ISSN 0280-5316
ISRN LUTFD2/TFRT--5747--SE

Logarithmic Concave Observers

Toivo Henningsson

Department of Automatic Control
Lund Institute of Technology
May 2005

Department of Automatic Control
Lund Institute of Technology
Box 118
SE-221 00 Lund Sweden

Document name
MASTER THESIS

Date of issue
May 2005

Document Number
ISRNLUTFD2/TFRT--5747--SE

Author(s)
Toivo Henningsson

Supervisor
Karl Johan Åström at Automatic Control in Lund

Sponsoring organization

Title and subtitle
Logarithmic Concave Observer. (Logaritmiskt konkava observerare).

Abstract
The problem of (online) state estimation for dynamical systems arises frequently in control. The well known Kalman Filter comprises a coherent theory for the case of linear systems with Gaussian noise, but as soon as either condition is relaxed the picture becomes much less clear. This thesis investigates the case when process disturbances and measurements are relaxed from Gaussian to log-concave. The range of systems that can be analyzed is broadened while still retaining enough structure that many desirable properties are preserved. Strongly log-concave functions are introduced as a means to quantify the Gaussian-like properties of log-concave functions. The main contribution of the thesis is two fundamental theorems, one giving a bound on covariance and the other describing how (strong) log-concavity is preserved and propagated. Applying the theorems to log-concave observers, they are found to have much in common with the Kalman Filter. It is shown that a Kalman Filter can be constructed that gives a conservative bound on the error covariance of the log-concave Bayesian Observer. Event based control is one case where measurements are far from (uncorrelated) Gaussian, but often log-concave. An example of control and state estimation for such a system is pursued throughout the thesis. Using proper consideration a Kalman Filter is found that gives a reasonable approximation of the optimal log-concave observer.

Keywords

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280-5316

ISBN

Language
English

Number of page
49

Recipient's notes

Security classification

Acknowledgements

I owe a number of people thanks for being able to complete this thesis. First I would like to thank Karl Johan Åström for great supervision and Tomas Olsson for many helpful comments and hints. I also would like to thank Leif Andersson for help with LaTeX. I would like to thank Magnus Fontes, Gunnar Sparr, Jan Holst, Tobias Rydén, Olivier Verdier and Anders P. Eriksson for giving their comments on the thesis during various stages. Finally I want to thank Maria for being there.

Contents

Introduction

In control it is desirable the know the state of the system being controlled; effective feedback depends on it. It is often not practical to measure the full state of a system directly, so one has to make do with the measurements that one has. The observer problem is the problem of estimating the state of a dynamical system given a model of the system, some of its input signals and some (usually noisy) measurements.

The optimal solution to the observer problem is the Bayesian Observer, which calculates the conditional probability of the state of a process given the measurements and any other available information. For the case of a linear system with Gaussian process disturbances and measurement noise, the optimal observer is the well known and widely adopted Kalman Filter.

When the conditions of linear dynamics or Gaussian noise are relaxed, matters become much more complicated. While the Bayesian Observer is still quite straightforward to derive, it is usually not practical to implement, and it is hard to assess the error made when trying to approximate it.

In this thesis the assumption of Gaussian functions is relaxed to logarithmic concave, or log-concave functions. The optimal observer is derived and analyzed, and some ways of approximating it are investigated. It turns out that observers for the log-concave case have much in common with the Kalman Filter, and are in many ways almost as well behaved. In particular, by stripping away the non-Gaussian parts of log-concavity, a Kalman Filter is found which gives a conservative bound on observer uncertainty.

One case where measurements are not accurately described by the model of a true value corrupted by additive Gaussian noise is the case of event based control, where measurements come in the form of events that are generated under certain conditions such as a signal passing over a threshold. In many cases, the measurements can still be shown to be log-concave, however. To illustrate and test the theory, a simple example of event based control is investigated throughout the thesis.

Notations

The symbol $\mathbb{R}$ will be used to refer to the set of real numbers. Vectors will be printed in boldface, for instance $\mathbf{x} \in \mathbb{R}^n$, and are considered as single-column matrices. Matrices will be represented using capital letters, for instance A. Stochastic variables will also be represented with capital letters, while the corresponding realizations will be represented with lowercase letters. For instance, x_k is the realization of X_k. Vector valued stochastic variables will be printed in boldface capital letters, for instance $\mathbf{X_k}$.

Integrals without subscripts will refer to integration over the entire range of the integration variable, for instance

$$\int f(\mathbf{x})\, dx$$

refers to the integral of $f(x)$ over $\mathbb{R}^n$. Relational operators applied to matrices will be interpreted in the sense of positive definiteness, for instance $A > B$ means that $A - B$ is positive definite and $A \geq B$ means that $A - B$ is positive semidefinite.

Statistical preliminaries

Probability densities will be a main topic of study. By a probability density is meant any function $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$ such that $f(\mathbf{x}) \geq 0$ and

$$\int f(\mathbf{x})\, dx = C = 1.$$

If the total probability integral is not equal to 1, but to some $C \in (0, \infty)$ then f is termed *normalizable*. A normalizable function can be turned into a probability density by the transformation $g(\mathbf{x}) = C^{-1} \cdot f(\mathbf{x})$, implying that positive constant factors can be safely disregarded when dealing with probability densities.

Probability densities will be denoted with the letter f, usually subscripted by either the stochastic variable that it concerns or some subscript of convenience for commonly used probability densities. For instance, $f_X(x)$ is the probability density of the stochastic variable X, meaning that

$$P(X \in \Omega) = \int_{x \in \Omega} f_X(x)\, dx,$$

where Ω is a subset of the set of values that X may assume. Conditional probability densities may be denoted like $f_{X|Y}(x|y)$, which is the conditional probability density of X given that $Y = y$.

The expectation or center of mass of the probability density $f(\mathbf{x})$ with respect to the variable $\mathbf{x}$ will be referred to as

$$\mathrm{E}_{\mathbf{x}}\big(f(\mathbf{x})\big) = \int x f(\mathbf{x})\, dx.$$

The covariance matrix of a probability density $f(\mathbf{x})$ with respect to the variable $\mathbf{x}$ will be referred to as

$$\mathrm{Cov}_{\mathbf{x}}\big(f(\mathbf{x})\big) = \int (\mathbf{x} - \mathbf{m}_{\mathbf{x}})(\mathbf{x} - \mathbf{m}_{\mathbf{x}})^T f(x)\, dx,$$

where $\mathbf{m}_{\mathbf{x}} = \mathrm{E}_{\mathbf{x}}\big(f(\mathbf{x})\big)$.

The observer problem

The problem of estimating the state of a dynamical system given some (noisy) measurements arises in many applications, among them control. Since the estimate is to be used in a feedback loop only online estimates (estimates based on information available at the current time) are of interest. A typical scenario is shown in figure 1. It is usually assumed that some stochastic model of the system dynamics, disturbances and measurement noise is available.

The optimal observer problem is the problem of finding the conditional probability density of the state of a system from the available information, including measurements. The optimal or Bayesian Observer represents in some sense the best estimate that can be found under any specific circumstances.

From a probability density often only the conditional mean state is used, because it is simple, and in the case of linear dynamics, Gaussian noise, and quadratic cost function can be shown to be optimal (LQG control). Control laws that for instance take uncertainty directly into account or are based on expected loss are conceivable however.

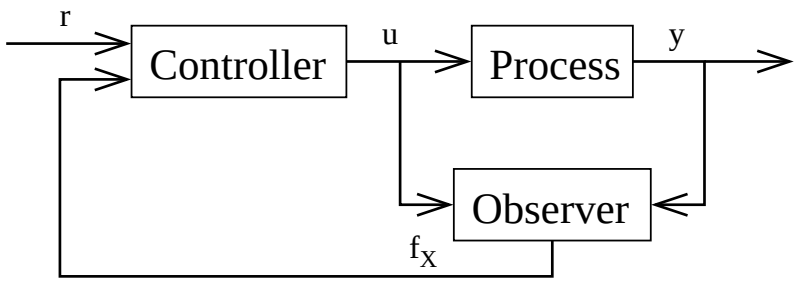

Figure 1 A typical scenario for using an observer. r is the reference signal, u the command signal, y the measurement signal and f_X is the system state probability density as estimated by the observer. The controller usually only extracts the mean state estimate from f_X, but could conceivably take for instance estimate uncertainty into account.

For most state space models, the form of the online Bayesian Observer is known. For continuous time models, the Fokker-Planck equation or Kolmogorov forward equation (see for instance [1]) describes the evolution of the probability density over the system state, and can be augmented to account for measurements. It usually takes the form of a partial differential equation in the probability density of the system state.

In discrete time, the Bayesian Observer usually takes the form of an integral equality that relates the probability density at one sampling instant to that of the previous instant and the latest measurement.

In a few special cases, there exists a closed form solution to the Bayesian Observer problem. In the case of linear dynamics and Gaussian disturbances and measurements, this is the well-known Kalman Filter. If the system has a finite number of possible states, the closed form solution is known as a grid-based filter. The Kalman Filter enjoys many desirable properties and has found widespread application.

When the actual Bayesian Observer is not tractable, or not practical to implement, it has to be modified. The Bayesian Observer should only be modified in order to simplify it, or give some other substantial benefit. For systems that are not very far from linear Gaussian, linearization (and Gaussianization) can often allow a successful application of the Kalman Filter. If the system is continually re-linearized around the current expected state, the result is known as an Extended Kalman Filter.

Sometimes a system is sufficiently non-linear or non-Gaussian that no Kalman Filter can give a reasonable approximation to the optimal observer. For instance, the Kalman Filter assumes that the probability density over the system state is Gaussian. If the conditional probability density is multimodal (having several distinct local maxima), no Gaussian approximation will be good.

The Particle Filter is a computationally intensive but very general way to approximate the Bayesian Observer, and can be applied in many cases where simpler observers fail. It employs a cloud of points (particles) to approximate an arbitrary probability density, which is updated using Monte Carlo simulation. Approximating grid-based filters can also be used, but are often even more computationally expensive for the same accuracy.

For a review of online observers in general and Particle Filters in particular, see [2].

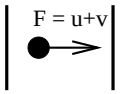

Figure 2 The Event Based Accelerometer. The sensor consists of a test mass confined between two walls, subject to a control acceleration u and an external acceleration v. An output pulse is generated when the mass hits either wall.

Process model

The processes that will be considered are systems in state space form with linear dynamics, state-independent but possibly non-Gaussian process noise and possibly non-Gaussian measurements. The main line of treatment will be limited to discrete time. With some additional assumptions, this structure will turn out to be very exploitable.

Let n be the dimension of the state space. Then $\mathbf{x_k} \in \mathbb{R}^n$ is the state of the process at time $t_k \in \mathbb{R}$. The time evolution of the process is described by

$$\mathbf{x_k} = \Phi \mathbf{x_{k-1}} + \Gamma \mathbf{u_{k-1}} + \mathbf{v_{k-1}}, \tag{1}$$

where $\Phi \in \mathbb{R}^{n \times n}$ describes the system dynamics, $\mathbf{u_k}$ is the input at time t_k, Γ describes how the input enters the system and $\mathbf{v_k}$ is the process noise. The process noise $\mathbf{v_k}$ is assumed to have zero expectation and be independent for different k. The distribution of $\mathbf{v_k}$ is described by the disturbance model

$$f_{\mathbf{V_k}}(\mathbf{v_k}) = f_N(\mathbf{v_k}). \tag{2}$$

Measurements $\mathbf{y_k}$ are generated at each instant with a probability distribution dependent only on $\mathbf{x_k}$,

$$f_{\mathbf{Y_k}|\mathbf{X_k}}(\mathbf{y_k}|\mathbf{x_k}) = f_M(\mathbf{y_k}, \mathbf{x_k}). \tag{3}$$

The matrices Φ and Γ as well as the probability densities f_N and f_M, though not explicitly indicated, may be functions of time. Together they define the process model.

Example: The Event Based Accelerometer

Consider an accelerometer in one dimension based on the following design. (See figure 2) A test mass is suspended freely as to move between two walls placed at ± 1. Let x_1 be the position and x_2 the velocity. The mass is subject to an unknown external acceleration v (to be estimated) and a control acceleration u. The external acceleration is modelled as (discrete time) white noise for simplicity.

The time evolution of the process is described by

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) &= u(t) + v(t). \end{aligned}$$

The output $y(t)$ is a function of position:

$$y(t) = \begin{cases} 1, & x_1(t) \geq 1 \\ 0, & |x_1(t)| < 1 \\ -1, & x_1(t) \leq -1. \end{cases}$$

This is the only non-classical property of the system.

It is not allowable to let the test mass pass either of the walls. Therefore it is assumed that any event, or nonzero reading of y, triggers a control response that tries to restore x_1 to the origin. The objective is to stabilize the process and estimate the external acceleration $v(t)$.

Similar process models describe other simple event based systems. The common case of measuring rotation angle with an encoder that generates events at regularly spaced angles would for instance benefit from much the same analysis as will be applied here.

To obtain a discrete time process model the model has to be sampled, which can be done in many ways. For a truly event based design one might make sure as to bracket each change of y with two closely placed sampling instants, otherwise maintaining a constant sample rate. For simplicity, the process will instead be sampled with constant sample rate h. Furthermore, it is assumed that h is so small that second order terms can be neglected.

Sampling yields the dynamics

$$\mathbf{x_k} = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \mathbf{x_{k-1}} + \begin{pmatrix} 0 \\ h \end{pmatrix} u_{k-1} + \begin{pmatrix} 0 \\ v_{k-1} \end{pmatrix} = \Phi \mathbf{x_{k-1}} + \Gamma u_{k-1} + \mathbf{v_{k-1}},$$

where

$$\mathbf{x_k} = \begin{pmatrix} x_{k,1} \\ x_{k,2} \end{pmatrix},$$

Φ, Γ and $\mathbf{v_k}$ can be read out from the equation and $V_k \in N(0, \sigma\sqrt{h})$ or equivalently

$$f_{V_k}(v_k) = f_{N_0}(v_k) = \frac{1}{\sigma\sqrt{2\pi h}} e^{-\frac{v_k^2}{2\sigma^2 h}},$$

where σ describes the strength of the process noise.

The measurements y_k are assumed to be available only at the sampling instants. The measurement function is

$$f_M(y_k, \mathbf{x_k}) = \begin{cases} \mathbf{I}_{(-\infty,-1]}(x_{k,1}), & y_k = 1 \\ \mathbf{I}_{(-1,1)}(x_{k,1}), & y_k = 0 \\ \mathbf{I}_{[1,\infty)}(x_{k,1}), & y_k = -1, \end{cases}$$

where $\mathbf{I}_I(x)$ is the indicator function for the interval I, taking the value 1 if $x \in I$ and 0 otherwise.

The Bayesian Observer

The optimal or Bayesian Observer for the chosen process model will now be derived by finding the conditional probability density of the process state given available measurements.

Since only online observers are considered, the information available will be the initial probability density $f_{\mathbf{X_0}}$ and the set of measurements up to the current time $\mathbf{y_{1:k}}$. The state of the observer is the probability density

$$f_k(\mathbf{x_k}) = f_{\mathbf{X_k}|\mathbf{y_{1:k}}, f_{\mathbf{X_0}}}(\mathbf{x_k}).$$

Since $\mathbf{X}$ is a Markov process, this state combined with the current input $\mathbf{u_k}$ and next measurement $\mathbf{y_{k+1}}$ is enough to calculate the a posteriori probability density f_{k+1}.

The initial state of the observer is given by

$$f_0(\mathbf{x_0}) = f_{\mathbf{X_0}}(\mathbf{x_0}).$$

The time evolution of the observer consists of three update steps: dynamics, noise and measurement. In the first two updates the process and disturbance models are used to predict the next state of the process, and in the third update the current measurement is used to refine the prediction.

The dynamics update computes the probability density

$$f_k^0(\mathbf{x_k}) = f_{\mathbf{X_k}|\mathbf{v_k}=0,\,\mathbf{y_{1:k-1}},\,f_{\mathbf{X_0}}}(\mathbf{x_k})$$

from f_{k-1}, that is, the probability density at instant k given the measurements up to $k-1$ and without considering process noise for the last step. Letting

$$\mathbf{z} = \mathbf{x_k} - \mathbf{v_{k-1}} = \Phi\mathbf{x_{k-1}} + \Gamma\mathbf{u_{k-1}}$$

and assuming Φ is nonsingular one finds that

$$\mathbf{x_{k-1}} = \Phi^{-1}(\mathbf{z} - \Gamma\mathbf{u_{k-1}})$$

and

$$dx_{k-1} = \det(\Phi)^{-1}\,dz,$$

so that

$$f_k^0(\mathbf{z}) = \det(\Phi)^{-1} \cdot f_{k-1}(\Phi^{-1}\mathbf{z} - \Phi^{-1}\Gamma\mathbf{u_{k-1}}),$$

where the factor $\det(\Phi)^{-1}$ is needed to preserve unit total probability. Apparently, the dynamics update corresponds to an affine transformation of the observer state.

The assumption of nonsingular Φ corresponds to the process having no deadbeat dynamics. The special case of deadbeat dynamics could be incorporated at the price of lengthier mathematical treatment, but this will not be done here.

The noise update computes the probability density

$$f_k^1(\mathbf{x_k}) = f_{\mathbf{X_k}|\mathbf{y_{1:k-1}},f_{\mathbf{X_0}}}(\mathbf{x_k})$$

from f_k^0, that is, the probability density at instant k given the measurements up to $k-1$. This can be done using the Chapman-Kolmogorov equation

$$f_k^1(\mathbf{x_k}) = \int f_{\mathbf{X_k}|\mathbf{Z}}(\mathbf{x_k}|\mathbf{z}) \cdot f_k^0(\mathbf{z})\,dz. \tag{4}$$

From (1) and (2) it is seen that the transition probability is described by

$$f_{\mathbf{X_k}|\mathbf{Z}}(\mathbf{x_k}|\mathbf{z}) = f_N(\mathbf{v_{k-1}}) = f_N(\mathbf{x_k} - \mathbf{z}),$$

which, inserted into (4) yields

$$f_k^1(\mathbf{x_k}) = \int f_N(\mathbf{x_k} - \mathbf{z})\, f_k^0(\mathbf{z})\,dz = \left(f_N * f_k^0\right)(\mathbf{x_k}),$$

the convolution of f_N and f_k^0.

The measurement update computes the probability density f_k from f_k^1, which can be done using Bayes' rule:

$$f_k(\mathbf{x_k}) = f_{\mathbf{X_k}|\mathbf{y_{1:k}},f_{\mathbf{X_0}}}(\mathbf{x_k}) = \frac{f_{\mathbf{Y_k}|\mathbf{X_k}}(\mathbf{y_k}|\mathbf{x_k}) \cdot f_{\mathbf{X_k}|\mathbf{y_{1:k-1}},f_{\mathbf{X_0}}}(\mathbf{x_k})}{f_{\mathbf{Y_k}|\mathbf{y_{1:k-1}}}(\mathbf{y_k})}.$$

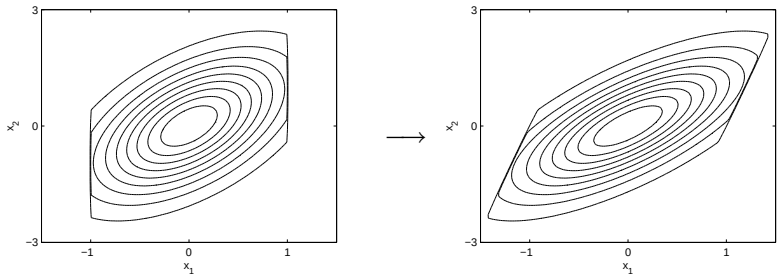

Figure 3 Illustration of the dynamics update of the Bayesian Observer for the Accelerometer: probability densities before and after. The affine transformation amounts to a shear in this case.

Identification of the probability densities yields

$$f_k(\mathbf{x_k}) = C^{-1} f_M(\mathbf{y_k}, \mathbf{x_k}) \cdot f_k^1(\mathbf{x_k}),$$

where $C = f_{\mathbf{Y_k}|\mathbf{y_{1:k-1}}}(\mathbf{y_k})$ is a positive constant that must be chosen so as to normalize f_k. The case when normalization can not be performed because $C = 0$ corresponds to that the observer has received contradictory information.

To summarize, the state of the Bayesian Observer is

$$f_k(\mathbf{x_k}) = f_{\mathbf{X_k}|\mathbf{y_{1:k}}, f_{\mathbf{X_0}}}(\mathbf{x_k}),$$

which is the probability density of the state of the process over its state space. The initial state is

$$f_0(\mathbf{x_0}) = f_{\mathbf{X_0}}(\mathbf{x_0}),$$

and the dynamics are described by the equations

$$\begin{aligned} f_k^0(\mathbf{z}) &= C_1 \cdot f_{k-1}(\Phi^{-1}\mathbf{z} - \Phi^{-1}\Gamma\mathbf{u_{k-1}}), && (5)\\ f_k^1(\mathbf{x_k}) &= \left(f_N * f_k^0\right)(\mathbf{x_k}), && (6)\\ f_k(\mathbf{x_k}) &= C_2 f_M(\mathbf{y_k}, \mathbf{x_k}) \cdot f_k^1(\mathbf{x_k}), && (7) \end{aligned}$$

where $C_1, C_2 > 0$ are constants that should be picked so as to normalize each step. The update is composed of an affine transformation for process dynamics, a convolution for process noise and a multiply for measurements. These three mathematical operations will play a major role in the forthcoming analysis.

Example: An observer for the Accelerometer

Here the Bayesian Observer for the Accelerometer will be derived. It will be simplified somewhat for easier treatment.

The initial state of the observer is assumed to be known. The determinant of Φ is one and

$$\Phi^{-1} = \begin{pmatrix} 1 & -h \\ 0 & 1 \end{pmatrix},$$

so that the dynamics update of the observer becomes simply

$$f_k^0(\mathbf{z}) = f_{k-1}(\mathbf{z_t}), \qquad (8)$$

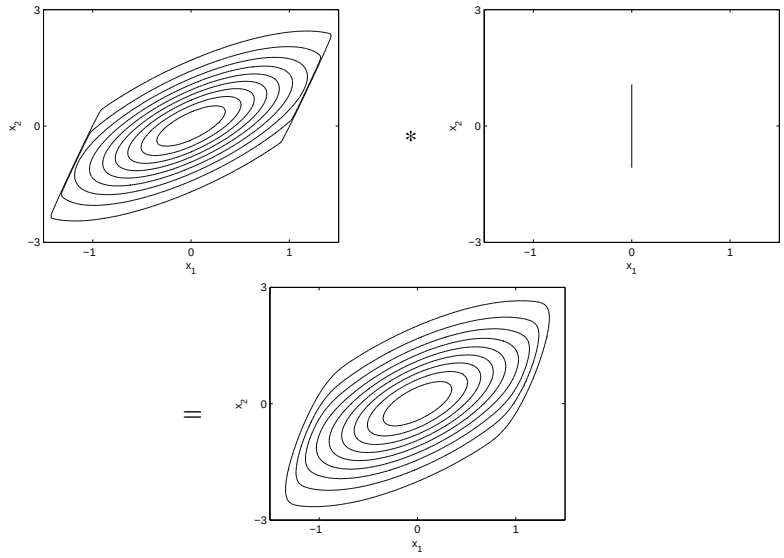

Figure 4 Illustration of the noise update of the Bayesian Observer for the Accelerometer: probability density before, noise density, and density after. The noise in this case only has a component in the x_2 direction, so its density is flat in the x_1 direction. The actual profile along x_2 is Gaussian. The effect is that of blurring the incoming density in the x_2 direction.

where

$$\mathbf{z_t} = \Phi^{-1}\mathbf{z} - \Phi^{-1}\Gamma\mathbf{u_{k-1}} = \begin{pmatrix} z_1 - hz_2 \\ z_2 \end{pmatrix} - \begin{pmatrix} -h^2 \\ h \end{pmatrix} u_{k-1}, \qquad (9)$$

which is a shear and translate, see figure 3.

The noise update is a convolution with the noise probability density

$$f_k^1(\mathbf{x_k}) = \left(f_N * f_k^0\right)(\mathbf{x_k}),$$

where

$$f_N(\mathbf{v_k}) = \delta(v_{k,1}) f_{N_0}(v_{k,2}),$$

and the delta function is needed since the process noise is assumed to affect only x_2, see figure 4.

When $y_k = \pm 1$, h is assumed to be so small that the crossing can be assumed to have taken place at $t = t_k$. To account for some measurement uncertainty, it is assumed that

$$x_{k,1} = y_k + m_k$$

where m_k is measurement noise, $M_k \in N(0, \sigma_m)$ with σ_m specifying the uncertainty. The measurement update will change to

$$f_k(\mathbf{x_k}) = \begin{cases} Ce^{-\frac{(x_{k,1}-y_k)^2}{2\sigma_m^2}} \cdot f_k^1(\mathbf{x_k}), & y_k \in \{-1, 1\}, \\ C\mathbf{I}_{(-1,1)}(x_{k,1}) \cdot f_k^1(\mathbf{x_k}), & y_k = 0 \end{cases}$$

where C is chosen so that f_k is normalized, see figure 5.

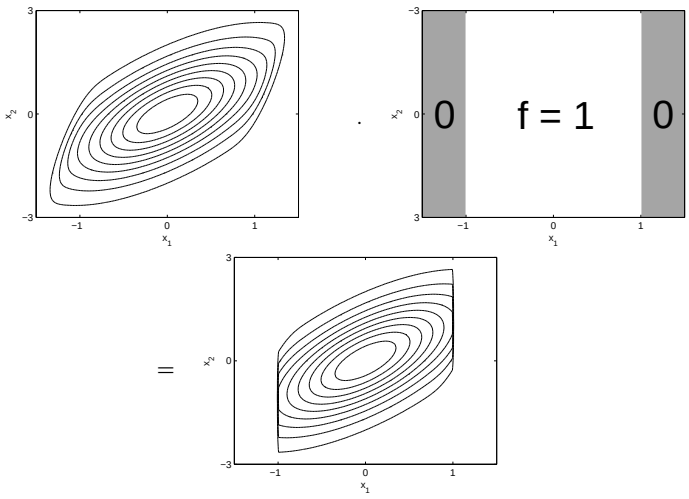

Figure 5 Illustration of the measurement update of the Bayesian Observer for the Accelerometer: probability density before, noise density, and density after. The current measurement $y = 0$ corresponds to multiplying with a window function that is 1 when $|x_1| < 1$ and 0 otherwise.

The Kalman Filter

When the process noise and measurements in the Bayesian Observer derived above are Gaussian, meaning that $f_N(\mathbf{v_k})$ is Gaussian and that $f_M(\mathbf{y_k}, \mathbf{x_k})$ is a Gaussian function of $\mathbf{x_k}$ for each $\mathbf{y_k}$, the Kalman Filter is obtained. The results on the Kalman Filter are not new; they are stated here for easy comparison.

Recall that a general Gaussian function can be written as

$$f(\mathbf{x}) = Ce^{-\frac{1}{2}(\mathbf{x}-\mathbf{m_x})^T Q(\mathbf{x}-\mathbf{m_x})},$$

where $\mathbf{x} \in \mathbb{R}^n$, $C \in \mathbb{R}$ describes the magnitude, $Q \in \mathbb{R}^{n\times n}$ is ≥ 0 and describes the spread, and $\mathbf{m_x} \in \mathbb{R}^n$ describes the point of symmetry.

Taking $Q = P^{-1}$, with $P > 0$, and $C = (2\pi)^{-n/2}\big(\det(P)\big)^{-1/2}$, a general Gaussian probability density is obtained. Letting

$$\varphi_P(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(P)}} e^{-\frac{1}{2}\mathbf{x}^T P^{-1}\mathbf{x}},$$

any Gaussian probability density can be written as $f(\mathbf{x}) = \varphi_P(\mathbf{x}-\mathbf{m_x})$ for some $\mathbf{m_x}$ and positive definite P.

Restricting observer states, process noise and measurement functions to be Gaussian simplifies many things. For one,

$$\begin{aligned} \mathrm{Cov}_{\mathbf{x}}\big(\varphi_P(\mathbf{x}-\mathbf{m_x})\big) &= P, \\ \mathrm{E}_{\mathbf{x}}\big(\varphi_P(\mathbf{x}-\mathbf{m_x})\big) &= \mathbf{m_x}, \end{aligned}$$

so the most important statistical measures are readily available. Furthermore, the set of Gaussian functions is closed under the operations of affine transformation, convolution and multiplication (when they are defined), which for Gaussian probability densities with $\mathbf{m_x} = 0$ amounts to

$$\det(A)\varphi_F(A\mathbf{x}+\mathbf{b}) = \varphi_{A^{-1}FA^{-T}}(\mathbf{x}), \tag{10}$$

$$(\varphi_F * \varphi_G)(\mathbf{x}) = \varphi_{F+G}(\mathbf{x}), \tag{11}$$

$$C\varphi_F(\mathbf{x}) \cdot \varphi_G(\mathbf{x}) = \varphi_{(F^{-1}+G^{-1})^{-1}}(\mathbf{x}), \tag{12}$$

where $C > 0$, $A \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$, and $A^{-T} = (A^{-1})^T = (A^T)^{-1}$. The covariances are calculated in the same way when $\mathbf{m_x} \neq 0$, and the composition rules for $\mathbf{m_x}$ are also simple.

The above formulas imply that the Kalman Filter is much easier both to implement and analyze than the general Bayesian Observer. If the initial state of the filter is a Gaussian density, so are all future states, and since a Gaussian density is completely described by its expectation and covariance, the state dimension of the Kalman Filter need not be more than $n + \frac{n(n+1)}{2}$, compared to the general Bayesian Observer that usually has infinite state dimension. If furthermore $\mathrm{Cov}_{\mathbf{x_k}}(f_M(\mathbf{y_k}, \mathbf{x_k}))$ is independent of $\mathbf{y_k}$ then the estimate covariance P can often be assumed to be stationary. In this case the Kalman Filter will be a linear time-invariant system with state dimension n.

A further advantage of using a Kalman filter is the great amount of knowledge that is already available on the subject.

Example: A conservative Kalman Filter approximation

The actual Bayesian Observer for the Accelerometer is not a Kalman Filter, because the measurements are not Gaussian. Here a Kalman Filter that approximates the Bayesian Observer will be derived from the approximate Bayesian Observer. It will be referred to as the Conservative Kalman Filter for reasons that will become clear later.

There is no easy way to approximate the rectangular window measurement function for $y = 0$ with a Gaussian function; in fact any approximation should probably take account of the interactions in the observer as a whole. A simple approximation which will turn out to have a special significance is to let $f_M(0, \mathbf{x}) = 1$, while keeping f_M for $y = \pm 1$, which happens to be Gaussian. This corresponds to disregarding the information contained in the measurement $y = 0$.

Let

$$P_k = P = \begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix}$$

be the covariance matrix of the current Gaussian observer state. Consider first the case when $y_k = 0$. Then only dynamics and process noise have to be considered. The dynamics update according to (5) consists of an affine transformation with transformation matrix $A = \Phi^{-1}$, and from (10) this corresponds to an update $P_1 = A^{-1}PA^{-T}$ so that

$$P_1 = \Phi P \Phi^T = \begin{pmatrix} p_{11} + 2hp_{12} + h^2 p_{22} & p_{12} + hp_{22} \\ p_{12} + hp_{22} & p_{22} \end{pmatrix}$$

gives the covariance after the dynamics update.

The noise update according to (6) consists of a convolution with the process noise probability density, which is Gaussian and has covariance

$$P_N = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 h \end{pmatrix}.$$

According to (11) this yields a new covariance matrix

$$P_{k+1} = P_2 = P_1 + P_N = \begin{pmatrix} p_{11} + 2hp_{12} + h^2 p_{22} & p_{12} + hp_{22} \\ p_{12} + hp_{22} & p_{22} + h\sigma^2 \end{pmatrix}.$$

This difference equation has an explicit solution; to simplify matters somewhat the equation will be solved only for the case $h \to 0$, yielding the differential equation

$$\dot{P}(t) = \begin{pmatrix} 2p_{12}(t) & p_{22}(t) \\ p_{22}(t) & \sigma^2 \end{pmatrix}.$$

The system is triangular and some calculations yield

$$P(t) = \begin{pmatrix} \frac{1}{3}\sigma^2 t^3 + p_{22}^0 t^2 + 2p_{12}^0 t & \frac{1}{2}\sigma^2 t^2 + p_{22}^0 t \\ \frac{1}{2}\sigma^2 t^2 + p_{22}^0 t & \sigma^2 t \end{pmatrix} + P(0). \qquad (13)$$

The covariance will first grow in the x_2 direction (linearly) corresponding to process noise and gradually shear and grow faster and faster in the x_1 direction (cubically) corresponding to uncertainties in velocity. This formula should give a good approximation of the observer state when $y = 0$ as long as the bulk of the probability mass stays between the borders $x_1 = \pm 1$.

The case when $y = \pm 1$ yields an instantaneous update, from (7) and (12) this becomes

$$P^+(t) = \left(\left(P^-(t)\right)^{-1} + Q_m\right)^{-1} = \left(I + P^-(t)Q_m\right)^{-1} P^-(t),$$

where

$$Q_m = \begin{pmatrix} \sigma_m^{-2} & 0 \\ 0 & 0 \end{pmatrix}$$

is the inverse covariance matrix for the measurement noise (as a limit when the noise along x_2 tends to infinity), $P^-(t)$ is the covariance just before the event, and $P^+(t)$ is the covariance just after.

Since everything is Gaussian, standard statistical theory can be easily applied. Consider the case of estimating the velocity x_2 from only the information in the two latest events. Let

$$\begin{aligned} Y_0 &= X_1(0) + M_0, \\ Y_1 &= X_1(t) + M_1, \end{aligned}$$

be the measurements obtained from events at time 0 and t respectively, where $M_i \in N(0, \sigma_m)$ is the measurement noise. Since X_2 is the velocity of X_1,

$$\begin{aligned} X_1(t) &= X_1(0) + \int_0^t X_2(\tau)\, d\tau = \\ &= X_1(0) + tX_2(t) + \int_0^t \left(X_2(\tau) - X_2(t)\right) d\tau = \\ &= X_1(0) + tX_2(t) + D, \end{aligned}$$

where D is the effect of the process noise plus any control acceleration acting between time 0 and t. Since the effect of control acceleration is purely deterministic, only the case $u = 0$ will be considered; the more general case can be obtained by superposition of the deterministic response to u and will not affect the estimation errors. Thus, using the fact that the process is reversible in time, D is equal to the variance of $X_1(t)$ starting from known initial conditions ($P(0) = 0$), and (13) yields

$$V(D) = \frac{1}{3}\sigma^2 t^3.$$

The velocity $X_2(t)$ can now be estimated by

$$\hat{X}_2(t) = \frac{1}{t}(Y_1 - Y_0),$$

where the estimation error is given by

$$\tilde{X}_2(t) = X_2(t) - \hat{X}_2(t) = -\frac{1}{t}(D + M_1 - M_0).$$

Since D, M_0 and M_1 are independent,

$$\begin{aligned} V(\tilde{X}_2(t)) &= \frac{1}{t^2}(V(D) + V(M_0) + V(M_1)) = \\ &= \frac{1}{t^2}\left(\frac{1}{3}\sigma^2 t^3 + 2\sigma_m^2\right) = \\ &= \frac{1}{3}\sigma^2 t + 2\sigma_m^2 t^{-2}. \end{aligned} \tag{14}$$

Since t is the time between events, this expression hints at the fact that at least up to a point set by the measurement error, state estimates will be more accurate with a shorter time between events. Thus the best control strategy is probably not to keep the state as close to zero as possible if the objective is to obtain accurate estimates. The apparent deterioration in estimate accuracy for small enough t might be alleviated by using information from older measurements.

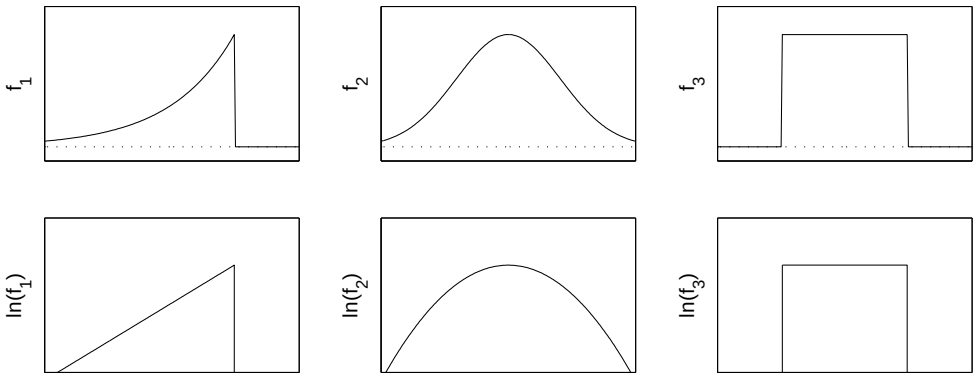

Figure 6 Some examples of log-concave functions in one dimension; the logarithm of each function is shown just below its graph. The dotted line corresponds to $f = 0$. $f_1(x) = e^x\theta(1-x)$: (truncated) exponential function (ordinary exponential is log-concave but not normalizable), $f_2(x) = e^{-x^2}$: Gaussian function, $f_3(x) = \theta(1+x)\theta(1-x)$: rectangular window. $\theta(x)$ is the Heaviside step function.

Log-concavity

While the assumptions of the Kalman Filter makes the analysis of the Bayesian Observer quite straightforward, it is not so easy to analyze the general case. For the case of log-concave disturbances and measurements, while being more general than the Gaussian case, many of the appealing properties of the Kalman Filter do however turn out to be preserved. First log-concavity will be introduced from a mathematical point of view.

Log-concave functions have found a use in many different areas, among them applied probability theory. Here we will mostly be concerned with log-concave probability densities and normalizable log-concave functions.

Log-concave functions

The results on general log-concavity are known since earlier, see for instance [3], [4], [5], and [6] for these results and many other. [7]

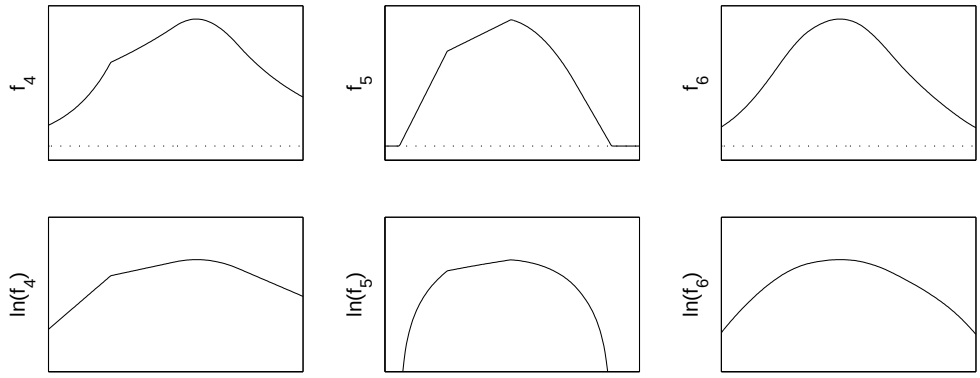

Figure 7 Some more examples of log-concave functions in one dimension; the logarithm of each function is shown just below its graph. The dotted line corresponds to $f = 0$. $f_4(x) = e^{g(x)}$, where g is concave and piecewise quadratic: exponential of concave function is log-concave, $f_5(x) = \max(g(x), 0)$: nonnegative function that is concave on a convex support is log-concave, f_6: randomly generated Gaussian-like function.

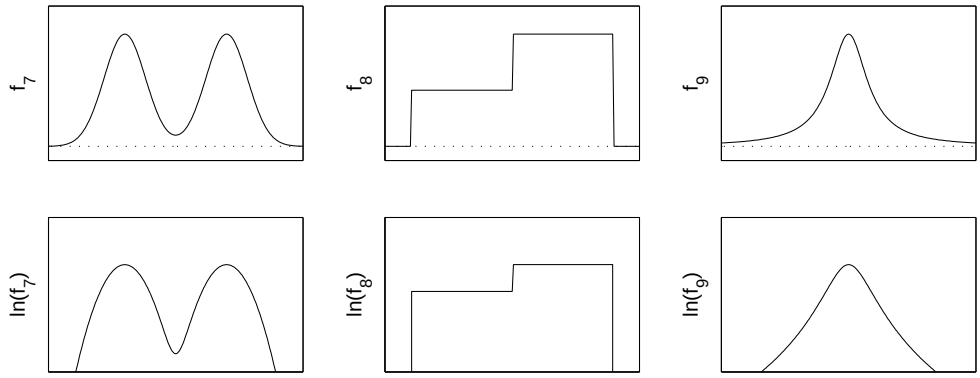

Figure 8 Some examples of *non-log-concave* functions in one dimension; the logarithm of each function is shown just below its graph. The dotted line corresponds to $f = 0$. $f_7(x) = e^{-3(x-1)^2} + e^{-3(x+1)^2}$: not unimodal, $f_8(x) = \theta(x+1) + \theta(x) - 2\theta(x-1)$: discontinuous outside boundary of support, $f_9(x) = \frac{1}{1+x^2}$: does not decay fast enough (too fat tails).

contains some chapters about convex and concave functions in general. Here, the results that are most useful in the current context will be summarized.

DEFINITION 1—LOG-CONCAVE FUNCTION
A function $f : \mathbb{R}^n \to \mathbb{R}$ is *log-concave* iff $f(\mathbf{x}) \geq 0$, f has convex support and $\ln\big(f(\mathbf{x})\big)$ is concave on the support of f. We denote this $f \in \mathcal{LC}$. If strict concavity holds for $\ln\big(f(\mathbf{x})\big)$ on the entire support of f then f($\mathbf{x}$) is *strictly log-concave*. □

Thus, a log-concave function is essentially any function f that can be expressed as $f(\mathbf{x}) = e^{-g(\mathbf{x})}$ where $g(\mathbf{x})$ is convex and may assume the value $+\infty$ as long as the set on which $g(\mathbf{x})$ is non-infinite is convex. For some examples of log-concave functions in one dimension, see figures 6 and 7. Figure 8 gives some counterexamples. For an excellent account of how to construct and recognize convex functions (which can easily be translated to the log-concave case), see [7];

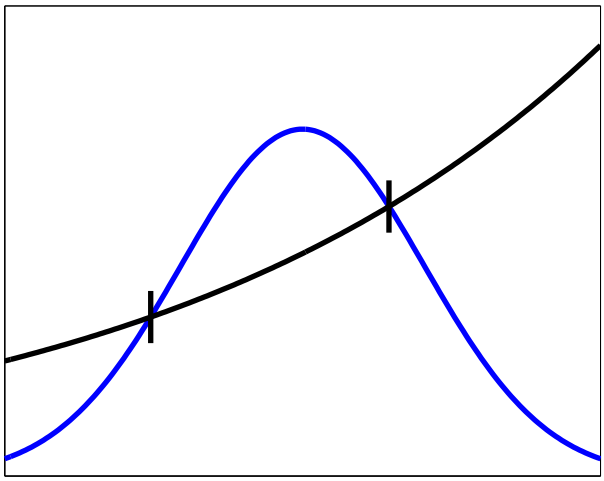

Figure 9 Illustration of the interpolating curve criterion for log-concavity. A function is log-concave iff it is nonnegative and for any exponential function drawn between two points on the graph (if one exists), the graph does not lie below the curve between the points. This also implies that the graph does not lie above the curve on the rest of the interval.

It could be argued that log-concavity is a much more reasonable form of convexity when concerned with probability densities than is either convexity or concavity; all non-constant convex and concave functions on $\mathbb{R}^n$ are unbounded. Furthermore log-concavity of a probability density makes the log-likelihood function concave, making for instance maximum likelihood estimation into a convex problem.

Just as for convex functions, there are several other ways to express the conditions for log-concavity:

- The function $f(\mathbf{x})$ is log-concave if and only if $f(\mathbf{x}) \geq 0$ and

$$f\left((1-\lambda)\mathbf{x_1} + \lambda\mathbf{x_2}\right) \geq f(\mathbf{x_1})^{1-\lambda} f(\mathbf{x_2})^{\lambda}, \lambda \in [0,1]$$

  for all $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^n$. If the inequality is strict when $\mathbf{x_1} \neq \mathbf{x_2}$ and $\lambda \in (0,1)$ then $f(\mathbf{x})$ is strictly log-concave. See figure 9.

- If $f(\mathbf{x}) \geq 0$ and f is twice differentiable with respect to $\mathbf{x}$, then f is log-concave if and only if

$$\nabla^2 \ln\big(f(\mathbf{x})\big) \leq 0,$$

  for all $\mathbf{x}$, where ∇^2 denotes the Hessian or second derivative matrix. If

$$\nabla^2 \ln\big(f(\mathbf{x})\big) < 0$$

  then f is strictly log-concave, but this is only a sufficient condition.

Many commonly used probability densities are log-concave, for instance exponential distributions and normal (Gaussian) distributions. For a more complete list, see [5].

Log-concave functions have much in common with convex functions. A function is log-concave if and only if it is log-concave on each line through its domain. Many useful properties of log-concave functions can be derived from the properties of convex functions, among them:

- Continuity. If f is log-concave then it is continuous on the interior of its support, and almost everywhere differentiable. It may however be discontinuous at the support boundary.
- Unimodality. Any log-concave function f is quasiconcave, meaning that all superlevel sets

  $$S_\alpha = \{\mathbf{x}; f(\mathbf{x}) \geq \alpha\}$$

  are convex. If the sets S_α are closed and bounded for all $\alpha \geq \alpha_0$, for some α_0 such that S_{α_0} is nonempty, then f assumes its maximum value on a convex set. If f is also strictly log-concave, it assumes its maximum value at a single point.

  The set of log-concave functions does in fact (at least in one dimension) coincide with the set of strongly unimodal functions; the set of functions that when convoluted with an arbitrary unimodal function remain unimodal.
- Global information available locally. If g($\mathbf{x}$) is a (multivariate) exponential function

  $$g(\mathbf{x}) = ae^{\mathbf{b}^T\mathbf{x}}$$

  such that $g(\mathbf{x}_0) = f(\mathbf{x}_0)$ and $g(\mathbf{x}) \geq f(\mathbf{x})$ in a small neighborhood of $\mathbf{x}_0$, then $g(\mathbf{x}) \geq f(\mathbf{x})$ for all $\mathbf{x}$, that is $g(\mathbf{x})$ is a global overestimator of $f(\mathbf{x})$.

Just like the plane is an extreme among convex functions in the sense that it is the only function that is both convex and concave, the exponential function is extreme among log-concave functions in the sense that it is both log-concave and log-convex. In many respects, any log-concave function is at least as well behaved as an exponential function.

Apart from what can be inferred from the properties of convex functions, log-concave functions have many other useful properties. The following theorem will be central in the forthcoming analysis.

THEOREM 1—PRÉKOPA

Let $f(\mathbf{x}, \mathbf{y})$ be a log-concave function of $\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^n$. Then the integral

$$g(\mathbf{x}) = \int f(\mathbf{x}, \mathbf{y})\, dy$$

is a log-concave function of $\mathbf{x}$. □

Proof. See [3] and [4].

Log-concavity is preserved under many operations:

- Multiplication. If f and g are log-concave then $f(\mathbf{x}) \cdot g(\mathbf{x})$ is log-concave. This can be derived from the fact that the sum of two convex functions is convex.
- Affine transformation. If f is log-concave, A is a matrix and $\mathbf{b}$ a vector of appropriate dimensions then $f(A\mathbf{x} + \mathbf{b})$ is log-concave.
- Convolution. If f and g are log-concave then the convolution

  $$(f * g)(\mathbf{x}) = \int f(\mathbf{x} - \mathbf{y}) g(\mathbf{y})\, dy$$

  is also log-concave. This can be derived from theorem 1 and the multiplication and affine transformation properties.

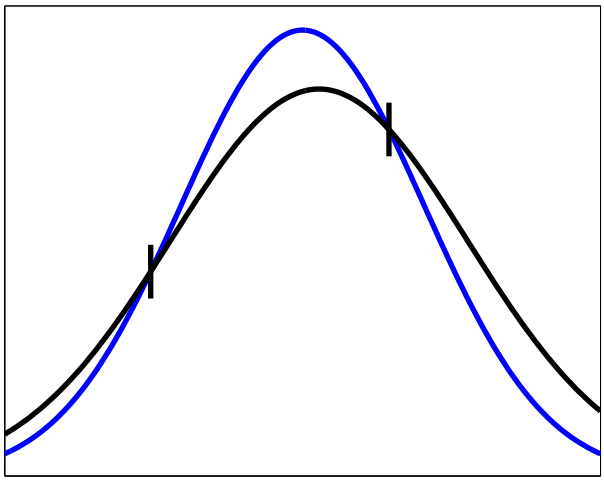

Figure 10 Illustration of the interpolating curve criterion for strong log-concavity. A function in one dimension is strongly log-concave of strength p iff it is nonnegative and for any Gaussian with covariance p drawn between two points on the graph (if one exists), the graph does not lie below the curve between the points. This also implies that the graph does not lie above the curve on the rest of the interval.

- Marginal densities. If f is a log-concave probability density then the marginal densities of f are log-concave. This is a direct consequence of theorem 1.

Log-concave functions in one dimensions have some properties that do not easily generalize to the multivariate case, for instance:

- Truncation. If $f(x), x \in \mathbb{R}$ is a log-concave probability density and $I \subset \mathbb{R}$ is a (possibly semi-infinite) interval, then the truncated probability density

$$g(x) = \begin{cases} \frac{f(x)}{\int_I f(x)\,dx}, & x \in I \\ 0 & \text{otherwise} \end{cases}$$

has a variance that is no greater than the variance of f. If f already has a support that is not infinite in both directions and truncation is used to cut away only a finite interval of the support, then the expectation can be shown to be unchanged or move away from the removed interval, a distance no longer than its length.

Strongly log-concave functions

It turns out that ordinary log-concavity is not a strong enough condition to make useful claims in the multivariate case, in particular there is need for a replacement for (or something better than) the truncation property. We therefore introduce the concept of strongly log-concave functions.

DEFINITION 2—STRONGLY LOG-CONCAVE FUNCTION
Let P be a positive definite matrix and define the class of functions

$$\mathcal{LC}(P) = \left\{ f; f(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{x}^T P^{-1}\mathbf{x}} f_0(\mathbf{x}), f_0 \in \mathcal{LC} \right\}.$$

Iff $f \in \mathcal{LC}(P)$ then f is *strongly log-concave* of strength P. □

That is, a strongly log-concave function is the product of a Gaussian function and a log-concave function; the covariance of the Gaussian determines the strength of log-concavity and shall be of prime interest in the following. Out of the example functions in figures 6 and 7, f_2, f_5, and possibly f_6 are strongly log-concave.

Any of the alternative conditions for log-concavity can be used to describe strong log-concavity, implying for instance that if f is positive and twice differentiable then $f \in \mathcal{LC}(P)$ is equivalent to

$$\nabla^2 \ln\big(f(\mathbf{x})\big) \leq -P^{-1}$$

for all $\mathbf{x}$, see also figure 10. Any strongly log-concave function f is strictly log-concave, and f goes to zero when $|\mathbf{x}| \to \infty$ at least as fast as the associated Gaussian. Also,

$$f \in \mathcal{LC}(P), P \leq P' \Longrightarrow f \in \mathcal{LC}(P').$$

If the limiting case of log-concavity is an exponential distribution, then the limiting case of strong log-concavity is a Gaussian distribution. We will now see that all functions in $\mathcal{LC}(P)$ have a great deal in common with Gaussian distributions with covariance P.

THEOREM 2—COVARIANCE BOUND
If $f \in \mathcal{LC}(P)$ is a probability density then

$$V = \mathrm{Cov}_x\big(f(\mathbf{x})\big) \leq P.$$

□

Thus any probability density in $\mathcal{LC}(P)$ is at least as well localized as the Gaussian density with covariance P. For the proof, see appendix B.

THEOREM 3—ENCAPSULATION PROPERTY
If $f \in \mathcal{LC}(F)$ and $g \in \mathcal{LC}(G)$ then

$$\begin{aligned} f(A\mathbf{x}+\mathbf{b}) &\in \mathcal{LC}(A^{-1}FA^{-T}) \\ (f*g)(\mathbf{x}) &\in \mathcal{LC}(F+G) \\ f(\mathbf{x})\cdot g(\mathbf{x}) &\in \mathcal{LC}\big((F^{-1}+G^{-1})^{-1}\big) \end{aligned}$$

where $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}$ and $f*g$ is the convolution of f and g. □

This compares well with the corresponding results for Gaussian distributions (10)–(12). Apparently, the covariance of strongly log-concave functions subject to these operations is bounded from above by the covariance of the corresponding Gaussian functions. No tighter bounds could be found under these assumptions since the Gaussian functions satisfy them exactly. Just as the set of Gaussian functions is closed under these operations, so is the set of log-concave functions. For the proof, see appendix B.

The log-concave Bayesian Observer

By a log-concave Bayesian Observer shall be meant the Bayesian Observer for a process with log-concave measurements and strongly log-concave disturbances, meaning that $f_M(\mathbf{y_k}, \mathbf{x_k})$ is a log-concave function of $\mathbf{x_k}$ for all $\mathbf{y_k}$ and that $f_N(\mathbf{v_k})$ is a strongly log-concave function of $\mathbf{v_k}$.

The Covariance Bound and Encapsulation Property theorems imply that the log-concave Bayesian Observer is much more well-behaved than the general case. In particular, if the original state of the observer is strongly log-concave, then so are all future states. This implies that the probability density over the process state will always be concentrated around a single point, which can simplify both controller and observer design. Although the observer state is still infinite dimensional, the strong regularity implied by strong log-concavity might allow for practical use of observers with quite low state dimension.

The theorems can also be used as a chain of inequalities to give an upper bound on the estimate covariance for the optimal observer. Such a bound can be interpreted as a certificate that there exists some observer that achieves the bound (or better). This could be useful for feasibility studies or as a benchmark.

With some additional results on the effects of measurements on the expectation of a log-concave probability density, the results may also be used to construct low-dimensional conservative observers for log-concave systems. For full usefulness, the theorems would have to be complemented, however. The mobility of the expectation of a probability density under the effect of measurements is directly related to its spread. Strong log-concavity in itself gives an upper bound on mobility but no lower bound.

Example: log-concavity and the Accelerometer

Since the process noise is Gaussian and the measurement function $f_M(y,x)$ is a log-concave function of x for each y, the Bayesian Observer for the Event Based Accelerometer is log-concave. This has many implications. In particular, the Conservative Kalman Filter approximation derived actually gives a conservative estimate of the observer covariance.

Since the estimate (14) is a conservative estimate based on the Conservative Kalman Filter, it is a conservative estimate of the velocity error for the Bayesian Observer just after an event.

Simulations

To test the theory and gain some insight into observer design for log-concave systems, computer simulations were made of the Bayesian Observer for the Event Based Accelerometer example. The process noise was chosen as $\sigma = 1$ in all simulations. Any change of σ would be equivalent to an appropriate rescaling of the t and x_2 axes.

Discretization

A grid-based filter is used to approximate the Bayesian Observer. A finite difference scheme is used to discretize the observer state $f_k(\mathbf{x_k})$ over the region $x_1 \in [-1,1], x_2 \in [-x_2^{max}, x_2^{max}]$, such that

$$p_k(i,j) = \Delta x_1 \Delta x_2 \cdot f_k(\Delta \mathbf{x_1} i + \Delta \mathbf{x_2} j),$$

where

$$\Delta \mathbf{x_1} = \begin{pmatrix} \Delta x_1 \\ 0 \end{pmatrix},$$
$$\Delta \mathbf{x_2} = \begin{pmatrix} 0 \\ \Delta x_2 \end{pmatrix},$$

Δx_1 and Δx_2 are discretization step sizes and $p_k(i,j)$ is assumed to be zero when $(\Delta \mathbf{x_1} i + \Delta \mathbf{x_2} j)$ is not within the region. It is assumed that x_2^{max} is large enough that the effects of disregarding states of greater velocity are negligible. Unit total probability means that

$$\sum_{i,j} p_k(i,j) = 1.$$

The three operations of affine transformation, convolution with noise probability density, and multiplication are discretized as follows. Assuming $u_k = 0$, the affine transformation update described by (8) and (9) is discretized as

$$p_k^0(i,j) = p_{k-1}(i,j) + \frac{h\Delta x_2}{\Delta x_1}\frac{j}{\Delta i}\big(p_{k-1}(i-\Delta i, j) - p_{k-1}(i,j)\big),$$

where $\Delta i = \pm 1$ is chosen so that $\frac{j}{\Delta i} \geq 0$ and it is assumed that h is chosen so small that $|\frac{h\Delta x_2}{\Delta x_1} j| \leq 1$ for all permissible j. This is a shear implemented by weighting the value of each element with its immediate upstream neighbor. Nonzero u_k is handled by integer displacement in the j direction.

Convolution with the noise probability density is discretized as

$$p_k^1(i,j) = (1-w)p_k^0(i,j) + \frac{w}{2}\Big(p_k^0(i,j-1) + p_k^0(i,j+1)\Big),$$

where $w = \frac{h\sigma^2}{\Delta x_2^2}$ and it is assumed that h is chosen so that $w \leq 1/2$. This approximates the Gaussian process noise with a three-point stochastic variable taking the values $\{-1, 0, 1\}$ and having the same incremental covariance.

The multiplication step for the case $y_k = 0$ is never actually performed, but is implicit in the fact that values of $p_k(i,j)$ outside the permitted region are treated as if they were zero. When $y_k = \pm 1$, all columns except the one on the proper border are set to zero, and the remaining column is multiplied by $\max(y_k j, 0)$ to approximate the probability mass flowing out during the affine transformation stage. The entire column is then displaced according to the assumed instant control action. Renormalization is carried out regularly.

Probability densities

Stationary state Simulations indicate that the observer is stable. When $y = 0$ it tends to a limit density as seen in figure 11. States with low velocity seem to be the most probable, and positive velocity seems to be correlated with positive position which is not surprising. The probability density has an overall appearance that comes quite close to a Gaussian, with the largest deviations occurring in the neighborhood of the points $x_1 = \pm 1, x_2 = 0$. All superlevel sets appear to be convex.

It is desirable to confirm whether this probability density is log-concave, and if so how strongly. Testing the (strength of) log-concavity for a function in one dimension is straightforward: given a function $f(x)$, find

$$q_f = \sup_q q; \frac{f(x)}{e^{-\frac{1}{2}qx^2}} \in \mathcal{LC}.$$

If the maximum exists and is non-negative, f is log-concave. If it is positive then $f \in \mathcal{LC}(p), p = q^{-1}$. The quantity $\sqrt{p}$ has the dimension

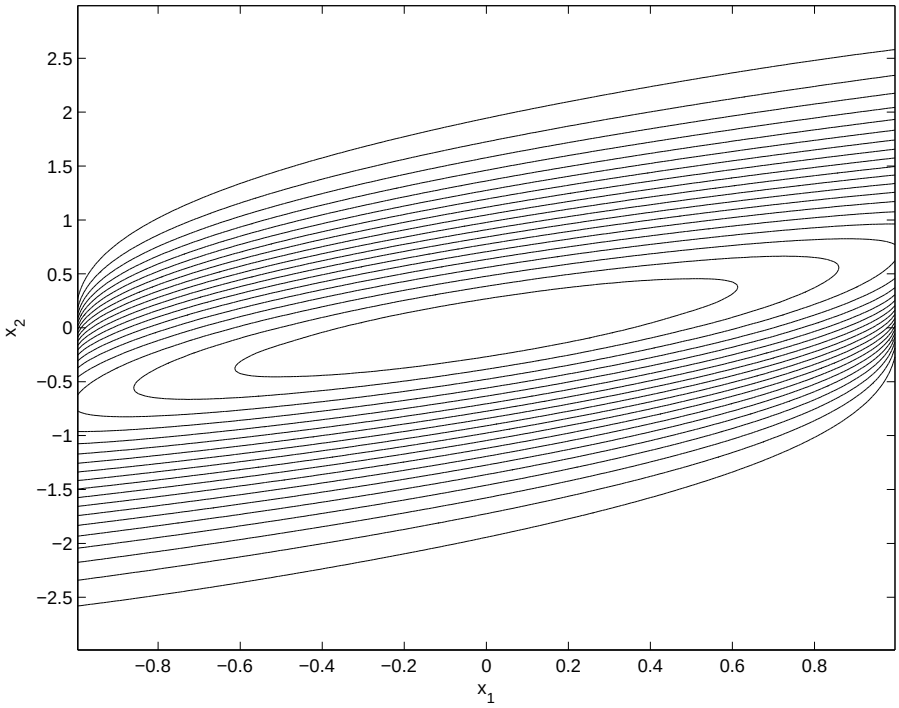

Figure 11 Stationary observer state for zero process output, $\sigma = 1$.

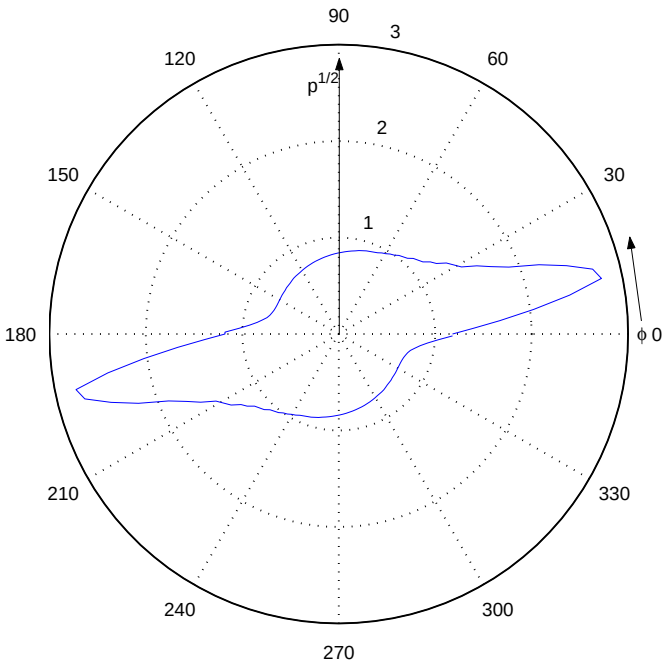

Figure 12 Estimated log-concavity strength as a function of angle for the stationary observer state. The square root of p is plotted since it has the dimension of length. The range is between $\sqrt{p_{min}} \approx 0.7$ and $\sqrt{p_{max}} \approx 2.8$, implying that the stationary state is strongly log-concave.

of length and is the standard deviation of the associated Gaussian. The test can be carried out numerically by finding a lower bound for a suitable approximation of $-\frac{d^2}{dx^2}\ln\big(f(x)\big)$.

Turning to higher dimensions, matters become a bit more complicated. Log-concavity of the function $f(\mathbf{x})$ can be tested by testing log-concavity of f on each line through its domain. There is however not guaranteed to be any single minimal P for which $f \in \mathcal{LC}(P)$; this is why strength of log-concavity has not been formulated as a function of f. The strength of log-concavity along each line does still give useful information, however.

In two dimensions one can look at $p(\phi)$, defined to be the maximum value of p found for all lines through f with angle ϕ, if one exists. If $p(\phi)$ exists for all ϕ has an upper bound p then f is strongly log-concave, for instance $f \in \mathcal{LC}(pI)$. A polar plot of $\sqrt{p(\phi)}$ for a Gaussian function yields an ellipse.

A plot of $\sqrt{p(\phi)}$ for the stationary observer state can be seen in figure 12. Apparently the stationary state is strongly log-concave, while only log-concavity was predicted from the theory. (All observer states are log-concave but strength of log-concavity may deteriorate without bounds with time.) The direction where $\sqrt{p(\phi)}$ spreads out the most coincides with the direction in which the inner level curves of the probability density are the most stretched.

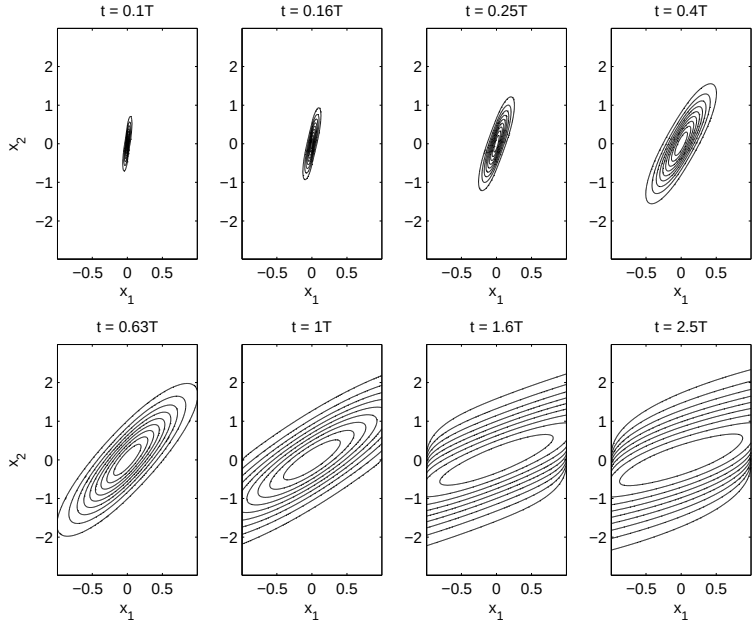

Figure 13 Time evolution of observer state starting from a Dirac pulse in the origin, with $y = 0$. The states appear to be approximately (truncated) Gaussian up to $t = T$, when the effects of the rectangular measurement window begin to be appreciable. Already at $t = 1.6T$, the state is very close to stationary. $T \approx 1.38$ is the estimated mean exit time in stationary state. The probability that no event occurs during an interval of length $2.5T$ is approximately 0.08 in this scenario.

Time evolution In figures 13 and 14, the time evolution of the observer state is plotted with two different initial conditions and no events. It is evident that the observer state converges quite quickly to the stationary state, which seems to be the same in both cases. In both

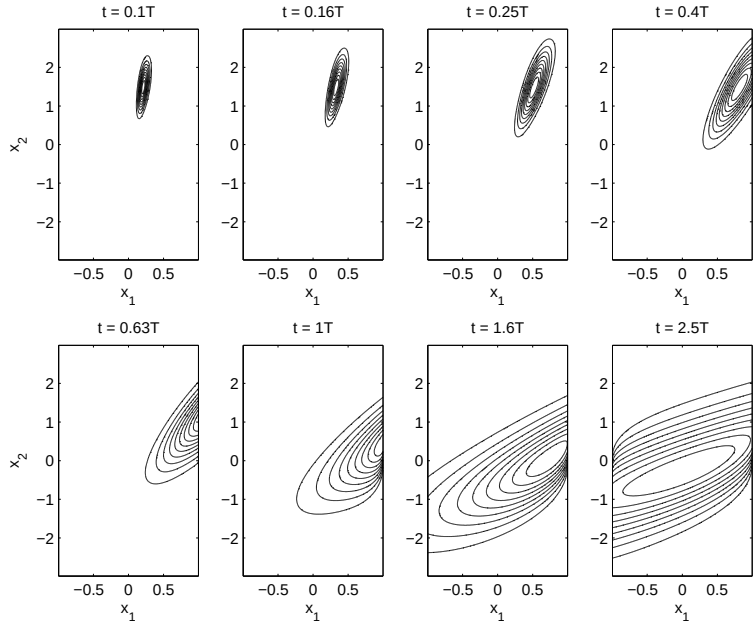

Figure 14 Time evolution of observer state starting from a Dirac pulse in $(x_1, x_2) = (0, 1.5)$, with $y = 0$. The states appear to be approximately (truncated) Gaussian up to $t = 0.63T$, when the effects of the rectangular measurement window begin to be appreciable. Already at $t = 2.5T$, the state is very close to stationary. $T \approx 1.38$ is the estimated mean exit time in stationary state.

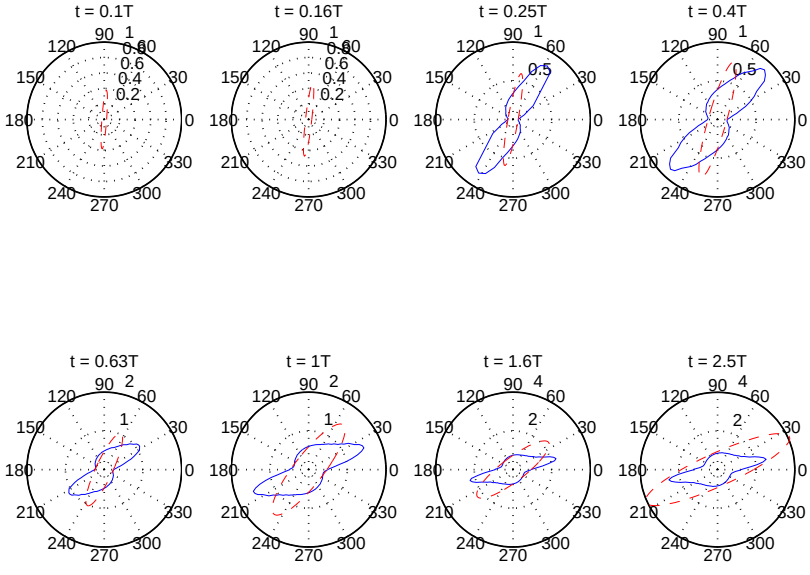

Figure 15 Estimated strength of log-concavity as a function of angle for the time evolution in figure 13. The dashed lines are Gaussian predictions calculated from (13) and are supposed to be conservative (enclosing the estimated curve); the fact that they in general are not might be explained by discretization errors in the simulation. The radial axis is $\sqrt{p}$. The first two observer states were too concentrated for a reliable estimate to be calculated.

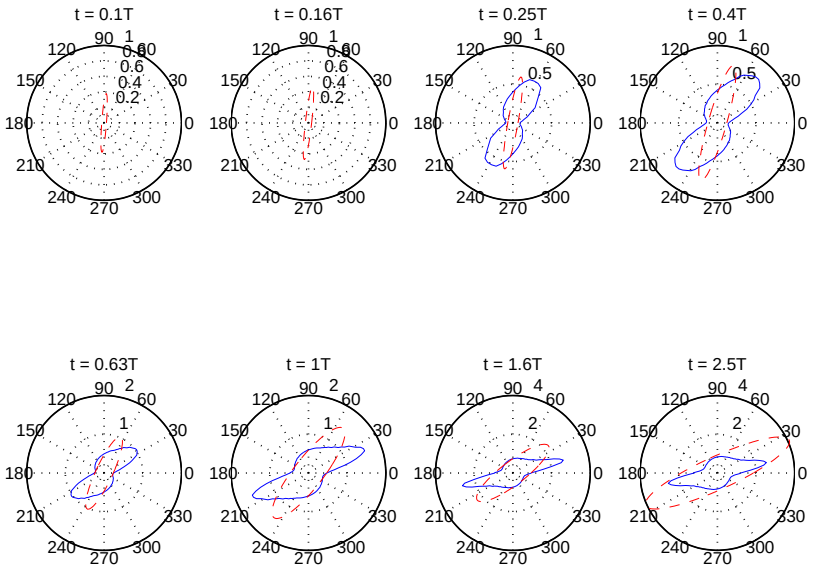

Figure 16 Estimated strength of log-concavity as a function of angle for the time evolution in figure 14. The difference to the log-concavity estimate in figure 15 is slight but noticeable.

cases the states seem to be approximately Gaussian until the effects of the rectangular measurement window start to become appreciable. At all times they are unimodal with convex superlevel sets.

It is worth noting that there is an intermediate phase when the state is actually truncated Gaussian. Not until an appreciable probability mass would have been flowing back through the borders, were they not there, does the state the within window $x_2 \in [-1, 1]$ begin to feel its effects and become noticeably non-Gaussian. Even then the probability densities seem to be quite well-behaved.

Figures 15 and 16 show log-concavity estimates for the two time series, combined with predictions calculated from the Conservative Kalman Filter using (13), which should give an upper bound on p.

Apparently the predictions do not give an upper bound, except for large t. What is more, the difference is big even long before the simulated states are affected by the rectangular measurement window and begin to deviate from the Gaussian. During this time the estimates should be accurate. The same predictions are, however, very accurate for the state covariances of the simulated observer.

The discrepancy could be the effect of discretization errors, indicating that explicit log-concavity is probably more fragile to discretization than is state covariance.

A non-log-concave example To illustrate the effects of a *non-log-concave* measurement situation, the measurement was modified to be

$$y = \begin{cases} 1, & x_1 \geq 1, \\ -1, & x_1 \leq -1, \\ 0, & |x_1| < 1/3, \\ 2, & \text{otherwise}, \end{cases}$$

where the difference is that y is 2 instead of 0 when $1/3 \leq |x_1| < 1$. This means that it is known when the position is between the walls

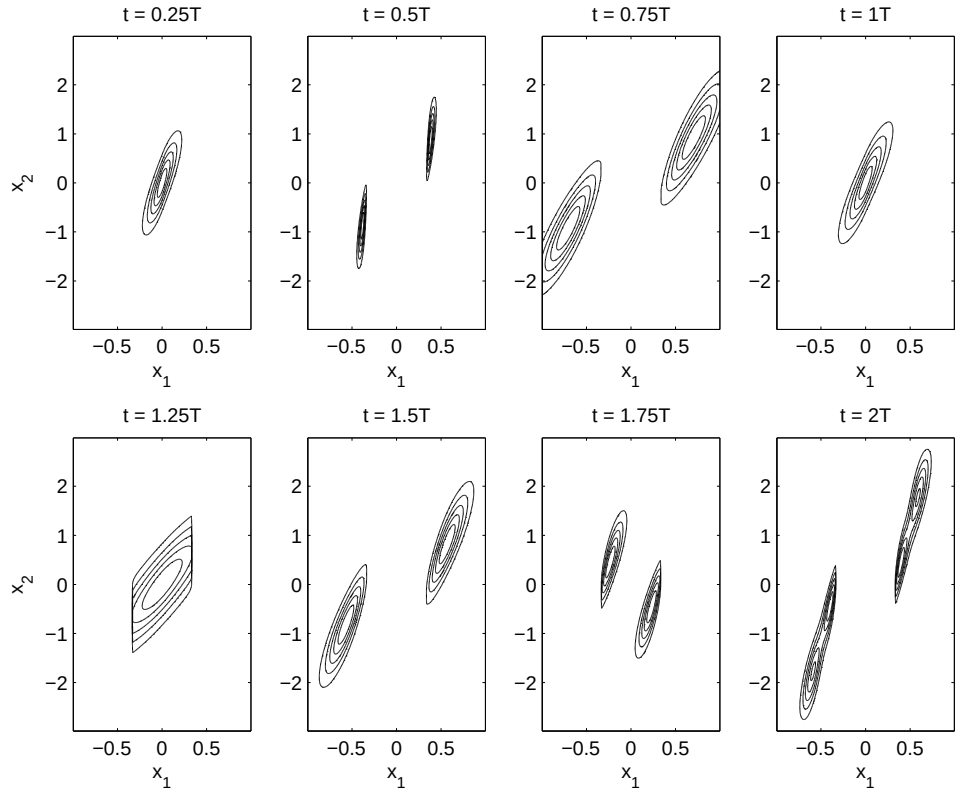

Figure 17 Snapshots from the Bayesian Observer with a *non-log-concave* measurement situation. Most states are clearly not log-concave; many are bimodal and the last one even tetramodal. The order of changes in y and snapshot times is: $y = 0, t = 0.25T$, $y = 2, t = 0.5T$, $t = 0.75T$, $y = 1$, $y = 0, t = T$, $t = 1.25T$, $y = 2, t = 1.5T$, $y = 0, t = 1.75T$, $y = 2, t = 2T$.

but farther away from the origin than $1/3$, but not on which side. The measurement function for $y = 2$ becomes non-log-concave in this case (two vertical stripes of 1 with a gap in between). For an example of a simulation of the Bayesian Observer for this process, see figure 17.

It is clear that violating the assumption of log-concave measurement function in this case lets the observer states violate the property of log-concavity; many are even multimodal. It is worth noting that the non-log-concavity enters the observer the first time the non-log-concave measurement function for $y = 2$ is used, but that it may persist even with log-concave measurements active as in the seventh snapshot.

In the fourth and fifth snapshots it seems that a period of log-concave measurements has managed to restore log-concavity. This is not unreasonable; consider for instance the case when y is held constant under log-concave conditions. If the state converges to a unique density under these conditions it must be log-concave, since log-concave initial states converge to it with log-concavity preserved.

The example uses a substantially non-log-concave measurement function to give a clear illustration. For measurement functions that are close to a log-concave function the effect is probably less pronounced, and may in some cases even be unnoticeable.

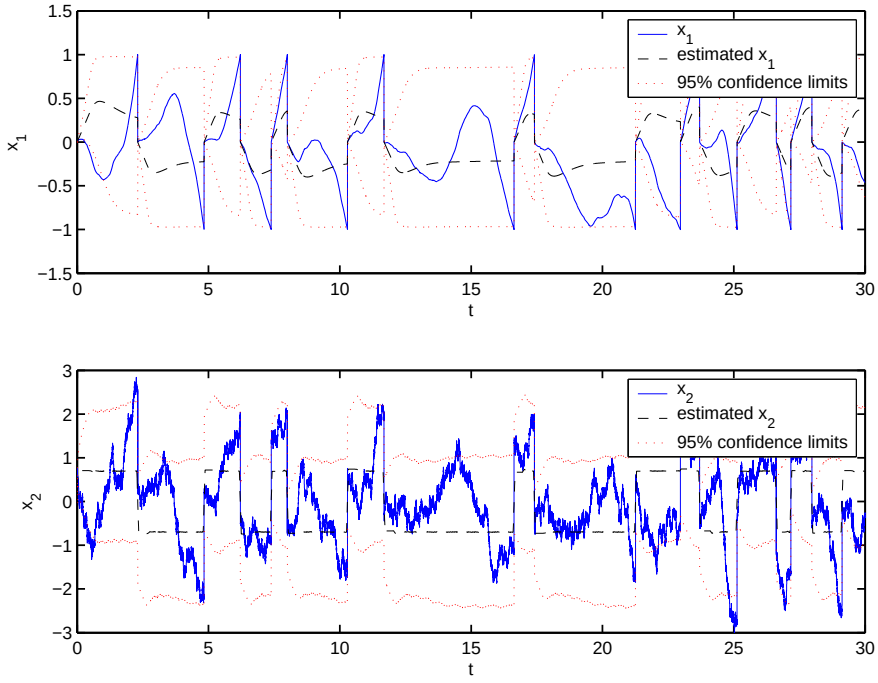

Figure 18 Actual and estimated state using the Discretized Bayesian Observer and $x_2^r = 0.7$. The square wave like appearance of the velocity estimate is a direct consequence of the ping pong control strategy. In both cases the actual state falls within the 95% confidence interval provided by the observer almost all of the time.

Filtering and control

The basic requirement for control of the Accelerometer is to keep the position x_1 from straying into the walls; behavior of the process once past the walls might typically be undefined or unpredictable in a real application. Thus the controller will have to apply some kind of strong restorative action to bring the position back toward the center as soon as an event is detected. This restorative action will be idealized as deadbeat control to bring the position to the origin.

A further requirement for any reasonable controller is to control the velocity x_2. Too high speeds are not desirable since they may require too much control authority for restoration and may demand too high bandwidth for the controller/observer. At the same time, it seems that higher speeds and a higher frequency of events may be desirable up to a point, to increase the amount of information available to the observer and thus yield better estimates. The prime interest has been in the estimate of x_2, since this is the state from which it is easiest to estimate the acceleration.

The best compromise between on the one hand low demands on control authority and controller/observer bandwidth and on the other hand high information content in the event sequence is realized if the speed can be closely controlled. To achieve this a simple scheme has been adopted which might be labeled *ping pong control*. The idea is to control x_2 to a reference velocity $\pm x_2^r$ that switches sign to always point away from the last wall that was reached. If the speed control is successful then a mean time between events of $t_{\text{mean}} \approx (x_2^r)^{-1}$ will be achieved. The process noise sets an upper limit on the mean time the process can be controlled to stay away from the walls.

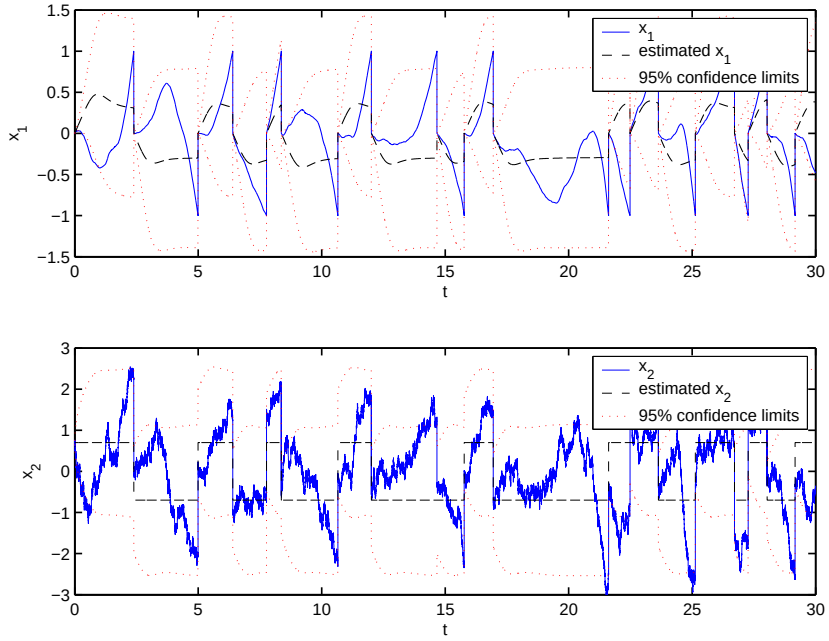

Figure 19 Actual and estimated state using the Full Kalman Filter and $x_2^r = 0.7$. The estimates and the resulting control (as can be seen from the actual states) are not very different from the Discretized Bayesian Observer in figure 18. The most noticeable difference is actually that the confidence intervals for position are wider in this case, often stretching beyond the $[-1, 1]$ range.

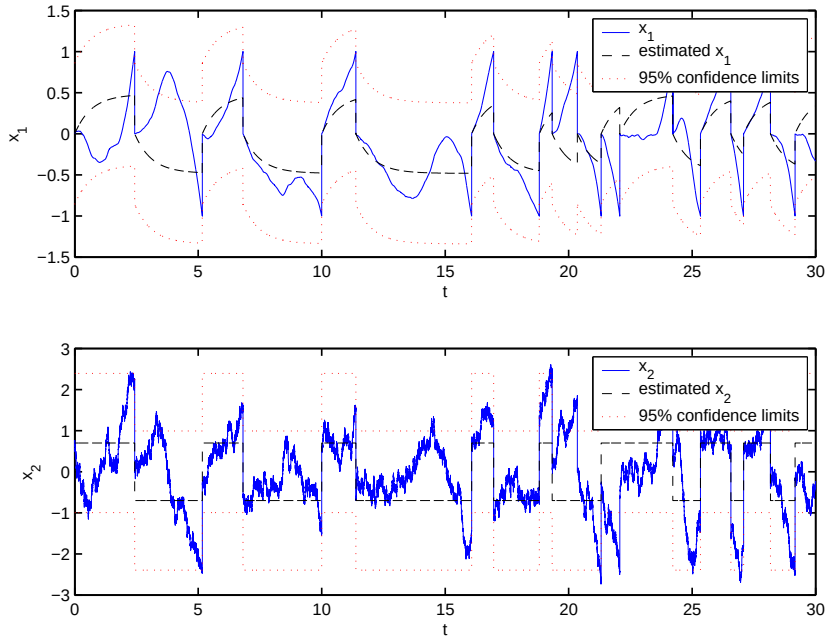

Figure 20 Actual and estimated state using the Simplified Kalman Filter and $x_2^r = 0.7$. The behavior is apparently different from figures 18 and 19. The confidence intervals still do seem to cover the actual states fairly well however.

Ping pong control has been simulated for different reference speeds using a deadbeat controller with three observers of varying complexity:

- The Discretized Bayesian Observer
- The Full Kalman Filter. This is a Kalman Filter approximation of the Bayesian Observer. The rectangular window measurement function for $y = 0$ has been replaced with a Gaussian function chosen to yield the same variance of x_1 in stationary state as with the Bayesian Observer. This was achieved by using the measurement variance $\sigma_o^2 h^{-1}$, where

  $$\sigma_o = 2^{-1/3}(p_{11}^{\text{stat}})^{2/3}\sigma^{-1/3},$$

  p_{11}^{stat} is the desired stationary variance in x_1 and σ^2 is the incremental variance of the process noise. The expression was derived from the analytical solution of the Riccati equation for the Full Kalman Filter.

  The measurement functions for $y = \pm 1$ are the same as for the Conservative Kalman Filter. The result is a Kalman Filter with time varying estimate covariance matrix P_k.
- The Simplified Kalman Filter. This is a Kalman Filter with constant covariance matrix. The measurement update rule for the filter is

  $$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + K_k\Big(y(k) - C\hat{\mathbf{x}}(k|k)\Big),$$

  where $C = (\,1 \quad 0\,)$, and $K_k = K_0$ when $y_k = 0$ and $K_k = K_e$ when $y_k = \pm 1$. K_0 and K_e are chosen as the mean effective K for the cases $y = 0$ and $y = \pm 1$ respectively of the Full Kalman Filter during a simulation run under the same conditions. The constant covariance matrix P is chosen in a similar way.

The same disturbance acceleration sequence was used in all simulations.

Figures 18, 19, and 20 show actual and estimated states from simulations of ping pong control with the three observers. In general it can be seen that the estimates do not follow the details of the evolution of the actual process (and neither could they since the details are mostly due to process noise, the effects of which only become visible at events) but that they almost always stay within a certain range of the actual values. The square wave like appearance of the velocity estimate is a direct consequence of the ping pong control strategy. It is not immediately evident from these plots whether one observer yields better estimates than another, though the greatest difference seems to be between the Simplified Kalman Filter and the other observers.

State estimate variances for the Discretized Bayesian Observer and the Full Kalman Filter are shown in figures 21 and 22 together with estimates from the Conservative Kalman Filter. (The Simplified Kalman Filter has constant estimate variance.) The conservative predictions are seen to always overestimate the actual variances, as they should. The variances of the Discretized Bayesian Observer and the Kalman Filter do not seem to differ by much, though the latter is a little more pessimistic.

The system was simulated with the three observers and with different reference speeds x_2^r, yielding different mean times between events t_{mean} and estimation errors. The results can be seen in figure 23. The Simplified Kalman Filter gave too unreliable control with

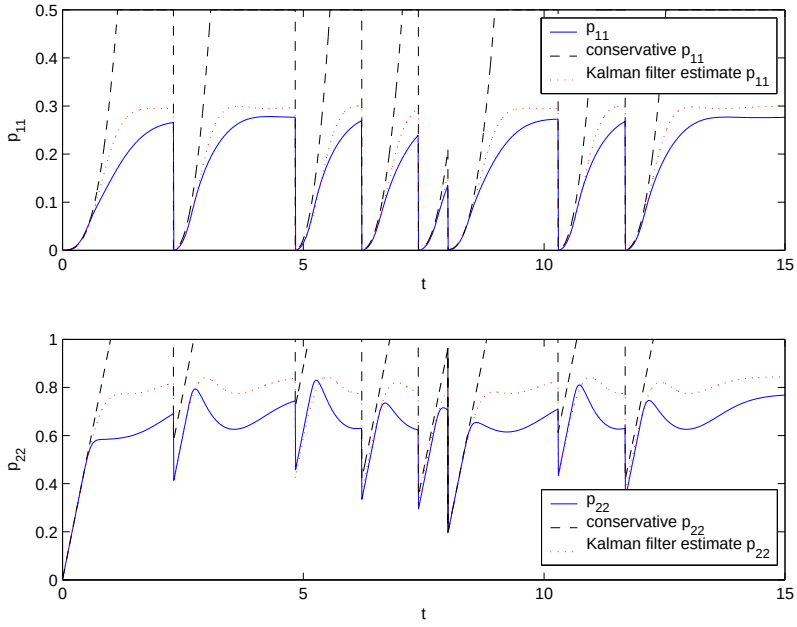

Figure 21 State estimate variances and predictions for the Discretized Bayesian Observer using $x_2^r = 0.7$. The predictions are very good for the position and quite good for the velocity a while after each event, after which the actual variances start to settle while the predictions do not. The Kalman Filter estimates come from the Full Kalman Filter and are seen to mimic the variance of the Discretized Bayesian Observer quite well, though settling at a somewhat higher level.

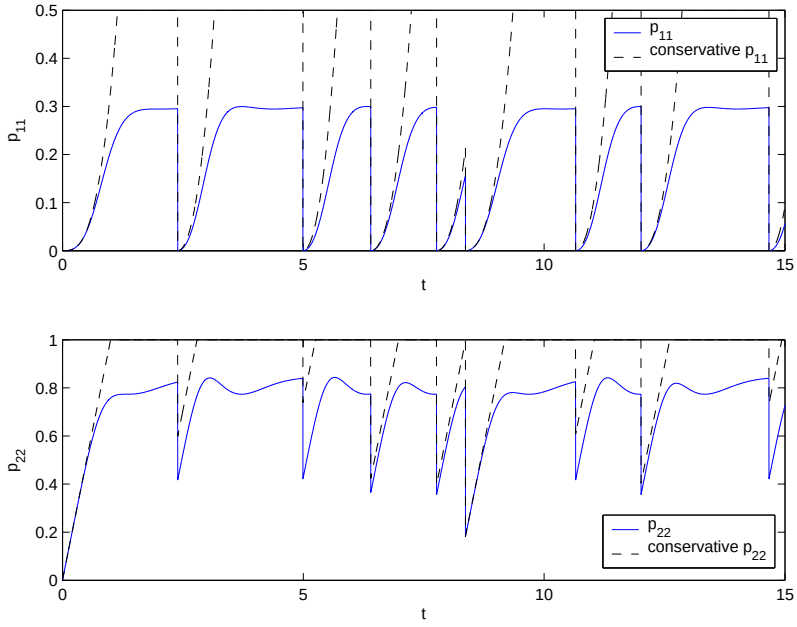

Figure 22 State estimate variances and predictions for the Full Kalman Filter using $x_2^r = 0.7$. The variances are very close to the Kalman Filter predictions in figure 21 and are always overestimated by the Conservative Kalman Filter.

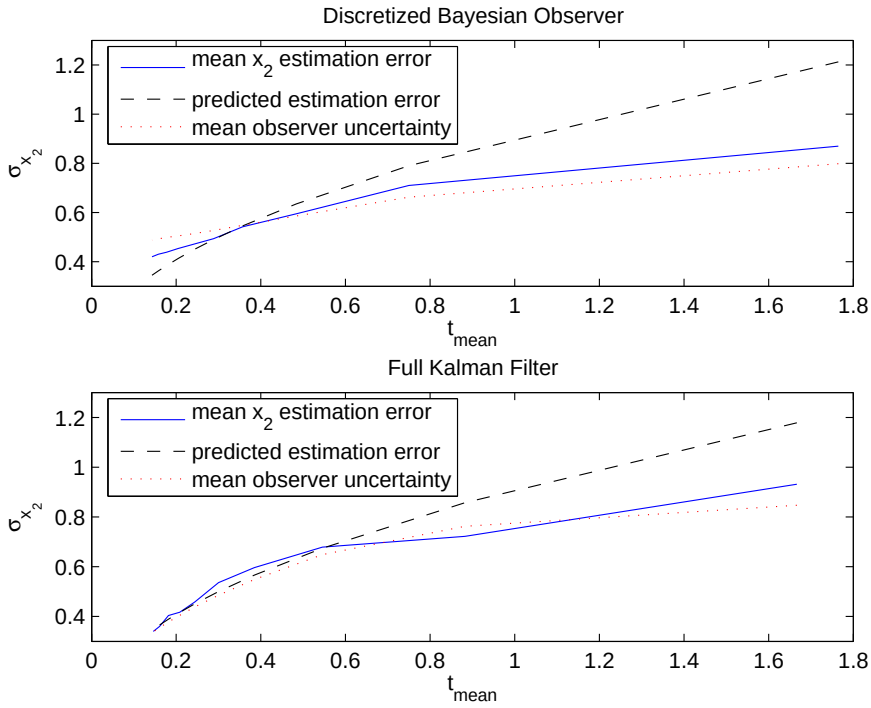

Figure 23 Estimation errors for x_2 as a function of mean time between events for the Discretized Bayesian Observer and the Full Kalman Filter. The conservative predictions seem to be conservative for the most part. The Discretized Bayesian Observer has a slight edge for low frequency of events while the Full Kalman Filter has a slight edge for high frequencies, probably because it is less susceptible to discretization errors which seem to be noticeable when $t_{\text{mean}} < 0.4$ for the former. The mean observer covariance seems to give a reasonable estimate of the actual estimation error, though for some reason the Discretized Bayesian Observer underestimates its own error slightly.

modest to high reference speeds and was not included in the comparison.

The figure confirms that a higher frequency of events gives a lower estimation error, up to a point. The conservative estimates obtained from (14) are mostly conservative, being accurate for high frequencies. The slack at low frequencies results because the Conservative Kalman Filter disregards the information contained in a long period without events.

It seems that the performance of the Discretized Bayesian Observer starts to deteriorate somewhat for high frequencies, probably due to discretization error. Simulations using a finer grid gave an improvement in this region. At the same time higher frequencies means a more concentrated observer state that is appreciably in touch with the walls only a small part of the time, meaning that a Kalman Filter has a good chance of approximating the Bayesian Observer well.

The most remarkable result is probably that the difference between the Discretized Bayesian Observer and the Kalman Filter is so small for a wide range of frequencies, the Discretized Bayesian Observer having a slight edge only for low frequencies where its greater complexity comes to use.

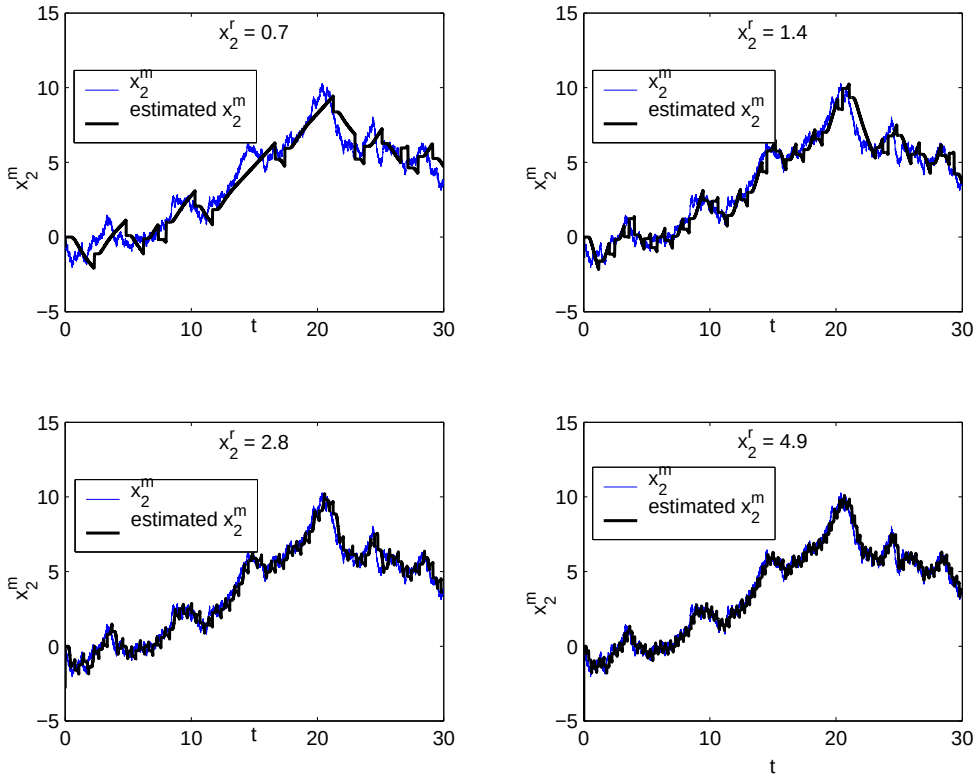

Figure 24 Estimated and actual integrals x_2^m of the disturbance acceleration for the Discretized Bayesian Observer with different reference speeds. Higher speeds and higher frequency of events yields better tracking and shorter time delays.

To examine the tracking performance more closely the signal

$$x_2^m(t) = \int_0^t v(\tau)\, d\tau$$

was adopted as the time integral of the disturbance acceleration. Estimates of x_2^m are readily available by combining an estimate of x_2 with the deterministic effects of the known control acceleration.

In figures 24, 25, and 26 actual and estimated x_2^m have been plotted for the three observers and different reference speeds. Here as well it is seen that the frequency of events is the main influence on tracking performance, and it also seems to have a direct influence on time delay. The tracking achieved by the three observers is very similar.

In figure 26 it can be seen how the Simplified Kalman Filter fails. When the error in the velocity estimate becomes too great, the ping pong control applies erroneous control signals that happen to bring the actual speed close to zero. The frequency of events drops and the estimate deteriorates even more.

When an event does not occur in the expected time, the velocity estimate of either of the first two observers decreases drastically in magnitude, causing control action to drive the velocity against the border. This seems to be enough to cause an actual event to occur quite soon. The Simplified Kalman Filter appears to have trouble appreciating the information contained in that no event has occurred for a while, since it does not keep track of the time since the last event.

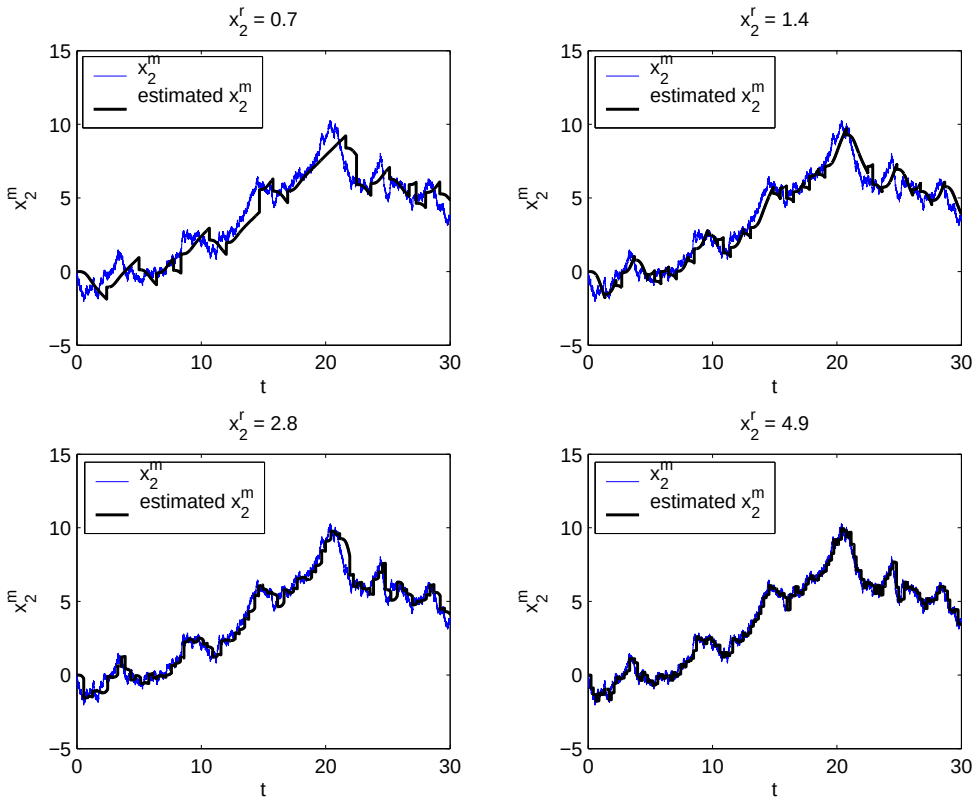

Figure 25 Estimated and actual integrals x_2^m of the disturbance acceleration for the Full Kalman Filter with different reference speeds. The tracking performance is very similar to the Discretized Bayesian Observer in figure 24, though possibly a bit worse for low reference speeds and a bit better for high reference speeds.

The robustness could probably be improved by letting it weight the measurements higher in general, at the cost of some accuracy.

It seems that while the approximation from the Bayesian Observer to the Full Kalman Filter is quite straightforward and does not involve very much compromise in this case, the approximation to the Simplified Kalman Filter is delicate and has to be tailored to the expected operating conditions.

Conclusions and future work

This thesis shows that the area of log-concave observers comprises an interesting field where powerful statements can be made while maintaining considerable generality. The beginnings of an analysis hints at the possibility that many more useful properties can be derived under similar assumptions.

The theory shows that observers for log-concave systems are in many ways very well behaved; almost as well behaved as Kalman Filters in some respects. This places the observer problem for log-concave systems much closer to the linear Gaussian case than the general nonlinear case, implying that log-concave observers should be easier to design and require less computing power than general nonlinear ones.

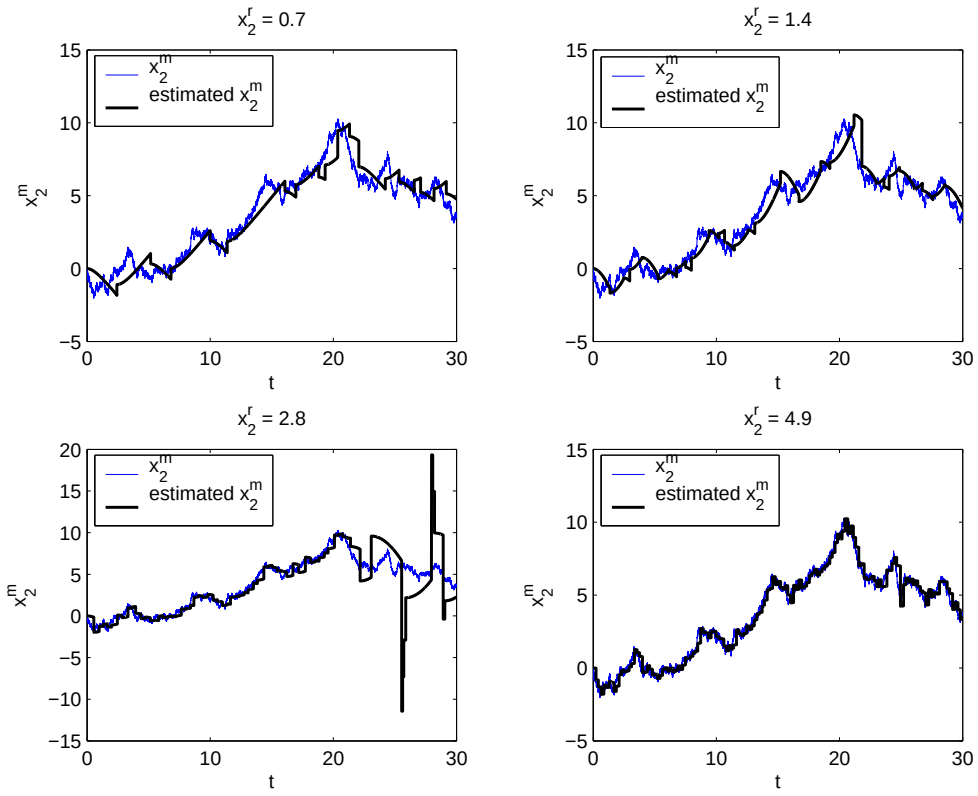

Figure 26 Estimated and actual integrals x_2^m of the disturbance acceleration for the Simplified Kalman Filter with different reference speeds. The tracking performance in general is close to but probably a little worse than for the Full Kalman Filter in figure 25. The main problem is tracking robustness, which can be seen in the bottom left corner where the observer estimate suddenly becomes too bad for reliable ping pong control.

The Event Based Accelerometer example shows that a Kalman Filter can provide a good approximate observer for a log-concave system even with a hard non-Gaussianity such as a discontinuous rectangular window measurement function.

In conclusion, it should be possible to find good approximate observers for many log-concave systems with moderate effort, and without using heavier artillery such as Particle Filters. Further developments in the theory of log-concave observers would certainly be an aid in such efforts. Additional assumptions of an entirely different nature from strong log-concavity might be necessary to form a powerful theory.

Much remains to be explored. Among some interesting points are

- Results on how to make approximations of log-concave observers that are good rather than conservative.
- Results on the behavior of expected values and other properties of probability densities in log-concave observers, apart from estimate covariances.
- Observers that exploit the property of log-concavity in a system.

A. Some properties of log-concave and strongly log-concave functions

In this appendix we derive some properties of (strongly) log-concave functions that will be needed for the proof of theorem 2.

Bounding functions

In this section we shall consider log-concave functions in one dimension (or the behavior of multivariate functions on a single line). As a definition for log-concavity, one can take the following conditions: $f(x)$ is log-concave iff $f(x) \geq 0$ and

$$\begin{aligned} f(x) &\geq f(x_1)^{1-\lambda} f(x_2)^{\lambda}, && (15) \\ x &= (1-\lambda)x_1 + \lambda x_2, \quad \lambda \in [0,1] && (16) \end{aligned}$$

for all $x_1, x_2 \in \mathbb{R}$. We will now show that the converse inequality holds when $\lambda \in \mathbb{R} \setminus (0,1)$.

Assuming $\lambda \neq 0$ we can solve (16) for x_2,

$$\begin{aligned} \lambda x_2 &= (\lambda - 1)x_1 + x \Longrightarrow \\ \Longrightarrow x_2 &= \left(1 - \frac{1}{\lambda}\right) x_1 + \frac{1}{\lambda} x = \\ &= (1-\mu)x_1 + \mu x, \end{aligned}$$

where $\mu = \frac{1}{\lambda}$. Putting $\lambda = \frac{1}{\mu}$ in (15) yields

$$\begin{aligned} f(x) &\geq f(x_1)^{1-\frac{1}{\mu}} f(x_2)^{\frac{1}{\mu}} \Longrightarrow \\ \Longrightarrow f(x)^{\mu} &\geq f(x_1)^{\mu-1} f(x_2) \Longrightarrow \\ \Longrightarrow f(x_1)^{1-\mu} f(x)^{\mu} &\geq f(x_2), \end{aligned}$$

assuming $f(x_1) > 0$. Putting $y = x_2, y_1 = x_1$ and $y_2 = x$ we find that

$$\begin{aligned} f(y) &\leq f(y_1)^{1-\mu} f(y_2)^{\mu}, \\ y &= (1-\mu)y_1 + \mu y_2, \quad \mu \in (1, \infty). \end{aligned}$$

When $\mu = 1$ the inequality reads

$$\begin{aligned} f(y) &\leq f(y_2), \\ y &= y_2, \end{aligned}$$

which is trivially true, so that the range of μ can be extended to $[1, \infty)$. Letting $\nu = 1 - \mu$ now yields

$$\begin{aligned} f(y) &\leq f(y_2)^{1-\nu} f(y_1)^{\nu}, \\ y &= (1-\nu)y_2 + \nu y_1, \quad \nu \in (-\infty, 0]. \end{aligned}$$

Summarizing we find that given a log-concave function $f(x)$, any $x_1, x_2 \in \mathbb{R}$ and letting

$$x = (1-\lambda)x_1 + \lambda x_2$$

$f(x)$ satisfies

$$\begin{aligned} f(x) &\geq f(x_1)^{1-\lambda} f(x_2)^{\lambda}, \quad \lambda \in [0,1], \\ f(x) &\leq f(x_1)^{1-\lambda} f(x_2)^{\lambda}, \quad \lambda \in \mathbb{R} \setminus (0,1), && (17) \end{aligned}$$

whenever the exponentiations are defined.

We now turn to the strongly log-concave case to derive a similar inequality. Letting $q = p^{-1}$ and

$$g(x) = e^{-\frac{1}{2}qx^2},$$

the condition on strong log-concavity for any function $f(x) = g(x)f_0(x)$ is that

$$f(x) \in \mathcal{LC}(p) \Longleftrightarrow f_0(x) \in \mathcal{LC}.$$

This means that if $f(x) \in \mathcal{LC}(p)$ then

$$f_0(x) = \frac{f(x)}{g(x)}$$

is log-concave. Thus, letting $\lambda \in [0,1]$ we can apply (15) to f_0, yielding

$$\begin{aligned} \frac{f(x)}{g(x)} &\geq \left(\frac{f(x)}{g(x_1)}\right)^{1-\lambda}\left(\frac{f(x_2)}{g(x_2)}\right)^{\lambda} \Longrightarrow \\ \Longrightarrow f(x) &\geq \frac{g(x)}{g(x_1)^{1-\lambda}g(x_2)^{\lambda}} f(x_1)^{1-\lambda} f(x_2)^{\lambda} \end{aligned} \tag{18}$$

where x is given by (16). Inserting (16) into

$$\begin{aligned} \ln\left(\frac{g(x)}{g(x_1)^{1-\lambda}g(x_2)^{\lambda}}\right) &= \frac{1}{2}q(-x^2 + (1-\lambda)x_1^2 + \lambda x_2^2) = \\ &= \frac{1}{2}q\Big(-\big((1-\lambda)x_1 + \lambda x_2\big)^2 + (1-\lambda)x_1^2 + \lambda x_2^2\Big) = \\ &= \frac{1}{2}q(-(1-\lambda)^2x_1^2 - 2\lambda(1-\lambda)x_1x_2 - \lambda^2x_2^2 + (1-\lambda)x_1^2 + \lambda x_2^2) = \\ &= \frac{1}{2}q\Big(\big((1-\lambda) - (1-\lambda^2)\big)x_1^2 - 2\lambda(1-\lambda)x_1x_2 + (\lambda - \lambda^2)x_2^2\Big) = \\ &= \frac{1}{2}q\lambda(1-\lambda)(x_1^2 - 2x_1x_2 + x_2^2) = \\ &= \frac{1}{2}q\lambda(1-\lambda)(x_1 - x_2)^2 \end{aligned}$$

we find that

$$\frac{g(x)}{g(x_1)^{1-\lambda}g(x_2)^{\lambda}} = e^{\frac{1}{2}q\lambda(1-\lambda)(x_1-x_2)^2}.$$

Furthermore

$$\begin{aligned} \lambda(x_1 - x_2) &= x_1 + (\lambda - 1)x_1 - \lambda x_2 = \\ &= x_1 - x, \\ (1-\lambda)(x_1 - x_2) &= (1-\lambda)x_1 + \lambda x_2 - x_2 = \\ &= x - x_2, \end{aligned}$$

so that we also can write

$$\frac{g(x)}{g(x_1)^{1-\lambda}g(x_2)^{\lambda}} = e^{\frac{1}{2}q(x-x_1)(x_2-x)}.$$

Inserting this into (18) we find that

$$\begin{aligned} f(x) &\geq e^{\frac{1}{2}q(x-x_1)(x_2-x)} f(x_1)^{1-\lambda} f(x_2)^{\lambda}, \\ x &= (1-\lambda)x_1 + \lambda x_2, \quad \lambda \in [0,1]. \end{aligned} \tag{19}$$

Looking at the derivation of (18) from (15) we find that it only involves multiplication by the same positive constant on both sides, so that the sign of the inequality does not matter. Hence we conclude that (19) holds with the converse inequality when $\lambda \in \mathbb{R} \setminus (0,1)$.

Summarizing we find that given a strongly log-concave function $f \in \mathcal{LC}(q^{-1})$, any $x_1, x_2 \in \mathbb{R}$ and letting

$$x = (1-\lambda)x_1 + \lambda x_2 \tag{20}$$

$f(x)$ satisfies

$$f(x) \geq e^{\frac{1}{2}q(x-x_1)(x_2-x)} f(x_1)^{1-\lambda} f(x_2)^{\lambda}, \quad \lambda \in [0,1], \tag{21}$$

$$f(x) \leq e^{\frac{1}{2}q(x-x_1)(x_2-x)} f(x_1)^{1-\lambda} f(x_2)^{\lambda}, \quad \lambda \in \mathbb{R} \setminus (0,1), \tag{22}$$

whenever the exponentiations are defined, that is, the exponential function bounding a log-concave function from above or below depending on the interval is replaced by a Gaussian function with the spread of $e^{-\frac{1}{2}qx^2}$.

Boundedness

All log-concave functions are ≥ 0 and hence a log-concave function is bounded iff it is bounded from above.

Consider any log-concave function $f(x)$ on a bounded interval I. If there exists less than three $x \in I$ such that $f(x) > 0$ then $f(x)$ is clearly bounded on I. Otherwise, let $x_1 < x_2 < x_3$ be three points in I such that $f(x_i) > 0$, $i \in \{1,2,3\}$. We can then apply (17) using x_1 and x_2 to show that $f(x)$ is bounded by an exponential function when $x \in I \setminus (x_1, x_2)$. Similarly, we can apply (17) using x_2 and x_3 to show that $f(x)$ is bounded by an exponential function when $x \in I \setminus (x_2, x_3)$. Since I is bounded, both exponential functions are bounded on I and hence $f(x)$ is bounded on I.

If $f(x)$ is strongly log-concave we can apply (22) instead of (17), which shows that $f(x)$ is bounded by at least one of two Gaussian functions for each x. Since any Gaussian function is bounded on $\mathbb{R}$ we can let $I = \mathbb{R}$ and the conclusion still holds. Also, as any Gaussian function goes to zero as $|x| \to \infty$, $f(x) \to 0$ as $|x| \to \infty$.

Summarizing, we find that any log-concave function is bounded on any bounded interval, and any strongly log-concave function is bounded on $\mathbb{R}$ and goes to zero as $|x| \to \infty$.

Continuity and the hypographic closure

Any convex function is continuous on the interior of its domain, implying that any log-concave function is continuous on the interior of its support. It is often convenient to assume that the continuity extends to the closure of the support. If this is not true then it can be remedied by changing the values of the function on a set of zero measure.

DEFINITION 3—HYPOGRAPHIC CLOSURE
The *hypographic closure* $f(x), x \in \mathbb{R}$ of a function $g(x)$ is defined as

$$f(x) = \lim_{t \to 0+} \sup_{|y-x| \leq t} g(y) \tag{23}$$

A function is *hypographically closed* iff it is the hypographic closure of itself. □

A function f is hypographically closed iff its hypograph $\{(x,y); y \leq f(x)\}$ is a closed set.

LEMMA 1—HYPOGRAPHIC CLOSURE
If $g(x), x \in \mathbb{R}$ is log-concave the hypographic closure $f(x)$ of $g(x)$ exists and is log-concave, and is continuous on the closure of its support. The hypographic closure satisfies

$$f(x) \geq g(x)$$

where $f(x) \neq g(x)$ for no more than two points $x \in \mathbb{R}$. Furthermore

$$\sup_{x \in \mathbb{R}} f(x) = \sup_{x \in \mathbb{R}} g(x).$$

□

Proof. Consider the definition (23). Since $g(x)$ is log-concave, it is bounded on any bounded interval and the supremum exists for all t. Furthermore $\sup_{|y-x| \leq t} g(y)$ is a decreasing function for decreasing t and is bounded from below by zero. Hence, the limit (23) exists for all x.

If g is continuous in a neighborhood of x, then $g(x) = f(x)$. We know that this is true for the interior of the support $S_g = \{x; g(x) \neq 0\}$, and trivially also for the exterior of the support, where $g(x)$ is zero.

Consider now the behavior of $f(x)$ on the boundary ∂S_g of S_g. Since $g(x)$ is log-concave S_g is a convex set, which in one dimension amounts to an interval. This means that either

- S_g has no boundary. Then $f(x) = g(x)$ and the lemma is true.
- S_g consists of a single point x_b. Then

  $$\sup_{|y-x_b| \leq t} g(y) = g(x_b)$$

  for all $t > 0$ so that $f(x_b) = g(x_b)$. S_g is closed and the lemma is true.
- S_g has positive measure and one or two boundary points x_b. Then $f(x_b)$ is the maximum of $g(x_b)$ and the limit of $g(y)$ when y goes to x_b from the interior. Let x_i be some point in the interior of S_g. Then we can apply (15) to find

  $$\begin{aligned} g(x) &\geq g(x_b)^{1-\lambda} g(x_i)^{\lambda}, \\ x &= (1-\lambda)x_b + \lambda x_i, \quad \lambda \in [0,1], \end{aligned}$$

  where x will be some point in the interior of S_g. When $\lambda \to 0$ we know that $x \to x_b$, and we see that the limit of $g(x)$ at x_b from the interior is bounded from below by $g(x_b)^{1-\lambda} g(x_i)^{\lambda}$ which goes to $g(x_b)$ as $\lambda \to 0$. Thus the limit of $g(x)$ when $x \to x_b$ from the interior is no less than $g(x_b)$ and we find that

  $$f(x_b) = \lim_{x \to x_b, x \in S_g \setminus \{x_b\}} g(x). \tag{24}$$

  Since $f(x_b)$ is the limit of $f(x)$ as x goes to x_b from the interior, $f(x)$ is continuous on $S_f \cup \partial S_f$, which is the closure of S_f. Since $f(x_b)$ is a limit of values of $g(x)$ we have that $f(x_b) \leq \sup_{x \in \mathbb{R}} g(x)$. At the same time, $f(x) \geq g(x)$ so that

  $$\sup_{x \in \mathbb{R}} f(x) = \sup_{x \in \mathbb{R}} g(x).$$

Finally, we must prove that f is log-concave, which means that we must verify that (15) holds. It is enough to verify this when $x_1, x_2 \in S_f$, since the right hand side evaluates to 0 otherwise. Using that (15) holds for g we can replace x_1 and x_2 with sequences that approach these points from the interior, taking limits and using that g is continuous in the interior and (24) we find that

$$g(x) \geq f(x_1)^{1-\lambda} f(x_2)^{\lambda},$$

under the conditions of (16). Using that $f(x) \geq g(x)$ completes the proof.

Unimodality of strongly log-concave functions

Consider any function $f \in \mathcal{LC}(p)$ defined on the real line, letting $q = p^{-1}$. Then (22) holds. Let $x_1, x_2 \in S_f$ be two distinct points such that $x_2 > x_1$. If two such points do not exist then $f(x)$ assumes it's maximum value either at a single point or (only if $f(x) \equiv 0$) on the entire real line.

Otherwise, since $f(x) \geq 0$ and $f(x) \to 0$ as $|x| \to \infty$ there exists some closed and bounded interval I such that

$$\sup_{x \in \mathbb{R}} f(x) = \sup_{x \in I} f(x).$$

Assuming that f is hypographically closed, I can be chosen as a compact subset of the closure of the support of f, and since f is continuous on this interval, $f(x)$ assumes its maximum value.

Let us now assume that $f(x)$ is not identically zero and assumes its maximum value at some point M_x,

$$f(x) \leq f(M_x) \quad \forall\, x \in \mathbb{R}.$$

Fixing some arbitrary x and letting $x_1 = M_x$ and

$$x_2 = x_1 + \frac{1}{\lambda}(x - x_1), \quad \lambda \geq 1$$

so that $x = (1-\lambda)x_1 + \lambda x_2$ we can apply (22) to find

$$\begin{aligned} f(x) &\leq e^{\frac{1}{2}q(x-M_x)(x_2-x)} f(M_x)^{1-\lambda} f(x_2)^{\lambda} = \\ &= e^{\frac{1}{2}q(x-M_x)(x_2-x)} f(M_x) \left(\frac{f(x_2)}{f(M_x)}\right)^{\lambda} \leq \\ &\leq e^{\frac{1}{2}q(x-M_x)(x_2-x)} f(M_x), \end{aligned}$$

and since the right hand side is a continuous function of x_2 we can let $\lambda \to \infty$ so that $x_2 \to M_x$, finding that

$$\begin{aligned} f(x) &\leq e^{\frac{1}{2}q(x-M_x)(M_x-x)} f(M_x) = \\ &= e^{-\frac{1}{2}q(x-M_x)^2} \sup_{y \in \mathbb{R}} f(y). \end{aligned}$$

We see that $f(x) < f(M_x)$ when $x \neq M_x$. Thus any strongly log-concave function that assumes its maximum value does so at a single point M_x, and the function is bounded from above by the inequality

$$f(x) \leq e^{-\frac{1}{2}q(x-M_x)^2} \sup_{y \in \mathbb{R}} f(y). \tag{25}$$

For any strongly log-concave function $f(x)$ the result can be applied to the hypographic closure of f, yielding the conclusion that f assumes or approaches its supremum at single point M_x, and the inequality still holds.

We shall now consider the monotonicity of log-concave functions. Let $f(x)$ be a log-concave function that assumes its maximum at $x = M_x$. Let $x_2 = M_x$ and let $x, x_1 \in \mathbb{R}$ be two points such that $|x_1 - M_x| < |x - M_x|$ and $x - M_x$ and $x_1 - M_x$ have the same sign. Then (17) can be applied to find

$$\begin{aligned} f(x) &\leq f(x_1)^{1-\lambda} f(M_x)^{\lambda} = \\ &= f(x_1) \left(\frac{f(M_x)}{f(x_1)} \right)^{\lambda} \leq \\ &\leq f(x_1) \end{aligned}$$

since $\lambda < 0$ and $f(M_x)$ is the maximum value of f. Thus f is increasing if $x \leq M_x$ and decreasing if $x \geq M_x$.

B. Proof of theorems on strongly log-concave functions

Multiplication

Let $f \in \mathcal{LC}(F)$ and $g \in \mathcal{LC}(G)$. Then

$$\begin{aligned} h(\mathbf{x}) &= f(\mathbf{x})g(\mathbf{x}) = \\ &= e^{-\frac{1}{2}\mathbf{x}^T F^{-1}\mathbf{x}} e^{-\frac{1}{2}\mathbf{x}^T G^{-1}\mathbf{x}} \cdot f_0(\mathbf{x})g_0(\mathbf{x}) = \\ &= e^{-\frac{1}{2}\mathbf{x}^T (F^{-1}+G^{-1})\mathbf{x}} \cdot f_0(\mathbf{x})g_0(\mathbf{x}) = \\ &= e^{-\frac{1}{2}\mathbf{x}^T H^{-1}\mathbf{x}} \cdot h_0(\mathbf{x}), \end{aligned}$$

where f_0 and g_0 are log-concave, $h_0(x) = f_0(x)g_0(x)$ and $H^{-1} = F^{-1} + G^{-1}$. Thus h_0 is log-concave and so $h \in \mathcal{LC}(H)$.

Affine transformation

Let $f \in \mathcal{LC}(F)$, $A \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Then

$$\begin{aligned} g(\mathbf{x}) &= f(A\mathbf{x} + \mathbf{b}) = \\ &= e^{-\frac{1}{2}(A\mathbf{x}+\mathbf{b})^T F^{-1}(A\mathbf{x}+\mathbf{b})} \cdot f_0(A\mathbf{x} + \mathbf{b}) = \\ &= e^{-\frac{1}{2}(\mathbf{x}^T A^T F^{-1} A\mathbf{x} + 2\mathbf{b}^T F^{-1} A\mathbf{x} + \mathbf{b}^T F^{-1}\mathbf{b})} \cdot f_0(A\mathbf{x} + \mathbf{b}) = \\ &= e^{-\frac{1}{2}\mathbf{x}^T (A^{-T} F A^{-1})^{-1}\mathbf{x}} \cdot \left(e^{-\frac{1}{2}\mathbf{b}^T F^{-1}\mathbf{b}} e^{-(A^T F^{-1}\mathbf{b})^T \mathbf{x}} f_0(\mathbf{x}) \right) = \\ &= e^{-\frac{1}{2}\mathbf{x}^T G^{-1}\mathbf{x}} \cdot g_0(\mathbf{x}), \end{aligned}$$

where

$$g_0(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{b}^T F^{-1}\mathbf{b}} e^{-(A^T F^{-1}\mathbf{b})^T \mathbf{x}} f_0(\mathbf{x})$$

and

$$G = A^{-T} F A^{-1}.$$

We see that g_0 is the product of a constant, an exponential function and a log-concave function. Since all these are log-concave, so is g_0. Thus $g \in \mathcal{LC}(G)$.

Convolution

For the following proof we shall need a matrix identity. Let A, B be positive definite matrices and $\mathbf{x}, \mathbf{y}$ be vectors. Furthermore, let

$$\begin{aligned} C &= (A^{-1} + B^{-1})^{-1} = A(A+B)^{-1}B, \\ \mathbf{z} &= \mathbf{y} - (A+B)^{-1}B\mathbf{x}. \end{aligned}$$

Then

$$\begin{aligned} & \mathbf{x}^T C\mathbf{x} + \mathbf{z}^T(A+B)\mathbf{z} = \\ = \; & \mathbf{x}^T C\mathbf{x} + \big(\mathbf{y}^T - \mathbf{x}^T B(A+B)^{-1}\big)(A+B)\big(\mathbf{y} - (A+B)^{-1}B\mathbf{x}\big) = \\ = \; & \mathbf{x}^T C\mathbf{x} + \mathbf{y}^T(A+B)\mathbf{y} - 2\mathbf{y}^T(A+B)(A+B)^{-1}B\mathbf{x} + \mathbf{x}^T B(A+B)^{-1}(A+B)(A+B)^{-1}B \\ = \; & \mathbf{x}^T C\mathbf{x} + \mathbf{y}^T(A+B)\mathbf{y} - 2\mathbf{y}^T B\mathbf{x} + \mathbf{x}^T B(A+B)^{-1}B\mathbf{x}. \end{aligned}$$

Rearranging and inserting the definition of C gives

$$\begin{aligned} & \mathbf{x}^T C\mathbf{x} + \mathbf{z}^T(A+B)\mathbf{z} = \\ = \; & \mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y} - 2\mathbf{y}^T B\mathbf{x} + \mathbf{x}^T B(A+B)^{-1}B\mathbf{x} + \mathbf{x}^T A(A+B)^{-1}B\mathbf{x} = \\ = \; & \mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y} - 2\mathbf{y}^T B\mathbf{x} + \mathbf{x}^T(A+B)(A+B)^{-1}B\mathbf{x} = \\ = \; & \mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y} - 2\mathbf{y}^T B\mathbf{x} + \mathbf{x}^T B\mathbf{x} = \\ = \; & \mathbf{y}^T A\mathbf{y} + (\mathbf{x}-\mathbf{y})^T B(\mathbf{x}-\mathbf{y}). \end{aligned}$$

We observe that both $A + B$ and $C = (A^{-1} + B^{-1})^{-1}$ are positive definite.

Let $f \in \mathcal{LC}(F)$ and $g \in \mathcal{LC}(G)$. Then the convolution of f and g is

$$\begin{aligned} h(\mathbf{x}) = (f * g)(\mathbf{x}) &= \int f(\mathbf{y}) g(\mathbf{x}-\mathbf{y})\, dy = \\ &= \int e^{-\frac{1}{2}\mathbf{y}^T F^{-1}\mathbf{y}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{y})^T G^{-1}(\mathbf{x}-\mathbf{y})} \cdot f_0(\mathbf{y}) g_0(\mathbf{x}-\mathbf{y})\, dy = \\ &= \int e^{-\frac{1}{2}\left(\mathbf{y}^T F^{-1}\mathbf{y} + (\mathbf{x}-\mathbf{y})^T G^{-1}(\mathbf{x}-\mathbf{y})\right)} \cdot f_0(\mathbf{y}) g_0(\mathbf{x}-\mathbf{y})\, dy. \end{aligned}$$

Letting $A = F^{-1}, B = G^{-1}$ and $C = H^{-1} = (F+G)^{-1}$ and using the matrix identity (26) yields

$$\begin{aligned} h(\mathbf{x}) &= \int e^{-\frac{1}{2}\left(\mathbf{x}^T H\mathbf{x} + \mathbf{z}^T(F^{-1}+G^{-1})\right)\mathbf{z}} \cdot f_0(\mathbf{y}) g_0(\mathbf{x}-\mathbf{y})\, dy = \\ &= e^{-\frac{1}{2}\mathbf{x}^T H\mathbf{x}} \int e^{-\frac{1}{2}\mathbf{z}^T(F^{-1}+G^{-1})\mathbf{z}} \cdot f_0(\mathbf{y}) g_0(\mathbf{x}-\mathbf{y})\, dy, \end{aligned}$$

where $\mathbf{z} = \mathbf{y} - FH^{-1}\mathbf{x}$. The integrand is the product of three log-concave functions and is thus log-concave, so that

$$h_0(\mathbf{x}) = \int e^{-\frac{1}{2}\mathbf{z}^T(F^{-1}+G^{-1})\mathbf{z}} \cdot f_0(\mathbf{y}) g_0(\mathbf{x}-\mathbf{y})\, dy$$

is a log-concave function of $\mathbf{x}$ and

$$h(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{x}^T H\mathbf{x}} \cdot h_0(\mathbf{x})$$

is a strongly log-concave function, $h \in \mathcal{LC}(H)$.

A useful lemma

The proof of this lemma requires application of many of the properties of log-concave functions derived in appendix A.

LEMMA 2—EXPECTATION BOUND
Let $f(x) \in \mathcal{LC}(p), p > 0$ be a probability density in one dimension and $M_x \in \mathbb{R}$ be the unique value such that $f(x)$ is increasing for $x < M_x$ and decreasing for $x > M_x$. Then

$$\int w(|x - M_x|) f(x)\, dx \leq \frac{1}{\sqrt{2\pi p}} \int w(|x|) e^{-\frac{1}{2}p^{-1}x^2}\, dx$$

where $w(x), x \geq 0$ is any nonnegative increasing function of x such that the right hand integral exists. □

Proof. Since the theorem only concerns integrals of f and a log-concave function coincides with its hypographic closure except for on a set of zero measure, f can be assumed to be hypographically closed without loss of generality. Thus $f(x)$ assumes its maximum value at some $x = M_x$. Using the definition of strong log-concavity, $f(x)$ can be written as

$$f(x) = e^{-\frac{1}{2}q(x-M_x)^2} f_0(x)$$

where $q = p^{-1}$, $f_0 \in \mathcal{LC}$, and the property (25) implies that $f_0(x)$ assumes its maximum for $x = M_x$. The function

$$g_0(x) = f_0(x - M_x) + f_0(-x - M_x), \qquad x \geq 0$$

defined only for $x \geq 0$ is then decreasing in x because of the unimodality of log-concave functions, as is

$$g(x) = f(x - M_x) + f(-x - M_x) = e^{-\frac{1}{2}qx^2} g_0(x), \qquad x \geq 0.$$

Let

$$C = \left(\int_0^\infty e^{-\frac{1}{2}qx^2} \right)^{-1} = \sqrt{\frac{2q}{\pi}}.$$

Since g_0 is decreasing in x, either

- There exists some $x_0 \geq 0$ such that

$$\begin{aligned} g_0(x) &\geq C, \quad x \leq x_0 \\ g_0(x) &\leq C, \quad x \geq x_0 \end{aligned}$$

- Either $g_0(x) > C$ for all x or $g_0(x) < C$ for all x. But

$$\int_0^\infty e^{-\frac{1}{2}qx^2} g_0(x) dx = \int_0^\infty g(x) dx = \int_{-\infty}^\infty f(x) dx = 1$$

and

$$\int_0^\infty e^{-\frac{1}{2}qx^2} C dx = 1,$$

which can not both be true if $g_0(x)$ is almost everywhere continuous and everywhere less than or everywhere greater than C.

Thus there must exist some x_0 that satisfies the conditions given above. Then

$$\begin{aligned} d &= \int w(|x-M_x|)f(x)dx - \int w(|x-M_x|)\frac{1}{2}Ce^{-\frac{1}{2}q(x-M_x)^2}dx = \\ &= \int w(|x-M_x|)e^{-\frac{1}{2}q(x-M_x)^2}(f_0(x)-\frac{1}{2}C)dx = \\ &= \int_0^\infty w(x)e^{-\frac{1}{2}qx^2}(g_0(x)-C)dx = \\ &= \int_0^{x_0} w(x)e^{-\frac{1}{2}qx^2}(g_0(x)-C)dx + \int_{x_0}^\infty w(x)e^{-\frac{1}{2}qx^2}(g_0(x)-C)dx. \end{aligned}$$

In the first integral the integrand is positive, so it can be overestimated by replacing $w(x)$ with its maximum value in the range, that is $w(x_0)$. In the second integral the integrand is negative, so it can be overestimated by replacing $w(x)$ with its minimum value in the range, which is also $w(x_0)$. Thus

$$\begin{aligned} d &\leq \int_0^{x_0} w(x_0)e^{-\frac{1}{2}qx^2}(g_0(x)-C)dx + \int_{x_0}^\infty w(x_0)e^{-\frac{1}{2}qx^2}(g_0(x)-C)dx = \\ &= w(x_0)\int_0^\infty e^{-\frac{1}{2}qx^2}(g_0(x)-C)dx = \\ &= w(x_0)\int_0^\infty (g(x)-Ce^{-\frac{1}{2}qx^2})dx = 0 \end{aligned}$$

since both terms in the integrand integrate to one by themselves. We find that

$$\begin{aligned} \int w(|x-M_x|)f(x)\,dx &\leq \int w(|x-M_x|)\frac{1}{2}Ce^{-\frac{1}{2}q(x-M_x)^2}\,dx = \\ &= \sqrt{\frac{q}{2\pi}}\int w(|x|)e^{-\frac{1}{2}qx^2}\,dx, \end{aligned}$$

which concludes the proof.

COROLLARY 1—ML / LEAST SQUARES DEPENDENCE
Let $f(x) \in \mathcal{LC}(p)$ be a probability density in one dimension that assumes its maximum at $x = M_x$ and let $m_x = \mathrm{E}_x(f(x))$. Then

$$|m_x - M_x| \leq \sqrt{\frac{2p}{\pi}}.$$

□

This is a bound on the difference between the maximum likelihood and the least squares estimate obtained from $f(x)$.

Proof. Let $w(x) = |x|$. Then lemma 2 can be applied to find

$$\begin{aligned} |m_x - M_x| &= \left|\int (x-M_x)f(x)\,dx\right| \leq \\ &\leq \int |x-M_x|f(x)\,dx \leq \\ &\leq \frac{1}{\sqrt{2\pi p}}\int |x|e^{-\frac{1}{2}p^{-1}x^2}\,dx = \sqrt{\frac{2p}{\pi}}. \end{aligned}$$

Bounded variance

We shall begin by proving the theorem in one dimension. Using lemma 2 we find that

$$\int (x - M_x)^2 f(x) dx \leq \frac{1}{\sqrt{2\pi p}} \int x^2 e^{-\frac{1}{2}p^{-1}x^2} \, dx = p.$$

At the same time the second moment of f assumes its minimum around $x = m_x$:

$$\begin{aligned}
\int (x - M_x)^2 f(x) dx &= \int \big((x - m_x) + (m_x - M_x)\big)^2 f(x) \, dx = \\
&= \int \big((x - m_x)^2 + 2(x - m_x)(m_x - M_x) + (m_x - M_x)^2\big) f(x) \, dx = \\
&= \left(\int (x - m_x)^2 f(x) \, dx\right) + 2(m_x - m_x)(m_x - M_x) + (m_x - M_x)^2 = \\
&= \mathrm{Cov}_x\big(f(x)\big) + (m_x - M_x)^2
\end{aligned}$$

so that

$$\mathrm{Cov}_x\big(f(x)\big) \leq \mathrm{Cov}_x\big(f(x)\big) + (m_x - M_x)^2 \leq p,$$

which proves the theorem in one dimension.

For the proof in $\mathbb{R}^n$ we shall need the following matrix inequality. Let Q be a positive definite matrix and $\mathbf{x}$ and $\mathbf{y}$ be vectors; then

$$(\mathbf{x}^T Q \mathbf{y})^2 \leq (\mathbf{x}^T Q \mathbf{x})(\mathbf{y}^T Q \mathbf{y}).$$

This is the Cauchy-Schwartz inequality for the inner product $(\mathbf{x}|\mathbf{y}) = \mathbf{x}^T Q \mathbf{y}$. Letting $\mathbf{y} = Q^{-1}\mathbf{z}$ the inequality is transformed into

$$\begin{aligned}
(\mathbf{x}^T Q Q^{-1} \mathbf{z})^2 &\leq (\mathbf{x}^T Q \mathbf{x})((Q^{-1}\mathbf{z})^T Q Q^{-1} \mathbf{z}) \Longrightarrow \\
\Longrightarrow (\mathbf{x}^T \mathbf{z})^2 &\leq (\mathbf{x}^T Q \mathbf{x})(\mathbf{z}^T Q^{-1} \mathbf{z}),
\end{aligned}$$

where Q^{-1} exists and is positive definite since Q is positive definite. Assuming $\mathbf{z} \neq 0$ we find

$$\begin{aligned}
(\mathbf{x}^T \mathbf{z})(\mathbf{z}^T \mathbf{x})(\mathbf{z}^T Q^{-1} \mathbf{z})^{-1} &\leq \mathbf{x}^T Q \mathbf{x} \Longrightarrow \\
\Longrightarrow \mathbf{x}^T \big(\mathbf{z}(\mathbf{z}^T Q^{-1} \mathbf{z})^{-1} \mathbf{z}^T\big) \mathbf{x} &\leq \mathbf{x}^T Q \mathbf{x}
\end{aligned}$$

which is the same as to say that

$$\mathbf{z}(\mathbf{z}^T Q^{-1} \mathbf{z})^{-1} \mathbf{z}^T \leq Q.$$

Now consider the probability density $f(\mathbf{x}) \in \mathcal{LC}(P)$. If the expectation $\mathbf{m_x}$ is not zero, it can be made zero by the simple transformation $g(\mathbf{x}) = f(\mathbf{x} - \mathbf{m_x})$, which does not alter the variance of f. Therefore, without loss of generality, let $\mathbf{m_x} = 0$. Let $\mathbf{e_z}$ be a unit vector. Then $f(\mathbf{x})$ can be written as

$$\begin{aligned}
f(\mathbf{x}) &= e^{-\frac{1}{2}\mathbf{x}^T P^{-1} \mathbf{x}} \cdot f_0(\mathbf{x}) = \\
&= e^{-\frac{1}{2}\mathbf{x}^T Q_r \mathbf{x}} \cdot \left(e^{-\frac{1}{2}\mathbf{x}^T (P^{-1} - Q_r)\mathbf{x}} f_0(\mathbf{x})\right) = \\
&= e^{-\frac{1}{2}\mathbf{x}^T Q_r \mathbf{x}} \cdot f_1(\mathbf{x})
\end{aligned}$$

where f_0 is log-concave and

$$Q_r = \mathbf{e_z}(\mathbf{e_z}^T P \mathbf{e_z})^{-1} \mathbf{e_z}^T \leq P^{-1}.$$

Since $P^{-1} - Q_r \geq 0$ the function $e^{-\frac{1}{2}\mathbf{x}^T(P^{-1}-Q_r)\mathbf{x}}$ is log-concave and hence f_1 is log-concave. Letting $V = \mathrm{Cov}_{\mathbf{x}}\big(f(\mathbf{x})\big)$ and looking at the variance along the $\mathbf{e_z}$ direction we see that

$$\mathbf{e_z}^T V \mathbf{e_z} = \int \mathbf{e_z}^T \mathbf{x}\mathbf{x}^T \mathbf{e_z} f(\mathbf{x})\, dx.$$

We can split the integral by letting $\mathbf{x} = \mathbf{y} + t\mathbf{e_z}$, where $\mathbf{y} \in [\mathbf{e_z}]^\perp = \{\mathbf{y} \in \mathbb{R}^n; \mathbf{y}^T\mathbf{e_z} = 0\}$ and $dx = dy\,dt$, finding

$$\begin{aligned}
\mathbf{e_z}^T V \mathbf{e_z} &= \int_{t\in\mathbb{R}} \int_{\mathbf{y}\in[\mathbf{e_z}]^\perp} \mathbf{e_z}^T(\mathbf{y}+t\mathbf{e_z})(\mathbf{y}+t\mathbf{e_z})^T\mathbf{e_z} f(\mathbf{y}+t\mathbf{e_z})\, dy\, dt = \\
&= \int_{t\in\mathbb{R}} t^2 \cdot e^{-\frac{1}{2}(\mathbf{y}+t\mathbf{e_z})^T Q_r (\mathbf{y}+t\mathbf{e_z})} \int_{\mathbf{y}\in[\mathbf{e_z}]^\perp} f_1(\mathbf{y}+t\mathbf{e_z})\, dy\, dt = \\
&= \int_{t\in\mathbb{R}} t^2 e^{-\frac{1}{2}(\mathbf{e_z}^T P \mathbf{e_z})^{-1} t^2} \left(\int_{\mathbf{y}\in[\mathbf{e_z}]^\perp} f_1(\mathbf{y}+t\mathbf{e_z})\, dy \right) dt,
\end{aligned}$$

since $Q_r\mathbf{y} = 0$. The inner integral is the marginal density of f_1 along $\mathbf{e_z}$ as a function of t; since f_1 is log-concave, so is the integral. Thus the outer integrand save t^2 is strongly log-concave with $p = \mathbf{e_z}^T P \mathbf{e_z}$. Furthermore it has zero expectation and the result in one dimension can be applied to find that

$$\begin{aligned}
\mathbf{e_z}^T V \mathbf{e_z} &\leq \mathbf{e_z}^T P \mathbf{e_z} \Longrightarrow \\
\Longrightarrow V &\leq P,
\end{aligned}$$

which proves the theorem.

References

[1] K. J. Åström, *Introduction to Stochastic Control Theory*. New York: Academic Press, 1970.

[2] S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, 50, 174-188, 2002.

[3] A. Prékopa, "Logarithmic concave measures with application to stochastic programming," *Acta Scientarium Mathematicarum*, 32, 301-315, 1971.

[4] A. Prékopa, "On logarithmic concave measures and functions," *Acta Scientarium Mathematicarum*, 33, 335-343, 1973.

[5] M. Bagnoli and T. Bergstrom, "Log-concave probability and its applications" (January 1, 2004). Department of Economics, UCSB. Ted Bergstrom. Paper 1989D. Available at http://repositories.cdlib.org/ucsbecon/bergstrom/1989D/.

[6] M.Y. An, "Log-concave probability distributions: theory and statistical testing." Technical report, Economics Department, Duke University, Durham, N.C. 27708-0097, 1995.

[7] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge: Cambridge University Press, 2004. Available at http://www.ee.ucla.edu/~vandenbe/cvxbook.html.