Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, pon@hin.no, fax: +47 76 96 68 10

Abstract: In this paper it is shown how a simple genetic algorithm can be utilized for tuning the parameters of an extended Kalman filter. Results from applying this genetic algorithm on a specific problem related to signal tracking is shown. Further, the performance index and how to obtain a simple objective function for the genetic algorithm are discussed. Finally it is shown that even if the search space is only two dimensional, the genetic algorithm is faster than a simple grid search method.

[1] In Proceedings of the 9th IEEE International Conference on Methods and Models in Automation and Robotics, 25 - 28 August 2003, Miedzyzdroje, Poland

2.1 Introduction

Since its introduction in 1960, the Kalman filter technique has grown to be perhaps one of the most often used in the area of process control. The results achieved from the last forty years of research have contributed to a continued expansion of this technique. In spite of this, the problems of parameter tuning are still not very well investigated. A very limited amount of paper and results seems to be published, but some methods for tuning the filter, including the extended Kalman filter (EKF), are reported see [6], [7].

In this paper the problem of tuning an EKF used for tracking a periodic signal from a noisy measurement is considered. The signal is assumed to be nearly sinusoidal, with varying frequency and amplitude. This filter is applied to the classical problem of controlling the balling drums used in the iron ore industry. The iron ore pelletizing process can roughly be divided into two parts, the cold part and the warm part. The cold part is the process where the green pellets is produced, and in the warm part the green pellets are dried and fired. The main components in the cold part are the balling drums in which the green pellets are made. This drums tend to give an slowly oscillating output when the moisture content in the pellets is below a certain limit, see [2], [9]. In this work it is assumed that the output is oscillating most of the time.

It is important that the flow from the cold part to the warm part is constant. An oscillating mass flow will normally result in uneven drying so that part of the pellets will be too wet when it enters the firing process. Too high moisture content will cause the pellet to crack during the firing process. An oscillating mass flow will therefore result in a lower production rate.

From an economical point of view, the moisture content should be kept at a minimum, due to the fact that it is more expensive to remove moisture by drying in the warm process, than removing it before the fines enters the drums. Moreover, the drums throughput will increase when the moisture level decrease. The product quality will also increase, see e.g. [2]. It is therefore desirable to keep the water content in the fines as low as possible. However, this will normally result in an oscillating drum output.

Several attempts have been made in order to stabilize an oscillating drum, see e.g. [3]. A major problem is that no reliable measurement device is available to measure on-line the moisture content in the green pellets. An indirect ”measurement” of the moisture content is the size of the oscillation. However, when the oscillation disappears, which is the goal of the stabilizing control, this information is lost. A possible solution is then to permit each drum to deliver an oscillating output, and design an automatic control for a cluster of drums in order to make the total mass flow constant, rather than stabilizing the output of each drum, see [8]. In the end, this will result in lower production costs, good product quality and high production rate. However, a corner stone in this automatic control is a filter which detects the oscillation and gives information about the amplitude and frequency. The specific filter chosen for this application is an Extended Kalman filter (EKF). Satisfactory control performance is dependent on a well designed filter. In the following, it is discussed how a simple genetic algorithm (GA) can be utilized to achieve this goal.

The outline of the paper is as follows: in Section 2 a very brief description of the up today reported results on automatic filter tuning is given. Some results regarding tuning parameters and performance indices are mentioned. In Section 3 and 4, the particular filter used in the signal tracking mentioned above is specified. The choice of performance index is explained in detail, and the results from the simulations are presented. In Section 5 concluding remarks are stated.

2.2 Preliminary results

2.2.1 Automatic tuning methods

As mentioned in the previous section, only a few methods for automatic parameter tuning are reported. In [7] some numerical methods are listed. In these methods the tuning problem is converted to a numerical optimization problem. In particular, Powell discusses and gives examples of how a simplex downhill algorithm can be used to solve the tuning problem when it is posed as a numerical optimization problem. In [6], the use of genetic algorithms are discussed. None of the above mentioned papers gives any time estimate of the tuning process. Clearly this will depend on how complex the problem is, but also on how the algorithm is implemented.

2.2.2 Tuning parameters

In [1] the following well known fact, which simplifies the tuning problem, is established and proved:

Let R be the measurement noise covariance and Q the process noise covariance. Let ϕ be the ratio between R and Q. Now in order to change the Kalman filter gain, only ϕ needs to be changed, if the following Jacobi matrix remains unchanged.

$$A(t) = \frac{\partial f(x)}{\partial x} \tag{2.1}$$

where $A(t)$ is the linearized state matrix obtained from the nonlinear function $f(x)$ of the given signal model:

$$\dot{x}(t) = f(x) + v(t) \tag{2.2}$$

$$y(t) = g(x) + z(t) \tag{2.3}$$

Then R or Q can be fixed while the other is used as tuning parameter.

In the case of R and Q being matrices, the idea of fixing one matrix is still valid, see [5] and [7]. Normally R is fixed as it is possible to determine its elements by testing or by statistics applied on real process measurements.

2.2.3 Performance index and evaluation function

Several performance indices are suggested in [6], [7]. Among these are:

1. "Whiteness" test of the residual.
2. The weighted value of the state estimation error.
3. The value of the measurement residual.

Test number one is suitable for linear filters only as it is not possible to guarantee the whiteness of the residual of a nonlinear filter, even if it is optimal tuned, see [7]. Test number two and three require the real signal to be known and are therefore only suitable when the system is simulated. One reasonable performance index, which correspond to alternative number two, is:

$$J(q_{11}, ..., q_{nn}) = \frac{1}{T - t_0 + 1} \sum_{t=t_0}^{T} \left(\widehat{e}_t^T W \widehat{e}_t \right) \tag{2.4}$$

where W is a weighting matrix, $\widehat{e}_i$ is the state estimation error vector and $[t_0,\ T]$ is the time interval over which the filter is tuned. This performance index is the one used later in this paper.

One basic element in numerical optimization by use of genetic algorithms is the evaluation (objective) function. This function provides information needed to judge which filter candidate among a set of candidates is performing best. To obtain a representative value from a performance index, like the weighted value of the state estimation error, Monte Carlo simulation may be used. The number of Monte Carlo runs should be sufficient to provide reliable information about the filter candidate's performance. This is often a time consuming part of the optimization algorithm.

2.3 The filter

2.3.1 Description

The signal model used by the EKF is a third order state space model, as follows:

$$x_1(k+1) = x_1(k) + v_1 \tag{2.5}$$

$$x_2(k+1) = x_1(k) + x_2(k) \tag{2.6}$$

$$x_3(k+1) = x_3(k) + v_2 \tag{2.7}$$

$$y(k) = x_3(k) \sin x_2(k) + z(k) \tag{2.8}$$

where x_1 is the phase increment (or frequency), x_2 is the phase and x_3 is the amplitude of the oscillation. $\mathbf{v} = [v_1, v_2]^T$ and z are white, zero mean processes with covariance matrices Q =diag(q_1, q_2) and R. Locally the state is uniquely determined by the output y, but owing to the factor $\sin x_2$ in output equation, this does not hold globally. In fact, a simultaneous change of sign in x_1 and x_2, or a shift of x_2 by any number of periods, does not change the output. However, with a reasonable initialization of the EKF this mild nonuniqueness does in general not cause problems. The choice of the matrix Q is a compromise between accuracy in steady state and capability to track a changing amplitude or frequency. R is set equal to the covariance of assumed measurement noise of the true, measured output.

2.3.2 Filter specification and performance index

As mentioned above, the filter is used for tracking a periodic signal with variable frequency and amplitude. Occasionally the amplitude may go to zero for some period. It is assumed that the amplitude is a function of the moisture and binder content; see [2] and [9]. Furthermore, it is assumed that the amplitude is controllable in some interval $a_1 < a < a_2$. If $a < a_1$, then the amplitude is out of the controllable area, and the oscillation is regarded to be nonexisting. By proper adjustment of the moisture and binder content the oscillation will be present again after some time. This reappearance is not necessarily very smooth, so the filter is required to have good transient response. Furthermore, the frequency tracking ability is important in order to obtain good control of the total flow. The same could be said about the amplitude and the phase. Therefore, the filter is required to estimate all three states with a high accuracy. For this reason the weighting matrix in the performance index is initially chosen equal to the unity matrix, i.e. $W = I$. As mentioned above, R is assumed known, and q_1 and q_2 are tuning parameters. The performance index is thus:

$$J(q_1, q_2) = \frac{1}{T+1} \sum_{0}^{T} \left(\widehat{e}_t^T W \widehat{e}_t \right) \tag{2.9}$$

where $\widehat{\mathbf{e}} = [(\widehat{x}_1 - x_1), (\widehat{x}_2 - x_2), (\widehat{x}_3 - x_3)]^T$ is the estimation error vector, T is the final time, which equals the number of samples and

$$W = \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{bmatrix} \tag{2.10}$$

is the 3 × 3 weighting matrix. Initially $w_1 = w_2 = w_3 = 1$. In all simulations $T = 2000$.

2.3.3 The evaluation function

The evaluation function gives the "fitness value" for each filter candidate to be considered. It is specific for each application, and the GA is designed to maximize it rather than minimizing it. A suitable evaluation function candidate, which is used further in this paper, is defined by:

$$E(Q) = N \left(\sum_{1}^{N} J(q_1, q_2) \right)^{-1} = N \left[\sum_{1}^{N} \left(\frac{1}{T+1} \sum_{t=0}^{T} \left(\widehat{e}_t^T W \widehat{e}_t \right) \right) \right]^{-1} \tag{2.11}$$

where N is the number of Monte Carlo runs used for evaluating one filter candidate. In the following $N = 50$. The best filter candidate will now be identified by the largest evaluation value (or fitness value), $E(Q)$.

During the optimization one should make sure that the same value $E(Q)$ is obtained at all evaluations of one and the same filter. If different test signals is used for each set of Monte Carlo runs, this may not be the case. It is then more difficult to select the best of two filters. This may be done, but the evaluation function should then include a statistical hypothesis test. This will result in a more complex and time consuming algorithm. To avoid doing this, one may generate a set of test signals when starting the GA, and use this set of signals all the way through. This means that if one want to use N Monte Carlo runs to evaluate each filter candidates performance, a set of N random signals are created, and this set of signals is used to evaluate all filter candidates in the optimization process.

2.3.4 The test signal

In order to obtain a reasonable value from the evaluation function after N Monte Carlo runs, a reasonable test signal is to be applied. This test signal should include both a period with steady signal value, a transient period and a period with a oscillating signal. The test signal applied in the simulations is shown in Figure 2.1, where the solid drawn line represents the real signal without noise. The time interval for which the signal is generated is $0 - 2000$ [s].

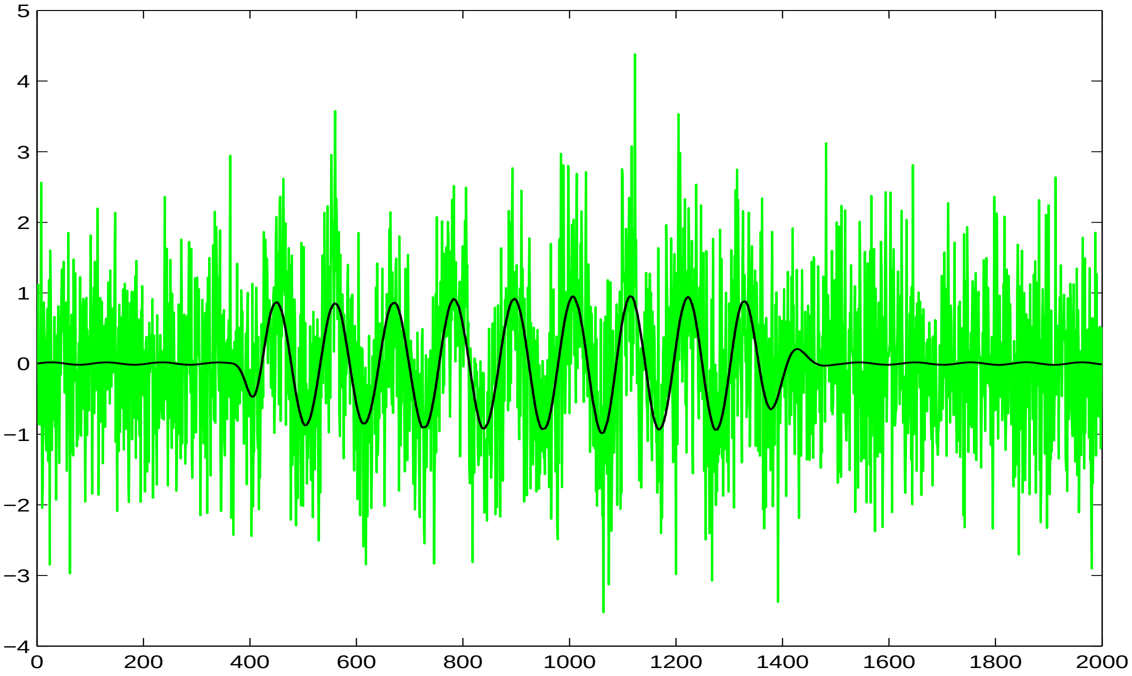

Figure 2.1: Testsignal with white noise

2.4 Results from simulations

In this section the results from running the GA with three different choices of the weighting matrix W are shown. Under each subsection the filter performance is illustrated by figures.

The maximization of the evaluation function is carried out with the free GAOT toolbox, see [4]. The initial population is set to 200 individuals, and it evolves over 100 generations. The typical time consumption for each GA simulation is 4-6 hours on a medium speed PC for the numbers of Monte Carlo runs, start population and number of generations specified above. However, it is interesting to note that even in this simple 2 dimensional optimization problem, the GA is much faster than a search over a grid, using the same evaluation function. For instance, the simulation in the following subsection equals in time consume a search over a 28x28 grid, which is very rough. If we scale the search area to start on 1, it will be $[1, 10^4] \times [1, 10^4]$. With only 28 points on the interval $[1, 10^4]$, one should not have any expectation at all to the performance of the resulting filter. On the other hand, a plot of the evaluation function based on a rough grid search may be very useful if one wishes to get an indication of the most promising area to search. It should be noted that the time consumption mentioned above holds when using MATLAB 6. It will be considerably reduced if the algorithm is implemented in e.g. C++ rather than in an interpreter. In this work however, the main task is not to make the algorithm time optimal.

2.4.1 Tuning with $W = I$

In section 2.3.2 the initial choice of the weighting matrix was the identity matrix. This is to reflect that all states should be estimated with as high accuracy as possible. In approximately 4 hours and 30 minutes the GA converged to an optimally tuned filter. In Figure 2.2, the amplitude diagram from this filter is shown. The dashed red line is the true amplitude and the blue line is the estimate. In Figure 2.3, the frequency diagram for the same filter is shown. The performance of the optimally tuned filter is validated through numerous Monte Carlo simulations, and it is found to work satisfactory.

Figure 2.4 and 2.5 shows a representative amplitude- and frequency diagram for a filter resulting from a search over a grid of 4096 points (64×64). The searching time was approximately 8 hours and 55 minutes. Clearly this filter performs weaker than the optimal filter found by the genetic algorithm, as it is far more noise sensitive. This is confirmed by numerous simulations, where the two filters performance are compared.

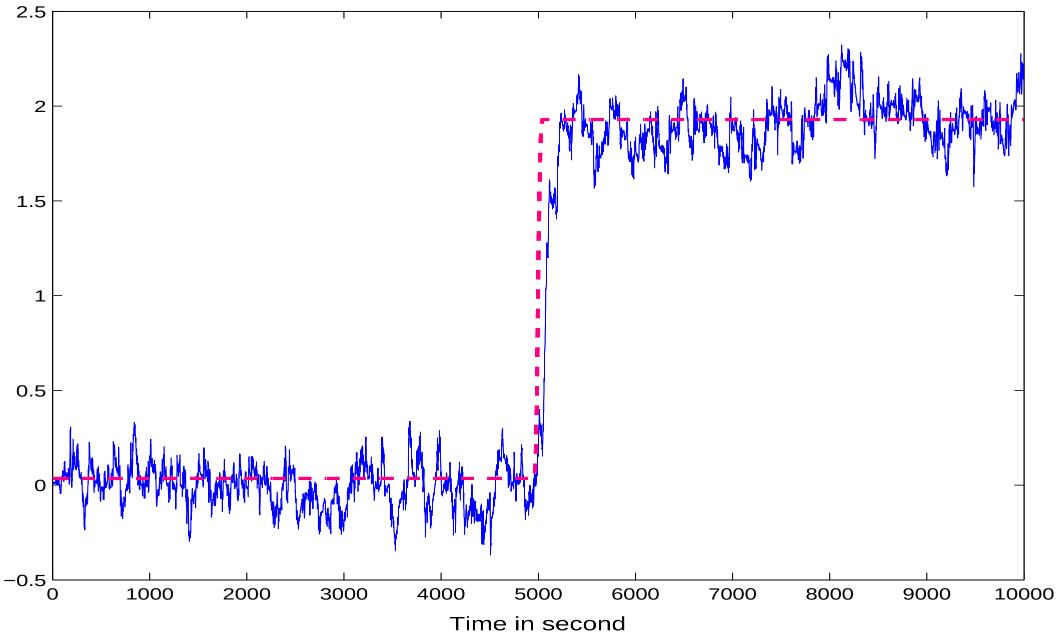

Figure 2.2: Amplitude diagram, W=I

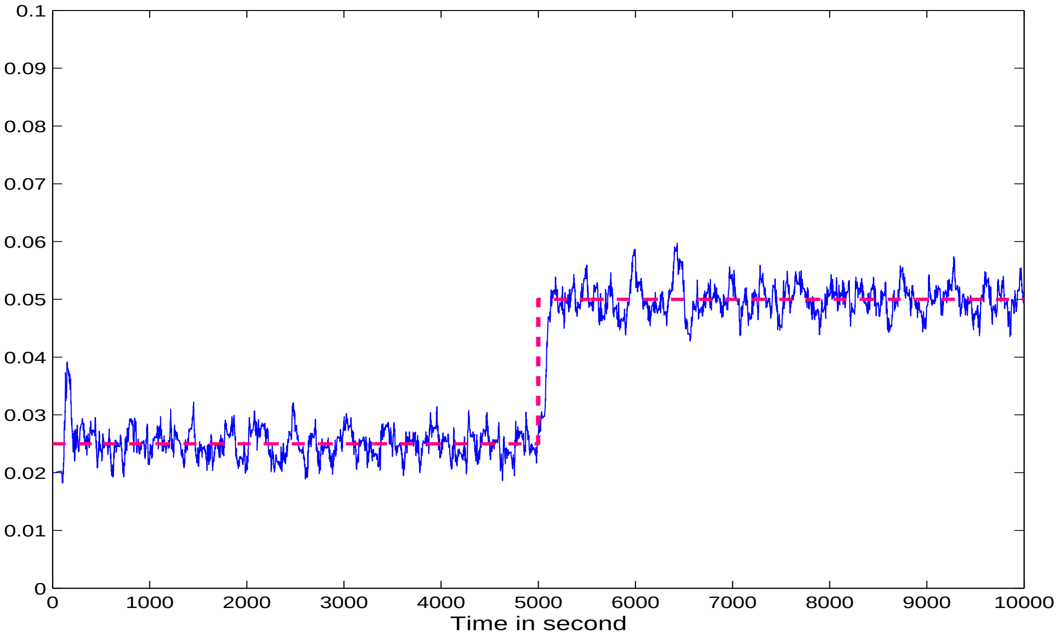

Figure 2.3: Frequency diagram, W=I

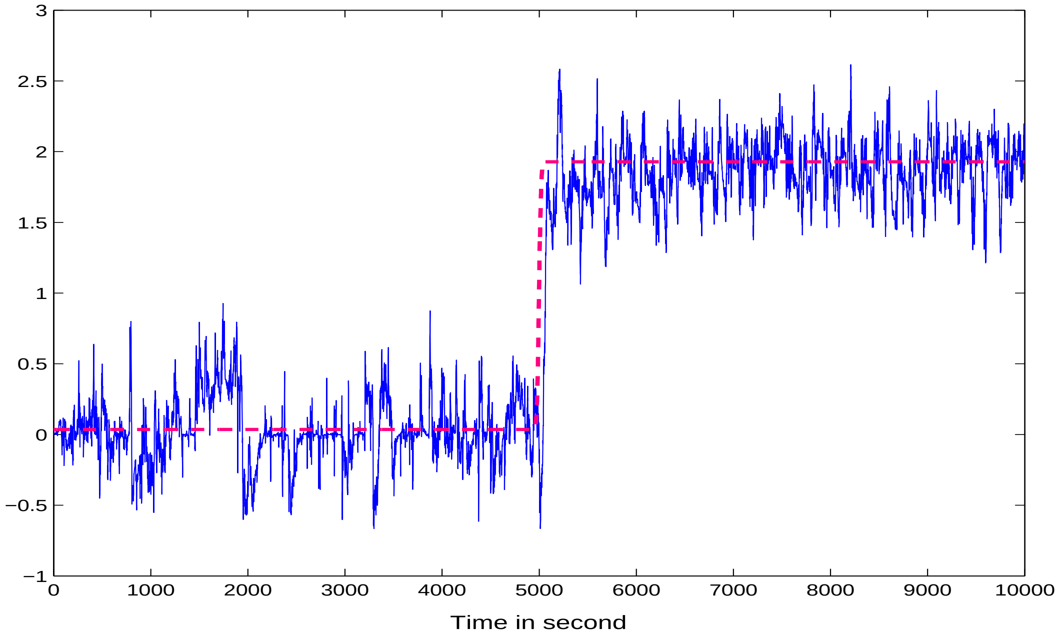

Figure 2.4: Amplitude diagram, grid search

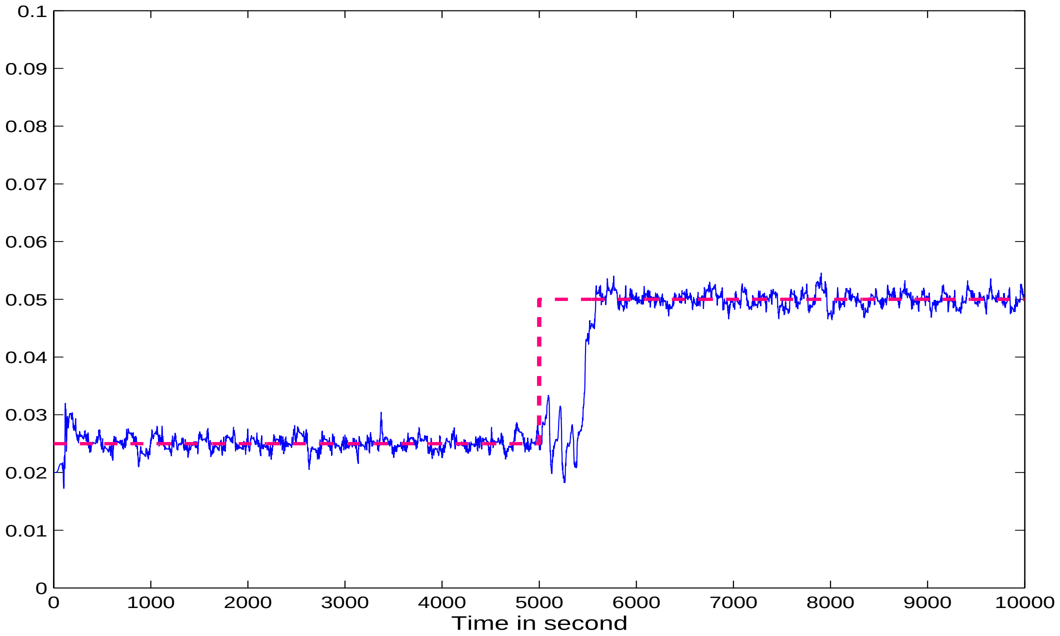

Figure 2.5: Frequency diagram, grid search

2.4.2 Tuning with $W \neq I$

It is of interest to investigate what happens if the weighting matrix is changed. Two different cases are considered. The first is with $W = \text{diag}[10^3, 1, 1]$ which puts a high weight on the frequency. The second case is with $W = \text{diag}[1, 1, 10^3]$ which puts a high weight on the amplitude. In Figure 2.6, 2.7, 2.8 and 2.9, representative amplitude and frequency diagrams for these filters are shown.

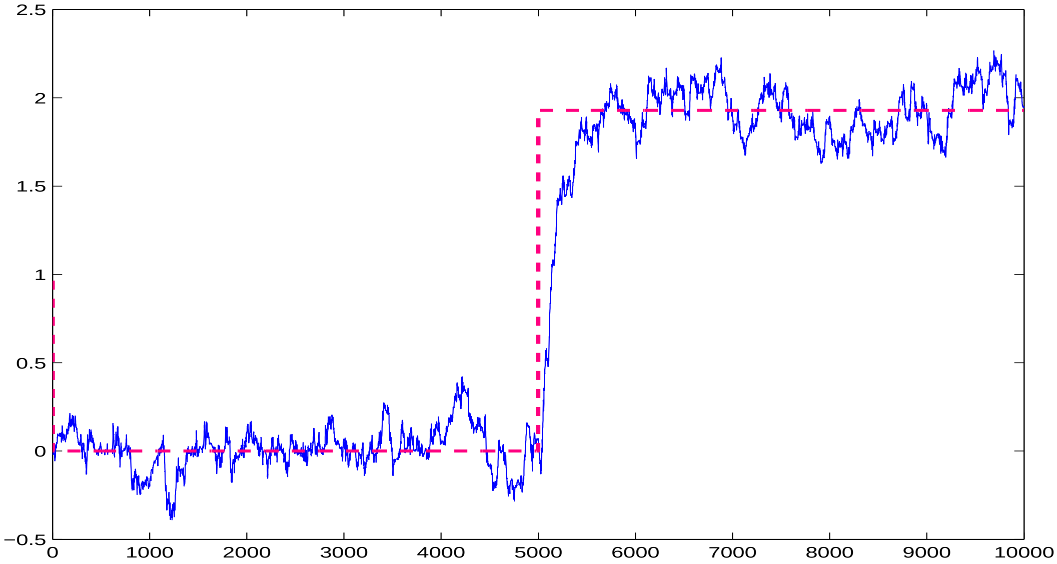

Figure 2.6: Amplitude diagram

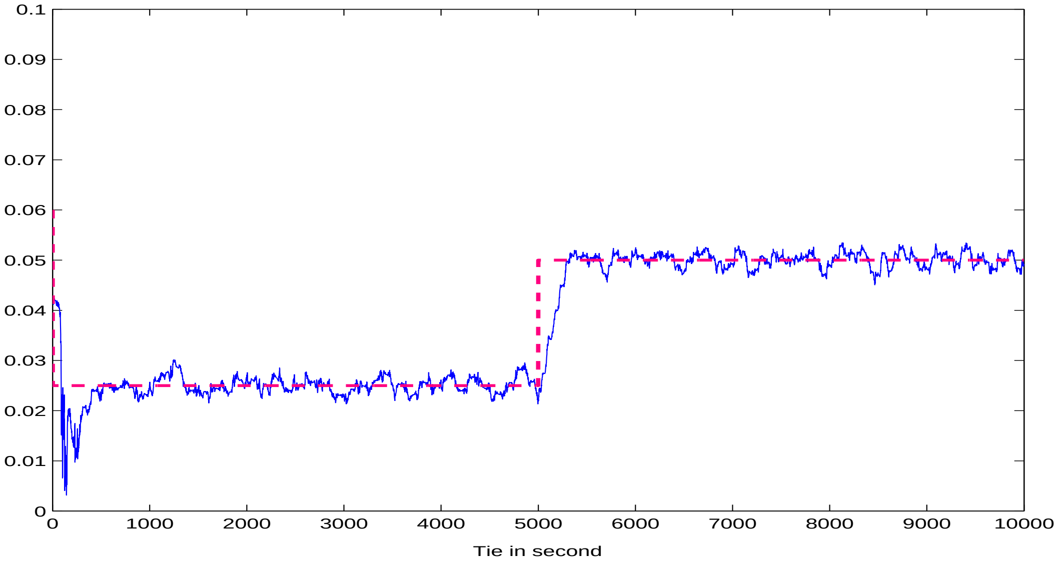

Figure 2.7: Frequency diagram

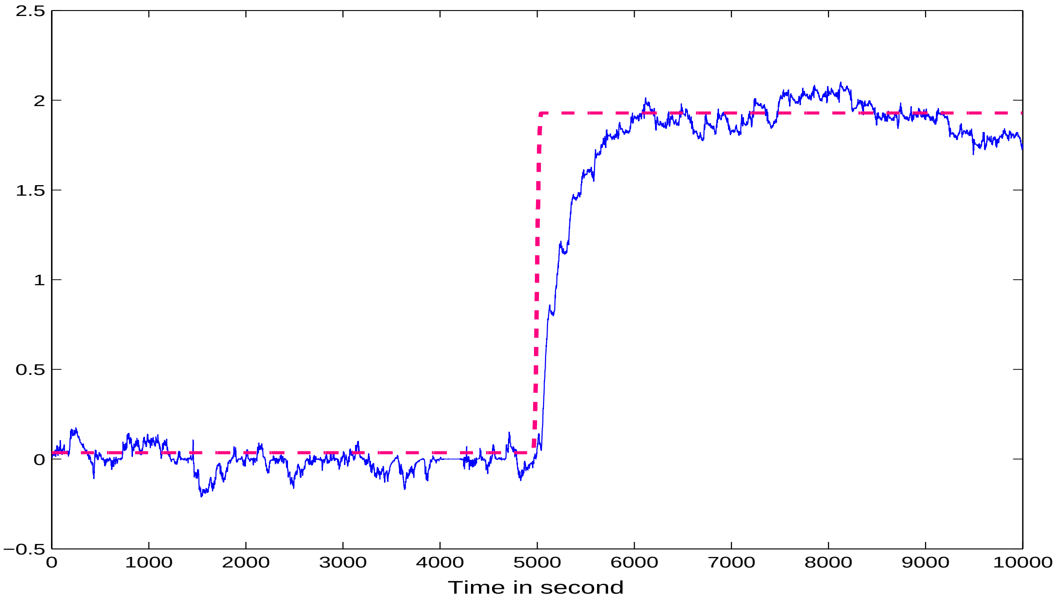

Figure 2.8: Amplitude diagram

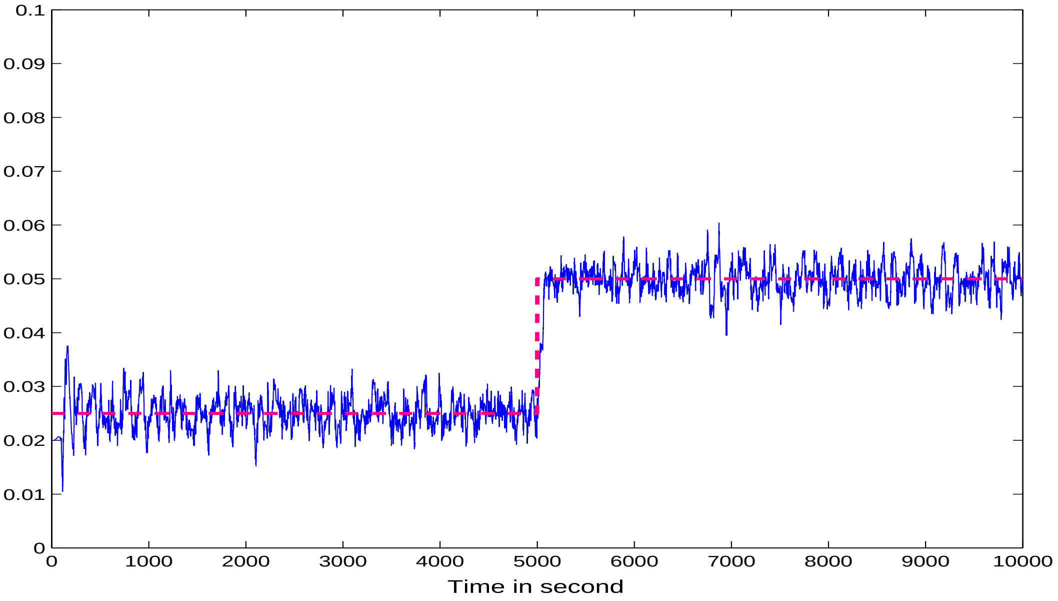

Figure 2.9: Frequency diagram

If we compare the responses in Fig. 2.2 and 2.3, with those shown in Fig. 2.6 and 2.7, we clearly see that the frequency estimate is better in Fig. 2.7. However, the response time for both the amplitude and frequency have increased. The elements of the covariance matrix, q_1 and q_2, have decreased by the factors 0,25 and 0,095 from Fig. 2.2 (2.3) to Fig. 2.6 (2.7), and this results in a filter which responds less to the measurements than the filter in Fig. 2.2 and 2.3. A consequence of this is that the estimate from the filter is slower and less noisy, as is seen in Fig. 2.6. and 2.7.

It we compare the responses in Fig. 2.2 and 2.3 with those shown in Fig. 2.8 and 2.9, we see a situation quite similar to the previous case when $W = diag[10^3, 1, 1]$. The estimate is less noisy and the response time is increased for the amplitude. In the covariance matrix q_1 is increased by a factor 8 and q_2 is decreased by a factor 0.072 from Fig. 2.2 (2.3) to Fig. 2.8 (2.9). With respect to the amplitude estimate, the filter will respond less to the measurements than the filter in Fig. 2.2 and 2.3, and pay more attention to the predicted state. The slow reaction on changes in amplitude is a consequence of this. For the frequency the opposite holds.

An advantage with filters who's output is heavily based on the prediction, is that it is more likely to work properly even if the signal to be estimated is very noisy. In the present application the most serious disadvantage of such a filter is the slower transient response. As in all application, one wishes to have a filter which can handle very noisy signals and at the same time have a quick transient response and excellent tracking ability. The best compromise seems to be somewhere in between the two filters which responses are shown in Fig. 2.6 and 2.7, and Fig. 2.8 and 2.9. A reasonable compromise is shown in Fig. 2.2 and 2.3, when all states are weighted equally.

2.5 Concluding remarks

In this paper some aspects regarding optimal tuning of an Extended Kalman filter by use of genetic algorithms has been discussed. The following main conclusions can be stated:

1. The genetic algorithm (GA) is a tool well suited for automatic filter tuning. It requires however, an evaluation function which gives a unique value for each filter candidate if the result is to be reliable. An alternative is to introduce statistical hypothesis testing in the evaluation function, but this will give a more complex function which require more time to be executed.

2. The time required to tune an EKF is considerably lower when using a simple genetic algorithm compared to searching over a grid. As shown in Section 2.4.1, a better result is obtain in shorter time when using the genetic algorithm as optimization tool, even for this simple 2D search problem.

ACKNOWLEDGEMENTS

The authors are grateful to LKAB, Sweden, for support both financial and practical during this research. In particular Mr. Magnus Rutfors deserves a great thanks for providing us with measurement data from the real process.

References

[1] S. Bittanti and S. Savaresi. On the parameterization and design of an extended Kalman filter frequency tracker. *IEEE Transaction on automatic control*, 45: 1718–1724, 2000.

[2] M. Cross. Mathematical model of balling-drum circuit of a pelletizing plant. *Ironmaking and steelmaking*, pages 159–169, 1977.

[3] M. Cross, R. W. Young, P. E. Wellstead, and R. D. Gibson. The mathematical modelling and control aspects of the pelletizing of iron ores. *Agglomeration 77, AIME New York*, pages 403–424, 1977.

[4] C. R. Houck, J. A. Joines, and M. G. Kay. The genetic algorithm optimization toolbox (gaot) for matlab 5. *http://www.ie.ncsu.edu/mirage/GAToolBox /gaot/*, 1996.

[5] Peter S. Maybeck. *Stochastic Models, Estimation and Control.* Academic Press, New York, 1979.

[6] Y. Oshman and Ilan G. Shaviv. Optimal tuning of a Kalman filter using genetic algorithms. *AIAA Paper 2000-4558*, 2000.

[7] T. D. Powell. Automated tuning of an extended Kalman filter using the downhill simplex algorithm. *Journal of guidance, control and dynamics*, 25:901–908, 2002.

[8] K. Rapp and P. -O. Nyman. Control of the amplitude in a surging balling drum circuit, a new approach to an old problem. *Presented at ECC 2003, Cambridge UK*, 2003a.

[9] K. V. S. Sastry and D. W. Fuerstenau. Kinetic and process analysis of the agglomeration of particulate materials by green pelletization. *Agglomeration 77, AIME New York*, pages 381–402, 1977.

PAPER 3

Optimization of extended Kalman filter for improved thresholding performance[1]

Knut Rapp*, **Per-Ole Nyman***

* Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, kr@hin.no, fax: +47 76 96 68 10
† Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, pon@hin.no, fax: +47 76 96 68 10

Abstract: In this paper it is shown how a simple genetic algorithm can be utilized for tuning the parameters of an extended Kalman filter. Results from applying this genetic algorithm on a specific problem related to signal tracking is shown. Further, the performance index and how to obtain a simple objective function for the genetic algorithm are discussed. Finally it is shown how the genetic algorithm can be modified in order to optimize the filters thresholding performance.

[1] In Proceedings of the 2nd IFAC conference on Control System Design, 7-10 September 2003, Bratislava, Slovak Republic

3.1 Introduction

Since its introduction in 1960, the Kalman filter technique has grown to be perhaps one of the most often used in the area of process control. The results achieved from the last forty years of research have contributed to a continued expansion of this technique. In spite of this, the problems of parameter tuning are still not very well investigated. A very limited amount of paper and results seems to be published, but some methods for tuning the filter, including the extended Kalman filter (EFK), are reported see [6], [5]. Normally it can be assumed that the process measurement noise covariance matrix is known, either from measured values, or from statistics applied on the real process data. However, if the noise covariance is not constant, the filter is to be designed to handle the variations which may occur.

In this paper the problems of tuning and optimizing an EKF used for tracking a periodic signal from a noisy measurement are considered. The signal to be tracked is assumed to be nearly sinusoidal, with varying frequency and amplitude. Further the problem with time varying measurement noise is treated. A simple design method is applied, and the results are discussed. This filter is applied to the classical problem of controlling the balling drums used in the iron ore industry. For this reason it is assumed that the signal to be tracked has quite low frequency. Normally it will be in the interval (0.015, 0.15)$[\frac{rad}{s}]$. For details see e.g [7].

The outline of the paper is as follows: In Section 2, the signal model is described and the EKF equations for this specific model are stated. In section 3, a very brief description of the up today reported results on automatic filter tuning is given. Some results regarding tuning parameters and performance indices are mentioned. It is shown how different choices of the weighting matrix will give filters with different properties. This is illustrated by three different cases. In Section 4, the problem of increasing the filters thresholding performance is discussed, and it is demonstrated how the genetic algorithm may be utilized to achieve this. In Section 5 concluding remarks are stated.

3.2 Signal model and EKF equations

3.2.1 Signal model

The signal model used by the EKF is a third order state space model, as follows:

$$x_{k+1}^{(1)} = x_k^{(1)} + v_k^{(1)} \tag{3.1}$$

$$x_{k+1}^{(2)} = x_k^{(1)} + x_k^{(2)} \tag{3.2}$$

$$x_{k+1}^{(3)} = x_k^3 + v_k^{(2)} \tag{3.3}$$

$$y_k = x_k^3 \sin x_k^2 + z_k \tag{3.4}$$

where $x^{(1)}$ is the phase increment (or frequency), $x^{(2)}$ is the phase and $x^{(3)}$ is the amplitude of the oscillation. $v = [v^{(1)} \ \ v^{(2)}]^T$ and z are white, zero mean processes with covariance matrices $Q = \mathrm{diag}(q_1, q_2)$ and R. Locally the state is uniquely determined by the output y, but owing to the factor $\sin(x^{(2)})$ in output equation, this does not hold globally. In fact, a simultaneous change of sign in $x^{(1)}$ and $x^{(2)}$, or a shift of $x^{(2)}$ by any number of periods, does not change the output. However, with a reasonable initialization of the EKF this mild nonuniqueness does in general not cause problems. The choice of the matrix Q is a compromise between accuracy in steady state and capability to track a changing amplitude or frequency. R is set equal to the covariance of assumed measurement noise of the true, measured output.

3.2.2 EKF equations

Written in a more compact form, equation (1)-(4) are given by

$$x_{k+1} = Ax_k + Bv_k \tag{3.5}$$

$$y_k = g_k(x_k) + z_k \tag{3.6}$$

The EKF equations derived from equation (1)-(4) are given by (see e.g. [1]):

$$\hat{x}_{k,k-1} = A\hat{x}_{k-1} \tag{3.7}$$

$$P_{k,k-1} = A \cdot P_{k-1,k-1} \cdot A^T + BQB^T \tag{3.8}$$

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k(y_k - g_k(\hat{x}_{k,k-1})) \tag{3.9}$$

$$P_{k,k} = P_{k,k-1} - K_k \left[\frac{\partial g_k}{\partial x_k} \hat{x}_{k,k-1}) \right] P_{k,k-1} \tag{3.10}$$

where $$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$G_k = \left[\frac{\partial g_k}{\partial x_k}\right]_{x=x_{k,k-1}} = \left[\begin{array}{ccc} 0 & x^{(3)}_{k,k-1}\cos(x^{(2)}_{k,k-1}) & \sin(x^{(2)}_{k,k-1}) \end{array}\right] \tag{3.11}$$

The filter gain matrix is given by

$$K_k = P_{k,k-1}\left[G_k(\hat{x}_{k,k-1})\right]^T \left[\left[G_k(\hat{x}_{k,k-1})\right] P_{k,k-1}\left[G_k(\hat{x}_{k,k-1})\right]^T + R\right]^{-1} \tag{3.12}$$

3.3 Automatic filter tuning

3.3.1 Automatic tuning methods

As mentioned in section 3.1, only a few methods for automatic parameter tuning are reported. In [6] some numerical methods are listed. In these methods the tuning problem is converted to a numerical optimization problem. In particular, Powell discusses and gives examples of how a simplex downhill algorithm can be used to solve the tuning problem when it is posed as a numerical optimization problem. In [5], the use of genetic algorithms are discussed. In this paper, an genetic algorithm is applied for the numerical optimization.

3.3.2 Filter specifications

This work focuses on the filters ability to track the signal frequency and amplitude. As mentioned in section 3.1, the signal is assumed to have a quite low frequency. Three properties are considered to be important:

a) The filter should be able to track the signal even if it is very noisy

b) The variance in the filter's estimate is to be as low as possible

c) The filter's transient response is to be sufficiently fast

As in almost all set of specifications, this set also include some contradictory wishes. It may be difficult, or even impossible, to obtain very quick transient response and very low variance in the estimate at the same time. An optimally tuned filter will therefore be a trade off between the specifications listed above.

3.3.3 Filter performance index

In this paper R is assumed known so q_1 and q_2 are the available filter tuning parameters (see [4]). In order to apply an numerical optimization algorithm, a suitable performance index is to be available, see [8] or [6]. A reasonable filter performance index is the (weighted) RMS of the estimation error. As R is constant the performance index will be a function of the elements in Q only. The performance index is given by:

$$J(q_1, q_2) = \left[\frac{1}{T+1} \sum_{t=0}^{T} \left(\hat{e}_t^T W \hat{e}_t \right) \right] \tag{3.13}$$

where $\hat{e} = [(\hat{x}_1 - x_1), (\hat{x}_2 - x_2), (\hat{x}_3 - x_3)]$ is the estimation error vector, T is the final time, which equals the number of samples and $W = \operatorname{diag}(w_1 \;\; w_2 \;\; w_3)$ is the 3x3 weighting matrix. In all simulations $T = 2000$.

3.3.4 The evaluation function

A basic element in a genetic algorithm is the evaluation function. The evaluation function gives the ”fitness value” for each filter candidate to be considered. It is specific for each application, and depends also of the genetic algorithm to be used. The GA used in this work is the free GAOT toolbox, which is design to maximize the evaluation function rather than minimizing it. See [2] for a complete description of this algorithm and how it is implemented. A suitable evaluation function candidate is given by

$$E(Q) = N \left[\sum_{1}^{N} \left(\frac{1}{T+1} \sum_{t=0}^{T} \left(\hat{e}_t^T W \hat{e}_t \right) \right) \right]^{-1} \tag{3.14}$$

where N is the number of Monte Carlo runs used for evaluating each filter candidate. See [8] for further details.

3.3.5 Results from simulations

The filter is tuned with a signal having amplitude 1 and frequency $6 \cdot 10^{-2} \, \frac{rad}{s}$. The signal noise variances are $R_s = 1$ and $Q_s = 10^{-8}$. Three cases are considered.

1. Equal weight on the frequency and amplitude error.

2. High weight on the frequency error

3. High weight on the amplitude error

In all three cases the initial population is set to 200 individuals, and it evolves over 100 generations.

In Figure 3.1, the amplitude and frequency diagram from this filter in case 1 is shown. The dashed line is the true amplitude and the solid-drawn line is the estimate.

In Figures 3.2 and 3.3 the amplitude- and frequency diagram for the filters in case 2 and 3 are shown.

As pointed out in section 3.3.2, the cost of a quick transient response will be a more noisy estimate. This is illustrated in Figure 3.1 and Figure 3.2.

In case 1 the filter responds quickly to the changes in both amplitude and frequency. However, the noise in both the amplitude and frequency estimate is considerable.

In case 2 the estimates, and in particular the frequency estimate, have lower variance. This is expected as the estimation error in the frequency is weighted quite heavily. The step response time is, however, increased.

In case 3 the variance in the amplitude estimate is decreased, but as in case two, the step response time is increased.

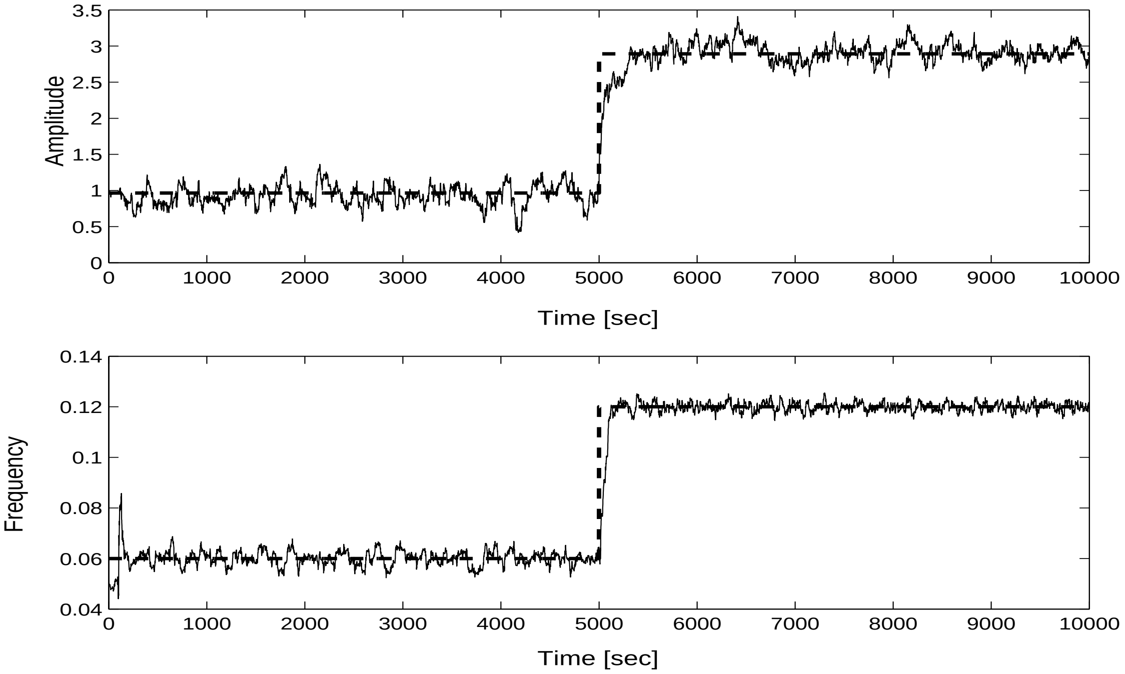

Figure 3.1: Amplitude and frequency diagram 1

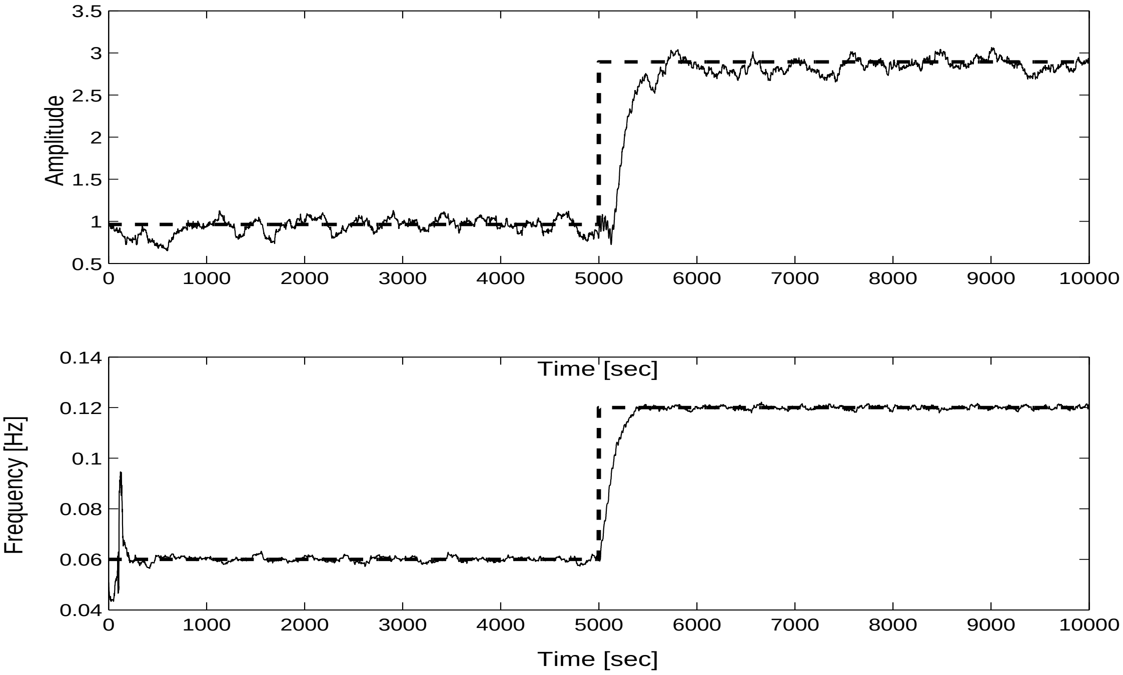

Figure 3.2: Amplitude and frequency diagram 2

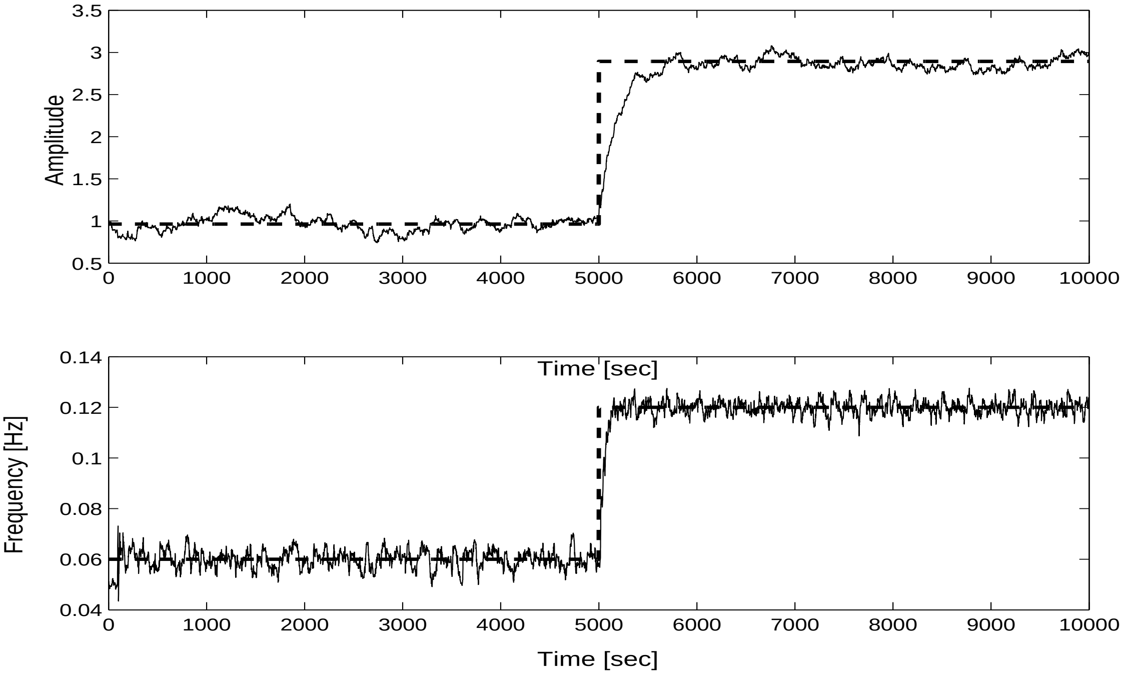

Figure 3.3: Amplitude and frequency diagram 3

These cases demonstrates that the genetic algorithm tool is well suited for filter tuning. If some of the properties are more important than others, e.g. low variance in one of the estimated states, then it is very easy to reflect this in the tuning process.

3.4 Optimizing the filters thresholding perfomance

3.4.1 The thresholding phenomenon

If the filter parameters (q_1 and q_2) are tuned for one value of the signals measurement noise R_s only, then it is impossible to guarantee the performance if it changes. If R_s decreases it may be expected that the performance is satisfactory. However, to obtain the optimal EKF for the given model, the filter is to be re-tuned. If R_s increases, the estimation error may become unacceptable large. In Figure 3.4 the relative frequency error versus signal-to-noise ratio (SNR) is shown.

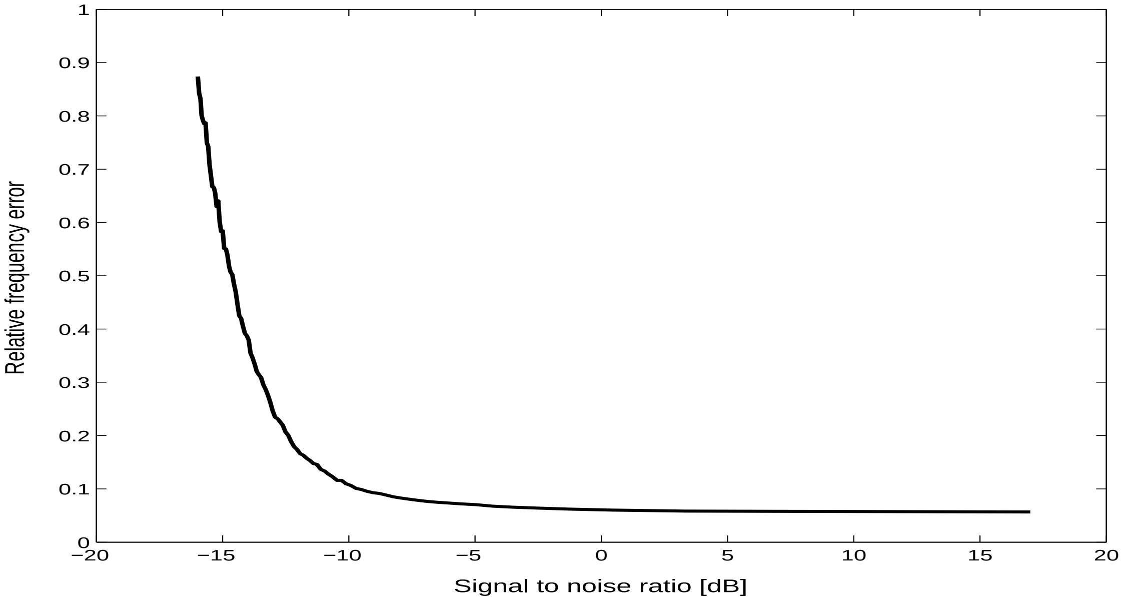

Figure 3.4: Frequency error vs. SNR

The signal-to-noise ratio is given by

$$SNR = 10 \log \frac{A^2}{2\mathrm{Var}(v)} \tag{3.15}$$

where A is the signals amplitude and $\mathrm{Var}(v)$ its variance. As clearly illustrated in Figure 3.4, the error in the filters frequency estimate starts to increase dramatically for a signal to noise ratio below approximately -12 dB. This phenomenon is refereed to as *the thresholding phenomenon*, see e.g. [3].

In this paper the term *thresholding performance* refers to the filters ability to estimate a state in a high noise environment, and *thresholding point* refers to the point where the estimation error starts to increase dramatically. Increasing a filters thresholding performance means to move its thresholding point to a lower SNR value.

3.4.2 Optimizing criteria

A optimizing criteria when using the GA should include the following two basic requirements

1) The thresholding point is to be placed at the lowest possible SNR level

2) The estimation error in the low noise area is to be sufficiently low

Even if it is an optimization of the thresholding performance, item number two in the above listing must be included to assure that the resulting filter is useful. Highest possible thresholding performance has no value if the performance in low noise is not satisfactory. These requirements may be expressed as follows:

1)

$$\min_{Q\in[Q_0,Q_1]} \left(\max_{R_s\in[R_0,R_1]} \parallel x - \hat{x} \parallel_2 \right) \tag{3.16}$$

where $[R_0, R_1]$ is the interval in which the optimization is carried out and $[Q_0, Q_1]$ is the interval for the filter tuning parameter.

2)

$$\max_{R_s\in[R_0,R_1]} \parallel x - \hat{x} \parallel_2 \leq M \tag{3.17}$$

where M is the largest allowed value for the 2-norm of the estimation error.

Under the assumption that the estimation error is a monotonous increasing function of the signal measurement noise variance R_s the evaluation function will be exactly the same as shown in (3.14). If this assumption is not valid, the evaluation function must be modified. In this paper the following function is applied:

$$E(Q) = N \left[\sum_{1}^{N} \sum_{i=1}^{K} \left(H_i \frac{1}{T+1} \sum_{t=0}^{T} E_{t,i} \right) \right]^{-1} \tag{3.18}$$

where K is the number of points to be considered, H_i is the weight for SNR point number i, N is the number of Monte Carlo runs used for evaluating each filter candidate, and $E_{t,i}$ equals $\hat{e}_t^T W \hat{e}_t$ at the i'th SNR point.

3.4.3 Results from simulations

In this section three different cases are considered.

1) the filter is tuned as in the previous section, i.e. with an signal-to-noise ratio SNR equal to -3 *dB*.

2) the filter is tuned for one single SNR value as in case 1, but now for SNR=-10 *dB*

3) the filter is tuned for three SNR values; SNR=-10 *dB*, SNR=-3*dB* and SNR=0 *dB*

Case 1

In Figures 3.5 and 3.6 the relative amplitude and frequency error vs. signal-to-noise ratio curves are shown. The solid-drawn line is for an amplitude $A = 1$ and the dashed line for $A = 2$.

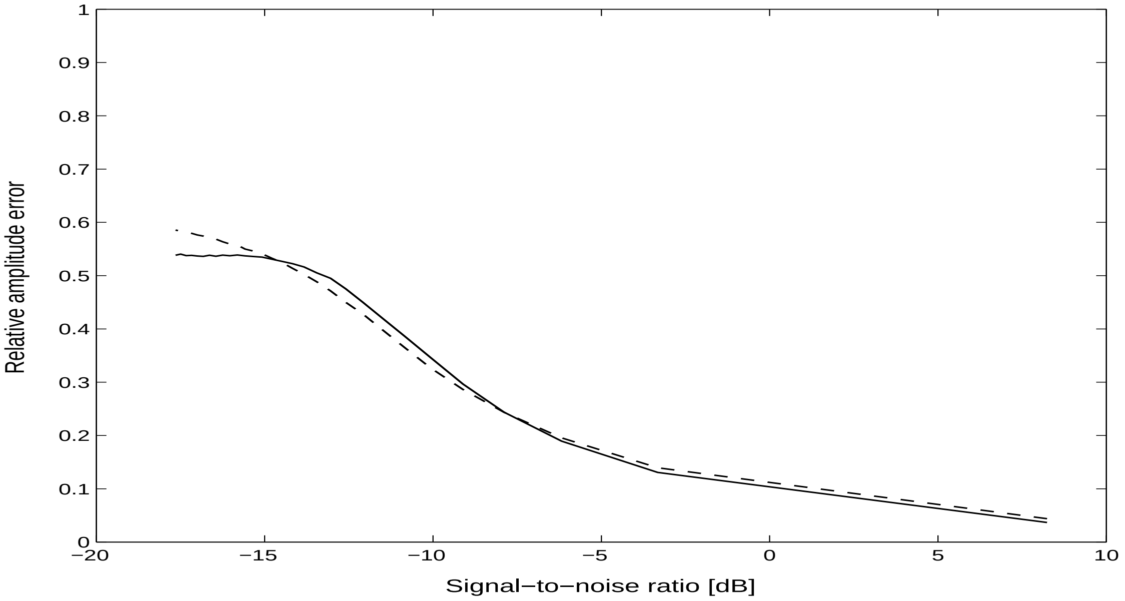

Figure 3.5: Amplitude error vs. SNR

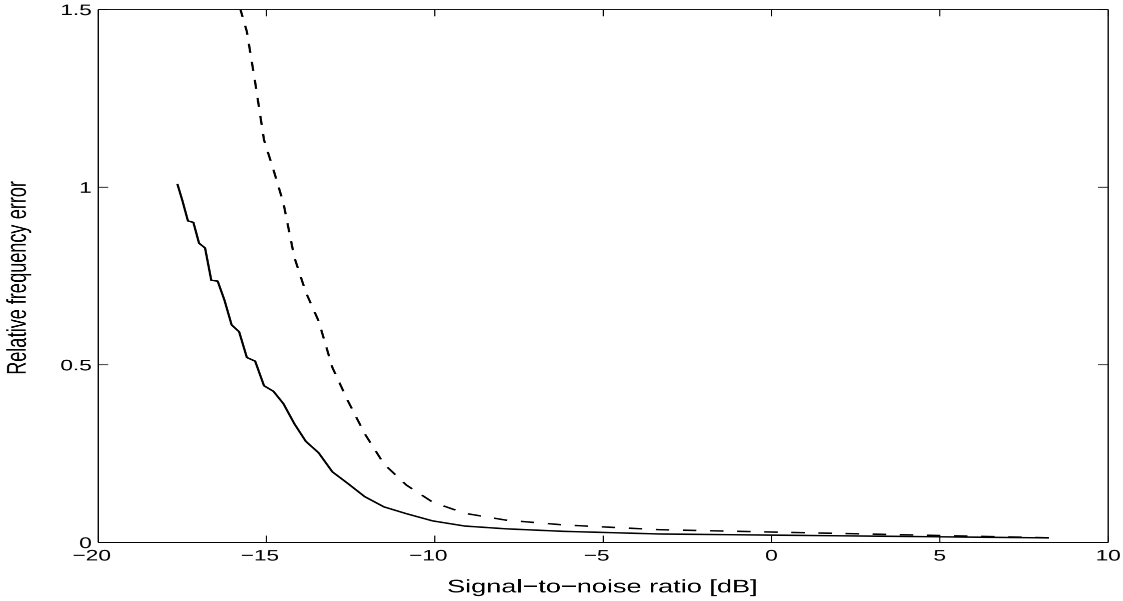

Figure 3.6: Frequency error vs. SNR

It should be noted that the relative error may very well exceed 1 (100 %) if e.g. the filter try to track a frequency which is more than twice the signals frequency. This is not very likely to happen in low noise, however, when the SNR is very low this is a common situation.

When comparing Figure 5 and Figure 6, it is observed that thresholding occure in the amplitude estimate on a SNR level where the frequency estimate is still very satisfactory. This suggests that it is the amplitude estimation which is the limiting factor for this filters thresholding performance.

Another interesting observation is that the SNR level where thresholding occure seems to depend on the amplitude.

Case 2

In Figures 3.7 and 3.8 the amplitude and frequency error vs. signal-to-noise ratio curves are shown.

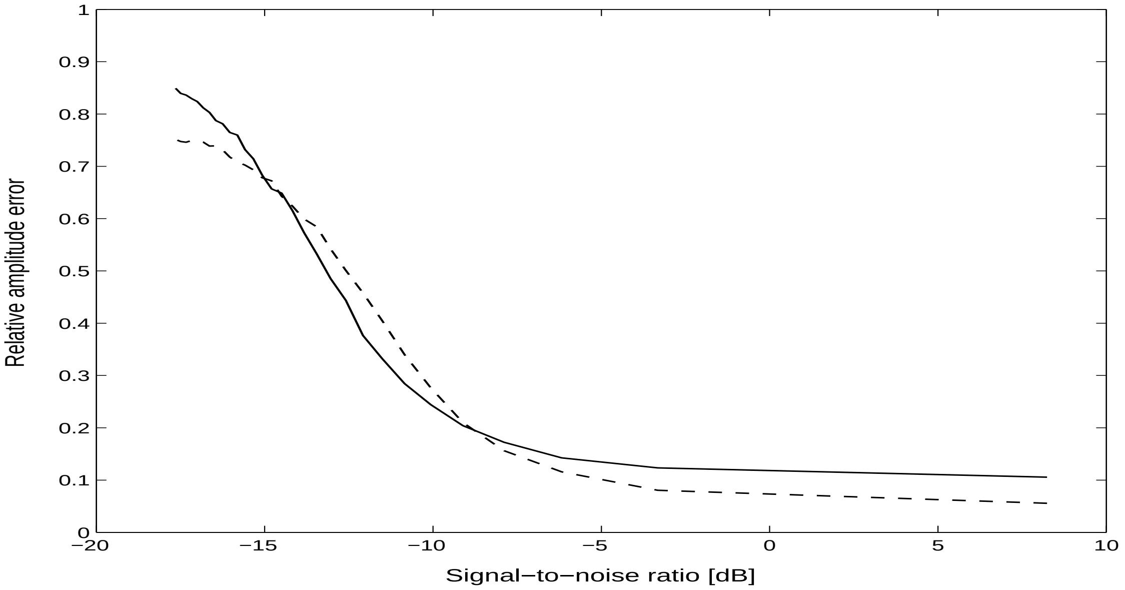

Figure 3.7: Amplitude error vs. SNR

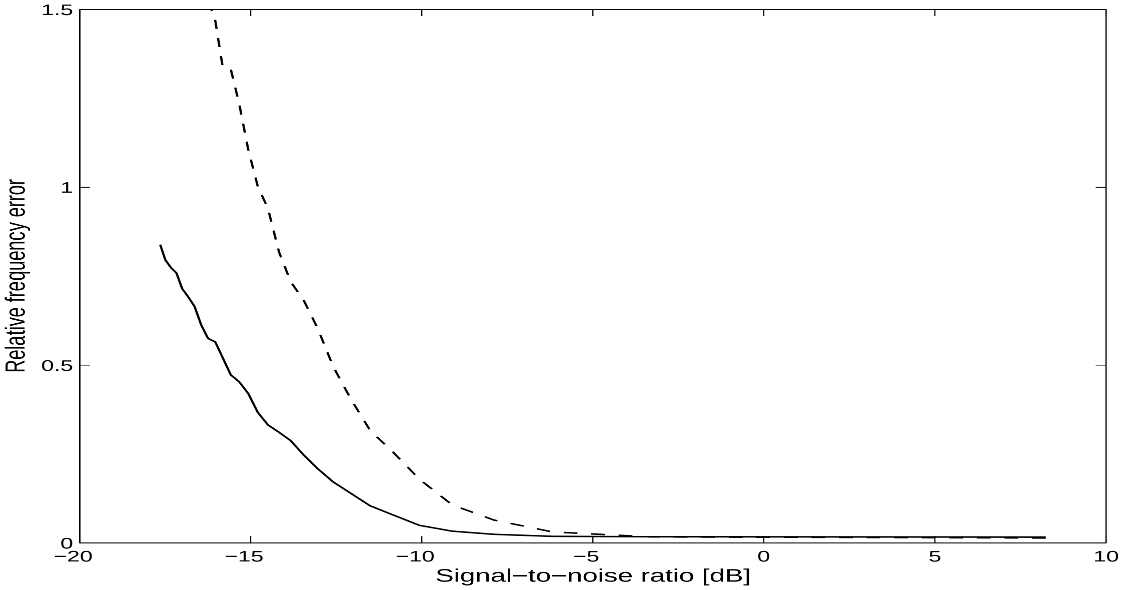

Figure 3.8: Frequency error vs. SNR

Compared with the previous case, it is clearly seen that for the amplitude estimation, the thresholding performance has increased considerably. However, the error in low noise area has also increased, in particular when the amplitude is low. For the frequency estimation the thresholding occur in fact earlier than compared with case 1. This difference is, however, not big.

Case 3

In Figures 3.9 and 3.10 the amplitude and frequency error vs. signal-to-noise ratio curves are shown.

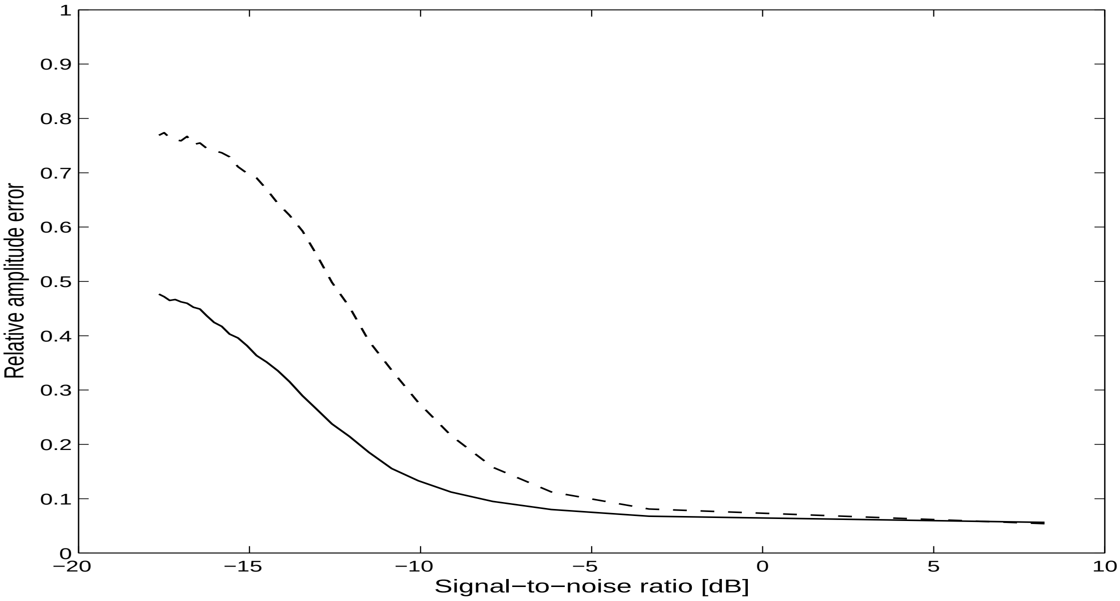

Figure 3.9: Amplitude error vs. SNR

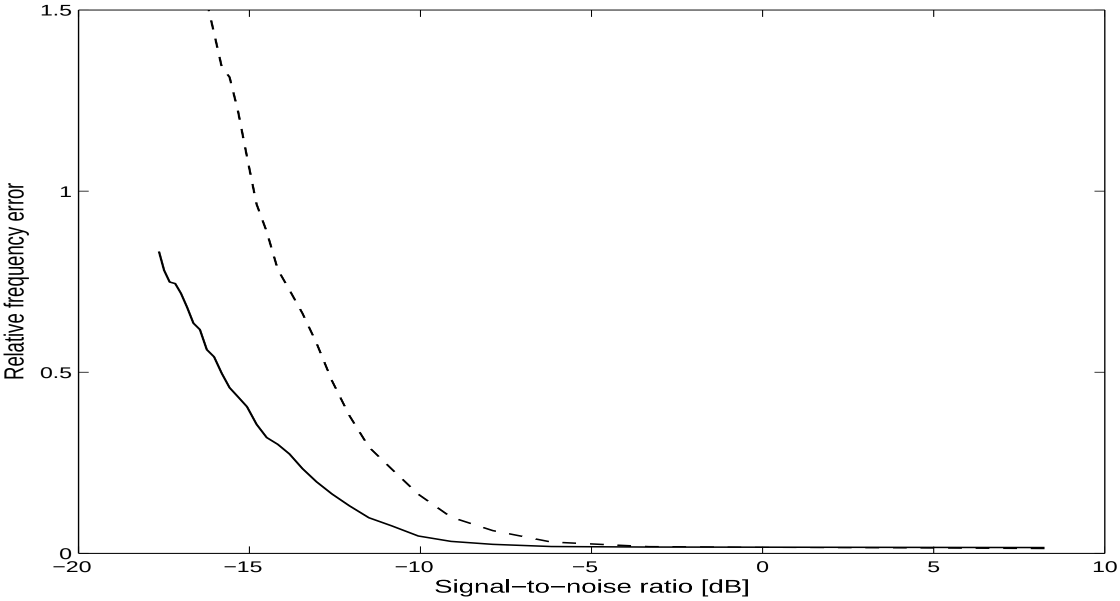

Figure 3.10: Frequency error vs. SNR

In this case the amplitude error in low noise area is reduced compared to the previous case. The cost for this is a slightly decreased thresholding performance. For the frequency estimation it is on the same level as before.

3.4.4 Discussion

As demonstrated by the previous cases, the GA may easily be modified to tune a filter with high thresholding performance. Due to its decreased noise sensitivity, it should be expected that the resulting filter will react slowly on a quick change in frequency or amplitude. For slowly varying amplitude and frequency the performance should be expected to be far better than for rapid changes. This is clearly illustrated in Figure 3.11 and Figure 3.12. The filter parameters are equal to those in section 3.4.3 case 3, and the noise variances are as given in section 3.3.5.

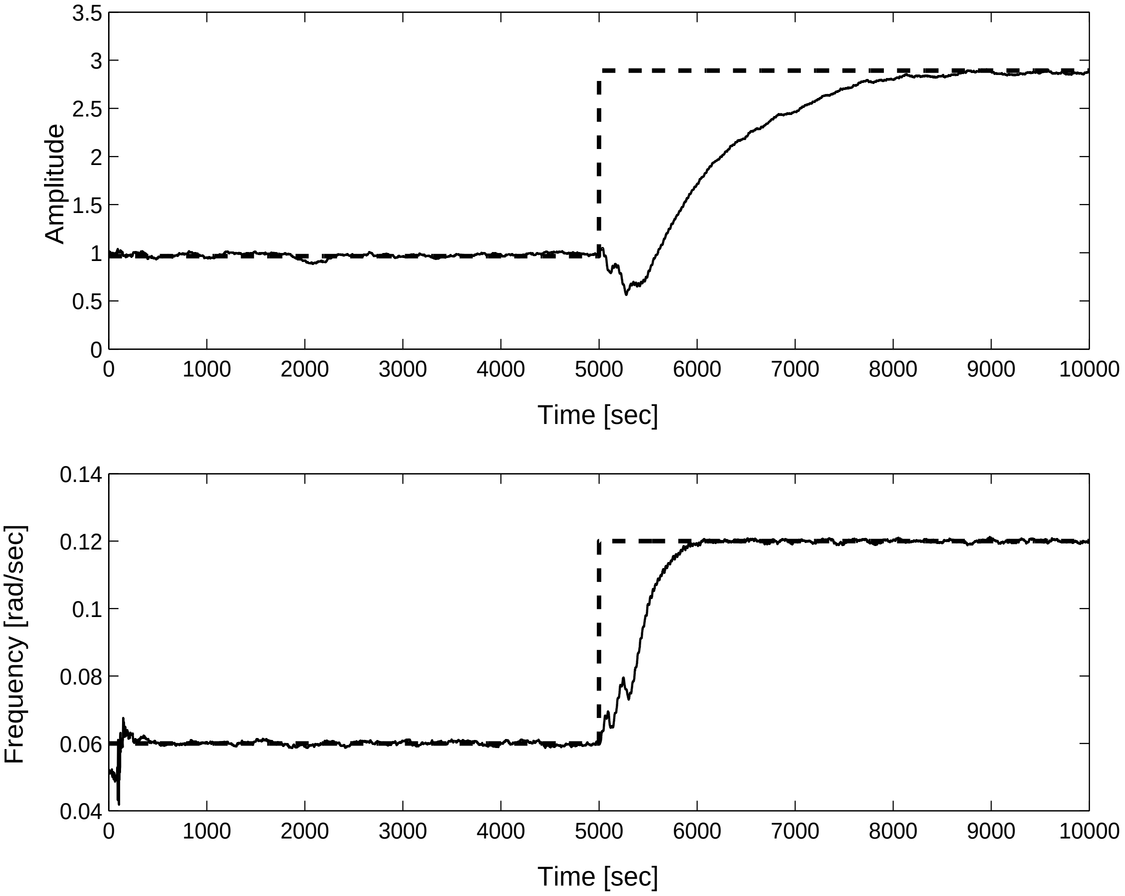

Figure 3.11: Amplitude and frequency diagram

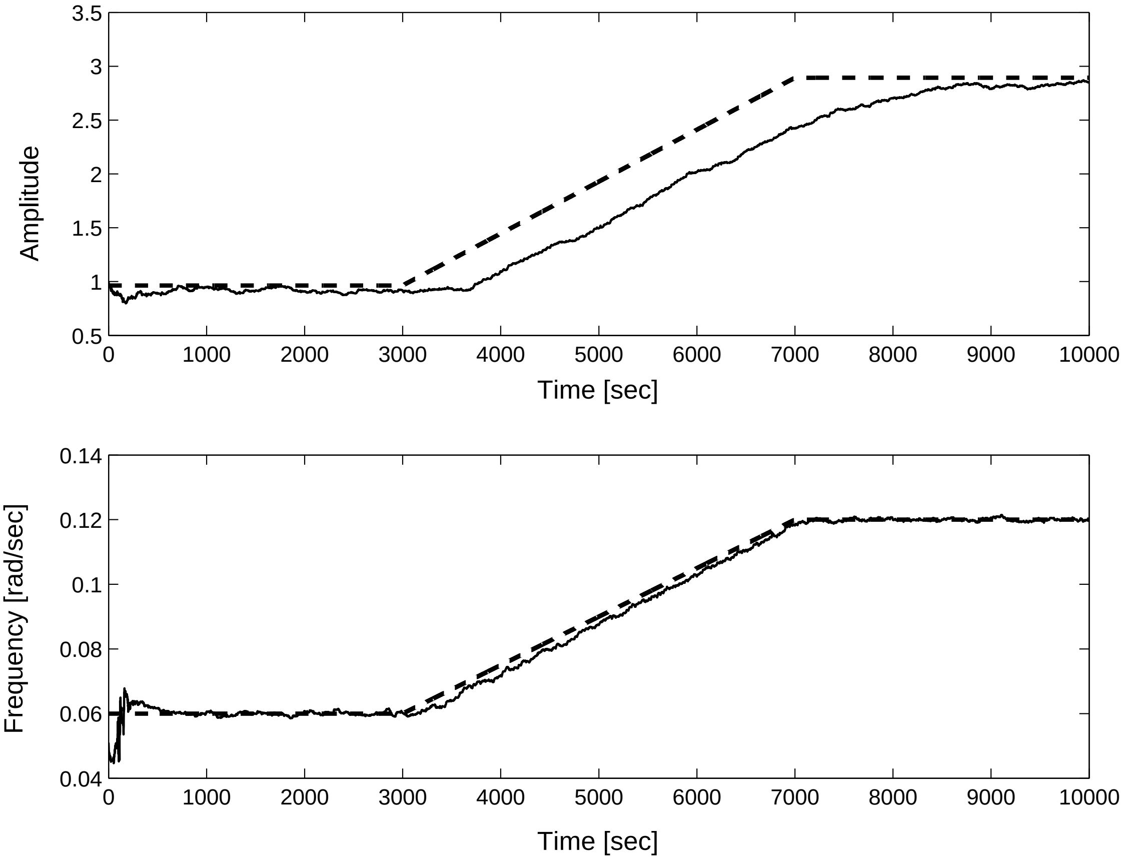

Figure 3.12: Amplitude and frequency diagram

3.5 Concluding remarks

In this paper some aspects regarding optimal tuning of an Extended Kalman filter by use of genetic algorithms have been discussed. It is demonstrated that genetic algorithms (GA) is a tool well suited for filter tuning. Further is is demonstrated how the GA can be modified in order to achieve better thresholding performance. This optimization is a trade off between low variance in the estimated states and a quick response. Filters optimized for maximum thresholding performance is only useful if the amplitude and frequency are almost constant, or changes very slowly.

ACKNOWLEDGEMENT

The authors are grateful to LKAB, Sweden, for financial support during this research.

References

[1] C. K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications.* Springer, Berlin, 1999.

[2] C. R. Houck, J. A. Joines, and M. G. Kay. The genetic algorithm optimization toolbox (gaot) for matlab 5. *http://www.ie.ncsu.edu/mirage/GAToolBox /gaot/*, 1996.

[3] Peter J. Kootsookos and Joanna M. Spanjaard. An extended Kalman filter for demodulation of polynomial phase signals. *IEEE Signal Processing Letters*, 1997.

[4] Peter S. Maybeck. *Stochastic Models, Estimation and Control.* Academic Press, New York, 1979.

[5] Y. Oshman and Ilan G. Shaviv. Optimal tuning of a Kalman filter using genetic algorithms. *AIAA Paper 2000-4558*, 2000.

[6] T. D. Powell. Automated tuning of an extended Kalman filter using the downhill simplex algorithm. *Journal of guidance, control and dynamics*, 25:901–908, 2002.

[7] K. Rapp and P. -O. Nyman. Control of the amplitude in a surging balling drum circuit, a new approach to an old problem. *Presented at ECC 2003, Cambridge UK*, 2003a.

[8] K. Rapp and P. -O. Nyman. Genetic algorithm based tuning of an extended Kalman filter. *Presented at MMAR 2003, Miedzyzdroje, Poland*, 2003b.

PAPER 4

On stability of the Extended Kalman Filter[1]

Knut Rapp*, **Per-Ole Nyman***

* Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, kr@hin.no, fax: +47 76 96 68 10
† Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, pon@hin.no, fax: +47 76 96 68 10

Abstract: In this paper the stability conditions for a time discrete extended Kalman filter (EKF) are considered. Only the special case where the signal model is linear in state and has a nonlinear output map is considered. The results published so far are very conservative, even though the results obtain by simulations are very promising. In this paper it is proved that by a proper choice of the covariance matrix Q_k, stability can be guaranteed for larger initial errors and noise processes than without considering the value of Q_k. A drawback is that the filter may become much more noise sensitive. The analysis is based on standard Lyapunov theory for discrete-time nonlinear systems and an explicit upper bound is given for the estimation error.

[1] In Proceedings of the 10th IEEE International Conference on Methods and Models in Automation and Robotics, 30 August - 2 September 2004 Miedzyzdroje, Poland

4.1 Introduction

In spite of the Extended Kalman filters popularity, there has been published surprisingly few results about its stability properties. It is well known that if the nonlinearities are moderate and if the initial error and noise processes are small, then the filter will be stable in most cases. However, proving this theoretically has been a rather hard task. During the last twenty years, this subject has been treated by several authors, see [6], [14], [9], [2], [12] and [8]. Unfortunately, these results are very conservative in the sense that stability can only be guaranteed if the initial error and the noise processes are extremely small. In this paper we show that this conservative results can be considerably improved by using tight upper bounds on certain matrices combined with a proper choice of the filter covariance matrices R_k and, in particular, Q_k.

The outline of this paper is as follows: In Section 2 the signal model and the EKF equations are stated together with the error dynamic. In Section 3 stability is proved and in Section 4 an alternative proof of the main theorem in Section 3 is given. In Section 5 an example to illustrate the new result is considered. In Section 6 some concluding remarks are stated.

4.2 EKF equations and error dynamics

In this paper we only consider a signal model which is linear in state and have a nonlinear output map. This signal model is given by:

$$x_{k+1} = A_k x_k + w_k \tag{4.1}$$

$$y_k = h(x_k) + v_k \tag{4.2}$$

where w_k and v_k are white, zero mean processes with covariance matrices $E[w_k w_k^T] = Q_k$ and $E[v_k v_k^T] = R_k$. For the remainder of this paper the function $h(x_k)$ is assumed to be continuous and differentiable.

The EKF associated with the above given signal model is given by the following set of coupled difference equations (see e.g. [3]):

Measurement update:

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k \left[y_k - h(\hat{x}_{k,k-1})\right] \tag{4.3}$$

$$P_{k,k} = \left[I - K_k H_k\right] P_{k,k-1} \tag{4.4}$$

where

$$H_k = \left[\frac{\partial h}{\partial x}\right]_{\hat{x}=\hat{x}_{k,k-1}} \tag{4.5}$$

Time update:

$$\hat{x}_{k,k-1} = A_{k-1}\,\hat{x}_{k-1,k-1} \tag{4.6}$$

$$P_{k,k-1} = A_{k-1} \cdot P_{k-1,k-1} \cdot A_{k-1}^T + Q_k \tag{4.7}$$

The filter gain matrix is given by

$$K_k = P_{k,k-1}H_k^T \left[H_k P_{k,k-1} H_k^T + R_k\right]^{-1} \tag{4.8}$$

For the remainder of this paper it is assume that R_k and Q_k are bounded from below by:

$$rI \leq R_k \quad \text{and} \quad qI \leq Q_k \tag{4.9}$$

for all $k \geq 0$, where $r, q > 0$.
Let $e_{k,k}$ and $e_{k,k-1}$ denote the error in the filtered state and predicted state respectively, that is,

$$e_{k,k} = x_k - \hat{x}_{k,k} \tag{4.10}$$

$$e_{k,k-1} = x_k - \hat{x}_{k,k-1} \tag{4.11}$$

Using the Taylor expansion

$$h(x_k) - h(\hat{x}_{k,k-1}) = H_k(x_k - \hat{x}_{k,k-1}) + \phi_h(x_k, \hat{x}_{k,k-1})$$

and the filter and signal model equations, it can be shown that the error dynamic is given by the following difference equation

$$e_{k,k} = \left[I - K_k H_k\right] A_{k-1} e_{k-1,k-1} + K_k \phi_h(x_k, \hat{x}_{k,k-1}) + \left[I - K_k H_k\right] w_{k-1} - K_k v_k \tag{4.12}$$

4.3 EKF stability

Assume that the noise processes are bounded in ∞-norm, i.e.

$$\|w_k\| \leq \bar{w} \quad \text{and} \quad \|v_k\| \leq \bar{v} \tag{4.13}$$

For the remainder of this paper the following assumptions are made:

$$\|A_k\| \leq a \tag{4.14}$$

$$p_1 I \leq P_{k,k} \leq p_2 I \tag{4.15}$$

$$p_1 I \leq P_{k,k-1} \leq p_2 I \tag{4.16}$$

$$\bar{\sigma}(H_k^T)/\underline{\sigma}^2(H_k^T) \leq h \tag{4.17}$$

Where $\bar{\sigma}(H_k^T)$ and $\underline{\sigma}(H_k^T)$ denoted the largest and the smallest singular value respectively.

Theorem 4.1. *Assume that the bounds given by (4.9) and (4.14)-(4.15) are fulfilled and that A_k is nonsingular for all $k \geq 0$. Assume further that there exist an $\bar{\epsilon}$ such that*

$$\|e_{k-1,k-1}\| \leq \bar{\epsilon} \tag{4.18}$$

which implies $\|x_k - \hat{x}_{k,k-1}\| \leq \epsilon_1(\bar{\epsilon})$, where

$$\epsilon_1(\bar{\epsilon}) = a\bar{\epsilon} + \bar{w}$$

Moreover, assume that

$$\|\phi(x_k, \hat{x}_{k,k-1})\| \leq \varphi \|x_k - \hat{x}_{k,k-1}\|^2 \tag{4.19}$$

holds for $\|x_k - \hat{x}_{k,k-1}\| \leq \epsilon_1(\bar{\epsilon}) = \epsilon_1$.

Then there exists an $\epsilon > 0$ such that the solution of the error model (4.12) is

1) *Locally exponential stable if the initial error satisfies $\|e_{0,0}\| \leq \epsilon$ and $\bar{w} = \bar{v} = 0$.*

2) *Bounded by*

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 - \frac{p_2}{\xi}\rho(\bar{w}, \bar{v}, \epsilon)$$

if the initial error satisfies $\|e_{0,0}\| \leq \epsilon$, and $\bar{w}$, $\bar{v}$ are sufficiently small. Here $\xi \in (-1, 0)$ is a constant and $\rho(\bar{w}, \bar{v}, \epsilon) > 0 \; \forall \; k \geq 0$ is a function to be defined later.

Proof: Denote in the following: $e_{k,k}$ by e_k, $e_{k-1,k-1}$ by e_{k-1}, $P_{k,k}$ by P_k and $P_{k-1,k-1}$ by P_{k-1}.

Let $V : R^n \rightarrow R$ be a positive function defined by

$$V(e_{k-1}) = e_{k-1}^T P_{k-1}^{-1} e_{k-1} \tag{4.20}$$

Then from (4.15)

$$\frac{1}{p_2}\|e_{k-1}\|^2 \leq V(e_{k-1}) \leq \frac{1}{p_1}\|e_{k-1}\|^2 \tag{4.21}$$

Define:

$$\tilde{A} = \big[A_{k-1} - K_k H_k A_{k-1}\big] \tag{4.22}$$

$$n_k = \big[I - K_k H_k\big] w_{k-1} - K_k v_k \tag{4.23}$$

$$l_k = K_k \phi_h(x, \hat{x}) \tag{4.24}$$

Then:

$$\begin{aligned}\Delta V :=& e_k^T P_k^{-1} e_k - e_{k-1}^T P_{k-1}^{-1} e_{k-1}\\ =& \big(\tilde{A} e_{k-1} + n_k + l_k\big)^T P_k^{-1} \big(\tilde{A} e_{k-1} + n_k + l_k\big) - e_{k-1}^T P_{k-1}^{-1} e_{k-1}\\ =& e_{k-1}^T \Big[\tilde{A}^T P_k^{-1} \tilde{A} - P_{k-1}^{-1}\Big] e_{k-1} + l_k^T P_k^{-1} \big(2\tilde{A} e_{k-1} + l_k\big) + n_k^T P_k^{-1} n_k\\ &+ 2 n_k^T P_k^{-1} \big(\tilde{A} e_{k-1} + l_k\big) \end{aligned} \tag{4.25}$$

In order to proceed further the following Lemma is needed:

Lemma 1. *If the conditions of Theorem 3.1 are fulfilled, there exist a real number* $0 < \gamma < 1$ *such that:*

$$\tilde{A}^T P_k^{-1} \tilde{A} \leq (1-\gamma) P_{k-1}^{-1} \tag{4.26}$$

Proof: Consider equation (4.4):

$$P_{k,k} = \big(I - K_k H_k\big) P_{k,k-1}$$

which can be written (see [5]):

$$P_{k,k} = \big(I - K_k H_k\big) P_{k,k-1} \big(I - K_k H_k\big)^T + K_k R_k K_k^T \tag{4.27}$$

Since $R_k > 0$, the following inequality can be established by use of (4.7)

$$P_k \geq \tilde{A} P_{k-1} \tilde{A}^T + (I - K_k H_k) Q_k (I - K_k H_k)^T \tag{4.28}$$

After some rearrangement of terms, this can be expressed:

$$P_k \geq \tilde{A} \cdot \Big[P_{k-1} + \tilde{A}^{-1} (I - K_k H_k) Q_k (I - K_k H_k)^T \tilde{A}^{-T}\Big] \cdot \tilde{A}^T \tag{4.29}$$

Multiplying from left and right with $\tilde{A}_k^{-1}$ and $\tilde{A}_k^{-T}$ and using (4.22) gives

$$\tilde{A}^{-1} P_k \tilde{A}^{-T} \geq P_{k-1} + A_{k-1}^{-1} Q_k A_{k-1}^{-T} \tag{4.30}$$

Taking the inverse of both sides and using (4.14)-(4.15) yields

$$\tilde{A}^T P_k^{-1} \tilde{A} \leq \left(1 + \frac{q}{p_2 f^2}\right)^{-1} P_{k-1}^{-1} \tag{4.31}$$

Setting

$$1 - \gamma = \left(1 + \frac{q}{p_2 a^2}\right)^{-1} \tag{4.32}$$

completes the proof. □

Now it can be showed that if the condition given by (4.17) holds, then a (rough) upper bound of the norm of the gain matrix is given by

$$\|K_k\| \leq h\frac{p_2}{p_1} \tag{4.33}$$

Equations (4.4), (4.15) and (4.16) give

$$\|I - K_k H_k\| \leq \|P_{k,k} P_{k,k-1}^{-1}\| \leq \frac{p_2}{q_1} \tag{4.34}$$

and by use of (4.14) and (4.22)

$$\|\tilde{A}_k\| \leq \|I - K_k H_k\| \|A_{k-1}\| \leq \frac{p_2}{q_1} f \tag{4.35}$$

By use of these results, the proof of Theorem 4.1 can be finished. From (4.25) and (4.26) it follows that

$$\Delta V \leq -\gamma V(e_{k-1}) + l_k^T P_k^{-1}\big(2\tilde{A}e_{k-1} + l_k\big) + n_k^T P_k^{-1} n_k + 2n_k^T P_k^{-1}\big(\tilde{A}e_{k-1} + l_k\big) \tag{4.36}$$

Considering the second term (see also [12] Lemma 3.2), it holds that

$$\begin{aligned} \|l_k^T P_k^{-1}\big(2\tilde{A}e_{k-1} + l_k\big)\| &\leq \|\phi_h(x,\hat{x})^T K_k^T P_k^{-1}\| \cdot \|2\tilde{A}e_{k-1} + K_k \phi_h(x,\hat{x})\| \\ &\leq \frac{h\varphi p_2^2}{p_1^3} \|x_k - \hat{x}_{k,k-1}\|^2 \cdot \Big(2a\|x_{k-1} - \hat{x}_{k-1,k-1}\| + h\varphi \|x_k - \hat{x}_{k,k-1}\|^2\Big) \end{aligned}$$

Using

$$\|e_{k,k-1}\|^2 \leq a^2 \|e_{k-1,k-1}\|^2 + 2a\bar{w}\|e_{k-1,k-1}\| + \bar{w}^2 \tag{4.37}$$

gives

$$\begin{aligned} \|l_k^T P_k^{-1}\big(2\tilde{A}e_{k-1} + l_k\big)\| \leq\ & h\varphi a^3 \frac{p_2^2}{p_1^3}\left(2 + ah\varphi\bar{\epsilon}\right)\|e_{k-1,k-1}\|^3 + h^2\varphi^2 \frac{p_2^2}{p_1^3}\bar{w}^4 + 4a\bar{\epsilon}h^2\varphi^2\frac{p_2^2}{p_1^3}\bar{w}^3 \\ & + 2ah\varphi\bar{\epsilon}\frac{p_2^2}{p_1^3}\left(1 + 3ah\varphi\bar{\epsilon}\right)\bar{w}^2 + 4a^2\bar{\epsilon}^2 h\varphi \frac{p_2^2}{p_1^3}\left(1 + ah\varphi\bar{\epsilon}\right)\bar{w} \end{aligned} \tag{4.38}$$

Thus

$$\Delta V \leq -\gamma V(e_{k-1}) + \bar{\varphi}\|e_{k-1,k-1}\|^3 + n_k^T P_k^{-1} n_k + 2n_k^T P_k^{-1}\big(\tilde{A}e_{k-1} + l_k\big) + \bar{w}W_1(\bar{w},\bar{\epsilon}) \tag{4.39}$$

for $\|e_{k-1,k-1}\| \leq \bar{\epsilon}$, where

$$\bar{\varphi} = h\varphi a^3 \frac{p_2^2}{p_1^3}\left(2 + ah\varphi\bar{\epsilon}\right) \tag{4.40}$$

and

$$W_1(\bar{w},\bar{\epsilon}) = h\varphi\frac{p_2^2}{p_1^3}\left[h\varphi\bar{w}^3 + 4a\epsilon h\varphi\bar{w}^2 + 2a\epsilon\left(1+3ah\varphi\epsilon\right)\bar{w} + 4a^2\epsilon^2\left(1+ah\varphi\epsilon\right)\right] \tag{4.41}$$

Choosing

$$\epsilon = \min\left(\bar{\epsilon}, \frac{\gamma}{\psi p_2\bar{\varphi}}\right) \tag{4.42}$$

where $\psi > 1$, gives for $\|e_{k-1,k-1}\| \leq \epsilon$

$$\bar{\varphi}\|e_{k-1,k-1}\|\|e_{k-1,k-1}\|^2 \leq \frac{\gamma}{\psi p_2\bar{\varphi}}\|e_{k-1,k-1}\|^2 \leq \frac{\gamma}{\psi}V(e_{k-1}) \tag{4.43}$$

Thus

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi}V(e_{k-1}) + n_k^TP_k^{-1}n_k + 2n_k^TP_k^{-1}\big(\tilde{A}e_{k-1}+l_k\big) + \bar{w}W_1(\bar{w},\epsilon) \tag{4.44}$$

for $\|e_{k-1,k-1}\| \leq \epsilon$.

Next consider the terms $2n_k^TP_k^{-1}\big(\tilde{A}e_{k-1}+l_k\big)$ and $n_k^TP_k^{-1}n_k$. Taking the norm and using the triangle inequality gives

$$\begin{aligned}&\|n_k^TP_k^{-1}n_k + 2n_k^TP_k^{-1}\big(\tilde{A}e_{k-1}+l_k\big)\| \leq \|n_k^TP_k^{-1}n_k\| + \|2n_k^TP_k^{-1}\big(\tilde{A}e_{k-1}+l_k\big)\|\\ &\leq \frac{1}{p_1}\Big(\|I-K_kH_k\|\bar{w} + \|K_k\|\bar{v}\Big)^2 + \frac{2}{p_1}\Big(\|I-K_kH_k\|\bar{w} + \|K_k\|\bar{v}\Big)\|\tilde{A}e_{k-1}+l_k\|\end{aligned} \tag{4.45}$$

Using inequalities (4.34), (4.35), and (4.33) together with equation (4.37), the following can be established

$$\|n_k^TP_k^{-1}n_k + 2n_k^TP_k^{-1}\big(\tilde{A}e_{k-1}+l_k\big)\| \leq (\bar{w}+h\bar{v})W_2(\bar{w},\bar{v},\epsilon) \tag{4.46}$$

where

$$W_2(\bar{w},\bar{v},\epsilon) = \frac{p_2^2}{p_1^3}\left[2h\varphi\bar{w}^2 + (1+4ah\varphi\epsilon)\bar{w} + h\bar{v} + 2a\epsilon(1+ah\varphi\epsilon)\right] \tag{4.47}$$

After some rearrangements the inequality (4.44) then yields

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi}V(e_{k-1}) + \bar{w}W_1(\bar{w},\epsilon) + (\bar{w}+h\bar{v})W_2(\bar{w},\bar{v},\epsilon) \tag{4.48}$$

Now define the function $\rho(\bar{w},\bar{v},\epsilon)$ to be

$$\rho(\bar{w},\bar{v},\epsilon) = \bar{w}W_1(\bar{w},\epsilon) + (\bar{w}+h\bar{v})W_2(\bar{w},\bar{v},\epsilon) \tag{4.49}$$

Thus

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi} V(e_{k-1}) + \rho(\bar{w}, \bar{v}, \epsilon) \tag{4.50}$$

Since $0 < \gamma < 1$ and $\psi > 1$

$$\xi := \frac{\gamma(1-\psi)}{\psi} \in (-1, 0) \tag{4.51}$$

Using (4.50) and (4.51) and starting at $k = 0$ gives

$$\begin{aligned}
V(e_{1,1}) &\leq (1+\xi)V(e_{0,0}) + \rho(\bar{w}, \bar{v}, \epsilon) \\
V(e_{2,2}) &\leq (1+\xi)V(e_{1,1}) + \rho(\bar{w}, \bar{v}, \epsilon) \\
&\leq (1+\xi)^2 V(e_{0,0}) + (1 + (1+\xi))\rho(\bar{w}, \bar{v}, \epsilon) \\
&\vdots \\
V(e_{k,k}) &\leq (1+\xi)^k V(e_{0,0}) + \sum_{n=0}^{n=k} (1+\xi)^n \rho(\bar{w}, \bar{v}, \epsilon)
\end{aligned} \tag{4.52}$$

Hence (4.15) and (4.52) implies

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 - \frac{p_2}{\xi}\rho(\bar{w}, \bar{v}, \epsilon) \tag{4.53}$$

In absence of noise $\rho(\bar{w}, \bar{v}, \epsilon) = 0$, and

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 \tag{4.54}$$

□

Remark 1: This result can be generalized to also cover the case of both nonlinear state and output map. See [11]. ■

Remark 2: A proof of statement 1 in Theorem 4.1 can be found in [13]. This proof require, however, that the matrix H_k is bounded from above, and is thus more restrictive than the result in the present paper. In [13] it is also shown that when an extended Kalman filter is used as a state observer for a deterministic nonlinear system, the rate of convergence of the observer can be assign in advance. ■

Even if this proves stability of the EKF, the results turn out to be very conservative, as also reported in [12]. One of the reasons for this is that γ as given by (4.32) will normally be very close to zero. However, a clever choice of the matrix Q_k yields a considerably better result.

Corollary 1. *If the conditions of Theorem 3.1 are fulfilled, then there exist a real and positive number δ such that γ in Lemma 1 is given by:*

$$\gamma = 1 - \frac{1}{1+\delta} \tag{4.55}$$

where δ can be chosen arbitrary in the interval $(0, N]$ and $N < \infty$.

Proof: Consider the inequality

$$P_k \geq \tilde{A} \cdot \Big[P_{k-1} + \tilde{A}^{-1}(I - K_k H_k) Q_k (I - K_k H_k)^T \tilde{A}^{-T}\Big] \cdot \tilde{A}^T \tag{4.56}$$

Now choose Q_k to be

$$\begin{aligned} Q_k &= \delta \cdot (I - K_k H_k)^{-1} \tilde{A} P_{k-1} \tilde{A}^T (I - K_k H_k)^{-T} \\ &= \delta A_{k-1} P_{k-1} A_{k-1}^T \end{aligned} \tag{4.57}$$

Which yields

$$P_k \geq \ \tilde{A} \cdot \Big[P_{k-1} + \delta P_{k-1}\Big] \cdot \tilde{A}^T \tag{4.58}$$

Thus

$$\tilde{A}^T P_k^{-1} \tilde{A} \leq \big(1 + \delta\big)^{-1} P_{k-1}^{-1} \tag{4.59}$$

Setting

$$1 - \gamma = \big(1 + \delta\big)^{-1} \tag{4.60}$$

completes the proof. □

Note that δ must be finite. Otherwise the conditions (4.15) and (4.16) will not be fulfilled as the upper bound will go to infinity if δ goes to infinity. Simulations indicates that δ should not be chosen too large, otherwise the noise sensitivity increases dramatically. This is due to the increase in the Kalman gain which, in general, make the filter act like the measurement noise has decreased.

In the proof of Theorem 4.1 a constant ψ appears. One might wonder what this constant represents, and how different choices affects the results of the analysis. Obviously, ψ has to be larger than 1, otherwise stability can not be proven. It should also be finite, as an infinite ψ gives $\epsilon = 0$, as seen from equation (4.42). Clearly ϵ decreases with increasing ψ, but on the other hand, the function $\rho(\bar{w}, \bar{v}, \epsilon)$ depend on ϵ^2, so the optimal value for ψ is not obvious at this first glance. Starting from a ψ slightly larger than 1, ψ may be regarded as a weighting constant between the influence of the initial error and the noise processes. In fact, in some interval, as ψ

is increased, ϵ decreases while the upper bound of the noise processes increases.

When estimating the maximum allowed initial error and the upper bounds for the noise processes by use of the above presented theory, the results may become very conservative. The reason for this is somewhat unclear, but there is a reason to believe that the following two items may be essential

- In the proof of Theorem 4.1, norms to estimate upper and lower bounds are widely used. This represents worst case scenarios and when using properties like the triangle inequality, the bounds may become rough. It is well known that this may yield conservative results (see e.g. [7]). In the following section, an alternative proof of Theorem 4.1 is given, in which the upper bounds of K_k, $[I - K_kH_k]$ and $\tilde{A}$ are tight.

- The conditions under which Theorem 4.1 is proved, are not strict enough, resulting in a too big class of nonlinear functions to be considered.

4.4 Alternative bounds on K_k, $[I - K_kH_k]$ and $\tilde{A}$

Assume that instead of using the upper bounds of K_k, $[I-K_kH_k]$ and $\tilde{A}$, as given by (4.33), (4.34) and (4.35), the following upper bounds, which can be verified during the estimation process, are applied

$$\|K_k\| \leq k_1 \tag{4.61}$$
$$\|I - K_kH_k\| \leq k_2 \tag{4.62}$$
$$\|\tilde{A}\| \leq \|I - K_kH_k\|\|A_{k-1}\| \leq ak_2 \tag{4.63}$$

As in Section 3, it is assumed that the following bounds hold

$$\|A_k\| \leq a \tag{4.64}$$
$$p_1 I \leq P_{k,k} \leq p_2 I \tag{4.65}$$
$$p_1 I \leq P_{k,k-1} \leq p_2 I \tag{4.66}$$

Now consider the inequality

$$\Delta V \leq -\gamma V(e_{k-1}) + l_k^T P_k^{-1}\big(2\tilde{A}e_{k-1} + l_k\big) + n_k^T P_k^{-1} n_k + 2n_k^T P_k^{-1}\big(\tilde{A}e_{k-1} + l_k\big) \tag{4.67}$$

for $\|e_{0,0}\| \leq \bar{\epsilon}$.

Using (4.61)-(4.66) gives for$\|e_{0,0}\| \leq \bar{\epsilon}$

$$\Delta V \leq -\gamma V(e_{k-1}) + \bar{\varphi}\|e_{k-1}\|^3 + \bar{w}\bar{W}_1(\bar{w},\bar{\epsilon}) + n_k^T P_k^{-1} n_k + 2n_k^T P_k^{-1}\big(\tilde{A}e_{k-1} + l_k\big) \tag{4.68}$$

where

$$\bar{\varphi} = \frac{a^3 k_1 \varphi}{p_1}\left(a k_1 \varphi \bar{\epsilon} + 2k_2\right) \tag{4.69}$$

and

$$\begin{aligned} &\bar{W}_1(\bar{w},\bar{\epsilon}) = \\ &\frac{1}{p_1}\bigg(\left(k_1 \varphi \bar{w}\right)^2 \left(\bar{w} + 4a\bar{\epsilon}\right) + 2ak_1\varphi\bar{\epsilon}\cdot(k_2 + 3ak_1\varphi\bar{\epsilon})\bar{w} + 4a^2 k_1 \varphi \bar{\epsilon}^2 (k_2 + ak_1\varphi\bar{\epsilon})\bigg) \end{aligned} \tag{4.70}$$

Furthermore,

$$\begin{aligned} \|n_k^T P_k^{-1} n_k + 2n_k^T P_k^{-1}\big(\tilde{A}e_{k-1} + l_k\big)\| &\leq \|n_k^T P_k^{-1} n_k\| + \|2n_k^T P_k^{-1}\big(\tilde{A}e_{k-1} + l_k\big)\| \\ &\leq (k_2\bar{w} + k_1\bar{v})\cdot \bar{W}_2(\bar{w},\bar{v},\bar{\epsilon}) \end{aligned} \tag{4.71}$$

where

$$\bar{W}_2(\bar{w},\bar{v},\bar{\epsilon}) = \frac{1}{p_1}\bigg(2a\bar{\epsilon}(ak_1\varphi\bar{\epsilon} + k_2) + k_1\bar{v}(1 + 2\varphi\bar{v})\bigg) \tag{4.72}$$

Therefore, for $\|e_{0,0}\| \leq \epsilon$, where ϵ is given by the equation (4.42)

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi} V(e_{k-1}) + \bar{\rho}(\bar{w},\bar{v},\epsilon) \tag{4.73}$$

where

$$\bar{\rho}(\bar{w},\bar{v},\epsilon) = \bar{w}\bar{W}_1(\bar{w},\epsilon) + (k_2\bar{w} + k_1\bar{v})\cdot\bar{W}_2(\bar{w},\bar{v},\epsilon) \tag{4.74}$$

These bounds will be tighter, however, a disadvantage by using such bounds is that they can not be determined in advance, but must be verified during the estimation process. On the other hand, tight bounds will of course yield less conservative results, which are highly desirable.

4.5 Numerical examples

4.5.1 Example 1

In this first example we consider a scalar signal model given by

$$x_{k+1} = ax_k + w_k \tag{4.75}$$

$$y_k = x_k + \eta x^2 + v_k \tag{4.76}$$

When $\eta = 0$ the system is linear and time-invariant, and convergence of the Kalman filter is guaranteed as the signal model is both controllable and observable (see e.g. [1] or [10]). An upper bound of the remainder term $\phi(x, \hat{x})$ is given by, (see e.g. [4])

$$|\phi(x, \hat{x})| \leq \frac{1}{2}\frac{\partial^2 h}{\partial x^2}\bigg|_{x=\tilde{x}} |x - \hat{x}|^2 = \eta|x - \hat{x}|^2 \tag{4.77}$$

Thus

$$\varphi \xrightarrow[\eta \to 0]{} 0 \Longrightarrow |\phi(x, \hat{x})| \xrightarrow[\eta \to 0]{} 0 \tag{4.78}$$

In this linear case inequality (4.50) can be written

$$\Delta V \leq -\gamma V(e_{k-1}) + \bar{\rho}(\bar{w}, \bar{v}, \epsilon) \tag{4.79}$$

where

$$\bar{\rho}(\bar{w}, \bar{v}, \epsilon) = \frac{p_2^2}{p_1^3}\left((\bar{w} + h\bar{v})^2 + 2a\epsilon(\bar{w} + h\bar{v})\right) \tag{4.80}$$

Now ϵ can be chosen arbitrarily, so stability can be ensured even if the noise processes are large. However, large noise processes will result in a larger bound on the error.

If η is chosen slightly larger than 0, ϵ will immediately be bounded by

$$\epsilon = \frac{\gamma}{\psi a^2 p_2 \bar{\varphi}(\eta)} \tag{4.81}$$

where

$$\bar{\varphi}(\eta) = ha^3 \frac{p_2^2}{p_1^3}\left(2 + \eta ha\bar{\epsilon}\right)\eta \tag{4.82}$$

Now consider the following case. Let $a = 0.99875$, $\eta = 1 \cdot 10^{-5}$, $\bar{w} = 3 \cdot 10^{-3}$, $\bar{v} = 0.95$ and $\psi = 3/2$. This system is almost linear, and the error is bounded, as shown in Figure 4.1 and 4.2. The initial error is $e_0 = 0.75$, so $\bar{\epsilon} = 0.75$. The filter tuning parameters are $R = 0.1$ and $Q = 1 \cdot 10^{-6}$. First we calculate γ by equation (4.32). By simulations it is found that the following bounds on the covariance apply

$$p_1 = 3.5 \cdot 10^{-4} \leq P \leq 1 \cdot 10^{-2} = p_2 \tag{4.83}$$

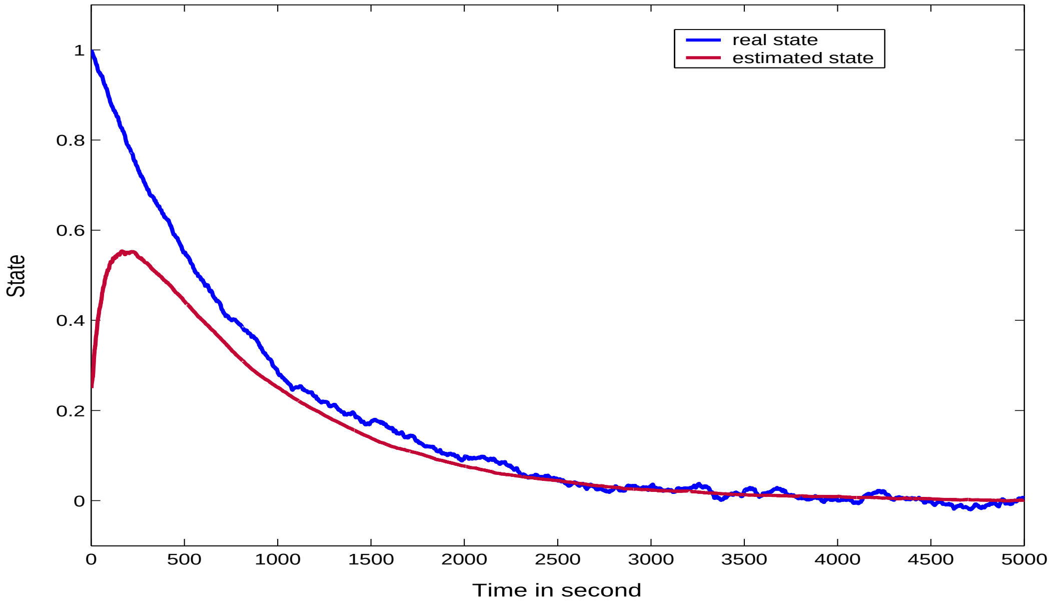

Figure 4.1: Real and estimated state

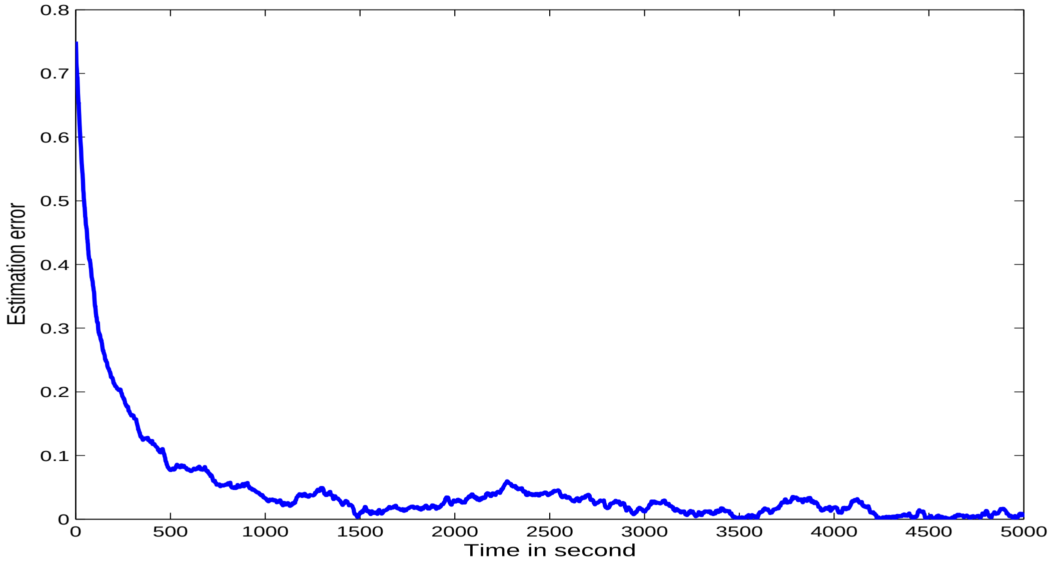

Figure 4.2: Estimation error

For these parameter values the following values of γ, $\bar{\varphi}(\eta)$ and ϵ are obtained

$$\gamma = 1 - \left(1 + \frac{q}{a^2 p_2}\right)^{-1} \approx 1 \cdot 10^{-4} \tag{4.84}$$

$$\bar{\varphi}(\eta) = ha^3 \frac{p_2^2}{p_1^3} \left(2 + \eta ha\bar{\epsilon}\right) \eta = 46.5 \tag{4.85}$$

$$\epsilon = \min\left(\bar{\epsilon}, \frac{\gamma}{\psi a^2 p_2 \bar{\varphi}(\eta)}\right) \approx 1.43 \cdot 10^{-4} \tag{4.86}$$

From this we conclude that stability can not be guaranteed for the simulated case. In order to obtain, say $\epsilon = \bar{\epsilon}$, η should be approximately $1.9 \cdot 10^{-9}$ or smaller, which is a value below any practical importance as long as the state is bounded from above by a quite small bound, as is the case in this example. If we try to estimate the maximum bound on the noise processes, we obtain $\bar{w} = \bar{v} \leq 7 \cdot 10^{-13}$ which further underlines that these estimates are very conservative.

Before we make use of Corollary 1, it may be of interest to see if the results from Section 4, yield larger estimates of ϵ, $\bar{w}$ and $\bar{v}$. By simulations it is found that the following bounds on K_k, $[I - K_k H_k]$ and $\tilde{A}$ apply

$$\|K_k\| \leq k_1 = 1 \cdot 10^{-2} \tag{4.87}$$

$$\|I - K_k H_k\| \leq k_2 = 1 \tag{4.88}$$

$$\|\tilde{A}\| \leq ak_2 = a = 0.99875 \tag{4.89}$$

Using these bounds and the upper and lower bounds on $P_{k,k}$ and $P_{k,k-1}$ given by (4.83), and assuming that $\bar{\epsilon} = 10$, gives the following values

$$\bar{\varphi} = \frac{a^3 k_1 \eta}{p_1} \left(ak_1 \eta \bar{\epsilon} + 2k_2\right) = 5.7 \cdot 10^{-4} \tag{4.90}$$

$$\epsilon = \min\left(\bar{\epsilon}, \frac{\gamma}{\psi a^2 p_2 \bar{\varphi}(\eta)}\right) = \bar{\epsilon} = 10 \tag{4.91}$$

In the interval $0 \leq e_{0,0} \leq 10$ the forward difference function ΔV is positive definite, which implies that there is not possible to guarantee that the error will be bounded for all $k \geq 0$. If the measurement noise is decreased to approximately the same level as the process noise, i.e. $\bar{v} = 3 \cdot 10^{-3}$, then the forward difference function ΔV is negative definite in the interval $6.3 \leq e_{k,k} \leq 10$ which implies that the error will remain bounded. In order to include $e_{0,0}$ in the interval in which the forward difference function is negative, the noise processes must be decreased to approximately $\bar{w} = \bar{v} = 3 \cdot 10^{-4}$, which make ΔV negative definite in the interval $0.62 \leq e_{k,k} \leq 10$, as shown in Figure 4.3.

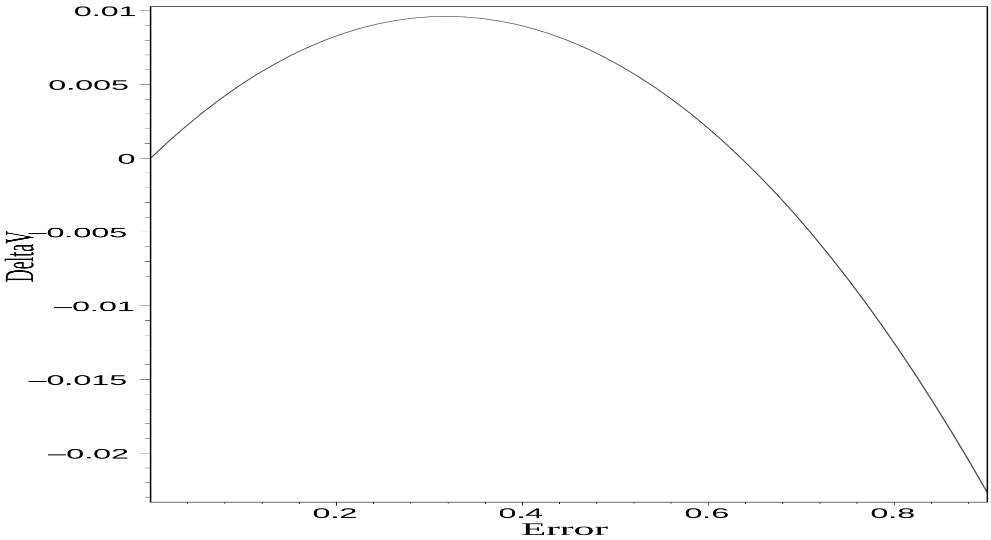

Figure 4.3: ΔV as function of $\|e_{k,k}\|$

Next we make use of Corollary 1, and we choose $\delta = 0.1$. By simulation it is found that the following bounds on the covariance apply

$$p_1 = 0.886 \leq P \leq 0.973 = p_2 \tag{4.92}$$

We assume that $\bar{\epsilon} = 2500$ is a reasonable value. Now the following values of γ, $\bar{\varphi}(\eta)$ and ϵ are obtained

$$\gamma = 1 - \frac{10}{11} = \frac{1}{11} \tag{4.93}$$

$$\bar{\varphi}(\eta) = ha^3 \frac{p_2^2}{p_1^3} (2 + \eta ha\bar{\epsilon})\, \eta = 2.75 \cdot 10^{-5} \tag{4.94}$$

$$\epsilon = \min\left(\bar{\epsilon}, \frac{\gamma}{\psi a^2 p_2 \bar{\varphi}(\eta)}\right) = \bar{\epsilon} = 2270 \tag{4.95}$$

Even if the initial error is no longer a problem, the measurement noise is. It turns out that ΔV is positive in the interval $0 \leq e_{k,k} \leq 160$ and negative for $160 < e_{k,k} \leq 2270$. The bound on the error, as shown in Figure 4.2, is therefore high, however, it can be guaranteed that it will remain bounded for all $k \geq 0$. In order to include $e_{0,0}$ in the interval in which the forward difference function is negative, the noise processes must be decreased to approximately $\bar{w} = \bar{v} = 2 \cdot 10^{-3}$, which make ΔV negative definite in the interval $0.67 \leq e_{k,k} \leq 2270$.

This case illustrates that even if the nonlinearity is very modest, it may be difficult to guarantee stability. When simulating this system, it is difficult to distinguish between the corresponding linear system (i.e. when $\eta = 0$) and this slightly nonlinear one, but even in this case the theoretical results are conservative. This conservatism is in fact rather surprising. It is natural to expect that the EKF associated with an almost linear signal model should have convergence properties close to the corresponding linear filter, however, this cannot be confirmed theoretically. This suggests that one should search for alternative proofs of EKF convergence which not make use of the Lyapunov function applied in this paper.

Consider now a more realistic case when $\eta = 0.1$, $\bar{w} = 3 \cdot 10^{-4}$, $\bar{v} = 9.5 \cdot 10^{-2}$ and the initial error $e_0 = 0.2$ $(= \bar{\epsilon})$. With $\delta = 0.05$, the bounds of the covariance is

$$p_1 = 0.327 \leq P \leq 0.474 = p_2 \tag{4.96}$$

Now we obtain

$$\bar{\varphi}(\eta) = ha^3\bar{\epsilon}\frac{p_2^2}{p_1^3}\left(2 + \eta ha\right)\eta = 0.269 \tag{4.97}$$

$$\epsilon = \min\left(\bar{\epsilon}, \frac{\gamma}{\psi a^2 p_2 \bar{\varphi}(\eta)}\right) = \bar{\epsilon} = 0.2 \tag{4.98}$$

In the interval $0.04 \leq e_{0,0} \leq 0.2$ the forward difference function ΔV is negative definite, which implies that the error will be bounded for all $k \geq 0$.

In Figure 4.4 the real and estimated state for this case is shown. In Figure 4.5 the transient period, which is remarkable short, is shown.

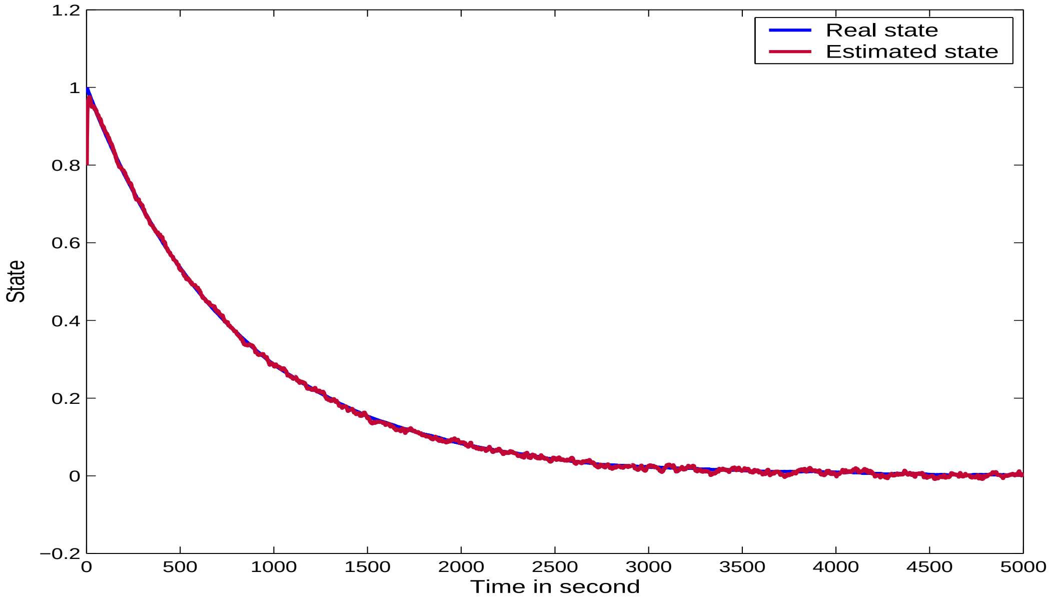

Figure 4.4: Real and estimated state

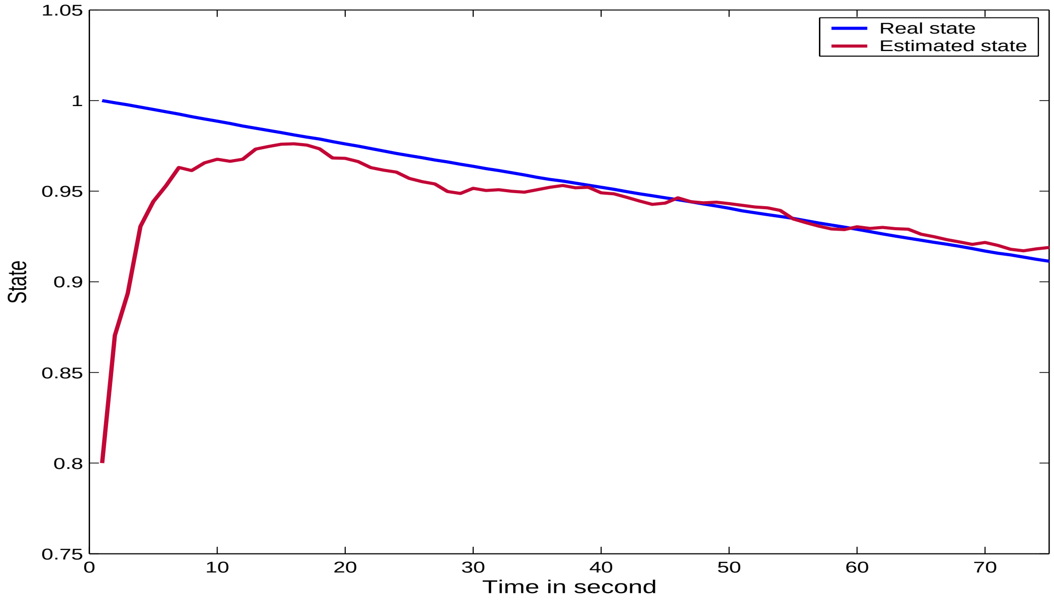

Figure 4.5: Transient period

4.5.2 Example 2

In this second example an EKF used for tracking the amplitude, phase and frequency of a low frequency signal is considered. The signal model is linear and time invariant in state, and has a nonlinear output map, and is given by:

$$x(k+1) = Ax(k) + Bw(k) \tag{4.99}$$
$$y(k) = x_3 \sin x_2 + v(k) \tag{4.100}$$

where x_1 is the phase increment (or frequency), x_2 is the phase and x_3 is the amplitude. The matrices A and B are given by:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

A reasonable choice of filter tuning parameters is $Q = 1 \cdot 10^{-5}$ and $R = 1$. This filter is by no means optimally tuned, however, the variance in the estimate and the transient period may be acceptable for some applications. By simulations it is found that $p_1 = 1.5 \cdot 10^{-4}$ and $p_2 = 1$. When estimating the maximum allowed initial error by use of (4.42) when not using Corollary 1, or the results from Section 4, the following value is obtained

$$\epsilon = 2.2 \cdot 10^{-17} \tag{4.101}$$

which is a value far below any practical significance.

In this case the upper bound on the gain matrix K_k and $[I - K_k H_k]$ is extremely high compared to the real bounds found by simulations. These bounds are

$$\begin{aligned} \|K_k\| &\leq k_1 = 0.34 \\ \|[I - K_k H_k]\| &\leq k_2 = 1.1 \end{aligned}$$

Using these values and the results from Section 4, we obtain the following bound on the initial error

$$\epsilon \leq 2.6 \cdot 10^{-9} \tag{4.102}$$

which is still very conservative.

If the results from Section 4 is combined with Corollary 1, the results are better. With $\delta = 0.025$ the following bounds on the gain and covariance apply

$$\begin{aligned} &\|K_k\| \leq k_1 = 0.28 \\ &\|[I - K_k H_k]\| \leq k_2 = 1.075 \\ &p_1 = 1.35 \cdot 10^{-5} \quad \text{and} \quad p_2 = 0.85 \end{aligned}$$

Using these values we obtain

$$\epsilon = 8.5 \cdot 10^{-7} \tag{4.103}$$

$$\bar{w} = \bar{v} = 1 \cdot 10^{-8} \tag{4.104}$$

These estimates are very conservative, even if the tighter estimates for the upper bound on the gain matrix K_k and the matrix $[I - K_k H_k]$ are applied. On reason for this is that the estimates of ϵ, $\bar{w}$ and $\bar{v}$ still depends on the ratio p_2/p_1, which is large for this specific example. The results obtained by simulation are, on the other hand, very satisfactory. The error will remain bounded, and the filter works quite well, if the initial error is bounded by $e_{0,0} \leq 0.75$ and the noise processes are bounded by $\bar{w} \leq 9.5 \cdot 10^{-5}$ and $\bar{v} \leq 3$.

4.6 Concluding remarks

In this paper the stability properties of an EKF for a signal model with linear state equation and nonlinear output map is considered. Our main conclusions are:

1) Previously published stability results for the EKF have been very conservative, i.e. stability could only be guaranteed when the initial error and the noise processes were extremely small. In this paper we show that for a certain choice of the matrix Q_k, and use of tight upper bounds for the Kalman gain matrix K_k and the matrix $[I - K_k H_k]$, these results can be considerably improved.

2) In the linear case the theorems presented in this paper recover stability unconditionally. However, if the signal model is modified to be only slightly nonlinear, the results may become conservative, even if the nonlinearity would be insignificant for any practical purpose. One reason for this conservatism is that the estimates of the maximum allowed initial error and the maximum upper bounds for the noise processes depends on the ratio p_1/p_2, which in some cases can become very large.

3) Stability can be guaranteed even if the signal model is unstable, provided that the Hessian matrix of the output map, $h(x_k)$, is finite for every $x \in R^n$. This extends the results given in [12], [13], [2] and [9].

ACKNOWLEDGEMENT

The authors are grateful to LKAB, Sweden, for financial support during this research.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentic-Hall, New Jersey, 1979.

[2] M. Boutayeb and D. Aubry. A strong tracking extended Kalman observer for nonlinear discrete-time systems. *IEEE Transaction on automatic control*, 44:1550–1556, 1999.

[3] C. K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications.* Springer, Berlin, 1999.

[4] J. E. Dennis Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, New Jersey, 1983.

[5] Arthur Gelb. *Applied Optimal Filtering.* MIT Press, Cambridge, Massachusetts, 1974.

[6] A. Bensoussan J. S. Baras and M. R. James. Dynamic observers as asymptotic limits of recursive filters: special cases. *SIAM Journal of Applied Mathematics*, 48:1147–1158, 1988.

[7] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, New Jersey, 2002.

[8] A. J. Krener. The convergence of the extended Kalman filter. *Directions in Mathematical Systems Theory and Optimization*, pages 173–182, 2003.

[9] B. F. La Scala, R. R. Bitmead, and M. R. James. Conditions for stability of the extended Kalman filter and their applications to the frequency tracking problem. *Mathematics of Control, Signals, and Systems*, 8:1–26, 1995.

[10] G. Minkler and J. Minkler. *Theory and Application of Kalman Filtering.* Magellan Book Company, Palm Bay, Florida, 1993.

[11] K. Rapp and P. -O. Nyman. Stability properties of the extended Kalman filter. *To be presented at NOLCOS 2004, Stuttgart, Germany.*

[12] K. Reif, Stefan Günter, Engin Yaz, and Rolf Unbehauen. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transaction on Automatic Control*, 44:714–728, 1999.

[13] K. Reif and R. Unbehauen. The extended Kalman filter as an exponential observer for nonlinear systems. *IEEE Transaction on Signal Processing*, 47:2324–2328, 1999.

[14] Y. Song and J. W. Grizzle. The extended Kalman filter as a local asymptotic observer for discrete-time nonlinear systems. *Journal of Mathematical Systems, Estimation, and Control*, 5:59–78, 1995.

PAPER 5

Stability properties of the Extended Kalman Filter[1]

Knut Rapp*, **Per-Ole Nyman***

* Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, kr@hin.no, fax: +47 76 96 68 10

† Narvik University College, P. O. Box 385, N-8505 Narvik, NORWAY, pon@hin.no, fax: +47 76 96 68 10

Abstract: In this paper the stability conditions for the discrete-time extended Kalman filter (EKF) are considered. Explicit upper bounds are given for the estimation error in both the noiseless and the general case. The results published so far generally requires that the state belongs to a compact subset of $\mathbb{R}^n$. In this paper it is proved that this can be relaxed to only require the state to belong to a open convex subset of $\mathbb{R}^n$, provided that the Hessian matrix of the output map is bounded in $\mathbb{R}^n$.

[1] In Proceedings of the 6th IFAC-Symposium on Nonlinear Control Systems, 01-03 September, 2004, Stuttgart, Germany (extended version)

5.1 Introduction

Extended Kalman filters have been widely used since the sixties and have gained large popularity since then. However, theoretical results for analyzing the design have not been available until recently, see [10].

Convergence of the extended Kalman filter has been treated by several authors, see [5, 12, 8, 10, 11, 7]. In [5] it is shown that the continuous-time EKF converges locally under some suitable, but rather strong, conditions. The results in [5] is extended by [7] to yield a much larger class of filters, and the conditions required for convergence are considerably relaxed compared to those given in [5], as the very strong uniform detectability requirement is shown to be superfluous for a broad class of filter. It should be mentioned that the uniform detectability condition may be very difficult to verify, and examples are given where even linear filters fail to meet this requirement (see [7]).

In [12] the discrete-time observer case is considered. The convergence of the EKF used as an observer is extended by [11], where it is shown that by a small modification of the filter algorithm, the rate of convergence can be prescribed by the designer. In [8] the discrete-time case where the signal model has a nonlinear state equation and a linear output map is treated, and the results from [12] are extended to also include the stochastic case. The general case, with both nonlinear state equation and output map, is treated in [10], which contains a proof that the EKF is stochastically stable under certain given conditions. In the present paper, the noise processes are assumed to have an absolute upper bound, rather than being normally distributed with a given variance. For this class of signal models it is shown that one of the assumption in the stability proofs presented earlier, namely that the Jacobi matrix of the output map must be bounded in norm, can be relaxed to only require a finite ratio between its largest and smallest singular value. This result can also be applied when considering stochastic stability.

The outline of this paper is as follows: In Section 2 the signal model and the EKF equations are stated together with the error dynamic. In Section 3 stability is proved and in Section 4 an example to illustrate the new result is considered. In Section 5 some concluding remarks are stated.

5.2 EKF equations and error dynamics

In this paper we consider a nonlinear signal model corrupted by noise in both the state and the measurement. The signal model is given by:

$$x_{k+1} = f(x_k) + w_k \tag{5.1}$$
$$y_k = h(x_k) + v_k \tag{5.2}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and w_k and v_k are white, zero mean processes with covariance matrices $E[w_k w_k^T] = Q_k$ and $E[v_k v_k^T] = R_k$. For the remainder of this paper the functions $f(x_k)$ and $h(x_k)$ are assumed to be continuous and differentiable. The EKF associated with the above given signal model is given by the following set of coupled difference equations (see e.g. [1]):

Measurement update:

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k \left[y_k - h(\hat{x}_{k,k-1})\right] \tag{5.3}$$
$$P_{k,k} = \left[I - K_k H_k\right] P_{k,k-1} \tag{5.4}$$

where

$$H_k = \left[\frac{\partial h}{\partial x}\right]_{\hat{x}=\hat{x}_{k,k-1}} \tag{5.5}$$

Time update:

$$\hat{x}_{k,k-1} = f(\hat{x}_{k-1,k-1}) \tag{5.6}$$
$$P_{k,k-1} = F_{k-1} \cdot P_{k-1,k-1} \cdot F_{k-1}^T + Q_k \tag{5.7}$$

where F_k is assumed to be nonsingular for all $k \geq 0$ and given by

$$F_k = \left[\frac{\partial f}{\partial x}\right]_{\hat{x}=\hat{x}_{k,k}} \tag{5.8}$$

The filter gain matrix is given by

$$K_k = P_{k,k-1} H_k^T \left[H_k P_{k,k-1} H_k^T + R_k\right]^{-1} \tag{5.9}$$

For the remainder of this paper it is assume that R_k and Q_k are bounded from below by:

$$\bar{r}I \leq R_k \quad \text{and} \quad \bar{q}I \leq Q_k \tag{5.10}$$

for all $k \geq 0$, where $\bar{r}, \bar{q} > 0$.

Let $e_{k,k}$ and $e_{k,k-1}$ denote the error in the filtered state and predicted state respectively, that is,

$$e_{k,k} = x_k - \hat{x}_{k,k} \tag{5.11}$$

$$e_{k,k-1} = x_k - \hat{x}_{k,k-1} \tag{5.12}$$

Equation (5.3) gives:

$$e_{k,k} = e_{k,k-1} - K_k \left[h(x_k) - h(\hat{x}_{k,k-1}) + v_k \right] \tag{5.13}$$

Because $h \in C^1$ it may be expanded as:

$$h(x_k) - h(\hat{x}_{k,k-1}) = H_k(x_k - \hat{x}_{k,k-1}) + \phi_h(x_k, \hat{x}_{k,k-1}) \tag{5.14}$$

where $\phi_h(x_k, \hat{x}_{k,k-1})$ is the remainder term, denoted $\phi_h(x, \hat{x})$ in the sequel.

Equation (5.13) can therefore be written:

$$e_{k,k} = \left[I - K_k H_k \right] e_{k,k-1} + K_k \phi_h(x, \hat{x}) - K_k v_k$$

Now:

$$x_k - \hat{x}_{k,k-1} = f(x_{k-1}) + w_{k-1} - f(\hat{x}_{k-1,k-1})$$

Using the Taylor expansion:

$$f(x_k) - f(\hat{x}_{k,k}) = F_k(x_k - \hat{x}_{k,k}) + \theta_f(x_k, \hat{x}_{k,k}) \tag{5.15}$$

gives

$$e_{k,k-1} = F_{k-1} e_{k-1,k-1} + w_{k-1} + \theta_f^-(x, \hat{x}) \tag{5.16}$$

where $\theta_f^-(x, \hat{x})$ denotes the remainder term at time $k-1$, i.e.
$\theta_f^-(x, \hat{x}) = \theta_f(x_{k-1}, \hat{x}_{k-1,k-1})$.

Thus

$$e_{k,k} = \tilde{F}_k e_{k-1,k-1} + n_k + l_k \tag{5.17}$$

where:

$$\tilde{F}_k = \left[I - K_k H_k \right] F_{k-1} \tag{5.18}$$

$$n_k = \Big[I - K_k H_k \Big] w_{k-1} - K_k v_k \tag{5.19}$$

$$l_k = \left[I - K_k H_k \right] \theta_f^-(x, \hat{x}) + K_k \phi_h(x, \hat{x}) \tag{5.20}$$

Note that $\tilde{F}_k$ exist because both F_k and $[I - K_k H_k]$ are nonsingular matrices.

5.3 Filter stability analysis

In the section stability of the EKF is proved. The proof is based on the Lyapunov method, see e.g. [13].
Assume that the noise processes are bounded in ∞-norm, i.e.

$$\|w_k\| \leq \bar{w} \quad \text{and} \quad \|v_k\| \leq \bar{v} \tag{5.21}$$

For the remainder of this paper the following assumptions are made:

$$\|F_k\| \leq f \tag{5.22}$$

$$p_1 I \leq P_{k,k} \leq p_2 I \tag{5.23}$$

$$q_1 I \leq P_{k,k-1} \leq q_2 I \tag{5.24}$$

$$\bar{\sigma}(H_k^T)/\underline{\sigma}^2(H_k^T) \leq h \tag{5.25}$$

Where $\bar{\sigma}(H_k^T)$ and $\underline{\sigma}(H_k^T)$ denoted the largest and the smallest singular value respectively.

Before we state the main Theorem of this paper, the following two preparatory Lemmas are stated and proved.

Lemma 1. *Assume that F_k is nonsingular for all $k \geq 0$ and that the conditions (5.22)-(5.24) are fulfilled. Then there exist a real number $0 < \gamma < 1$ such that:*

$$\tilde{F}_k^T P_k^{-1} \tilde{F}_k \leq (1-\gamma) P_{k-1}^{-1} \tag{5.26}$$

Proof: Consider equation (5.4):

$$P_{k,k} = \big(I - K_k H_k\big) P_{k,k-1} \tag{5.27}$$

which can be written (see [3]):

$$P_{k,k} = \big(I - K_k H_k\big) P_{k,k-1} \big(I - K_k H_k\big)^T + K_k R_k K_k^T \tag{5.28}$$

Since $R_k > 0$, the following inequality can be established by use of (5.7)

$$P_k \geq \tilde{F}_k P_{k-1} \tilde{F}_k^T + (I - K_k H_k) Q_k (I - K_k H_k)^T \tag{5.29}$$

After some rearrangement of terms, this can be expressed:

$$P_k \geq \tilde{F}_k \cdot \Big[P_{k-1} + \tilde{F}_k^{-1}(I - K_k H_k) Q_k (I - K_k H_k)^T \tilde{F}_k^{-T}\Big] \cdot \tilde{F}_k^T \tag{5.30}$$

Multiplying from left and right with $\tilde{F}_k^{-1}$ and $\tilde{F}_k^{-T}$ and using (5.18) gives

$$\tilde{F}_k^{-1} P_k \tilde{F}_k^{-T} \geq P_{k-1} + F_{k-1}^{-1} Q_k F_{k-1}^{-T} \tag{5.31}$$

Taking the inverse of both sides and using (5.22)-(5.23) yields

$$\tilde{F}_k^T P_k^{-1} \tilde{F}_k \leq \left(1 + \frac{q}{p_2 f^2}\right)^{-1} P_{k-1}^{-1} \tag{5.32}$$

Setting $1 - \gamma = \left(1 + \frac{q}{p_2 f^2}\right)^{-1}$ completes the proof. □

The next Lemma gives a upper bound of the norm of the Kalman gain matrix, which does not require the matrix H_k to be bounded in norm.

Lemma 2. *If the condition given by (5.25) holds then a (rough) upper bound of the norm of the Kalman gain matrix is given by*

$$\|K_k\| \leq h \frac{q_2}{q_1} \tag{5.33}$$

Proof:

$$\begin{aligned}\bar{\sigma}(K_k) &= \bar{\sigma}\Big(P_{k,k-1} H_k^T \big(H_k P_{k,k-1} H_k^T + R_k\big)^{-1}\Big) \\ &\leq \bar{\sigma}\Big(P_{k,k-1} H_k^T\Big) \bar{\sigma}\Big(\big(H_k P_{k,k-1} H_k^T + R_k\big)^{-1}\Big) \\ &= \bar{\sigma}\Big(P_{k,k-1} H_k^T\Big) \Big(\underline{\sigma}\big(H_k P_{k,k-1} H_k^T + R_k\big)\Big)^{-1}\end{aligned} \tag{5.34}$$

The matrices in the second factor are positive definite so the singular values are equal to the eigenvalues. By the Rayleigh-Ritz characterization (see e.g. [4]) the following is obtained

$$\begin{aligned}\underline{\sigma}\big(H_k P_{k,k-1} H_k^T + R_k\big) &= \lambda_{min}\big(H_k P_{k,k-1} H_k^T + R_k\big) \\ &= \min_{\|x\|=1} \big(x^T (H_k P_{k,k-1} H_k^T) x + x^T (R_k) x\big) \\ &\geq \min_{\|x\|=1} \big(x^T (H_k P_{k,k-1} H_k^T) x\big) + \min_{\|x\|=1} \big(x^T (R_k) x\big) \\ &= \lambda_{min}\big(H_k P_{k,k-1} H_k^T\big) + \lambda_{min}\big(R_k\big) = \underline{\sigma}\big(H_k P_{k,k-1} H_k^T\big) + \underline{\sigma}\big(R_k\big)\end{aligned} \tag{5.35}$$

which implies

$$\bar{\sigma}(K_k) \leq \bar{\sigma}\Big(P_{k,k-1} H_k^T\Big) \Big(\underline{\sigma}\big(H_k P_{k,k-1} H_k^T\big) + \underline{\sigma}\big(R_k\big)\Big)^{-1} \tag{5.36}$$

Using similar arguments it can be shown that

$$\underline{\sigma}\big(H_k P_{k,k-1} H_k^T\big) \geq q_1 \underline{\sigma}^2(H_k^T) \tag{5.37}$$

Moreover

$$\bar{\sigma}(P_{k,k-1} H_k) \leq q_2 \bar{\sigma}(H_k^T) \tag{5.38}$$

Hence

$$\bar{\sigma}(K_k) \leq \frac{q_2 \bar{\sigma}(H_k^T)}{q_1 \underline{\sigma}^2(H_k^T) + \underline{\sigma}(R_k)} \tag{5.39}$$

Thus (5.25) gives

$$\|K_k\| \leq \frac{q_2 \bar{\sigma}(H_k^T)}{q_1 \underline{\sigma}^2(H_k^T) + \underline{\sigma}(R_k)} \leq h \frac{q_2}{q_1} \tag{5.40}$$

□

Furthermore, the equations (5.4), (5.23) and (5.24) gives

$$\|I - K_k H_k\| \leq \|P_{k,k} P_{k,k-1}^{-1}\| \leq \frac{p_2}{q_1} \tag{5.41}$$

and by use of (5.22) and (5.18)

$$\|\tilde{F}_k\| \leq \|I - K_k H_k\| \|f_{k-1}\| \leq \frac{p_2}{q_1} f \tag{5.42}$$

Now the main theorem of this paper can be stated and proved.

Theorem 5.1. *Assume that the bounds given by (5.10) and (5.22)-(5.25) are fulfilled and that f_k is nonsingular for all $k \geq 0$. Assume further that there exist an $\bar{\epsilon}$ such that*

$$\|e_{k-1,k-1}\| \leq \bar{\epsilon} \tag{5.43}$$

which implies $\|x_k - \hat{x}_{k,k-1}\| \leq \epsilon_1(\bar{\epsilon})$, where

$$\epsilon_1(\bar{\epsilon}) = a\bar{\epsilon} + \bar{w}$$

Moreover, assume that

$$\|\phi(x_k, \hat{x}_{k,k-1})\| \leq \varphi \|x_k - \hat{x}_{k,k-1}\|^2 \tag{5.44}$$

and

$$\|\theta(x_k, \hat{x}_{k,k})\| \leq \vartheta \|x_k - \hat{x}_{k,k}\|^2 \tag{5.45}$$

holds for $\|x_k - \hat{x}_{k,k-1}\| \leq \epsilon_1(\bar{\epsilon}) = \epsilon_1$ and $\|x_k - \hat{x}_{k,k}\| \leq \epsilon_1(\bar{\epsilon}) = \epsilon_1$ respectively.

Then there exists an $\epsilon > 0$ such that the solution of the error model (5.17) is:

1) *Locally exponential stable if the initial error satisfies $\|e_{0,0}\| \leq \epsilon$ and $\bar{w} = \bar{v} = 0$.*

2) *Bounded by*

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 - \frac{p_2}{\xi}\rho(\bar{w}, \bar{v}, \epsilon)$$

if the initial error satisfies $\|e_{0,0}\| \leq \epsilon$, and $\bar{w}$, $\bar{v}$ are sufficiently small. Here $\xi \in (-1, 0)$ is a constant and $\rho(\bar{w}, \bar{v}, \epsilon) > 0 \; \forall \; k \geq 0$ is a function to be defined later.

Proof: Denote in the following: $e_{k,k}$ by e_k, $e_{k-1,k-1}$ by e_{k-1}, $P_{k,k}$ by P_k and $P_{k-1,k-1}$ by P_{k-1}.

Let $V : R^n \rightarrow R$ be a positive function defined by

$$V(e_{k-1}) = e_{k-1}^T P_{k-1}^{-1} e_{k-1} \tag{5.46}$$

Such that from (5.23)

$$\frac{1}{p_2}\|e_{k-1}\|^2 \leq V(e_{k-1}) \leq \frac{1}{p_1}\|e_{k-1}\|^2 \tag{5.47}$$

Then:

$$\begin{aligned}\Delta V :&= e_k^T P_k^{-1} e_k - e_{k-1}^T P_{k-1}^{-1} e_{k-1}\\ &= \big(\tilde{F}_k e_{k-1} + n_k + l_k\big)^T P_k^{-1}\big(\tilde{F}_k e_{k-1} + n_k + l_k\big) - e_{k-1}^T P_{k-1}^{-1} e_{k-1}\\ &= e_{k-1}^T\Big[\tilde{F}_k^T P_k^{-1}\tilde{F}_k - P_{k-1}^{-1}\Big]e_{k-1} + n_k^T P_k^{-1} n_k + l_k^T P_k^{-1}\big(2\tilde{F}_k e_{k-1} + l_k\big)\\ &+ 2n_k^T P_k^{-1}\big(\tilde{F}_k e_{k-1} + l_k\big)\end{aligned} \tag{5.48}$$

By use of Lemma 1 it follows that

$$\Delta V \leq -\gamma V(e_{k-1}) + l_k^T P_k^{-1}\big(2\tilde{F}_k e_{k-1} + l_k\big) + 2n_k^T P_k^{-1}\big(\tilde{F}_k e_{k-1} + l_k\big) + n_k^T P_k^{-1} n_k \tag{5.49}$$

Considering the second term (see also [10] Lemma 3.2), it holds that

$$\begin{aligned}\|l_k^T P_k^{-1}\big(2\tilde{F}_k e_{k-1} + l_k\big)\| \leq \|P_k^{-1}\|&\Big(\|\theta_f^-(x,\hat{x})^T\left[I - K_k H_k\right]^T\| + \|\phi_h(x,\hat{x})^T K_k^T\|\Big)\\ &\Big(\|2\tilde{F}_k e_{k-1}\| + \|\left[I - K_k H_k\right]\theta_f^-(x,\hat{x})\| + \|K_k\phi_h(x,\hat{x})\|\Big)\end{aligned} \tag{5.50}$$

Using

$$\|e_{k,k-1}\| \leq \vartheta\|e_{k-1,k-1}\|^2 + f\|e_{k-1,k-1}\| + \bar{w}^2$$

gives

$$\|l_k^T P_k^{-1}\big(2\tilde{F}_k e_{k-1} + l_k\big)\| \leq \bar{\varphi}\|e_{k-1,k-1}\|^3 + \bar{w}W_1\big(\bar{w},\bar{\epsilon}\big) \tag{5.51}$$

where

$$\begin{aligned}\bar{\varphi} =& \frac{1}{q_1^2 p_1}\Bigg[h^2\varphi^2\vartheta^2 q_2^2\left(\vartheta\bar{\epsilon}^5 + 4\bar{\epsilon}^4\right) + 2h\varphi\vartheta^2 q_2\bar{\epsilon}^3\left(3h\varphi f^2 q_2 + \vartheta p_2\right)\\ &+ \big(2\left(p_2\vartheta + hf^2 q_2\varphi\right) + q_1\big)2hfq_2\vartheta\varphi\bar{\epsilon}^2 + \Big(4hf^2 q_1 q_2\vartheta\varphi + \left(p_2\varphi + hf^2 q_2\varphi\right)^2\Big)\bar{\epsilon}\Bigg]\\ &+ \frac{2f}{q_1 p_1}\left(p_2\vartheta + hf^2 q_2\varphi\right)\end{aligned} \tag{5.52}$$

and

$$\begin{aligned} W_1\big(\bar{w},\bar{\epsilon}\big) = \frac{1}{q_1p_1}\Bigg[& 2h\varphi\bar{\epsilon}\frac{q_2}{q_1}\bar{w}\Big(3hq_2\varphi\vartheta(2f\bar{\epsilon}^2+\vartheta\bar{\epsilon}^3)+fq_1+\big(p_2\vartheta+3hq_2f^2\varphi\big)\,\bar{\epsilon}\Big) \\ & + h^2\varphi^2\frac{q_2^2}{q_1}\bar{w}^2\,\big(\bar{w}+4(\vartheta\bar{\epsilon}^2+f\bar{\epsilon})\big)+4h\varphi\bar{\epsilon}^2\frac{q_2}{q_1p_1}\Big[f^2+f\vartheta\bar{\epsilon}\left(1+\frac{p_2}{q_1}\right) \\ & + h\varphi\bar{\epsilon}\frac{q_2}{q_1}\,\big(f\vartheta\bar{\epsilon}(2+f)+\vartheta^2\bar{\epsilon}^2+f^3\big)\Big]\Bigg] \end{aligned} \tag{5.53}$$

Thus

$$\Delta V \leq -\gamma V(e_{k-1})+\bar{\varphi}\|e_{k-1,k-1}\|^3+n_k^TP_k^{-1}n_k+2n_k^TP_k^{-1}\big(\tilde{f}_ke_{k-1}+l_k\big)+\bar{w}W_1\big(\bar{w},\bar{\epsilon}\big) \tag{5.54}$$

for $\|e_{k-1,k-1}\| \leq \bar{\epsilon}$.

Choosing

$$\epsilon = \min\left(\bar{\epsilon},\frac{\gamma}{\psi p_2\bar{\varphi}}\right) \tag{5.55}$$

where $\psi > 1$, gives for $\|e_{k-1,k-1}\| \leq \epsilon$

$$\bar{\varphi}\|e_{k-1,k-1}\|\|e_{k-1,k-1}\|^2 \leq \frac{\gamma}{\psi p_2\bar{\varphi}}\|e_{k-1,k-1}\|^2 \leq \frac{\gamma}{\psi}V(e_{k-1}) \tag{5.56}$$

Thus

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi}V(e_{k-1})+n_k^TP_k^{-1}n_k+2n_k^TP_k^{-1}\big(\tilde{F}_ke_{k-1}+l_k\big)+\bar{w}W_1\big(\bar{w},\epsilon\big) \tag{5.57}$$

for $\|e_{k-1,k-1}\| \leq \epsilon$.

Next consider the terms $n_k^TP_k^{-1}n_k$ and $2n_k^TP_k^{-1}\big(\tilde{F}_ke_{k-1}+l_k\big)$. Using inequalities (5.41), (5.42), and (5.33), the following can be established

$$\|n_k^TP_k^{-1}n_k\| \leq \|P_k^{-1}\|\|n_k\|^2 \leq \frac{1}{p_1}\Big(\|I-K_kH_k\|\bar{w}+\|K_k\|\bar{v}\Big)^2 \leq \frac{1}{q_1^2p_1}\big(p_2\bar{w}+hq_2\bar{v}\big)^2 \tag{5.58}$$

and

$$\begin{aligned} & \|2n_k^TP_k^{-1}\big(\tilde{F}_ke_{k-1}+l_k\big)\| \leq 2\|n_k^TP_k^{-1}\|\|\tilde{F}_ke_{k-1}+l_k\| \\ & \leq \frac{2}{q_1p_1}\big(p_2\bar{w}+q_2h\bar{v}\big)\times\left(f\frac{p_2}{q_1}\|e_{k-1}\|+\vartheta\frac{p_2}{q_1}\|e_{k-1}\|^2+h\varphi\frac{q_2}{q_1}\|e_{k,k-1}\|^2\right) \end{aligned} \tag{5.59}$$

Substituting for $\|e_{k,k-1}\|$ and adding (5.58) to (5.59) yields

$$\|2n_k^T P_k^{-1}\big(\tilde{f}_k e_{k-1} + l_k\big)\| + \|n_k^T P_k^{-1} n_k\| \leq (p_2\bar{w} + hq_2\bar{v})\, W_2\big(\bar{w}, \bar{v}, \epsilon\big) \tag{5.60}$$

where

$$W_2\big(\bar{w}, \bar{v}, \epsilon\big) = \frac{1}{q_1^2 p_1} \times \\ \Big[2\Big(fq_2\epsilon + p_2\vartheta\epsilon^2 + hq_2\varphi\big(\vartheta^2\epsilon^4 + 2f\vartheta\epsilon^3 + \big(2\vartheta + f^2\big)\,\epsilon^2 + 2f\bar{w}\epsilon + \bar{w}^2\big)\Big) + p_2\bar{w} + hq_2\bar{v}\Big] \tag{5.61}$$

Therefore

$$\Delta V \leq \frac{\gamma(1-\psi)}{\psi} V(e_{k-1}) + \rho\big(\bar{w}, \bar{v}, \epsilon\big) \tag{5.62}$$

where

$$\rho\big(\bar{w}, \bar{v}, \epsilon\big) = \bar{w}W_1\big(\bar{w}, \epsilon\big) + (p_2\bar{w} + hq_2\bar{v})\, W_2\big(\bar{w}, \bar{v}, \epsilon\big) \tag{5.63}$$

Since $0 < \gamma < 1$ and $\psi > 1$

$$\xi := \frac{\gamma(1-\psi)}{\psi} \in (-1, 0) \tag{5.64}$$

Using (5.62) and (5.64) and starting at $k = 0$ gives

$$\begin{aligned} V(e_{1,1}) &\leq (1+\xi)V(e_{0,0}) + \rho(\bar{w}, \bar{v}, \epsilon) \\ V(e_{2,2}) &\leq (1+\xi)V(e_{1,1}) + \rho(\bar{w}, \bar{v}, \epsilon) \\ &\leq (1+\xi)^2 V(e_{0,0}) + (1 + (1+\xi))\rho(\bar{w}, \bar{v}, \epsilon) \\ &\vdots \\ V(e_{k,k}) &\leq (1+\xi)^k V(e_{0,0}) + \sum_{n=0}^{n=k} (1+\xi)^n \rho(\bar{w}, \bar{v}, \epsilon) \end{aligned} \tag{5.65}$$

Hence (5.23) and (5.65) implies

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 - \frac{p_2}{\xi}\rho(\bar{w}, \bar{v}, \epsilon) \tag{5.66}$$

In absence of noise $\rho(\bar{w}, \bar{v}, \epsilon) = 0$, and

$$\|e_{k,k}\|^2 \leq \frac{p_2}{p_1}(1+\xi)^k \|e_{0,0}\|^2 \tag{5.67}$$

□

Remark 1: The proof of this theorem can be based on the Input-to-state stability (ISS) concept for discrete-time nonlinear systems, see e.g. [6]. However, as far as the authors are aware of, the definition of a discrete-time version local ISS system does not seem to be easily available in the literature. One definition of a local ISS system, and a proposition regarding local ISS of discrete-time systems, are given below[2].

Definition 5.1. *The discrete-time system*

$$x(k+1) = f(x(k), u(k)), \quad x(0) = x_0 \tag{5.68}$$

where $f : D \times D_u \to \mathbb{R}^n$ is continuous, with $D = \{x \in \mathbb{R}^n : |x| \leq r\}$ and $D_u = \{u \in \mathbb{R}^m : |u(k)| \leq r_u\}$, is said to be locally input-to-state stable *(ISS) if there exist a class $\mathcal{KL}$ function β, a class $\mathcal{K}$ function γ and constants $k_1 > 0$, $k_2 > 0$ such that for each solution $x(t, \xi, u)$ of (5.68) corresponding to initial state ξ with $|\xi| \leq k_1$ and input u with $\|u\|_\infty \leq k_2$ we have*

$$\|x(k, \xi, u)\| \leq \beta(\|\xi\|, t) + \gamma(\|u\|_\infty) \tag{5.69}$$

It is said to be input-to-state stable*, or* globally *ISS if $D = \mathbb{R}^n$, $D_u = \mathbb{R}^m$, and (5.69) holds for all initial states and all bounded inputs u.*

It is assumed that the unforced system

$$x(k+1) = f(x(k), 0)$$

has an asymptotically stable equilibrium at $x = 0$.

Proposition 5.1. *If the system (5.68) admits an ISS-Lyapunov function[3] on D, then it is locally ISS with*

$$\gamma = \alpha_1^{-1} \circ \chi \tag{5.70}$$

$$k_1 = \alpha_2^{-1}(\alpha_1(r)) \tag{5.71}$$

$$k_2 = \min\{r_u, \chi^{-1}(\alpha_1(k_1))\} \tag{5.72}$$

■

Remark 2: The condition (5.44) can be relaxed to

$$\|\phi(x_k, \hat{x}_{k,k-1})\| \leq \varphi \|x_k - \hat{x}_{k,k-1}\| \tag{5.73}$$

[2] Per-Ole Nyman, 2004, to be published

[3] See appendix B for a definition of an ISS-Lyapunov function

This will considerably simplify the functions W_1 and W_2, however, a term $p_2 f \vartheta (2 + h\varphi q_2/q_1)\epsilon^2$ will then be included in (5.62). This term will only be small if both the nonlinearity in state and in the measurement are very modest. This corresponds to the intuitive conclusion that if the nonlinearities are modest, the EKF will be stable if reasonably initialized. ■

Remark 3: In this proof the bound on the noise processes required to obtain stability are not quantified. Unfortunately, this turns out be to be quite difficult in this case. When considering stochastic stability of the EKF formulated in terms of the a-priori variables, this is more easy, see [10]. ■

5.4 Example

5.4.1 Example 1

This first example is taken from [10]. The signal model is given by

$$f(x_k) = \begin{bmatrix} x_1 + \tau x_2 \\ x_2 + \tau\left(-x_1 + (x_1^2 + x_2^2 - 1)x_2\right) \end{bmatrix} \tag{5.74}$$

$$h(x_k) = x_1 \tag{5.75}$$

By simulations (predictor-corrector) it is found that the following bounds on the covariance matrices apply

$$0.5 = p_1 \leq P_{k,k} \leq p_2 = 1.6 \tag{5.76}$$

$$0.5 = q_1 \leq P_{k,k-1} \leq q_2 = 1.6 \tag{5.77}$$

Furthermore, $f = h = 1$. Using $\psi = 3/2$ and assuming $\bar{\epsilon} = 1$ yields

$$\epsilon = 8.1 \cdot 10^{-3} \quad \text{and} \quad \bar{w} = \bar{v} = 2 \cdot 10^{-8} \tag{5.78}$$

This results are conservative, as also reported in [10]. By simulations it is found that this filter performs satisfactory if the initial error and noise processes are below the following bounds

$$\epsilon = 0.4, \quad \bar{w} = 1 \cdot 10^{-4} \quad \text{and} \quad \bar{v} = 10 \tag{5.79}$$

In Figure 5.1 and 5.2 the real and estimated states are shown for the above mentioned case.

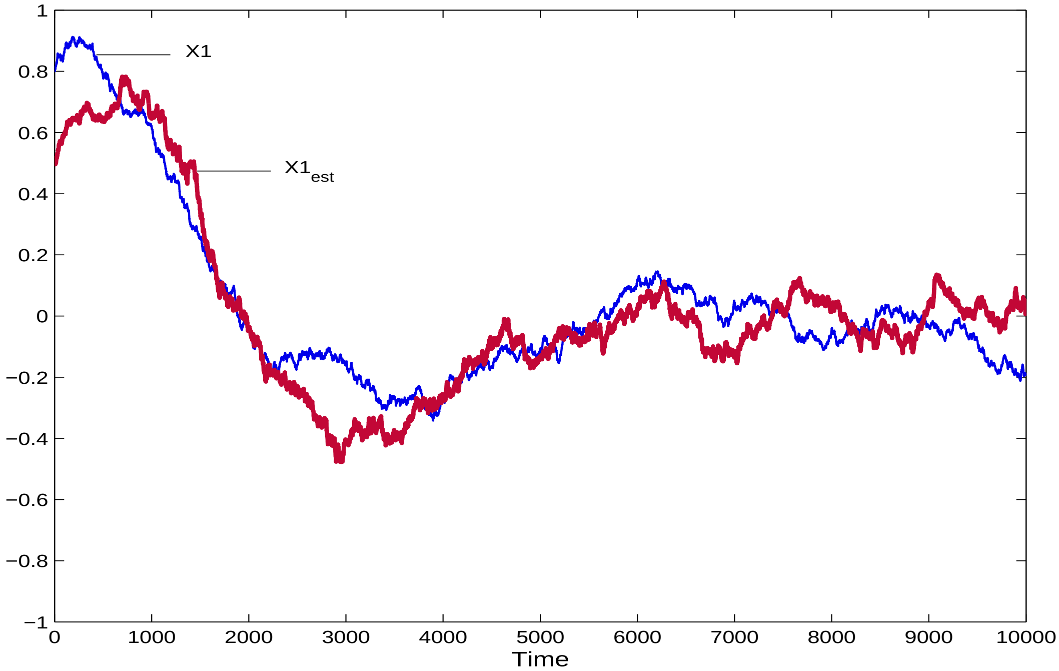

Figure 5.1: Real and estimated state, x_1

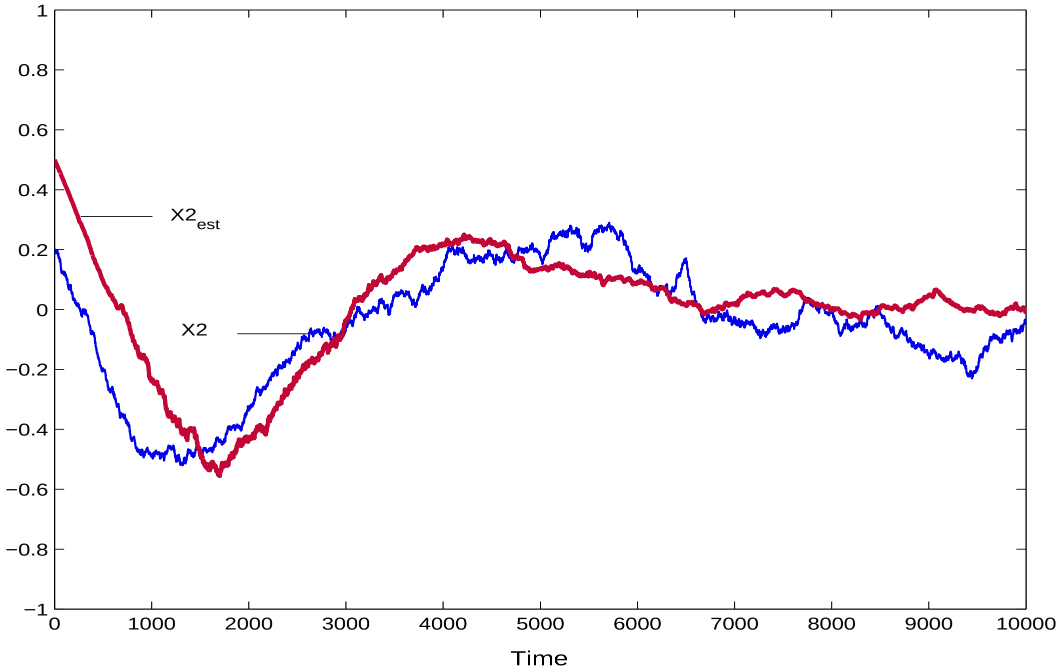

Figure 5.2: Real and estimated state, x_2

5.4.2 Example 2

In this second example an EKF used for tracking the amplitude, phase and frequency of a low frequency signal is considered. The signal model is linear and time invariant in state, and is given by:

$$x(k+1) = Ax(k) + Bw(k) \tag{5.80}$$
$$y(k) = x_3 \sin x_2 + v(k) \tag{5.81}$$

where x_1 is the phase increment (or frequency), x_2 is the phase and x_3 is the amplitude. The matrices A and B are given by:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Proving stability by Theorem 5.1 is straight forward. The bounds on the matrices are:

$$\|A\| = 1 \tag{5.82}$$
$$\bar{\sigma}(H_k^T)/\underline{\sigma}^2(H_k^T) = 1 \tag{5.83}$$

By simulations it is found that the ratio $p_2/p_1 = p_2/q_1 = q_2/q_1 = 830$ are sufficient.

Now it must be shown that condition (5.44) holds for some constant φ. The remainder term in the Taylor expansion is given by (see e.g. [2])

$$\begin{aligned}&\phi(x_k, \hat{x}_{k,k-1}) = \\ &h(x_k) - h(\hat{x}_{k,k-1}) - H_k(x_k - \hat{x}_{k,k-1}) = \\ &\frac{1}{2}\left(x_k - \hat{x}_{k,k-1}\right)^T \frac{\partial^2 h}{\partial x^2}(\tilde{x}_k)\left(x_k - \hat{x}_{k,k-1}\right)\end{aligned} \tag{5.84}$$

such that

$$\varphi = \max_{1\le i\le n} \sup_{x\in\mathcal{M}} \left\| \frac{1}{2}\frac{\partial^2 h_i}{\partial x^2}(x) \right\| \tag{5.85}$$

provided that $x_k, \tilde{x}_k, \hat{x}_{k,k-1} \in \mathcal{M}$ where $\mathcal{M} \in \mathbb{R}^n$ is a convex and open set. It must be required that x_3 is bounded, as can be seen from equation (5.81). However, this does not require the state space to be compact, as required in [12], since no bounds are put on x_1 and x_2.

Choosing $\bar{\epsilon} = 1 \cdot 10^{-5}$ and using (5.55) with $\varphi = 1/2$, gives $\epsilon = 1.28 \cdot 10^{-9}$. To obtain a bounded $e_{k,k}$, the noise processes must be bounded by approximately

$\bar{w} = \bar{v} = 1.17 \cdot 10^{-19}$. These values are far below the limit of any practical significance, and the results are therefore only of theoretical interest.

Consider now a slightly modified signal model. Let the matrix B be given by

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.86}$$

It is not possible now to guarantee that the state x_3 is bounded. A small modification is therefore necessary to prove stability in a stringent manner. Specifically, let the matrix A be replaced by

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1-\delta \end{bmatrix} \tag{5.87}$$

Choosing delta to be small and positive will now result in a bounded x_3, and stability can be proved.

The results obtained by simulations are in fact very satisfactory with regard to both the convergence speed and the size of the initial error and the noise processes, even though this has not been confirmed by the theoretical analysis. By simulations it is found that the error will remain bounded and that the filter yields a satisfactory estimate, even when an initial error of up to $\|e_0\| < 0.75$ is allowed, and the noise processes are bounded by up to $\bar{w} = 9.5 \times 10^{-5}$ and $\bar{v} = 3$. This situation is illustrated in Figure 5.3. It should be noted that the process noise, although bounded by $\bar{w}$, yields a rather fluctuating amplitude. In many real applications $\bar{w}$ could therefore be taken even smaller.

5.5 Conclusion

In this paper the stability properties of an EKF is considered. The main conclusions are:

1) The assumption in [10] that the matrix H_k is bounded in norm is relaxed to only requiring a finite ratio between its largest and smallest singular value, provided that the norm of the Hessian matrix of the function $h(x_k)$ is finite for any $x \in \mathbb{R}^n$.

2) The results obtained are very conservative (see also [10], section V, Numerical Simulations). One of the reasons for this is that the ratios p_2/p_1 and q_2/q_1,

which plays an important role in the analysis, normally are large. This suggests that either some important properties are disregarded when using either Lyapunov analysis or the total stability theorem to prove EKF stability, or that the problem is formulated to generally, i.e. the conditions under which stability is proved should be stronger.

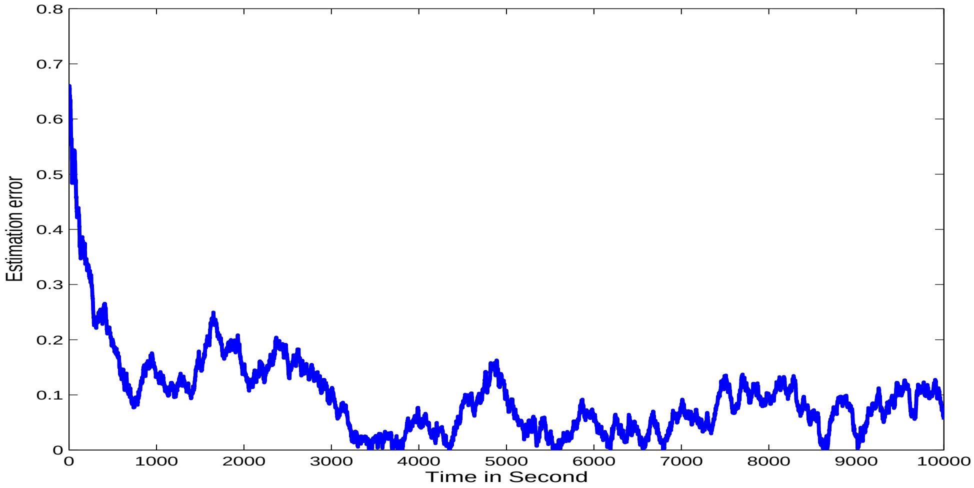

Figure 5.3: Estimation error example 2

3) In the linear case the Theorems presented in this paper recover stability unconditionally. However, if the signal model is modified to be only slightly nonlinear, the results may become conservative. It can be shown that in the special case of linear state map, a certain choice of the matrix Q_k yields better results, see [9].

4) If the matrix H_k is required to be bounded in norm, which is the case for a broad class of filters, a more tight bound can be applied for the gain matrix K_k. Such a bound is applied in [10], but it turns out that the results are still conservative. Therefore, using the bound given by (5.33), will extend the set of filters for which stability can be proved.

ACKNOWLEDGEMENT

The authors are grateful to LKAB, Sweden, for financial support during this research, and to the anonymous referees for their valuable comments and suggestions.

References

[1] C. K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications.* Springer, Berlin, 1999.

[2] J. E. Dennis Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, New Jersey, 1983.

[3] Arthur Gelb. *Applied Optimal Filtering.* MIT Press, Cambridge, Massachusetts, 1974.

[4] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1985.

[5] A. Bensoussan J. S. Baras and M. R. James. Dynamic observers as asymptotic limits of recursive filters: special cases. *SIAM Journal of Applied Mathematics*, 48:1147–1158, 1988.

[6] Z. P. Jiang and Y. Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 37:857–869, 2001.

[7] A. J. Krener. The convergence of the extended Kalman filter. *Directions in Mathematical Systems Theory and Optimization*, pages 173–182, 2003.

[8] B. F. La Scala, R.R. Bitmead, and M.R. James. Conditions for stability of the extended Kalman filter and their applications to the frequency tracking problem. *Mathematics of Control, Signals, and Systems*, 8:1–26, 1995.

[9] K. Rapp and P. -O. Nyman. On stability of the extended Kalman filter. *To be presented at MMAR 2004, Poland*, 2004.

[10] K. Reif, Stefan Günter, Engin Yaz, and Rolf Unbehauen. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transaction on automatic control*, 44:714–728, 1999.

[11] K. Reif and R. Unbehauen. The extended Kalman filter as an exponential observer for nonlinear systems. *IEEE Transaction on Signal Processing*, 47:2324–2328, 1999.

[12] Y. Song and J. W. Grizzle. The extended Kalman filter as a local asymptotic observer for discrete-time nonlinear systems. *Journal of Mathematical Systems, Estimation, and Control*, 5:59–78, 1995.

[13] M. Vidyasagar. *Nonlinear Systems Analysis, 2nd ed.* Prentice Hall, New Jersey, 1993.

Appendix A

Genetic algorithms, a simple and useful tool

A.1 Introduction

When solving problems which includes searching numerically for some maximum or minimum, some efficient search algorithm has to be chosen. In this work Genetic Algorithms (GA) has been used for the filter tuning problem. The only prerequisite for using an automatic search method like GA's is that the tuning problems can be converted into max/min problems. The reason for choosing GA is mostly related to the filter tuning problem, were the relation between the tuning parameters (the covariance matrices) and the filters performance is very complicated. Unlike more conventional methods for optimization of functions, genetic algorithms does not require the function to have properties like differentiability or continuity, and therefore allows more complicated functions to be considered. One of the very desirable features of GA's is that there is no problems including additional tests, like stability test for instance, and this make this algorithms suitable for a wide class of searching and optimizing problems. In this work the free GAOT toolbox (Genetic Algorithm for Optimization Toolbox) for MATLAB is used.

A.2 Genetic algorithms in general

In this section a very short description of Genetic Algorithms is given. The purpose of this section is to briefly introduce the main parts of a GA.

A simple genetic algorithm can be described by the following scheme ([2]):

1) Given a initial population P_0 of N individuals

```
2) $i \leftarrow 1$
3) $\bar{P}_i \leftarrow$ selection-function $(P_{i-1})$
4) $P_i \leftarrow$ reproduction-function $(\bar{P}_i)$
5) evaluate$(P_i)$
6) $i \leftarrow i + 1$
7) Repeat from step 3 until termination
```

While searching the functions solution space, the GA simulates evolution and uses the best individuals (possible solutions) in one population to produce the new population of possible solutions. Each individual in the population is described by use of a *chromosome representation*. The next generation is made by letting *genetic operators* create new individuals from a randomly selected set of old individuals. The selection requires a *selection function*, and a good individual has normally a larger probability to be picked than a bad individual. The *evaluation function* is used to assign a *fitness value* to each individual in the population, and this fitness value is used to judge wheatear one individual is good or bad. The GA will need a criterion for when to stop. Normally this is given in number of generations, such that when the last generation is created, the solution is the best individual, judged by its fitness value, in the final population.

A.3 The GAOT toolbox for MATLAB

The representation of individuals can be either floating point or binary. Normally the floating point representation will be the most effective representation when measured in terms of CPU time, (see [2]). When starting GAOT the size of the initial population and number of generations (used as stop criterion) must be given together with the following two inputs:

- Upper and lower bounds of each variable
- The evaluation function

In addition to these two inputs, a number of optional inputs may be given. By default, GAOT selects randomly an initial population in the search space, restricted by the upper and lower bounds for the parameters. The output is a string containing the best solution, and the following optional output are provided

- The end population (endPop)

- A matrix of the best individuals and their corresponding generation (bPop)
- A matrix of maximum and mean functional values of the population for each generation (traceInfo)

As all Genetic algorithms are based on that some individuals (the best individuals) are selected and transferred to the next generation, the selection function is a corner stone in each GA. In GAOT the selection is carried out after all new individuals have been evaluated. The selection is based on the principle that an individual with high fitness value is more likely to be selected than an individual with low fitness value. Three different functions are implemented in this toolbox, which are

- Roulette wheel selection
- Normalized geometric selection
- Tournament selection

Different types of selection function are thoroughly discussed in [1].

There are two basic types of genetic operators[1], and both are used in GAOT. These are *mutation* and *crossover*. Mutation takes one individual and alter it to create a new individual $(parent) \rightarrow (children)$, while crossover takes two individuals and make two new $(parent1, parent2) \rightarrow (child1, child2)$. A number of different mutations and crossovers exist, see e.g. [1], [3] or [2] for more details. In GAOT the number of mutations and crossovers performed on each generation can be chosen. Four different types of mutations and three different types of crossovers are implemented. The number of each to be performed can be chosen. This is convenient for problems where one type of mutation or crossover yields better results than others.

This toolbox search for a maximum function value, so the evaluation function must be designed in such a way that an individual with high fitness value is superior to one with a lower value.

References

[1] Thomas Bäck. *Evolutionary Algorithms in Theory and Practice.* Oxford University Press, Oxford, 1996.

[2] C. R. Houck, J. A. Joines, and M. G. Kay. The genetic algorithm optimization toolbox (gaot) for matlab 5. *http://www.ie.ncsu.edu/mirage/GAToolBox /gaot/*, 1996.

[1] In [3] also the selection function is classified as a genetic operator, while in [2] it is not

[3] Mo Jamshidi, Leandros dos Santos Coelho, Renato A. Krohling, and Peter J. Fleming. *Robust Control Systems with Genetic Algorithms*. CRC Press, Boca Raton London New York Washington DC, 2003.

Appendix B

Local ISS discrete-time systems[1]

Consider the discrete time system

$$x(k+1) = f(x(k), u(k))\,, \quad x(0) = x_0 \tag{B.1}$$

where $f : D \times D_u \to \mathbb{R}^n$ is continuous, with $D = \{x \in \mathbb{R}^n : |x| \leq r\}$ and $D_u = \{u \in \mathbb{R}^m : |u(k)| \leq r_u\}$.

It is assumed that the unforced system

$$x(k+1) = f(x(k), 0)$$

has an asymptotically stable equilibrium at $x = 0$.

Definition B.1. *The system (B.1) is said to be* locally input-to-state stable *(ISS) if there exist a class $\mathcal{KL}$ function β, a class $\mathcal{K}$ function γ and constants $k_1 > 0$, $k_2 > 0$ such that for each solution $x(t, \xi, u)$ of (B.1) corresponding to initial state ξ with $|\xi| \leq k_1$ and input u with $\|u\|_\infty \leq k_2$ we have*

$$\|x(k, \xi, u)\| \leq \beta(\|\xi\|, t) + \gamma(\|u\|_\infty) \tag{B.2}$$

It is said to be input-to-state stable*, or* globally *ISS if $D = \mathbb{R}^n$, $D_u = \mathbb{R}^m$, and (B.2) holds for all initial states and all bounded inputs u.*

Definition B.2. *A continuous function $V : D \to \mathbb{R}$ is said to be an ISS-Lyapunov function on D for the system (B.1) if there exist class $\mathcal{K}_\infty$ functions α_1 and α_2 on D such that*

$$\alpha_1(\|\xi\|) \leq V(\xi) \leq \alpha_2(\|\xi\|) \quad \forall \xi \in D \tag{B.3}$$

and there exist a $\mathcal{K}_\infty$ function α_3 and a $\mathcal{K}$ function σ such that

$$V(f(\xi, \mu)) - V(\xi) \leq -\alpha_3(\|\xi\|) + \sigma(|\mu|) \tag{B.4}$$

for all ξ, μ with $|\xi| \leq r$ and $|\mu| \leq r_u$.

[1] Per-Ole Nyman, 2004, to be published

Note that (B.4) implies

$$-\alpha_3(|\xi|) = -\alpha_3(\alpha_2^{-1}(\alpha_2(|\xi|)) \leq -\alpha_3(\alpha_2^{-1}(V(\xi))) = -\alpha_4(V(\xi))$$

where $\alpha_4 := \alpha_3 \circ \alpha_2^{-1}$. Thus

$$V(f(\xi,\mu)) - V(\xi) \leq -\alpha_4(V(\xi)) + \sigma(|\mu|) \tag{B.5}$$

for all ξ, μ with $|\xi| \leq r$ and $|\mu| \leq r_u$. Moreover, there exists a $\mathcal{K}_\infty$ function $\hat{\alpha}_4 \leq \alpha_4$ such that $Id - \hat{\alpha}_4 \in \mathcal{K}$. Consequently,

$$V(f(\xi,\mu)) - V(\xi) \leq -\hat{\alpha}_4(V(\xi)) + \sigma(|\mu|) \tag{B.6}$$

for all ξ, μ with $|\xi| \leq r$ and $|\mu| \leq r_u$.
Let $0 < c < 1$, and define the function $\chi : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\chi(s) = \hat{a}_4^{-1}\left(\frac{\sigma(s)}{c}\right) \tag{B.7}$$

This is a $\mathcal{K}$ function.

Proposition B.1. *If the system (B.1) admits an ISS-Lyapunov function on D, then it is locally ISS with*

$$\gamma = \alpha_1^{-1} \circ \chi \tag{B.8}$$

$$k_1 = \alpha_2^{-1}(\alpha_1(r)) \tag{B.9}$$

$$k_2 = \min\{r_u, \chi^{-1}(\alpha_1(k_1))\} \tag{B.10}$$