Model Predictive Control

Lecture: Moving Horizon Estimation

Panos Patrinos

ESAT-STADIUS, KU Leuven

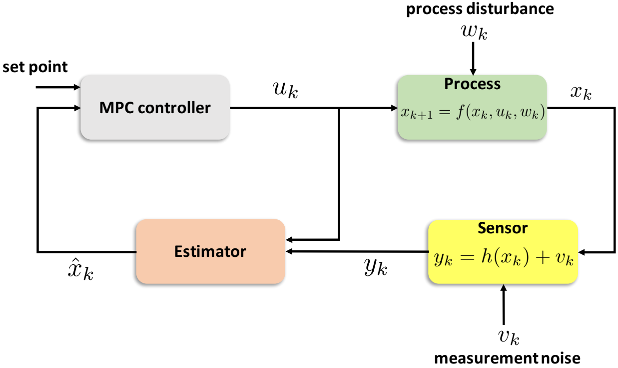
MPC and state estimation

- MPC is a state feedback control law. It needs state measurements.
- Typically not all states are measured either because it is too costly or simply because it is impossible.
- What is measured is the system output.
- Disturbances enter the system at two places.
- Process disturbances, denoted by w, account for modelling errors as well as for process variations.
- Output disturbances v account for modelling errors and for (random) sensor errors

state estimation

- Retrieve the states that have generated the observed outputs, given the dynamical model.
- Recover from a wrong initial guess
- Find good state estimates despite noisy measurements.

Control and estimation



- Given the data $\mathbf{y} = \{y_0, \ldots, y_T\}$ find the "best" estimate $\hat{x}_{T+m}$ of x_{T+m}
  - $m = 0$: filtering
  - $m > 0$: prediction
  - $m < 0$: smoothing

Class of systems

- Since the inputs are exogenous to the estimator, we can "hide" them by considering

$$x_{k+1} = f_k(x_k, w_k)$$

  where $f_k(x_k, w_k) = f(x_k, u_k, w_k)$.

- In case the model also contains unknown parameters

$$x_{k+1} = f(x_k, u_k, w_k, p)$$

  we can consider an augmented state vector $\tilde{x}_k = (x_k, p_k)$ and

$$f_k(\tilde{x}_k, w_k) = \begin{bmatrix} f(x_k, u_k, w_k, p_k) \\ p_k \end{bmatrix}$$

  such that $p_{k+1} = p_k$.

State estimation and MAP

- Given noisy output measurements $\mathbf{y} = (y_0, \dots, y_{T-1})$ estimate most likely state $\mathbf{x} = (x_0, \dots, x_T)$ and disturbance $\mathbf{w} = (w_0, \dots, w_{T-1})$ sequences

$$
\begin{aligned}
\underset{\mathbf{x},\mathbf{w}}{\text{maximize}} \;& p(\mathbf{x}, \mathbf{w} \mid \mathbf{y}) \\
\text{subject to } \;& x_{k+1} = f_k(x_k, w_k), && k \in \mathbb{N}_{[0,T-1]} \\
& y_k = h_k(x_k) + v_k, && k \in \mathbb{N}_{[0,T-1]}
\end{aligned}
$$

- Also known as **maximum a posteriori (MAP)** estimation.
- $p(\mathbf{x}, \mathbf{w} \mid \mathbf{y})$ is called the posterior probability.

Stochastic assumptions

- x_0, $w_0, \dots, w_{T-1}$, $v_0, \dots, v_{T-1}$ are iid random vectors with PDFs

$$
\begin{aligned}
p_v(v_k) &= c_v \exp(-\ell_v(v_k)) \\
p_w(w_k) &= c_w \exp(-\ell_w(w_k)) \\
p_x(x_0) &= c_x \exp(-\ell_x(x_0))
\end{aligned}
$$

State estimation and MAP

- Applying Bayes theorem, posterior is expressed as

$$p(\mathbf{x}, \mathbf{w} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) p(\mathbf{x}, \mathbf{w})}{p(\mathbf{y})} = C \overbrace{p(\mathbf{y} \mid \mathbf{x}, \mathbf{w})}^{\text{likelihood}} \overbrace{p(\mathbf{x}, \mathbf{w})}^{\text{prior}}$$

  ($C = p(\mathbf{y})^{-1}$ is constant)
- The prior density is

$$\begin{aligned} p(\mathbf{x}, \mathbf{w}) &= p(x_0, \dots, x_T, w_0, \dots, w_{T-1}) \\ &= \begin{cases} 0 & \text{if not } x_{k+1} = f_k(x_k, w_k) \text{ for all } k \in \mathbb{N}_{[0, T-1]} \\ p_x(x_0) \prod_{k=0}^{T-1} p_w(w_k) & \text{otherwise} \end{cases} \end{aligned}$$

- The likelihood density is

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) &= p(y_0, \dots, y_{T-1} \mid x_0, \dots, x_T, w_0, \dots, w_{T-1}) \\ &= \prod_{k=0}^{T-1} p(y_k \mid x_k, w_k) = \prod_{k=0}^{T-1} p_v(y_k - h_k(x_k)) \end{aligned}$$

State estimation and MAP

- MAP estimator solves

$$
\begin{aligned}
\operatorname*{argmax}_{\mathbf{x},\mathbf{w}} p(\mathbf{x}, \mathbf{w} \mid \mathbf{y}) &= \operatorname*{argmax}_{\mathbf{x},\mathbf{w}} \log p(\mathbf{x}, \mathbf{w} \mid \mathbf{y}) \\
&= \operatorname*{argmax}_{\mathbf{x},\mathbf{w}} \log p(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) + \log p(\mathbf{x}, \mathbf{w}) \\
&= \operatorname*{argmin}_{\mathbf{x},\mathbf{w}} - \log p(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) - \log p(\mathbf{x}, \mathbf{w}) \\
&= \operatorname*{argmin}_{\mathbf{x},\mathbf{w}} - \log \prod_{k=0}^{T-1} p_v(y_k - h_k(x_k)) - \log p_x(x_0) \prod_{k=0}^{T-1} p_w(w_k) \\
&= \operatorname*{argmin}_{\mathbf{x},\mathbf{w}} \ell_x(x_0) + \sum_{k=0}^{T-1} \ell_v(y_k - h_k(x_k)) + \ell_w(w_k)
\end{aligned}
$$

- We arrive at the **full information estimation** problem

$$
\begin{aligned}
&\text{minimize } \ell_x(x_0) + \sum_{k=0}^{T-1} \ell_v(y_k - h_k(x_k)) + \ell_w(w_k) \\
&\text{subject to } x_{k+1} = f_k(x_k, w_k) \qquad\qquad k \in \mathbb{N}_{[0, T-1]}
\end{aligned}
$$

State estimation and MAP

$$V_T^\star = \min \ell_x(x_0) + \sum_{k=0}^{T-1} \overbrace{\ell_v(y_k - h_k(x_k)) + \ell_w(w_k)}^{\ell_k(x_k, w_k)}$$

$$\text{s.t. } x_{k+1} = f_k(x_k, w_k) \qquad k \in \mathbb{N}_{[0, T-1]}$$

- This is called the **Full Information Estimation (FIE) problem**
- It yields estimates $(\hat{x}_0, \ldots, \hat{x}_T)$ given measurements $(y_0, \ldots, y_{T-1})$
- We are typically interested in $\hat{x}_T$ (prediction of state at time T)
- $\ell_x(x_0)$, $\ell_k(x_k, w_k)$ are called the **stage costs**
- Similar to **OCP** but no initial condition, process noise replaces control input
- Problem size increases with T

State estimation and MAP

$$
\begin{aligned}
V_T^\star = \min\ & \ell_x(x_0) + \sum_{k=0}^{T-1} \overbrace{\ell_v(y_k - h_k(x_k)) + \ell_w(w_k)}^{\ell_k(x_k, w_k)} \\
\text{s.t. } & x_{k+1} = f_k(x_k, w_k) \qquad\qquad k \in \mathbb{N}_{[0, T-1]}
\end{aligned}
$$

Choices for stage costs

- Typically quadratic $\ell_x(x_0) = \frac{1}{2}\|x_0 - \bar{x}_0\|^2_{(P_0^-)^{-1}}$
- Corresponds to $x_0 \sim \mathcal{N}(\bar{x}_0, P_0^-)$ (normally distributed with mean $\bar{x}_0$ and covariance matrix P_0^-)
- Recall for $z \sim \mathcal{N}(\bar{z}, \Sigma)$ (normally distributed with mean $\bar{z}$ and covariance Σ)

$$
p_z(z) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left[-\frac{1}{2}(z - \bar{z})^\top \Sigma^{-1}(z - \bar{z})\right]
$$

- Good estimate $\bar{x}_0 \rightarrow$"small" covariance$\rightarrow$large penalty for deviating from $\bar{x}_0$

State estimation and MAP

$$V_T^\star = \min\ \ell_x(x_0) + \sum_{k=0}^{T-1} \overbrace{\ell_v(y_k - h_k(x_k)) + \ell_w(w_k)}^{\ell_k(x_k, w_k)}$$

$$\text{s.t. } x_{k+1} = f_k(x_k, w_k) \qquad\qquad k \in \mathbb{N}_{[0,T-1]}$$

Choices for stage costs

- $w_0, \dots, w_{T-1},\ v_0, \dots, v_{T-1}$ are **normally distributed** with means and covariance matrices

$$\mathbf{E}(w_k) = 0, \qquad Q = \mathbf{E}(w_k w_k^\top)$$
$$\mathbf{E}(v_k) = 0, \qquad R = \mathbf{E}(v_k v_k^\top)$$

- Then

$$\ell_w(w_k) = \tfrac{1}{2}\|w\|_{Q^{-1}}^2, \qquad \ell_v(v_k) = \tfrac{1}{2}\|v_k\|_{R^{-1}}^2$$

State estimation and MAP

$$V_T^\star = \min \ell_x(x_0) + \sum_{k=0}^{T-1} \overbrace{\ell_v(y_k - h_k(x_k)) + \ell_w(w_k)}^{\ell_k(x_k, w_k)}$$

$$\text{s.t. } x_{k+1} = f_k(x_k, w_k) \qquad\qquad k \in \mathbb{N}_{[0, T-1]}$$

Choices for stage costs

$$\ell_v(v_k) = \beta_v \|v_k\|_1$$

- Models v_k noise whose components follow Laplace distribution
- Recall that $z \sim \text{Laplace}(\bar{z}, \beta)$ if it has PDF

$$p_z(z) = \tfrac{1}{2\beta} \exp\left[-\tfrac{|z-\bar{z}|}{\beta}\right]$$

- Large residuals are penalized less–robustness against outliers

State estimation and MAP

$$
\begin{aligned}
V_T^\star = \min\ & \ell_x(x_0) + \sum_{k=0}^{T-1} \overbrace{\ell_v(y_k - h_k(x_k)) + \ell_w(w_k)}^{\ell_k(x_k, w_k)} \\
\text{s.t. }\ & x_{k+1} = f_k(x_k, w_k) \qquad\qquad k \in \mathbb{N}_{[0, T-1]}
\end{aligned}
$$

Choices for stage costs

$$
\ell_v(v) = \begin{cases} \frac{1}{2} v^2 & |v| \le \delta \\ \alpha\left(|v| - \frac{1}{2}\delta\right) & |v| \ge \delta \end{cases} \qquad \text{Huber Loss}
$$

- Models v_k with fat-tailed distribution
- Large residuals are penalized less–robustness against outliers

Full information Estimation (FIE)

$$\begin{aligned} \text{minimize} \; & \ell_x(x_0) + \sum_{k=0}^{T-1} \ell_k(x_k, w_k) \\ \text{subject to} \; & x_{k+1} = f_k(x_k, w_k) \qquad\qquad k \in \mathbb{N}_{[0,T-1]} \end{aligned}$$

- Problem size grows as T increases!
- We apply principle of **nested minimization** going **forward** in time.
- Also known as **forward dynamic programming**.

Full information Estimation (FIE)

$$
\begin{aligned}
\hat{V}_T^\star &= \min_{\{x_k, w_k\}_{k=0}^{T-1}, x_T} \left\{ \ell_x(x_0) + \sum_{k=0}^{T-1} \ell_k(x_k, w_k) \mid x_{k+1} = f_k(x_k, w_k), k \in \mathbb{N}_{[0, T-1]} \right\} \\
&= \min_{\{x_k, w_k\}_{k=1}^{T-1}, x_T} \left\{ \overbrace{\min_{x_0, w_0} \{\ell_x(x_0) + \ell_0(x_0, w_0) \mid x_1 = f_0(x_0, w_0)\}}^{V_1^\star(x_1)} \right. \\
&\left. + \sum_{k=1}^{T-1} \ell_k(x_k, w_k) \mid x_{k+1} = f_k(x_k, w_k), k \in \mathbb{N}_{[1, T-1]} \right\} \\
&= \min_{\{x_k, w_k\}_{k=1}^{T-1}, x_T} \{ V_1^\star(x_1) + \sum_{k=1}^{T-1} \ell_k(x_k, w_k) \mid x_{k+1} = f_k(x_k, w_k), k \in \mathbb{N}_{[1, T-1]} \} \\
&\qquad \vdots \\
&= \min_{x_T} \overbrace{\min_{x_{T-1}, w_{T-1}} \{ V_{T-1}^\star(x_{T-1}) + \ell_{T-1}(x_{T-1}, w_{T-1}) \mid x_T = f_k(x_{T-1}, w_{T-1}) \}}^{V_T^\star(x_T)} \\
&= \min_{x_T} V_T^\star(x_T)
\end{aligned}
$$

Full information Estimation (FIE)

- Leads to the **forward DP** algorithm

$$
\begin{aligned}
V_0^\star(x_0) &= \ell_x(x_0) \\
V_{k+1}^\star(x_{k+1}) &= \min_{x_k, w_k} \{ V_k^\star(x_k) + \ell(x_k, w_k) \mid x_{k+1} = f(x_k, w_k) \}, \quad k \in \mathbb{N}_{[0, T-1]} \\
\hat{V}_T^\star &= \min_{x_T} V_T^\star(x_T)
\end{aligned}
$$

- $V_k^\star$, $k = 0, \dots, T$ are called **arrival costs**
- $V_k^\star$ summarizes the effect of the data $(y_0, \dots, y_{k-1})$ on the state x_k
- Can be viewed as negative logarithm of $p(x_k \mid y_0, \dots, y_{k-1})$

Full information Estimation and Kalman Filter

$$\begin{aligned} V_0^\star(x_0) &= \ell_x(x_0) \\ V_{k+1}^\star(x_{k+1}) &= \min_{x_k, w_k} \{ V_k^\star(x_k) + \ell_k(x_k, w_k) \mid x_{k+1} = f(x_k, w_k) \}, \quad k \in \mathbb{N}_{[0, T-1]} \\ \hat{V}_T^\star &= \min_{x_T} V_T^\star(x_T) \end{aligned}$$

- If we know arrival cost $V_T^\star$, to estimate the state at time T we compute

$$\hat{x}_T = \operatorname*{argmin}_{x_T} V_T^\star(x_T)$$

- As in optimal control, it is very hard to compute arrival costs explicitly
- There is one case where this is possible: The **Kalman Filter**.

Kalman Filter and Least Squares

Assume

- Linear dynamics (typically $b_k = Bu_k$)

$$x_{k+1} = Ax_k + b_k + w_k \qquad\qquad y_k = Cx_k + v_k$$

- Quadratic costs

$$\ell_x(x) = \tfrac{1}{2}\|x - \bar{x}_0\|^2_{(P_0^-)^{-1}} \qquad \ell_w(w) = \tfrac{1}{2}\|w\|^2_{Q^{-1}} \qquad \ell_v(v) = \tfrac{1}{2}\|v\|^2_{R^{-1}}$$

- Full information problem becomes linear least squares

$$\underset{x_0,\ldots,x_T}{\text{minimize}}\ \tfrac{1}{2}\|x_0 - \bar{x}_0\|^2_{(P_0^-)^{-1}} + \tfrac{1}{2}\sum_{k=0}^{T-1} \|y_k - Cx_k\|^2_{R^{-1}} + \|x_{k+1} - (Ax_k + b_k)\|^2_{Q^{-1}}$$

- We used state dynamics to eliminate process noise w_k.
- Can be solved by batch optimization.
- Each time we receive new measurement we need to solve a larger problem.
- Can avoid this using forward DP and smart linear algebra $\rightarrow$ Kalman Filter

Kalman Filter and Forward DP

$$V_0^\star(x_0) = \tfrac{1}{2}\|x_0 - \bar{x}_0\|^2_{(P_0^-)^{-1}}$$

$$V_{k+1}^\star(x_{k+1}) = \min_{x_k} V_k^\star(x_k) + \tfrac{1}{2}\|y_k - Cx_k\|^2_{R^{-1}} + \tfrac{1}{2}\|x_{k+1} - (Ax_k + b_k)\|^2_{Q^{-1}}, \ \ k \in \mathbb{N}_{[0,T-1]}$$

$$V_{T+1}^\star = \min_{x_T} V_T^\star(x_T)$$

Quadratic Sum

Let H_1, H_2 be symmetric positive definite and consider

$$q(x) = \tfrac{1}{2}\|x - \bar{x}\|^2_{H_1^{-1}} + \tfrac{1}{2}\|\Gamma x - b\|^2_{H_2^{-1}}.$$

Then

1. $q(x) = \frac{1}{2}\|x - x^\star\|^2_{M^{-1}} + \frac{1}{2}\|\Gamma\bar{x} - b\|^2_{S^{-1}}$

$$\text{where} \quad S = H_2 + \Gamma H_1 \Gamma^\top \qquad M = H_1 - H_1\Gamma^\top S^{-1}\Gamma H_1$$
$$x^\star = \bar{x} + L(b - \Gamma\bar{x}) \qquad L = H_1\Gamma^\top S^{-1}.$$

2. $x^\star = \operatorname{argmin}_{x\in\mathbb{R}^n} q(x)$ and $\min q(x) = \frac{1}{2}\|\Gamma\bar{x} - b\|^2_{S^{-1}}$.

Kalman Filter and Forward DP

Proof

Assume $x^\star$ is a minimizer of q. It satisfies $\nabla q(x^\star) = 0$ or

$$H_1^{-1}(x^\star - \bar{x}) + \Gamma^\top H_2^{-1}(\Gamma x^\star - b) = 0$$

Solving in terms of $x^\star$

$$\begin{aligned}
x^\star &= [H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]^{-1}(H_1^{-1}\bar{x} + \Gamma^\top H_2^{-1} b) \\
&= \bar{x} + [H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]^{-1}(H_1^{-1}\bar{x} - [H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]\bar{x} + \Gamma^\top H_2^{-1} b) \\
&= \bar{x} + [H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]^{-1}\Gamma^\top H_2^{-1}(b - \Gamma\bar{x}) \\
&= \bar{x} + [H_1 - H_1\Gamma^\top(H_2 + \Gamma H_1 \Gamma^\top)^{-1}\Gamma H_1]\Gamma^\top H_2^{-1}(b - \Gamma\bar{x}) \\
&= \bar{x} + H_1\Gamma^\top[I - (H_2 + \Gamma H_1\Gamma^\top)^{-1}\Gamma H_1 \Gamma^\top]H_2^{-1}(b - \Gamma\bar{x}) \\
&= \bar{x} + H_1\Gamma^\top[I - S^{-1}(S - H_2)]H_2^{-1}(b - \Gamma\bar{x}) \\
&= \bar{x} + H_1\Gamma^\top S^{-1}(b - \Gamma\bar{x})
\end{aligned}$$

where the second equality follows by adding and subtracting $\bar{x} = [H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]^{-1}[H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]\bar{x}$ and rearranging, the fourth equality follows by the matrix inversion lemma

$$[H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]^{-1} = H_1 - H_1\Gamma^\top(H_2 + \Gamma H_1\Gamma^\top)^{-1}\Gamma H_1$$

and the sixth equality by definition of S.

Kalman Filter and Forward DP

We will next calculate $q(x^\star)$:

$$\begin{aligned}
q(x^\star) &= \tfrac{1}{2}\|x^\star - \bar{x}\|^2_{H_1^{-1}} + \tfrac{1}{2}\|\Gamma x^\star - b\|^2_{H_2^{-1}} \\
&= \tfrac{1}{2}\|H_1\Gamma^\top S^{-1}(b - \Gamma\bar{x})\|^2_{H_1^{-1}} + \tfrac{1}{2}\|(I - \Gamma H_1 \Gamma^\top S^{-1})(b - \Gamma\bar{x})\|^2_{H_2^{-1}} \\
&= \tfrac{1}{2}\|H_1\Gamma^\top S^{-1}(b - \Gamma\bar{x})\|^2_{H_1^{-1}} + \tfrac{1}{2}\|(I - (S - H_2)S^{-1})(b - \Gamma\bar{x})\|^2_{H_2^{-1}} \\
&= \tfrac{1}{2}\|H_1\Gamma^\top S^{-1}(b - \Gamma\bar{x})\|^2_{H_1^{-1}} + \tfrac{1}{2}\|H_2 S^{-1}(b - \Gamma\bar{x})\|^2_{H_2^{-1}} \\
&= \tfrac{1}{2}(b - \Gamma\bar{x})^\top S^{-1}\Gamma H_1 \Gamma^\top S^{-1}(b - \Gamma\bar{x}) + \tfrac{1}{2}(b - \Gamma\bar{x})^\top S^{-1} H_2 S^{-1}(b - \Gamma\bar{x}) \\
&= \tfrac{1}{2}(b - \Gamma\bar{x})^\top S^{-1}(\Gamma H_1 \Gamma^\top + H_2) S^{-1}(b - \Gamma\bar{x}) \\
&= \tfrac{1}{2}(b - \Gamma\bar{x})^\top S^{-1}(b - \Gamma\bar{x}) = \tfrac{1}{2}\|\Gamma\bar{x} - b\|^2_{S^{-1}}
\end{aligned}$$

Since q is quadratic, the second-order Taylor expansion around any point is exact, globally. Expanding around $x^\star$

$$\nabla q(x^\star) = 0$$

$$\begin{aligned}
q(x) &= q(x^\star) + \cancel{\nabla q(x^\star)^\top (x - x^\star)} + \tfrac{1}{2}(x - x^\star)^\top \nabla^2 q(x^\star)(x - x^\star) \\
&= q(x^\star) + \tfrac{1}{2}(x - x^\star)^\top \nabla^2 q(x^\star)(x - x^\star) \\
&= q(x^\star) + \tfrac{1}{2}(x - x^\star)^\top M^{-1}(x - x^\star) \\
&= \tfrac{1}{2}\|x - x^\star\|^2_{M^{-1}} + \tfrac{1}{2}\|\Gamma\bar{x} - b\|^2_{S^{-1}}
\end{aligned}$$

where we made use of $[H_1^{-1} + \Gamma^\top H_2^{-1}\Gamma]^{-1} = H_1 - H_1\Gamma^\top S^{-1}\Gamma H_1 = M$.

Kalman Filter and Forward DP

- Going back to Kalman.
- Consider first step in forward DP

$$V_1^\star(x_1) = \min_{x_0} \tfrac{1}{2}\|x_0 - \bar{x}_0\|^2_{(P_0^-)^{-1}} + \tfrac{1}{2}\|y_0 - Cx_0\|^2_{R^{-1}} + \tfrac{1}{2}\|x_1 - (Ax_0 + b_0)\|^2_{Q^{-1}}$$

- Use item 1 in "Quadratic Sum" to express first two terms as

$$\tfrac{1}{2}\|x_0 - \bar{x}_0\|^2_{(P_0^-)^{-1}} + \tfrac{1}{2}\|y_0 - Cx_0\|^2_{R^{-1}} = \tfrac{1}{2}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \tfrac{1}{2}\|y_0 - C\bar{x}_0\|^2_{S_0^{-1}}$$

$$\begin{aligned} \text{where} \quad S_0 &= R + CP_0^- C^\top & P_0 &= P_0^- - P_0^- C^\top S_0^{-1} C P_0^- \\ \hat{x}_0 &= \bar{x}_0 + L_0(y_0 - C\bar{x}_0) & L_0 &= P_0^- C^\top S_0^{-1} \end{aligned}$$

- Thus

$$V_1^\star(x_1) = \tfrac{1}{2}\|y_0 - C\bar{x}_0\|^2_{S_0^{-1}} + \min_{x_0} \tfrac{1}{2}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \tfrac{1}{2}\|x_1 - (Ax_0 + b_0)\|^2_{Q^{-1}}$$

Kalman Filter and Forward DP

$$V_1^\star(x_1) = \tfrac{1}{2}\|y_0 - C\bar{x}_0\|^2_{S_0^{-1}} + \min_{x_0} \tfrac{1}{2}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \tfrac{1}{2}\|x_1 - (Ax_0 + b_0)\|^2_{Q^{-1}}$$

- Use item 2 in "Quadratic Sum" to express the min as

$$\min_{x_0} \tfrac{1}{2}\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \tfrac{1}{2}\|x_1 - (Ax_0 + b_0)\|^2_{Q^{-1}} = \tfrac{1}{2}\|x_1 - (A\hat{x}_0 + b_0)\|^2_{(P_1^-)^{-1}}$$

  where $P_1^- = Q + AP_0A^\top$
- Overall we obtain the arrival cost $V_1^\star$

$$V_1^\star(x_1) = \tfrac{1}{2}\|x_1 - \bar{x}_1\|^2_{(P_1^-)^{-1}} + c_1$$

  where $\bar{x}_1 = A\hat{x}_0 + b_0$, $c_1 = \tfrac{1}{2}\|y_0 - C\bar{x}_0\|^2_{S_0^{-1}}$
- $V_1^\star(x_1)$ has exactly the same form as $V_0^\star(x_0) = \tfrac{1}{2}\|x_0 - \bar{x}_0\|^2_{(P_0^-)^{-1}}$.
- Repeat same steps to arrive at recursive formulas for state estimates $\hat{x}_k$ and matrices P_k^-

Kalman Filter

Measurement update (after y_k is received)

$$P_k = P_k^- - P_k^- C^\top (C P_k^- C^\top + R)^{-1} C P_k^-$$
$$L_k = P_k^- C^\top (C P_k^- C^\top + R)^{-1}$$
$$\hat{x}_k = \bar{x}_k + L_k (y_k - C\bar{x}_k)$$

Time update (before y_{k+1} is received)

$$\bar{x}_{k+1} = A\hat{x}_k + b_k$$
$$P_{k+1}^- = A P_k A^\top + Q$$

- L_k: the Kalman gain.
- P_k: *a posteriori* estimate covariance.
- $\hat{x}_k$: *a posteriori* state estimate (after y_k is known).
- $\bar{x}_{k+1}$: *a priori* state estimate (before y_{k+1} is known).
- P_{k+1}^-: *a priori* estimate covariance.
- A posteriori estimate covariance can be expressed as $P_k = (I - L_k C) P_k^-$.

Kalman Filter

Pros

- Optimal estimator for linear systems.
- Solves full information estimation problem.
- Processes output measurements **recursively**.
- Computationally very efficient.

Cons

- Not applicable to nonlinear systems
- Not applicable to systems with constraints

Extended Kalman Filter (EKF)

- Heuristic extension of Kalman Filter to nonlinear systems $x_{k+1} = f_k(x_k, w_k)$
- Linearizes dynamics around $(\hat{x}_k, 0)$, output around $\bar{x}_k$

$$A_k = \frac{\partial f_k}{\partial x}(\hat{x}_k, 0),\, G_k = \frac{\partial f_k}{\partial w}(\hat{x}_k, 0),\, C_k = \frac{\partial h_k}{\partial x}(\bar{x}_k)$$

- Mean propagates through nonlinear model.
- Covariance propagates through linearization.

Measurement update (after y_k is received)

$$\begin{aligned} P_k &= P_k^- - P_k^- C_k^\top (C_k P_k^- C_k^\top + R)^{-1} C_k P_k^- \\ L_k &= P_k^- C_k^\top (C_k P_k^- C_k^\top + R)^{-1} \\ \hat{x}_k &= \bar{x}_k + L_k(y_k - h_k(\bar{x}_k)) \end{aligned}$$

Time update (before y_{k+1} is received)

$$\begin{aligned} \bar{x}_{k+1} &= f_k(\hat{x}_k, 0) \\ P_{k+1}^- &= A_k P_k A_k^\top + G_k Q G_k^\top \end{aligned}$$

Comments on EKF

Pros

- Simple and often works well.
- Probably the most widely used estimation algorithm for nonlinear systems.

Cons

- Difficult to tune, works well only for nearly linear systems.
- Essentially a local method. Can diverge.
- Proven to work only under restrictive conditions.
- Cannot handle constraints.

Constrained Estimation

$$\begin{aligned} \text{minimize } & \ell_x(x_0) + \sum_{k=0}^{T-1} \ell_v(y_k - h_k(x_k)) + \ell_w(w_k) \\ \text{subject to } & x_{k+1} = f_k(x_k, w_k) & k \in \mathbb{N}_{[0,T-1]} \\ & x_k \in X, w_k \in W, v_k \in V & k \in \mathbb{N}_{[0,T-1]} \end{aligned}$$

- Constraints on states, process and measurement noise easily handled
- Constrained nonlinear optimization problem that grows with time!
- Idea of MHE: use only a window of past measurements

MHE

- Using forward DP can write FIE problem as

$$\begin{aligned} \text{minimize } & V^\star_{T-N}(x_{T-N}) + \sum_{k=T-N}^{T-1} \ell_k(x_k, w_k) \\ \text{subject to } & x_{k+1} = f_k(x_k, w_k) && k \in \mathbb{N}_{[T-N,T-1]} \\ & x_k \in X, w_k \in W, v_k \in V && k \in \mathbb{N}_{[T-N,T-1]} \end{aligned}$$

- Arrival cost $V^\star_{T-N}$ is hard (if not impossible) to compute.
- Replace $V^\star_{T-N}$ with a user-defined **prior weighting Γ_{T-N}**

$$\begin{aligned} \text{minimize } & \Gamma_{T-N}(x_{T-N}) + \sum_{k=T-N}^{T-1} \ell_k(x_k, w_k) \\ \text{subject to } & x_{k+1} = f_k(x_k, w_k) && k \in \mathbb{N}_{[T-N,T-1]} \\ & x_k \in X, w_k \in W, v_k \in V && k \in \mathbb{N}_{[T-N,T-1]} \end{aligned}$$

MHE

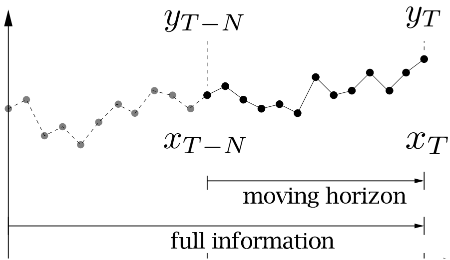
- MHE considers only N most recent measurements $(y_{T-N}, \ldots, y_{T-1})$ and solves

$$\hat{V}_T = \min \Gamma_{T-N}(x_{T-N}) + \sum_{k=T-N}^{T-1} \ell_k(x_k, w_k)$$

$$\text{s.t. } x_{k+1} = f_k(x_k, w_k) \qquad k \in \mathbb{N}_{[T-N, T-1]}$$

$$x_k \in X, w_k \in W, v_k \in V \qquad k \in \mathbb{N}_{[T-N, T-1]}$$

to compute prediction $\hat{x}_T$



Choices for prior weighting

Zero prior weighting

$$\boxed{\Gamma_{T-N}(z) = 0}$$

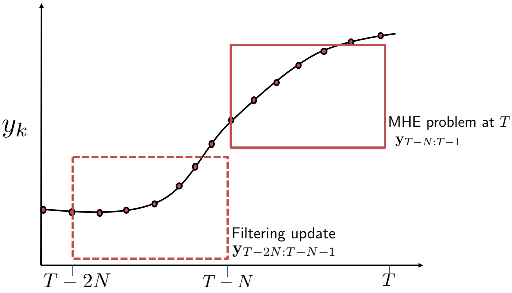
- Under observability condition and other technical assumptions can show the following:
  For large enough N, estimation error converges to zero whenever disturbances converge to zero.

Choices for prior weighting

Prior weighting based on filtering update

$$\boxed{\Gamma_{T-N}(z) = \frac{1}{2}\|z - \bar{x}_{T-N}\|^2_{(P^-_{T-N})^{-1}} + \hat{V}_{T-N}}$$

- $\bar{x}_{T-N}$ comes from the solution of MHE problem at time $T - N$
- P^-_{T-N} comes from (E)KF

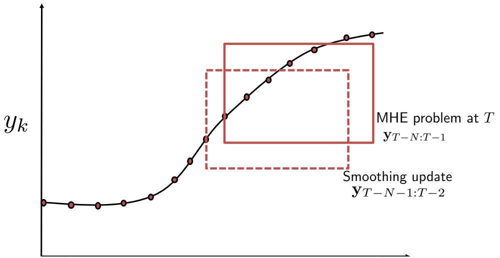


Choices for prior weighting

Prior weighting based on smoothing update

$$\boxed{\Gamma_{T-N}(z) = \frac{1}{2}\|z - \hat{x}_{T-N}\|^2_{(P_{T-N|T-2})^{-1}} - \frac{1}{2}\|\mathbf{y}_{T-N:T-2} - \mathcal{O}_{N-1}z\|^2_{W_{N-1}^{-1}} + \hat{V}_{T-1}}$$

- $\hat{x}_{T-N}$ comes from the solution of MHE problem at time $T-1$
- This $\hat{x}_{T-N}$ is based on $\mathbf{y}_{T-N-1:T-2} = \{y_{T-N-1}, \ldots, y_{T-2}\}$
- The cost of MHE at time T depends on $\mathbf{y}_{T-N:T-1} = \{y_{T-N}, \ldots, y_{T-1}\}$
- $P_{T-N|T-2}$ and the $-\frac{1}{2}\|\mathbf{y}_{T-N:T-2} - \mathcal{O}_{N-1}z\|^2_{W_{N-1}^{-1}}$ term adjust prior weighting so we do not double count data $\mathbf{y}_{T-N:T-2}$



Choices for prior weighting

Prior weighting based on smoothing update

$$\boxed{\Gamma_{T-N}(z) = \frac{1}{2}\|z - \hat{x}_{T-N}\|^2_{(P_{T-N|T-2})^{-1}} - \frac{1}{2}\|\mathbf{y}_{T-N:T-2} - \mathcal{O}_{N-1}z\|^2_{W_{N-1}^{-1}} + \hat{V}_{T-1}}$$

- $P_{T-N|T-2}$ is obtained by iterating backwards $N-1$ times

  $$P_{k|T-2} = P_k + P_k A^\top (P^-_{k+1})^{-1}(P_{k+1|T-2} - P^-_{k+1})(P^-_{k+1})^{-1} A P_k,\; k = T-2, \ldots, T-N$$

  starting from $P_{T-1|T-2} = P^-_{T-1}$
- P^-_{k+1}, P_k, $k = T-2, \ldots, T-N$ come from (E)KF
- $W_{N-1} = \tilde{R} + \mathcal{O}_{N-1}\tilde{Q}\mathcal{O}^\top_{N-1}$
  ($\tilde{R} = \mathrm{diag}(R, R, \ldots, R)$, $\tilde{Q} = \mathrm{diag}(Q, Q, \ldots, Q)$)

$$\mathcal{O}_{N-1} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & C & 0 & \cdots & 0 \\ 0 & CA & C & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & CA^{N-3} & CA^{N-4} & \cdots & C \end{bmatrix}$$

Choices for prior weighting-properties

Filtering vs. smoothing prior weighting

- For unconstrained linear systems they are the same and equal to arrival cost.
- For constrained linear systems they are different.
- For constrained linear systems: Estimation error converges to zero whenever disturbances converge to zero.
- For nonlinear systems they are heuristics.

Constrained nonlinear estimation

Example: Batch chemical reactor

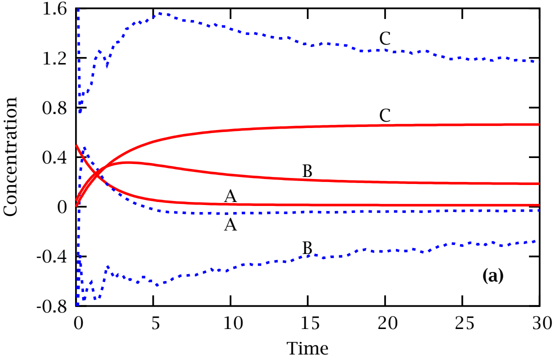
$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} c_A \\ c_B \\ c_C \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa_1 c_A - \kappa_{-1} c_B c_C \\ \kappa_2 c_B^2 - \kappa_{-2} c_C \end{bmatrix}$$

- state $x = (c_A, c_B, c_C)$, measurement $y = 32.84(x_1 + x_2 + x_3)$
- $P_0^- = (0.5)^2 I$, $Q = (0.001)^2 I$, $R = (0.25)^2$
- $\bar{x}_0 = (1, 0, 4)$, $x_0 = (0.5, 0.05, 0)$ (poor initial guess)
- state constraints $x \geq 0$.

Constrained nonlinear estimation

Example: Batch chemical reactor

EKF results

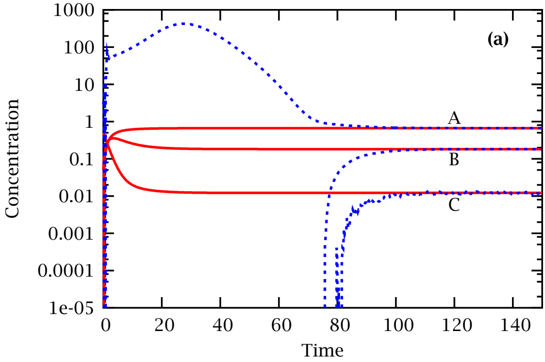


- negative concentrations

Constrained nonlinear estimation

Example: Batch chemical reactor

Clipped EKF results



- large estimate errors

Constrained nonlinear estimation

Example: Batch chemical reactor

MHE results (smoothing prior)