Chapter 3

State Estimation

3.1 Kalman Filtering

In this section, we study the Kalman filter. First we state the problem and its solution. In particular, we discuss some of the senses in which the Kalman filter is optimal. After that, we give a relatively straightforward proof of the Kalman filter.

Problem Formulation

We assume the process model is described by a linear time-varying (LTV) model in discrete time

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + N_k w_k \\ y_k &= C_k x_k + D_k u_k + v_k, \end{aligned} \tag{3.1}$$

where $x_k \in \mathbb{R}^n$ is the state, $u_k \in \mathbb{R}^m$ is the input, $y_k \in \mathbb{R}^p$ is the output, $w_k \in \mathbb{R}^o$ is the process noise, and $v_k \in \mathbb{R}^p$ is the measurement noise.

The problem is as follows: Given a measurement sequence and an input sequence up to time k,

$$Y_k = \{y_i, u_i : 0 \leq i \leq k\},$$

what is the best possible estimate $\hat{x}_{k+q}$ of x_{k+q}? And what is meant by the "best estimate"? If we are have $q = 0$, we call it a *filtering problem*, if $q > 0$ we call it a *prediction problem*, and if $q < 0$ we call it a *smoothing problem*. We will here only dicuss filtering and prediction, which are the most common applications in control enginieering. The smoothing problem is of interest in signal processing where time delays usually are a minor concern.

To solve the estimation problem, a model of the noise v_k and w_k are needed. We here adopt a stochastic model for the noise. It should be noted, however, that it is also possible to develop a deterministic worst-case theory, see, for example, H_∞ filtering in [6].

Structure and Optimality of the Kalman Filter

We now give the form of the Kalman filter, and dicuss under what assumptions and it what senses it constitutes the best possible filter.

Assume that the noise has zero mean, is *white* (the noise is uncorrelated in *time*), and the covariances

$$\mathbb{E}\left\{\begin{bmatrix} w_k \\ v_k \end{bmatrix}\begin{bmatrix} w_l^T & v_l^T \end{bmatrix}\right\} = \begin{bmatrix} Q_k & S_k \\ S_k^T & R_k \end{bmatrix}\delta_{kl} = \Sigma_k \delta_{kl},$$

where δ_{kl} is the Kronecker delta ($\delta_{kl} = 1$ if $k = l$, and else $\delta_{kl} = 0$) are known. Furthermore, we assume some knowledge about the initial state x_0. We assume that

$$\mathbb{E}\{x_0\} = \bar{x}_0, \quad \text{and} \quad \mathbb{E}\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T\} = P_0$$

are known.

The Kalman filter has an iterative structure. It has two types of states: $\hat{x}_{k|k-1}$ denoting the estimate of the state at time k, given Y_{k-1}, and $\hat{x}_{k|k}$ denoting the estimate of the state at time k, given Y_k. There is a corrector step where the most recent measurement is taken into account, and there is a prediction step for the next time instant.

Algorithm 1 (The Kalman Filter).

0. Initialization:

$$\begin{aligned} k &:= 0 \\ \hat{x}_{0|-1} &:= \bar{x}_0 \\ P_{0|-1} &:= P_0 \end{aligned}$$

1. Correction:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - C_k\hat{x}_{k|k-1} - D_k u_k) \\ K_k &= P_{k|k-1}C_k^T(C_k P_{k|k-1} C_k^T + R_k)^{-1} \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1}C_k^T(C_k P_{k|k-1} C_k^T + R_k)^{-1} C_k P_{k|k-1} \end{aligned}$$

2a. One-step prediction ($S_k = 0$):

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + B_k u_k$$
$$P_{k+1|k} = A_k P_{k|k} A_k^T + N_k Q_k N_k^T$$

2b. One-step prediction ($S_k \neq 0$):

$$\hat{x}_{k+1|k} = (A_k - N_k S_k R_k^{-1} C_k)\hat{x}_{k|k} + B_k u_k + N_k S_k R_k^{-1} y_k$$
$$P_{k+1|k} = (A_k - N_k S_k R_k^{-1} C_k) P_{k|k} (A_k - N_k S_k R_k^{-1} C_k)^T + N_k (Q_k - S_k R_k^{-1} S_k^T) N_k^T$$

3. Put $k := k + 1$ and goto 1

$P_{k|k}$ and $P_{k|k-1}$ are the covariance of the estimation error:

$$P_{k|k} = \mathbb{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T\}$$
$$P_{k|k-1} = \mathbb{E}\{(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T\}$$

and are measures of the uncertainty of the estimate. The covariance of estimation error can also be updated as

$$P_{k+1|k} = A_k P_{k|k-1} A_k^T + N_k Q_k N_k^T - A_k K_k C_k P_{k|k-1} A_k^T.$$

The first two terms in the right-hand side represent natural evolution of uncertainty. The last term shows how much uncertainty the Kalman filter removes.

Remark 1 ($S_k \neq 0$). *Notice that for a sampled continuous-time model, the cross-covariance S_k can be nonzero even though the process and measurement noise in continuous time are uncorrelated. See [2]*

Theorem 1 (Optimality of the Kalman Filter 1). *Assume that the noise is white and Gaussian and uncorrelated with x_0, which is also Gaussian:*

$$\begin{bmatrix} w_k \\ v_k \end{bmatrix} \in \mathcal{N}(0, \Sigma_k), \quad x_0 \in \mathcal{N}(\bar{x}_0, P_0).$$

Then the Kalman filter gives the minimum-variance estimate of x_k. That is, the covariances $P_{k|k}$ and $P_{k|k-1}$ are the smallest possible. We also have that the estimates are the conditional expectations

$$\hat{x}_{k|k} = \mathbb{E}\{x_k | Y_k\}$$
$$\hat{x}_{k|k-1} = \mathbb{E}\{x_k | Y_{k-1}\}.$$

Proof. In the case $S_k = 0$, we show the result below. For the general case, see [1]. □

Remark 2. *A small covariance matrix is a covariance matrix with a small trace. The Kalman filter minimizes the trace of the covariance matrix. Hence, trace* $P_{k|k} \leq$ *trace* $\tilde{P}_k$ *where* $\tilde{P}_k$ *is the covariance of the estimation error of any other filter.*

The theorem says that there is no other filter that does a better job, not even a nonlinear one. Furthermore, one can show that the estimates are not only the minimum variance estimates, but also the maximum Bayesian a posteriori (MAP) estimate, see Section 3.2, which loosely speaking means that the estimate is the most likely value of x_k, given the information available.

Even if the noise and initial state are not Gaussian, we have the following result:

Theorem 2 (Optimality of the Kalman Filter 2). *Assume that the noise is white and uncorrelated with* x_0*. Then the Kalman filter is the optimal* linear *estimator in the sense that no other* linear *filter gives a smaller variance on the estimation error.*

Proof. See [1]. □

Of course, in the non-Gaussian case, nonlinear filters can do a much better job. One option is to use moving horizon estimators, see Section 3.2, or particle filters.

Example 1. *Consider the following simple example where we would like to estimate a scalar state* x *that is constant (no process noise):*

$$\begin{aligned} x_{k+1} &= x_k, \quad x_0 = x, \\ y_k &= x_k + v_k, \quad \mathbb{E}\{v_k^2\} = 1. \end{aligned}$$

We use the Kalman filter and compare to a Luenberger-type observer in the form

$$\hat{x}_{k+1} = \hat{x}_k + K(y_k - \hat{x}_k)$$

for some different (constant) values of K*. One such run is shown in Figure 3.1. As can be seen, the Kalman filter performs best. For large* K *the observer is sensitive to the measurement noise, and for small* K *it updates its estimate very slowly. The Kalman filter changes the gain over time and uses an optimal trade off.*

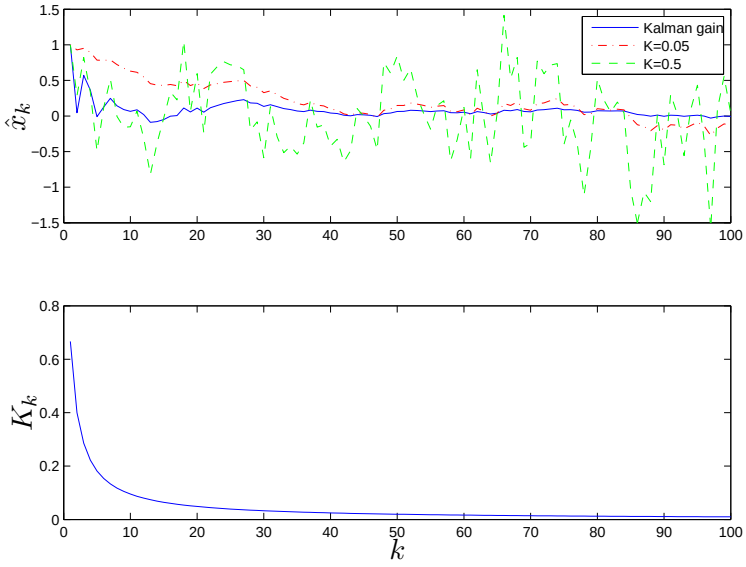

Figure 3.1: A run of the Kalman filter in Example 1 together with two other "standard" observers. Here $x = 0$, $\bar{x}_0 = 1$, and $P_0 = 2$.

Sensor Fusion

Assume that there are $p > 1$ sensors. Then the Kalman filter automatically weights the measurements and fuses the data into a single optimal estimate of the state. Assume for simplicity that the measurement noise of the sensors are uncorrelated. This means a diagonal R_k. Then the gain K_k is given by

$$K_k = P_{k|k-1}C_k^T \left(C_k P_{k|k-1} C_k^T + \begin{bmatrix} R_{k,11} & & \\ & \ddots & \\ & & R_{k,pp} \end{bmatrix} \right)^{-1}.$$

A large $R_{l,ii}$ (sensor l has a lot of measurement noise) leads to low influence of sensor l on the estimate.

Proof of Optimality of the Kalman Filter

We need some lemmas from probability theory to derive the Kalman filter.

Lemma 1. *Assume that the stochastic variables x and y are jointly distributed. Then the minimum-variance estimate $\hat{x}$ of x, given y is the conditional expectation*

$$\hat{x} = \mathbb{E}\{x|y\}.$$

That is

$$\mathbb{E}\{\|x - \hat{x}\|^2|y\} \leq \mathbb{E}\{\|x - f(y)\|^2|y\}$$

for any other estimate $f(y)$.

Proof. See, for example, [1]. □

Lemma 2. *Assume that x and y have a joint Gaussian distribution with mean and covariance*

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

Then the stochastic variable x, conditioned on the information y, is Gaussian with mean and covariance

$$\bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}) \quad \textit{and} \quad \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}.$$

That is,

$$\mathbb{E}\{x|y\} = \bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}).$$

Proof. See, for example, [1]. □

Another property we need is that if x, y, and z are jointly Gaussian, and the mean and covariances in Lemma 2 are conditioned on z, then the above formulas still holds given that y and z are known. Another property of the Gaussian distribution that we will use is that if x and y are Gaussian, then $x + y$ is also Gaussian.

The next lemma shows how means and covariances propagate.

Lemma 3. *Assume that*

$$x_{k+1} = A_k x_k + N_k w_k$$

and that $\mathbb{E}\{x_k\} = \bar{x}_k$, $\mathbb{E}\{(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T\} = P_k$, *and that* $\mathbb{E}\{w_k\} = 0$, $\mathbb{E}\{w_k w_k^T\} = Q_k$ *and* $\mathbb{E}\{w_k x_k^T\} = 0$. *Then*

$$\begin{aligned}
\bar{x}_{k+1} &= \mathbb{E}\{x_{k+1}\} = A_k \bar{x}_k \\
P_{k+1} &= \mathbb{E}\{(x_{k+1} - \bar{x}_{k+1})(x_{k+1} - \bar{x}_{k+1})^T\} = A_k P_k A_k^T + N_k Q_k N_k^T \\
P_{k+1,k} &= \mathbb{E}\{(x_{k+1} - \bar{x}_{k+1})(x_k - \bar{x}_k)^T\} = A_k P_k.
\end{aligned}$$

Proof. See, for example, [1]. □

Proof of Theorem 1

We prove now Theorem 1 in the case $S_k = 0$ and $u_k = 0$. We prove it in a number of steps. Essentially we are iteratively updating the mean and the covariance of the state and the measurement signal using Lemmas 1–3.

1. ("Correct") Start at time $k = 0$. By the assumptions and Lemma 3 the stochastic variable $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ is Gaussian with mean and covariance

$$\begin{bmatrix} \bar{x}_0 \\ C_0 \bar{x}_0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_0 & P_0 C_0^T \\ C_0 P_0 & C_0 P_0 C_0^T + R_0 \end{bmatrix}.$$

   Hence, x_0 conditioned on y_0 gives the minimum variance estimate

$$\begin{aligned}
\hat{x}_{0|0} &= \mathbb{E}\{x_0 | Y_0\} = \bar{x}_0 + P_0 C_0^T (C_0 P_0 C_0^T + R_0)^{-1} (y_0 - C_0 \bar{x}_0) \\
P_{0|0} &= P_0 - P_0 C_0^T (C_0 P_0 C_0^T + R_0)^{-1} C_0 P_0.
\end{aligned}$$

2. ("One-step predictor for state") Conditioning x_1 on y_0 and using that $S_k = 0$, gives a Gaussian distribution with mean and covariance

$$\hat{x}_{1|0} = \mathbb{E}\{x_1 | Y_0\} = A_0 \hat{x}_{0|0} \quad \text{and} \quad P_{1|0} = A_0 P_{0|0} A_0^T + N_0 Q_0 N_0^T.$$

3. ("One-step predictor for output") It follows that y_1 conditioned on y_0 is Gaussian with mean and covariance

$$\hat{y}_{1|0} = C_1\hat{x}_{1|0} \quad \text{and} \quad C_1P_{1|0}C_1^T + R_1$$

and

$$\mathbb{E}\{(y_1 - \hat{y}_{1|0})(x_1 - \hat{x}_{1|0})^T\} = C_1P_{1|0}.$$

4. ("Collect the elements") so that the stochastic variable $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ conditioned on y_0 is Gaussian and has mean and covariance

$$\begin{bmatrix} \hat{x}_{1|0} \\ C_1\hat{x}_{1|0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_{1|0} & P_{1|0}C_1^T \\ C_1P_{1|0} & C_1P_{1|0}C_1^T + R_1 \end{bmatrix}.$$

5. ("Correct") Now we are essentially back to Step 1, and we can repeat the steps with obvious changes of time indices.

3.2 Moving Horizon Estimation

In this section, we consider the nonlinear estimation problem. That is, the dynamics is given by the nonlinear equations

$$\begin{aligned} x_{k+1} &= f_k(x_k, w_k) \\ y_k &= h_k(x_k) + v_k \end{aligned} \tag{3.2}$$

with the constraints

$$x_k \in \mathbb{X}_k, \quad w_k \in \mathbb{W}_k, \quad w_k \in \mathbb{V}_k. \tag{3.3}$$

One can interpret the constraints $\mathbb{W}_k$ and $\mathbb{V}_k$ as w_k and v_k have some truncated (usually Gaussian) probability distribution. The interpretation of $\mathbb{X}_k$ is more complicated since its enforcement may result in acausality, see [5]. Nevertheless, in practice it can be useful. We have not included an input u_k in (3.2), but it is easy to include it by putting $f_k(x_k, w_k) := \bar{f}_k(x_k, u_k, w_k)$.

To estimate the state x_k in (3.2) from the measurements Y_k is much more difficult than for linear problems (3.1), that is solved with the Kalman filter. This is both due to the nonlinear dynamics and the constraints. We will here discuss two different methods: the extended Kalman filter and moving horizon estimation.

The Extended Kalman Filter

Often one of the simplest ways to get a good estimate of the state x_k in (3.2) is to linearize the equations and to apply the Kalman filter. This approach is called the *Extended Kalman Filter* (EKF). We use the following approximations of (3.2) (assuming that f_k and h_k are smooth):

$$\begin{aligned} f_k(x_k, w_k) &\approx f_k(\hat{x}_{k|k}, 0) + A_k(x_k - \hat{x}_{k|k}) + N_k w_k \\ h_k(x_k) &\approx h_k(\hat{x}_{k|k-1}) + C_k(x_k - \hat{x}_{k|k-1}) \end{aligned}$$

where

$$A_k = \frac{\partial}{\partial x} f_k(x,0)\Big|_{x=\hat{x}_{k|k}}, N_k = \frac{\partial}{\partial w} f_k(\hat{x}_{k|k}, w)\Big|_{w=0} \quad C_k = \frac{\partial}{\partial x} h_k(x,0)\Big|_{x=\hat{x}_{k|k-1}}.$$

Just as for the normal Kalman filter we assume that the noise is white, has zero mean, and

$$\mathbb{E}\{w_k w_l^T\} = Q_k \delta_{kl}, \quad \mathbb{E}\{v_k v_l^T\} = R_k \delta_{kl}, \quad \mathbb{E}\{w_k v_l^T\} = 0.$$

For simplicity we only consider the case $S_k = 0$ here. Just as before, we also assume that the initial state x_0 has a known mean $\bar{x}_0$ and covariance P_0.

Algorithm 2 (The Extended Kalman Filter).

0. Initialization:

$$\begin{aligned} k &:= 1 \\ \hat{x}_{0|-1} &:= \bar{x}_0 \\ P_{0|-1} &:= P_0 \end{aligned}$$

1. Corrector:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - h_k(\hat{x}_{k|k-1})) \\ K_k &= P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + R_k)^{-1} \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + R_k)^{-1} C_k P_{k|k-1} \end{aligned}$$

2. One-step predictor:

$$\begin{aligned} \hat{x}_{k+1|k} &= f_k(\hat{x}_{k|k}, 0) \\ P_{k+1|k} &= A_k P_{k|k} A_k^T + N_k Q_k N_k^T \end{aligned}$$

3. Put $k := k + 1$ and goto 1

As can be seen, these equations do not take the constraints (3.3) into account. Furthermore, it is hard to prove how and when the EKF works well, but in practice it often does. Next, we study the moving horizon estimator which can take the constraints (3.3) into account.

Bayesian MAP Estimates

Assume that the variables x and y are jointly distributed. Then the Bayesian *maximum a posteriori* (MAP) estimate of x given y is defined as

$$\hat{x} = \arg\max_x p(x|y).$$

Essentially, this means that we want to find the most likely value of x, given y. This is the criteria we will use to motivate the estimates of x_k that moving horizon estimation delivers. Notice that the estimates the standard Kalman filter delivers in the linear Gaussian case coincides with the MAP estimate.

We consider a restricted class of systems (3.2) given by

$$\begin{aligned} x_{k+1} &= f_k(x_k) + w_k \\ y_k &= h_k(x_k) + v_k. \end{aligned} \tag{3.4}$$

We assume that the noise is mutually independent, and the initial state and noise have (truncated) Gaussian distributions. That is

$$\begin{aligned} p_{w_k}(w) &\propto \exp\left(-\frac{1}{2} w^T Q_k^{-1} w\right), \quad w \in \mathbb{W}_k \\ p_{v_k}(v) &\propto \exp\left(-\frac{1}{2} v^T R_k^{-1} v\right), \quad v \in \mathbb{V}_k \\ p_{x_0}(x) &\propto \exp\left(-\frac{1}{2} (x - \bar{x}_0)^T P_0^{-1} (x - \bar{x}_0)\right), \quad x \in \mathbb{X}_0, \end{aligned}$$

where the constraints are closed and convex. We want to find the MAP estimates of the states $\{x_0, \ldots, x_T\}$ given the measurements $\{y_0, \ldots, y_{T-1}\}$. Using Bayes' rule one can derive that

$$\begin{aligned} p(x_0, \ldots, x_T | y_0, \ldots, y_{T-1}) &\propto p_{x_0}(x_0) \prod_{k=0}^{T-1} p_{v_k}(y_k - h_k(x_k)) p(x_{k+1}|x_k) \\ p(x_{k+1}|x_k) &= p_{w_k}(x_{k+1} - f_k(x_k)), \end{aligned}$$

see [4] for details. Then, using the MAP estimate criterion

$$\begin{aligned}
&\arg \max_{\{x_0,\dots,x_T\}} p(x_0,\dots,x_T|y_0,\dots,y_{T-1}) \\
&= \arg \max_{\{x_0,\dots,x_T\}} \sum_{k=0}^{T-1} \log p_{v_k}(y_k - h_k(x_k)) + \log p(x_{k+1}|x_k) + \log p_{x_0}(x_0) \\
&= \arg \min_{\{x_0,\dots,x_T\}} \sum_{k=0}^{T-1} \|y_k - h_k(x_k)\|^2_{R_k^{-1}} + \|x_{k+1} - f(x_k)\|^2_{Q_k^{-1}} + \|x_0 - \bar{x}_0\|^2_{P_0^{-1}} \\
&= \arg \min_{\{x_0,\dots,x_T\}} \sum_{k=0}^{T-1} \|v_k\|^2_{R_k^{-1}} + \|w_k\|^2_{Q_k^{-1}} + \|x_0 - \bar{x}_0\|^2_{P_0^{-1}},
\end{aligned}$$

where the minimization is done subject to the dynamics and all constraints, and where $\|x\|^2_Q = x^T Q x$. It is common to re-parameterize the problem with an initial state and a process noise sequence. Once these are found, it is easy to find the state sequence by just using the model (3.2). Hence, the minimization problem is often formulated as

$$\min_{x_0,\{w_k\}_{k=0}^{T-1}} \sum_{k=0}^{T-1} L_k(w_k, v_k) + \Gamma(x_0), \tag{3.5}$$

for some positive functions L_k and Γ. In the above example, the functions are

$$L_k(w,v) = \|v\|^2_{R_k^{-1}} + \|w\|^2_{Q_k^{-1}}, \quad \Gamma(x_0) = \|x_0 - \bar{x}_0\|^2_{P_0^{-1}}.$$

In moving horizon estimation, a criterion in the form (3.5) is usually the starting point. We have here motivated that the solution to such a minimization problem can lead to MAP estimates, given that the functions L_k and Γ are chosen properly. Even when the system is not in the form (3.4), it is still often reasonable to choose an estimate based on (3.5) with quadratic cost functions, even though it will not give the exact MAP estimates.

We note that

- to obtain the estimate $\{x_0,\dots,x_T\}$, or $x_0, \{w_k\}_{k=0}^{T-1}$, we need to solve a constrained optimization problem;
- the optimization problem may have multiple local minima;
- the problem complexity grows at least linearly with the horizon T;
- for linear systems (3.1) without constraints, the minimization problem can be solved recursively and leads to the Kalman filter, see [5].

The Moving Horizon Estimator

(This presentation is based on [5], and the notation is essentially the same.)

Based on the discussion in the previous section, it makes sense to minimze the cost function

$$\Phi_T^* := \min_{x_0, \{w_k\}_{k=0}^{T-1}} \sum_{k=0}^{T-1} L_k(w_k, v_k) + \Gamma(x_0), \tag{3.6}$$

subject to (3.2)–(3.3) for a given stage cost function $L_k(\cdot) \geq 0$ and an initial penalty function $\Gamma(\cdot) \geq 0$. The initial penalty summarizes the a priori knowledge of the state, and $\Gamma(\bar{x}_0) = 0$ and $\Gamma(x) > 0$ for $x \neq \bar{x}_0$. The solution to (3.6) is denoted by

$$\hat{x}_{0|T-1}, \quad \{\hat{w}_{k|T-1}\}_{k=0}^{T-1},$$

and the estimations of the state at times $0 \leq k \leq T$ are given by

$$\hat{x}_{k|T-1} = x(k; \hat{x}_{0|T-1}, 0, \{\hat{w}_k\}_{k=0}^{T-1}),$$

computed by iterating (3.2), and we define

$$\hat{x}_T := \hat{x}_{T|T-1}.$$

For simplicity, we do not perform the corrector step in this very short introduction of MHE.

The idea of MHE is to reapeatedly solve (3.6) for increasing T as new measurements arrive. But to reduce computation cost, we use a moving window (horizon) of length N:

$$\begin{aligned}
\Phi_T^* &= \min_{x_0, \{w_k\}_{k=0}^{T-1}} \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \sum_{k=0}^{T-N-1} L_k(w_k, v_k) + \Gamma(x_0) \\
&= \min_{z \in \mathcal{R}_{T-N}, \{w_k\}_{k=T-N}^{T-1}} \left(\sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \mathcal{Z}_{T-N}(z) \right),
\end{aligned}$$

where $\mathcal{Z}_{T-N}(\cdot)$ is called the *arrival cost* at time $T-N$ and $\mathcal{R}_{T-N}$ is the reachable set of the state space subject to all constraints and dynamics of the system. The arrival cost at z at time T is given by

$$\begin{aligned}
\mathcal{Z}_T(z) = &\min_{x_0, \{w_k\}_{k=0}^{T-1}} \sum_{k=0}^{T-1} L_k(w_k, v_k) + \Gamma(x_0) \\
&\text{subject to constraints, dynamics, and} \quad x_T = z.
\end{aligned}$$

Hence, we keep the number of decision variables constant as T increases, and store the knowledge from the measurements outside the moving horizon in the arrival cost function. The problem is of course to compute $\mathcal{Z}_T$.

Except in the linear unconstrained and quadratic cost case, the arrival cost is hard to obtain in algebraic expressions. In the linear case, however, we have

$$\mathcal{Z}_T(z) = (z - \hat{x}_T)^T P_T^{-1}(z - \hat{x}_T) + \Phi_T^*, \tag{3.7}$$

where $P_k := P_{k|k-1}$ is the covariance from the Kalman filter. That is, it satisfies

$$P_{k+1} = A_k P_k A_k^T + N_k Q_k N_k^T - A_k P_k C_k^T (R_k + C_k P_k C_k^T)^{-1} C_k P_k A_k^T.$$

As can be seen from (3.7), chosing $z = \hat{x}_T$ minimizes the arrival cost. Hence, to change the estimate of the state at time T, the following measurements must give sufficient reason to do so.

Approximations

Since it is often not tractable to compute $\mathcal{Z}_T$ exactly, one can (and must) often use approximations. We here discuss two types of approximations. After that, we discuss the stability properties associated with them.

The approximate MHE problem is given by

$$\hat{\Phi}_T = \min_{z \in \mathcal{R}_{T-N}, \{w_k\}_{k=T-N}^{T-1}} \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \hat{\mathcal{Z}}_{T-N}(z) \tag{3.8}$$

for some approximation $\hat{\mathcal{Z}}$ of the arrival cost. We denote a solution to (3.8) by

$$z^*, \quad \{\hat{w}_{k|T-1}^{mh}\}_{k=T-N}^{T-1}$$

and the estimate of the state are computed as

$$\hat{x}_{k|T-1}^{mh} = x(k; z^*, T-N, \{\hat{w}_{j|T-1}^{mh}\})$$

and define $\hat{x}_k^{mh} := \hat{x}_{k|k-1}^{mh}$.

The simplest possible approximation of $\mathcal{Z}_{T-N}$ is to pick

$$\hat{\mathcal{Z}}_{T-N}(z) = \hat{\Phi}_{T-N}, \tag{3.9}$$

meaning that in the estimation at time T, we do not take the meaurements before time $T-N$ into account. This is because $\hat{\Phi}_{T-N}$ is a constant. This

may be of interest for stability reasons, see below, but generally we would like to include some of the previous knowledge for performance reasons.

Another choice is to use EKF around the estimated trajectory $\{\hat{x}_k^{mh}\}$ and define

$$\hat{\mathcal{Z}}_{T-N}(z) = (z - \hat{x}_{T-N}^{mh})^T P_{T-N}^{-1}(z - \hat{x}_{T-N}^{mh}). \tag{3.10}$$

If the stage cost L_k is not quadratic, one picks

$$R_k^{-1} = \left.\frac{\partial L_k(0,v)}{\partial v \partial v^T}\right|_{\hat{x}_k^{mh}}, \quad Q_k^{-1} = \left.\frac{\partial L_k(w,v)}{\partial w \partial w^T}\right|_{w=0,\hat{x}_k^{mh}}$$

in the predictor and corrector steps.

Some Stability Results

To prove that (approximate) MHE is asymptotically stable is not trivial. We will here give an introduction to the available results, and discuss some sufficient conditions. The real proofs can be found in [4, 5] and rely on Lyapunov function arguments.

Asymptotic stability here means that when there is no noise in the system, i.e., $w_k = v_k = 0$, then $\|x_k - \hat{x}_k^{mh}\| \to 0$ when $k \to \infty$ if the initial estimate is good enough. When bounded noise is present, we also want that $\|x_k - \hat{x}_k^{mh}\|$ is uniformly bounded for all k.

The fist very natural assumption is that the system should be *uniformly observable* over the chosen horizon N. This means that different initial states give different outputs over the horizon N. More precisely, we require that there is a K*-function* φ such that

$$\varphi(\|x_1 - x_2\|) \leq \sum_{j=k}^{k+N-1} \|y(j; x_1, k) - y(j; x_2, k)\| \tag{3.11}$$

for all k and initial states x_1 and x_2.

Definition 1 (K-function [5]). *A function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is a K-function if it is continuous, strictly monotone increasing, $\varphi(x) > 0$, $x \neq 0$, $\varphi(0) = 0$, and $\lim_{x\to\infty} \varphi(x) = \infty$.*

We also require that the approximate arrival cost satisfies

$$0 \leq \hat{\mathcal{Z}}_k(z) - \hat{\Phi}_k \leq \gamma(\|z - \hat{x}_k^{mh}\|),$$

for some K-function γ. We call this condition (**C1**). It means that there is a cost to change the initial best guess ($\hat{x}_k^{mh}$). Unless we have some good

more recent measurements, there should be nothing to gain by changing the estimate.

The second condition we put on the arrival cost is that it should not add any new information to the problem. We require that

$$\hat{\mathcal{Z}}_T(p) \leq \min_{z \in \mathcal{R}_{T-N}, \{w_k\}_{k=T-N}^{T-1}} \left(\sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \hat{\mathcal{Z}}_{T-N}(z) \right)$$

subject to $x_T = p$, for all T. We call this condition (**C2**). Essentially it means that there is some forgetting of the previous estimation. For instance, (3.9) satisfies this (we have complete forgetting). If we violate this inequality, we are essentially saying that the estimate $\hat{x}_T = p$ is worse than we actually have reason the believe.

The condition (**C2**) is hard to prove in many cases, for instance in the EKF approximation (3.10). However, the EKF approximation satisfies (**C2**) when the system dynamics is linear, L_k is quadratic, and the constraints are convex. How (**C2**) can be relaxed for more general cases is shown in [5].

Under the assumptions that the system is uniformly observable with the chosen N, and **(C1)** and **(C2)** are satisfied (along with some assumptions on f, h, L, Γ, see [5]) then MHE is asymptotically stable. Notice, however, that these conditions are only sufficient conditions for stability. The MHE can be stable even if they are violated. Still, the conditions are natural and it makes sense to check them if possible.

Bibliography

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Dover Publications, Inc., 2005.

[2] K. J. Åström. *Introduction to Stochastic Control Theory*. Dover Publications, Inc., 2006.

[3] R. M. Murray, editor. *Control in an Information Rich World: Report of the Panel on Future Direcitons in Control, Dynamics and Systems*. SIAM, 2003. Available at `http://www.cds.caltech.edu/~murray/cdspanel`.

[4] C. V. Rao. *Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete-Time Systems*. PhD thesis, University of Wisconsin-Madison, Feb 2000.

[5] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transaction on Automatic Control*, 48(2):246–258, 2003.

[6] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, Upper Saddle River, New Jersey, 1996.