University of Pisa
Department of Civil and Industrial Engineering

University of Wisconsin - Madison
Department of Chemical and Biological Engineering

Moving Horizon Estimation: Confrontation of Filtering and Smoothing Covariance Updating

Marco Vaccari   James B. Rawlings   Gabriele Pannocchia

Systems Seminar Fall 2017 - Spring 2018

September 22, 2017

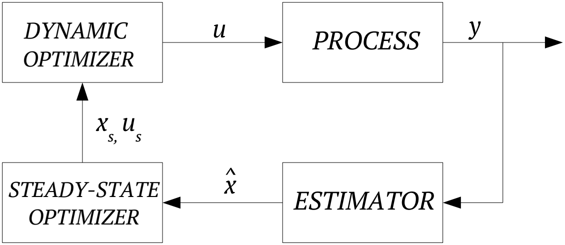
Overview

1. Introduction
2. Moving Horizon Estimation (MHE)
3. Prior Weight Updating
   - Filtering
   - Smoothing
4. Smoothing covariance derivation
5. Results
   - Linear Time-Varying (LTV) case: MHE vs. KF
   - Non Linear (NL) case
6. Conclusion

Introduction: Model Predictive Control (MPC)

Most MPC algorithms are divided into three modules.

- **State estimator**: receive current output measurement (y), and updates state (x) predictions.
- **Steady-state optimizer**: computes the state, input (u) and output targets to match the desired external setpoints while respecting the imposed constraints.
- **Dynamic optimizer**: finds optimal trajectory from current state to target.



Introduction: A useful analogy

Similitude

The basic MPC idea is similar to a chess game:

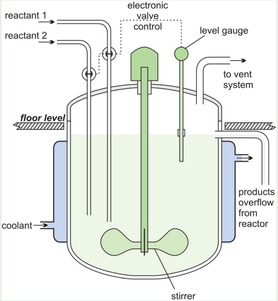
- The player uses a "model" (i.e., available moves for each piece) to forecast a finite sequence of planned moves
- Chooses the best sequence.
- Makes the first move of this optimal sequence and waits for the opponent response
- Repeats the whole prediction and optimization process.

Garry Kasparov vs IBM Deep Blue. 1996

Introduction: State estimation - A challenging task

Insufficient number of measurements

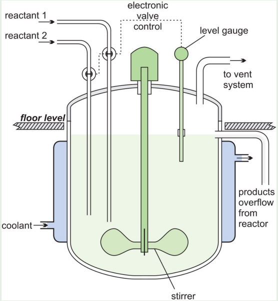
Usually measurable variables (y) are a small subset of the variables used to build the system model (x)



Introduction: State estimation - A challenging task

Insufficient number of measurements

Usually measurable variables (y) are a small subset of the variables used to build the system model (x)



Noises

- Measurements are corrupted with sensor noise (signal noise, external agents, etc.)
- State evolution is corrupted with process noise (changes into the system model, external disturb, etc.)

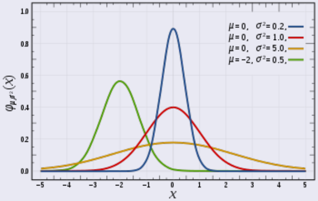
Introduction: State estimation - The simplest case

Linear discrete time model

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + Gw_k$$
$$\hat{y}_k = C\hat{x}_k + Du_k + v_k$$

Normally distributed noises

$$w_k \sim \mathcal{N}(0, Q)$$
$$v_k \sim \mathcal{N}(0, R)$$



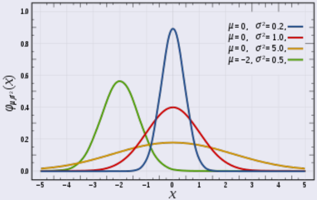
Introduction: State estimation - The simplest case

Linear discrete time model

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + Gw_k$$
$$\hat{y}_k = C\hat{x}_k + Du_k + v_k$$

Normally distributed noises

$$w_k \sim \mathcal{N}(0, Q)$$
$$v_k \sim \mathcal{N}(0, R)$$



The solution

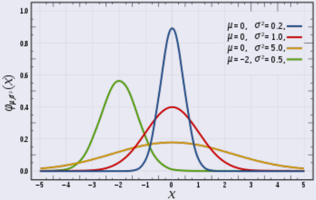
Introduction: State estimation - The simplest case

Linear discrete time model

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + Gw_k$$
$$\hat{y}_k = C\hat{x}_k + Du_k + v_k$$

Normally distributed noises

$$w_k \sim \mathcal{N}(0, Q)$$
$$v_k \sim \mathcal{N}(0, R)$$



The solution

This optimal state estimator is known as the **Kalman filter**(KF) (Kalman, 1960a).

Introduction: State estimation - The Non Linear case

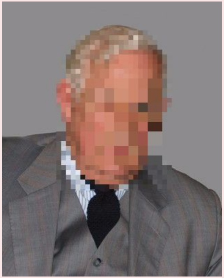
KF variants for non linear cases

- Extended KF (EKF)
- Unscented KF (UKF)
- Kalman-Bucy filter
- Hybrid Kalman filter
- etc..

Introduction: State estimation - The Non Linear case

KF variants for non linear cases

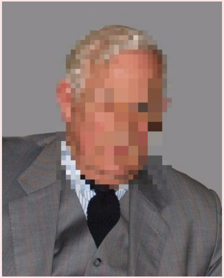
- Extended KF (EKF)
- Unscented KF (UKF)
- Kalman-Bucy filter
- Hybrid Kalman filter
- etc..



Introduction: State estimation - The Non Linear case

KF variants for non linear cases

- Extended KF (EKF)
- Unscented KF (UKF)
- Kalman-Bucy filter
- Hybrid Kalman filter
- etc..



State Estimation as Optimal Control of Estimate Error

- Full Information Estimation (FIE)
- Moving Horizon Estimation (MHE)

MHE: Nomenclature

System considered

$$\hat{\mathbf{x}}_{k+1} = F(\hat{\mathbf{x}}_k, \mathbf{u}_k, \mathbf{w}_k)$$
$$\hat{\mathbf{y}}_k = H(\hat{\mathbf{x}}_k, \mathbf{u}_k) + \mathbf{v}_k$$
$$\mathbf{w}_k \sim \mathcal{N}(0, Q_k)$$
$$\mathbf{v}_k \sim \mathcal{N}(0, R_k)$$

State: $\hat{\mathbf{x}} \in \mathbb{R}^n$

Input: $\mathbf{u} \in \mathbb{R}^m$

Output: $\hat{\mathbf{y}} \in \mathbb{R}^p$

Process noise: $\mathbf{w} \in \mathbb{R}^n$

Measurement noise: $\mathbf{v} \in \mathbb{R}^p$

Constraints

$$E\hat{\mathbf{x}} - e \leq 0 \qquad S\mathbf{w} - s \leq 0 \qquad J\mathbf{v} - j \leq 0$$

MHE: Full Information Estimation

Problem formulation

$$\min_{\mathbf{x},\mathbf{w},\mathbf{v}} \quad \Gamma(\gamma) + \sum_{i=0}^{N} \ell(w_i, v_i)$$

subject to:

$$\gamma = \hat{x}_0 - \bar{x}_0$$
$$\hat{x}_{i+1} = F(\hat{x}_i, u_i, w_i)$$
$$\boldsymbol{y}_i = H(\hat{x}_i, u_i) + v_i$$
$$Ex_i - e \leq 0, \qquad Sw_i - s \leq 0, \qquad Jv_i - j \leq 0.$$

where $\boldsymbol{y}_i$ represents the measurement at time i

Pro and cons

Best theoretical properties in terms
of stability and optimality

MHE: Full Information Estimation

Problem formulation

$$\min_{\mathbf{x},\mathbf{w},\mathbf{v}} \quad \Gamma(\gamma) + \sum_{i=0}^{N} \ell(w_i, v_i)$$

subject to:

$$\gamma = \hat{x}_0 - \bar{x}_0$$
$$\hat{x}_{i+1} = F(\hat{x}_i, u_i, w_i)$$
$$\boldsymbol{y}_i = H(\hat{x}_i, u_i) + v_i$$
$$Ex_i - e \leq 0, \qquad Sw_i - s \leq 0, \qquad Jv_i - j \leq 0.$$

where $\boldsymbol{y}_i$ represents the measurement at time i

Pro and cons

Best theoretical properties in terms of stability and optimality *BUT..* Computationally intractable except for the simplest cases

MHE: Moving Horizon Estimation

Method for practical estimator design

To come as close as possible to the FIE properties while maintaining a tractable online computation

MHE: Moving Horizon Estimation

Method for practical estimator design

To come as close as possible to the FIE properties while maintaining a tractable online computation

Problem formulation

$$\min_{\mathbf{x},\mathbf{w},\mathbf{v}} \quad \Gamma(\gamma) + \sum_{i=0}^{N_T} \ell(w_i, v_i)$$

subject to:

$$\gamma = \hat{x}_0 - \bar{x}_0$$
$$\hat{x}_{i+1} = F(\hat{x}_i, u_i, w_i)$$
$$\boldsymbol{y}_i = H(\hat{x}_i, u_i) + v_i$$
$$Ex_i - e \leq 0, \qquad Sw_i - s \leq 0, \qquad Jv_i - j \leq 0.$$

where $\boldsymbol{y}_i$ represents the measurement at time i

MHE: Prior Weight

$\Gamma(\gamma)$: What is it?

approximates the arrival cost, which summarizes the effects of past information before $k = 0$

Initially:

$$\Gamma(\gamma) = \frac{1}{2}(\hat{x}_0 - \bar{x}_0)' P_0^{-1}(\hat{x}_0 - \bar{x}_0)$$

P_0 chosen by the user and usually represents the covariance of the *a priori* state estimate $\bar{x}_0$.

MHE: Prior Weight

$\Gamma(\gamma)$: What is it?

approximates the arrival cost, which summarizes the effects of past information before $k = 0$

Initially:

$$\Gamma(\gamma) = \frac{1}{2}(\hat{x}_0 - \bar{x}_0)' P_0^{-1} (\hat{x}_0 - \bar{x}_0)$$

P_0 chosen by the user and usually represents the covariance of the *a priori* state estimate $\bar{x}_0$.

Updating?

Full Information Estimation $\rightarrow$ NO
Moving Horizon Estimation $\rightarrow$ YES

MHE: Prior Weight

$\Gamma(\gamma)$: What is it?

approximates the arrival cost, which summarizes the effects of past information before $k = 0$

Initially:

$$\Gamma(\gamma) = \frac{1}{2}(\hat{x}_0 - \bar{x}_0)' P_0^{-1} (\hat{x}_0 - \bar{x}_0)$$

P_0 chosen by the user and usually represents the covariance of the *a priori* state estimate $\bar{x}_0$.

Updating?

Full Information Estimation $\rightarrow$ NO
Moving Horizon Estimation $\rightarrow$ YES
when $k = N_T$

- $\bar{x}_1$ has to reflect the estimate of $\hat{x}_1$
- P_1 has to reflect the covariance of $\hat{x}_1$

Γ Updating: MHE as a filter

Filtering

The common implementation of MHE is as a filter.
At $k = N_T + 1$:

- $\bar{x}_1 = \hat{x}_{1|0}$ is the x_1 estimation given measurements at time 0
- $P_1 = P_{1|0}$

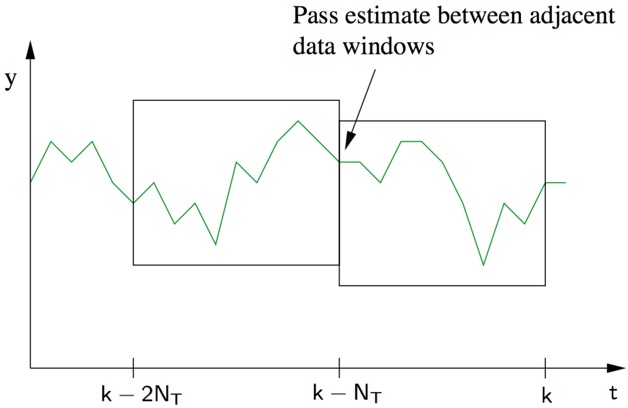
Note

When shifting the horizon, the *a priori* estimate is updated based at first on **only one** data point.

Note 2

This method for unconstrained linear systems is equivalent to Kalman filter

Γ Updating: MHE as a filter



Γ Updating: A smoother solution

Smoothing

Takes advantage of more information by including more data
At $k = N_T + 1$:

- $\bar{x}_1 = \hat{x}_{1|N_T}$ is the x_1 estimation given the measurement at time N_T
- $P_1 =$ **??**

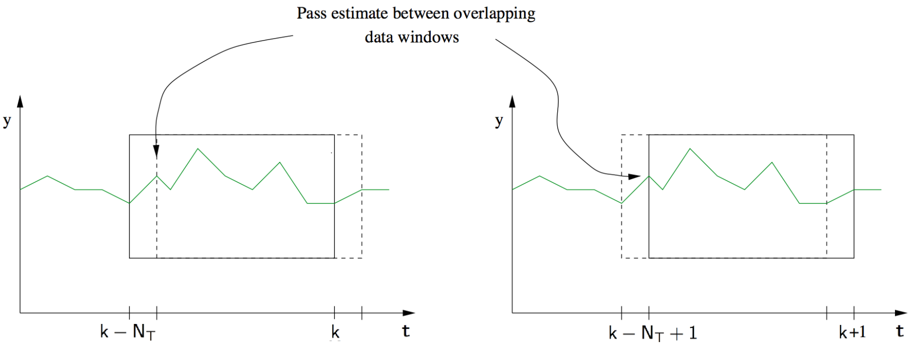
Literature

- Time invariant linear systems [Rao et al. (2001)]
- Nonlinear dynamical systems based on approximating the nonlinear model as a time-varying linear function [Tenny and Rawlings (2002)]

Note 2

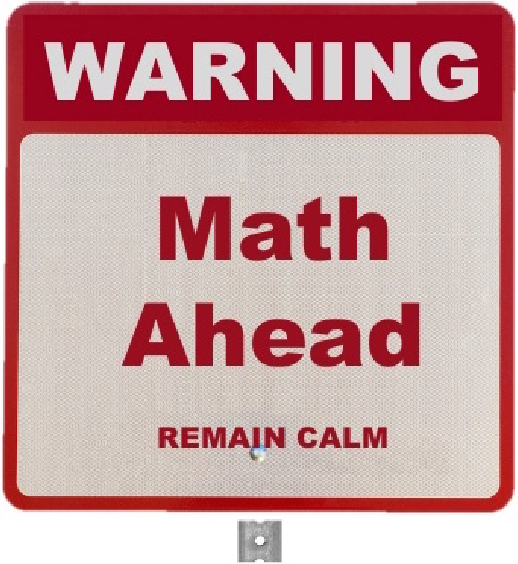
This method for unconstrained linear systems is equivalent to Kalman filter

Γ Updating: A smoother solution



Γ_s derivation

Γ_s derivation



Γ_s derivation: Nomenclature

Linearized state and output map

$$A_k = \frac{\partial F}{\partial x}, \qquad B_k = \frac{\partial F}{\partial u}, \qquad G_k = \frac{\partial F}{\partial w}, \qquad C_k = \frac{\partial H}{\partial x}, \qquad D_k = \frac{\partial H}{\partial u}$$

and two constant term f_k and h_k expressed as:

$$\hat{x}_{k+1|N_T} = A_k \hat{x}_{k|N_T} + B_k u_k + G_k w_k + f_k$$
$$y_k = C_k \hat{x}_{k|N_T} + D_k u_k + v_k + h_k$$

Linearized quadratic stage cost

$$\tilde{\ell}(w_k, v_k) = \frac{1}{2} \begin{bmatrix} w_k \\ v_k \end{bmatrix}' \begin{bmatrix} Q_k & S_k \\ S_k' & R_k \end{bmatrix}^{-1} \begin{bmatrix} w_k \\ v_k \end{bmatrix}$$

where:

$$Q_k = \frac{\partial^2 \ell}{\partial w_k^2}, \qquad R_k = \frac{\partial^2 \ell}{\partial v_k^2}, \qquad S_k = \frac{\partial^2 \ell}{\partial w_k \partial v_k}$$

Γ_s derivation: Definitions

Random variables

At time $k \geq N_T$, define

$$x_{k-N_T+1} \sim \mathcal{N}(\bar{x}_{k-N_T+1}, P_{k-N_T+1})$$
$$w_{k-N_T+1} \sim \mathcal{N}(0, Q_{k-N_T+1})$$
$$v_{k-N_T+1} \sim \mathcal{N}(0, R_{k-N_T+1})$$

where $x_{k-N_T+1} := \hat{x}_{k-N_T+1|k-1}$

Assumption

- x_{k-N_T+1} independent from $w_{k-N_T+1}, v_{k-N_T+1}, k \geq 0$, i.e:
$$\mathbb{E}[x_{k-N_T+1}, w_{k-N_T+1}] = 0, \qquad \mathbb{E}[x_{k-N_T+1}, v_{k-N_T+1}] = 0$$
- $w_{k-N_T+1}, v_{k-N_T+1}, k \geq 0$ are normally jointed, i.e.:
$$\begin{bmatrix} w_k \\ v_k \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} Q_{k-N_T+1} & S_{k-N_T+1} \\ S'_{k-N_T+1} & R'_{k-N_T+1} \end{bmatrix}\right)$$

Γ_s derivation: Definitions

Express the state $\mathbf{x}_k$ and output trajectory $\mathbf{y}_k$ in terms of the starting random variables with known probability densities.

$$\mathbf{x}_k = \begin{bmatrix} x_{k-N_T+1} \\ x_{k-N_T+2} \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} \qquad \mathbf{y}_k = \begin{bmatrix} y_{k-N_T+1} \\ y_{k-N_T+2} \\ \vdots \\ y_{k-1} \\ y_k \end{bmatrix} \qquad \mathbf{w}_k = \begin{bmatrix} w_{k-N_T+1} \\ v_{k-N_T+1} \\ w_{k-N_T+2} \\ v_{k-N_T+2} \\ \vdots \\ w_{k-1} \\ v_{k-1} \\ v_k \end{bmatrix}$$

In compact form..

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathcal{I}_k & \mathcal{A}_k \\ \mathcal{O}_k & \mathcal{G}_k \end{bmatrix} \begin{bmatrix} x_{k-N_T+1} \\ \mathbf{w}_k \end{bmatrix} + \begin{bmatrix} \mathcal{F}_k \\ \mathcal{H}_k \end{bmatrix}$$

Γ_s derivation: What do we need?

From [Rao (2000)]

Ratio of probability densities

$$\Gamma_s = -\ln(p_{x_{k-N_T+1}|\mathbf{y}_k} / p_{\mathbf{y}_k|x_{k-N_T+1}})$$

Focus on x_{k-N_T+1}

The problem is reduced to a smaller one:

$$\begin{bmatrix} x_{k-N_T+1} \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathcal{O}_k & \mathcal{G}_k \end{bmatrix} \begin{bmatrix} x_{k-N_T+1} \\ \mathbf{w}_k \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{H}_k \end{bmatrix}$$

The joint density of $(x_{k-N_T+1}, \mathbf{y}_k)$

$$\begin{bmatrix} x_{k-N_T+1} \\ \mathbf{y}_k \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{x}_{k-N_T+1} \\ \mathcal{O}_k \bar{x}_{k-N_T+1} + \mathcal{H}_k \end{bmatrix}, \begin{bmatrix} P_{k-N_T+1} & P_{k-N_T+1}\mathcal{O}_k' \\ \mathcal{O}_k P_{k-N_T+1} & \mathcal{G}_k \mathcal{Q}_k \mathcal{G}_k' + \mathcal{O}_k P_{k-N_T+1} \mathcal{O}_k' \end{bmatrix} \right)$$

Γ_s derivation: Conditional densities

$p(x_{k-N_T+1}|\mathbf{y}_k)$

$$p(x_{k-N_T+1}|\mathbf{y}_k) = \mathcal{N}(m_{x|\mathbf{y}_k}, P_{x|\mathbf{y}_k})$$

$$m_{x|\mathbf{y}_k} = \bar{x}_{k-N_T+1} + P_{k-N_T+1}\mathcal{O}_k'(\mathcal{G}_k\mathcal{Q}_k\mathcal{G}_k' + \mathcal{O}_kP_{k-N_T+1}\mathcal{O}_k')^{-1}(\mathbf{y} - \mathcal{O}_k m_{k-N_T+1} - \mathcal{H}_k)$$

$$P_{x|\mathbf{y}_k} = P_{k-N_T+1} - P_{k-N_T+1}\mathcal{O}_k'(\mathcal{G}_k\mathcal{Q}_k\mathcal{G}_k' + \mathcal{O}_kP_{k-N_T+1}\mathcal{O}_k')^{-1}\mathcal{O}_kP_{k-N_T+1}$$

$p(\mathbf{y}_k|x_{k-N_T+1})$

$$p(\mathbf{y}_k|x_{k-N_T+1}) = \mathcal{N}(m_{\mathbf{y}_k|x}, P_{\mathbf{y}_k|x})$$

$$m_{\mathbf{y}_k|x} = \mathcal{O}_k x_{k-N_T+1} + \mathcal{H}_k \qquad P_{\mathbf{y}_k|x} = \mathcal{G}_k\mathcal{Q}_k\mathcal{G}_k'$$

$\Gamma_s(x_{k-N_T+1})$

$$\Gamma_s(x_{k-N_T+1}) := (1/2)(x_{k-N_T+1} - m_{x|\mathbf{y}_k})'P_{x|\mathbf{y}_k}{}^{-1}(x_{k-N_T+1} - m_{x|\mathbf{y}_k}) - (1/2)(\mathbf{y}_k - m_{\mathbf{y}_k|x})'P_{\mathbf{y}_k|x}{}^{-1}(\mathbf{y}_k - m_{\mathbf{y}_k|x})$$

Γ_s derivation: Conditional densities

$p(x_{k-N_T+1}|\mathbf{y}_k)$

$$p(x_{k-N_T+1}|\mathbf{y}_k) = \mathcal{N}(m_{x|\mathbf{y}_k}, P_{x|\mathbf{y}_k})$$

$$m_{x|\mathbf{y}_k} = \bar{x}_{k-N_T+1} + P_{k-N_T+1}\mathcal{O}_k'(\mathcal{G}_k\mathcal{Q}_k\mathcal{G}_k' + \mathcal{O}_kP_{k-N_T+1}\mathcal{O}_k')^{-1}(\mathbf{y} - \mathcal{O}_km_{k-N_T+1} - \mathcal{H}_k)$$

$$P_{x|\mathbf{y}_k} = P_{k-N_T+1} - P_{k-N_T+1}\mathcal{O}_k'(\mathcal{G}_k\mathcal{Q}_k\mathcal{G}_k' + \mathcal{O}_kP_{k-N_T+1}\mathcal{O}_k')^{-1}\mathcal{O}_kP_{k-N_T+1}$$

$p(\mathbf{y}_k|x_{k-N_T+1})$

$$p(\mathbf{y}_k|x_{k-N_T+1}) = \mathcal{N}(m_{\mathbf{y}_k|x}, P_{\mathbf{y}_k|x})$$

$$m_{\mathbf{y}_k|x} = \mathcal{O}_kx_{k-N_T+1} + \mathcal{H}_k \qquad P_{\mathbf{y}_k|x} = \mathcal{G}_k\mathcal{Q}_k\mathcal{G}_k'$$

$\Gamma_s(x_{k-N_T+1})$

$$\Gamma_s(x_{k-N_T+1}) := (1/2)(x_{k-N_T+1} - m_{x|\mathbf{y}_k})'P_{x|\mathbf{y}_k}{}^{-1}(x_{k-N_T+1} - m_{x|\mathbf{y}_k}) - (1/2)(\mathbf{y}_k - m_{\mathbf{y}_k|x})'P_{\mathbf{y}_k|x}{}^{-1}(\mathbf{y}_k - m_{\mathbf{y}_k|x})$$

Γ_s calculation: A more efficient way

$\Gamma_s(x_{k-N_T+1})$

$$\Gamma_s(x_{k-N_T+1}) := (1/2)(x_{k-N_T+1} - m_{x|\mathbf{y}_k})' P_{x|\mathbf{y}_k}{}^{-1}(x_{k-N_T+1} - m_{x|\mathbf{y}_k}) - (1/2)(\mathbf{y}_k - m_{\mathbf{y}_k|x})' P^{-1}_{\mathbf{y}_k|x}(\mathbf{y}_k - m_{\mathbf{y}_k|x})$$

- $m_{x|\mathbf{y}_k} = \hat{x}_{k-N_T|k}$ from the previous optimal sequence
- $P_{x|\mathbf{y}_k}$ from Riccati backward recursion [Rauch et al. (1965)]:

$$P_{k|N_T} = P_{k|k} + P_{k|k} A_k' P^{-1}_{k+1|k}(P_{k+1|N_T} - P_{k+1|k}) P^{-1}_{k+1|k} A_k P_{k|k}$$

where

$$P_{k|k} = P_{k|k-1} - P_{k|k-1} C_k' (R_k + C_k P_{k|k-1} C_k')^{-1} C_k P_{k|k-1}$$

and

$$P_{k+1|k} = G_k Q_k G_k' + A_k P_{k|k} A_k'$$

stops at $P_{1|N_T}$

Results: MHE *vs*. KF on LTV

Full Linear Time-Varying case

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + G_k w_k$$
$$\hat{y}_k = C_k \hat{x}_k + D_k u_k + v_k$$

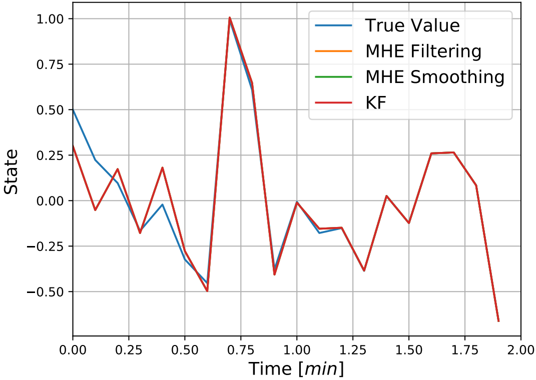
$$\ell(w_k, v_k) = \frac{1}{2} \left(w_k' Q_k^{-1} w_k + v_k' R_k^{-1} v_k \right)$$

Results: MHE *vs.* KF on LTV

Full Linear Time-Varying case

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + G_k w_k$$
$$\hat{y}_k = C_k \hat{x}_k + D_k u_k + v_k$$

$$\ell(w_k, v_k) = \frac{1}{2}\left(w_k' Q_k^{-1} w_k + v_k' R_k^{-1} v_k\right)$$



NL case: Gas phase reactor

Description

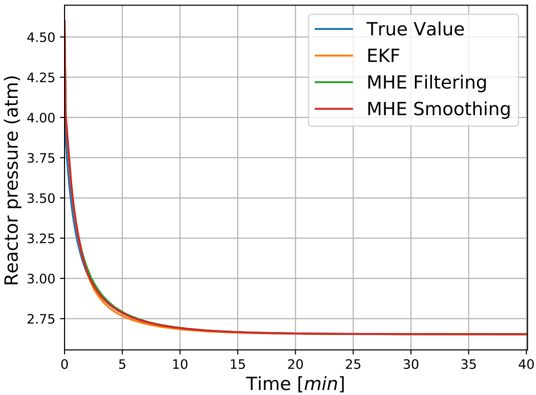
Isothermal gas-phase reactor where reversible reaction $2A \rightleftharpoons B$ is taking place.
An initial amount of A and B are charged to the reactor, but the composition of the original mixture is not known accurately.
A pressure gauge measures the total pressure of the system as the species react.
The simulation is run in open-loop.

Nomenclature

$$\dot{p_A} = -2k_1 p_A^2 + 2k_2 p_B$$
$$\dot{p_B} = k_1 p_A^2 - k_2 p_B$$

$$x = \begin{bmatrix} p_A \\ p_B \end{bmatrix} \quad \mathrm{u} = 0 \quad y = \begin{bmatrix} p_A + p_B \end{bmatrix}$$
$$\bar{x}_0 \neq x_0$$

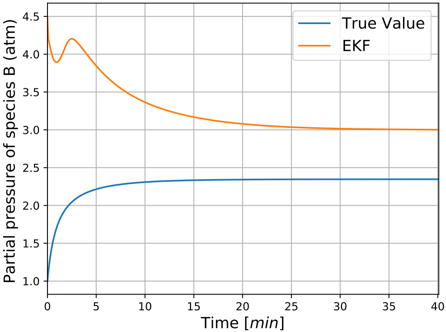
Results: Output

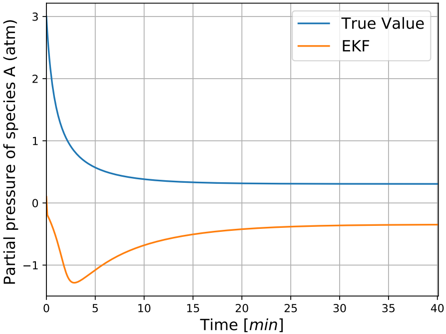


Note

All the three methods calculate the same total reactor pressure

Results: MHE *vs.* EKF on NL

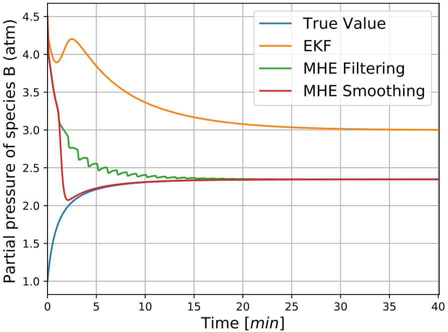


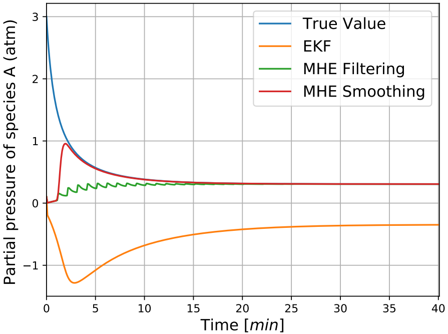


Why does EKF fails? [Haseltine (2005)]

- system model and measurement are such that multiple states satisfy the steady-state measurement
- the estimator is given a poor initial guess of the state

Results: Smoothing *vs.* Filtering on NL





Note

Periodic behavior of the estimates obtained using the filter update ($T = N_T \times h$) due to propagation of estimation error from a poor a priori estimate.

Conclusion

Wrapping up

- We calculated the prior weighting covariance updating in the filter and smoothing MHE problem
- We demonstrated the MHE problem is equal to the KF problem for LTV systems
- We showed the efficiency of MHE against EKF in case where the last fails
- Smoothing updated can be more suitable for nonlinear control applications

Conclusion

Wrapping up

- We calculated the prior weighting covariance updating in the filter and smoothing MHE problem
- We demonstrated the MHE problem is equal to the KF problem for LTV systems
- We showed the efficiency of MHE against EKF in case where the last fails
- Smoothing updated can be more suitable for nonlinear control applications

Thank you for your attention

E. L. Haseltine. *Systems Analysis of Stochastic and Population Balance Models for Chemically Reacting Systems*. PhD thesis, University of Wisconsin–Madison, 2005.

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

C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear state estimation – a moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.

H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA J.*, 3(8):1445–1450, 1965.

M. J. Tenny and J. B. Rawlings. Efficient moving horizon estimation and nonlinear model predictive control. In *Proceedings of the American Control Conference*, pages 4475–4480, Anchorage, Alaska, May 2002.