Current status and future challenges of moving horizon state estimation

James B. Rawlings, Luo Ji, Giulio Mancuso

Dept. of Chemical and Biological Engineering,
Univ. of Wisconsin-Madison, WI, USA

Center for Control, Dynamical Systems, and Computation
College of Engineering
UC Santa Barbara
March 1, 2013
1. Overview of our research
2. On-line optimization and state estimation
3. Current status of MHE theory
4. Nonlinear applications, i.e., How do we make this work?
5. Comparison of some nonlinear estimation methods
6. Conclusions and future work
\[
\frac{dx}{dt} = f(x, u) \quad \hat{y} \\
y = g(x, u) \quad \hat{y}
\]
What Our Research Group Does

Systems Engineering—Especially Process Control

What Our Research Group Does

Systems Engineering—Especially Process Control

- Algorithms for implementation. We work in real time.
What Our Research Group Does

Systems Engineering—Especially Process Control

- Algorithms for implementation. We work in real time.
- Modeling: The process.
What Our Research Group Does

Systems Engineering—Especially Process Control

- Algorithms for implementation. We work in real time.
- Modeling: The process. And the uncertainty.
What Our Research Group Does

Systems Engineering—Especially Process Control

- Algorithms for implementation. We work in real time.
- Modeling: The process. And the uncertainty.
- Industrial collaborations. ExxonMobil, Shell, Eastman, Aspentech.
What Our Research Group Does

Systems Engineering—Especially Process Control

- Algorithms for implementation. We work in real time.
- Modeling: The process. And the uncertainty.
- Industrial collaborations. ExxonMobil, Shell, Eastman, Aspentech. Identify and solve research problems that have impact on practice.
The model predictive control framework

Reconcile the past

Forecast the future

- MH Estimate
- Measurement

- Forecast
- MPC control

sensors

\( y \)

actuators

\( u \)

t

time

Moving horizon estimation

Rawlings/Ji/Mancuso
Predictive control

\[
\begin{align*}
\min_{u(t)} & \int_0^T |y_{sp} - g(x, u)|^2_Q + |u_{sp} - u|^2_R \, dt \\
\dot{x} & = f(x, u) \\
x(0) & = x_0 \quad \text{(given)} \\
y & = g(x, u)
\end{align*}
\]
State estimation

\[
\min_{x_0, w(t)} \int_{-T}^{0} \left| y - g(x, u) \right|^2_R + |\dot{x} - f(x, u)|^2_Q \, dt
\]

\[
\dot{x} = f(x, u) + w \quad \text{(process noise)}
\]

\[
y = g(x, u) + v \quad \text{(measurement noise)}
\]
Large industrial success story!

Linear MPC and ethylene manufacturing

- Number of MPC applications in ethylene: 800 to 1200
Linear MPC and ethylene manufacturing

- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
Large industrial success story!

Linear MPC and ethylene manufacturing

- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
- Achieved primarily by increased on-spec product, decreased energy use
Large industrial success story!

**Linear MPC and ethylene manufacturing**
- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
- Achieved primarily by increased on-spec product, decreased energy use

**Eastman Chemical experience with MPC**
- First MPC implemented in 1996

Rawlings/Ji/Mancuso
Large industrial success story!

Linear MPC and ethylene manufacturing

- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
- Achieved primarily by increased on-spec product, decreased energy use

Eastman Chemical experience with MPC

- First MPC implemented in 1996
- Currently 55-60 MPC applications of varying complexity
Large industrial success story!

Linear MPC and ethylene manufacturing

- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
- Achieved primarily by increased on-spec product, decreased energy use

Eastman Chemical experience with MPC

- First MPC implemented in 1996
- Currently 55-60 MPC applications of varying complexity
- 30-50 M$/year increased profit due to increased throughput (2008)
Large industrial success story!

Linear MPC and ethylene manufacturing
- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
- Achieved primarily by increased on-spec product, decreased energy use

Eastman Chemical experience with MPC
- First MPC implemented in 1996
- Currently 55-60 MPC applications of varying complexity
- 30-50 M$/year increased profit due to increased throughput (2008)

Praxair experience with MPC
- Praxair currently has more than 150 MPC installations
Large industrial success story!

**Linear MPC and ethylene manufacturing**
- Number of MPC applications in ethylene: 800 to 1200
- Credits 500 to 800 M$/yr (2007)
- Achieved primarily by increased on-spec product, decreased energy use

**Eastman Chemical experience with MPC**
- First MPC implemented in 1996
- Currently 55-60 MPC applications of varying complexity
- 30-50 M$/year increased profit due to increased throughput (2008)

**Praxair experience with MPC**
- Praxair currently has more than 150 MPC installations
- 16 M$/year increased profit (2008)
State estimation — Why use optimization?

The statistical viewpoint

- System model

\[ x^+ = f(x, w) \quad y = h(x) + v \]
State estimation — Why use optimization?

The statistical viewpoint

- System model

\[ x^+ = f(x, w) \quad y = h(x) + v \]

- Disturbances, \( w \) and \( v \), initial state, \( x(0) \), modeled as random variables. The central limit theorem justifies using (zero mean) normal distributions for \( w, v \). Obtaining variances part of the modeling/identification problem!
State estimation — Why use optimization?

The statistical viewpoint

- System model

\[ x^+ = f(x, w) \quad \quad y = h(x) + v \]

- Disturbances, \( w \) and \( v \), initial state, \( x(0) \), modeled as random variables. The central limit theorem justifies using (zero mean) normal distributions for \( w, v \). Obtaining variances part of the modeling/identification problem!

- We observe \( y(k) := y(0), y(1), \ldots, y(T) \) and wish to estimate \( x(k) \).
State estimation — Why use optimization?

The statistical viewpoint

- **System model**

  \[ x^+ = f(x, w) \quad y = h(x) + v \]

- Disturbances, \( w \) and \( v \), initial state, \( x(0) \), modeled as random variables. The central limit theorem justifies using (zero mean) normal distributions for \( w, v \). Obtaining variances part of the modeling/identification problem!

- We observe \( y(k) := y(0), y(1), \ldots, y(T) \) and wish to estimate \( x(k) \).

- Statistically optimal estimate; maximize conditional density

  \[
  \max_{x(k)} p(x(k) \mid y(k))
  \]
State estimation — Why use optimization?

The statistical viewpoint

- **System model**
  \[ x^+ = f(x, w) \quad y = h(x) + v \]

- Disturbances, \( w \) and \( v \), initial state, \( x(0) \), modeled as random variables. The central limit theorem justifies using (zero mean) normal distributions for \( w \), \( v \). Obtaining variances part of the modeling/identification problem!

- We observe \( y(k) := y(0), y(1), \ldots, y(T) \) and wish to estimate \( x(k) \).

- Statistically optimal estimate; maximize conditional density
  \[ \max_{x(k)} p(x(k) \mid y(k)) \]

- Computing \( p \) usually intractable for nonlinear models
State estimation — Why use optimization?

The engineering viewpoint

- Reduce goal from statistical optimality to practical real time algorithm.

\[
V_T(x(0), w) = \ell_x(x(0) - x_0) + T - 1 \sum_{i=0}^{T-1} \ell_i(w(i), v(i))
\]

subject to model

\[
x + f(x, w) \\
y = h(x) + v
\]

Full information estimator is

\[
\min_{x(0), w} V_T(x(0), w)
\]

Can easily add knowledge (constraints) on \(w, v, x\) to the formulation.
State estimation — Why use optimization?

The engineering viewpoint

- Reduce goal from statistical optimality to practical real time algorithm.
- Choose a merit function

\[
V_T(x(0), w) = \ell_x(x(0) - \bar{x}_0) + \sum_{i=0}^{T-1} \ell_i(w(i), v(i))
\]

subject to model

\[
x^+ = f(x, w) \quad y = h(x) + v
\]
State estimation — Why use optimization?

The engineering viewpoint

- Reduce goal from statistical optimality to practical real time algorithm.
- Choose a merit function

\[ V_T(x(0), w) = \ell_x(x(0) - \bar{x}_0) + \sum_{i=0}^{T-1} \ell_i(w(i), v(i)) \]

subject to model

\[ x^+ = f(x, w) \quad y = h(x) + v \]

- Full information estimator is

\[ \min_{x(0), w} V_T(x(0), w) \]
State estimation — Why use optimization?

The engineering viewpoint

- Reduce goal from statistical optimality to practical real time algorithm.
- Choose a merit function

\[
V_T(x(0), w) = \ell_x(x(0) - \bar{x}_0) + \sum_{i=0}^{T-1} \ell_i(w(i), v(i))
\]

subject to model

\[
x^+ = f(x, w) \quad y = h(x) + v
\]

- Full information estimator is

\[
\min_{x(0), w} V_T(x(0), w)
\]

- Can easily add knowledge (constraints) on \(w, v, x\) to the formulation.
Current status of full information theory

- Compare to the linear theory.

\[ x + w = Ax + w, \]
\[ y = Cx + v. \]

If the system is detectable and \( Q, R > 0 \), then the estimate error, \( \tilde{x} := x - \hat{x} \), satisfies:

\[ \tilde{x} + \tilde{A} \tilde{x} = w - Lv \]

and \( \tilde{A} := A - LC \) is stable.

Provides this bound on estimate error:

\[ |\tilde{x}(k)| \leq c \lambda_k |\tilde{x}(0)| + c w \|w\|_0 + c w \|w\|_0 \cdot (k - 1) + c v \|v\|_0 \cdot (k - 1), \]

using the norm over the sequence, \( \|w\|_0 := \max_{j \leq k-1} |w(j)| \).
Current status of full information theory

- Compare to the linear theory.
- Linear model, \( x^+ = Ax + w, \quad y = Cx + v. \)
• Compare to the linear theory.

• Linear model, \( x^+ = Ax + w \), \( y = Cx + v \).

• If the system is detectable and \( Q, R > 0 \), then the estimate error, \( \tilde{x} := x - \hat{x} \), satisfies
  \[
  \tilde{x}^+ = \tilde{A} \tilde{x} + w - Lv
  \]

  and \( \tilde{A} := A - LC \) is stable.
Current status of full information theory

- Compare to the linear theory.
- Linear model, $x^+ = Ax + w, \quad y = Cx + v$.
- If the system is detectable and $Q, R > 0$, then the estimate error, $\tilde{x} := x - \hat{x}$, satisfies
  \[
  \tilde{x}^+ = \tilde{A}\tilde{x} + w - Lv
  \]
  and $\tilde{A} := A - LC$ is stable.
- Provides this bound on estimate error
  \[
  |\tilde{x}(k)| \leq c\lambda^k |\tilde{x}(0)| + c_w \|w\|_{0:k-1} + c_v \|v\|_{0:k-1}
  \]
  using the norm over the sequence, $\|w\|_{0:k-1} := \max_{j \in 0:k-1} |w(j)|$.
What happens in the nonlinear theory?

- Linear model $\rightarrow$ nonlinear model, $x^+ = f(x, w)$, $y = h(x) + v$
What happens in the nonlinear theory?

- Linear model $\rightarrow$ nonlinear model, $x^+ = f(x, w)$  $y = h(x) + v$
- Detectability $\rightarrow$ i-IOSS (Sontag and Wang, 1997)
What happens in the nonlinear theory?

- Linear model → nonlinear model, \( x^+ = f(x, w) \quad y = h(x) + v \)
- Detectability → i-IOSS (Sontag and Wang, 1997)
- \( w' Q^{-1} w + v' R^{-1} v \rightarrow \ell(w, v) \)
What happens in the nonlinear theory?

- Linear model $\rightarrow$ nonlinear model, $x^+ = f(x, w)$, $y = h(x) + v$
- Detectability $\rightarrow$ i-IOSS (Sontag and Wang, 1997)
- $w'Q^{-1}w + v'R^{-1}v \rightarrow \ell(w, v)$
- $Q, R > 0 \rightarrow$ stage cost: $\ell(w, v)$ underbounded by $K$-function
What happens in the nonlinear theory?

- Linear model $\rightarrow$ nonlinear model, $x^+ = f(x, w)$, \( y = h(x) + v \)
- Detectability $\rightarrow$ i-IOSS (Sontag and Wang, 1997)
- \( w'Q^{-1}w + v'R^{-1}v \rightarrow \ell(w, v) \)
- $Q, R > 0$ $\rightarrow$ stage cost: $\ell(w, v)$ underbounded by $K$-function
- Establishing the linear system’s bound on estimate error $\rightarrow$
  Establishing the RGAS property for the nonlinear system
Nonlinear detectability

Definition (i-IOSS (Sontag and Wang, 1997))

The system $x^+ = f(x, w), y = h(x)$ is incrementally input/output-to-state stable (i-IOSS) if there exist functions $\alpha(\cdot) \in KL$ and $\gamma_1(\cdot), \gamma_2(\cdot) \in K$ such that for every two initial states $z_1$ and $z_2$, and any two disturbance sequences $w_1$ and $w_2$ generating state sequences $x_1(z_1, w_1)$ and $x_2(z_2, w_2)$, the following holds for all $k \in \mathbb{I}_{\geq 0}$

$$|x(k; z_1, w_1) - x(k; z_2, w_2)| \leq \alpha_1(|z_1 - z_2|, k) + \gamma_1(\|w_1 - w_2\|_{0:k-1}) + \gamma_2(\|h(x_1) - h(x_2)\|_{0:k-1}) \tag{1}$$
The system $\dot{x} = f(x, w), y = h(x)$ is incrementally input/output-to-state stable (i-IOSS) if there exist functions $\alpha(\cdot) \in \mathcal{KL}$ and $\gamma_1(\cdot), \gamma_2(\cdot) \in \mathcal{K}$ such that for every two initial states $z_1$ and $z_2$, and any two disturbance sequences $w_1$ and $w_2$ generating state sequences $x_1(z_1, w_1)$ and $x_2(z_2, w_2)$, the following holds for all $k \in \mathbb{I}_{\geq 0}$

$$|x(k; z_1, w_1) - x(k; z_2, w_2)| \leq \alpha_1(|z_1 - z_2|, k) + \gamma_1(\|w_1 - w_2\|_{0:k-1}) + \gamma_2(\|h(x_1) - h(x_2)\|_{0:k-1}) \quad (1)$$

In words: if the disturbances are close and the measurements are close, the states have to become close.
Robust global asymptotic stability

**Definition (Robust global asymptotic stability (RGAS))**

The estimate is based on the *noisy* measurement \( y = h(x(x_0, w)) + v \). The estimate is RGAS if for all \( x_0 \) and \( \bar{x}_0 \), and bounded \((w, v)\), there exist functions \( \alpha(\cdot) \in \mathcal{KL} \) and \( \delta_w(\cdot), \delta_v(\cdot) \in \mathcal{K} \) such that the following holds for all \( k \in \mathbb{I}_{\geq 0} \)

\[
\left|x(k; x_0, w) - x(k; \hat{x}(0|k), \hat{w}_k)\right| \leq \alpha(|x_0 - \bar{x}_0|, k) + \delta_w(\|w\|_{0:k-1}) + \delta_v(\|v\|_{0:k-1})
\]  

(2)

Recall the linear system result

\[
\left|\tilde{x}(k)\right| \leq c\lambda^k \left|\tilde{x}(0)\right| + c_w \|w\|_{0:k-1} + c_v \|v\|_{0:k-1}
\]
Robust global asymptotic stability

Definition (Robust global asymptotic stability (RGAS))

The estimate is based on the noisy measurement \( y = h(x(x_0, w)) + v \). The estimate is RGAS if for all \( x_0 \) and \( \bar{x}_0 \), and bounded \( (w, v) \), there exist functions \( \alpha(\cdot) \in KL \) and \( \delta_w(\cdot), \delta_v(\cdot) \in K \) such that the following holds for all \( k \in \mathbb{I}_{\geq 0} \)

\[
|x(k; x_0, w) - x(k; \hat{x}(0|k), \hat{w}_k)| \leq \alpha(|x_0 - \bar{x}_0|, k) + \delta_w(\|w\|_{0:k-1}) + \delta_v(\|v\|_{0:k-1}) \quad (2)
\]

Recall the linear system result

\[
|\ddot{x}(k)| \leq c\lambda^k |\ddot{x}(0)| + c_w \|w\|_{0:k-1} + c_v \|v\|_{0:k-1}
\]
Theorem (RGAS of full information estimate)

Consider an i-IOSS (detectable) system and measurement sequence generated by the nonlinear model subject to bounded, convergent disturbances, and stage cost satisfying appropriate bounds. Then the full information estimator is RGAS.

Precise statement and proof are provided by Rawlings and Ji (2012) and (Rawlings and Mayne, 2009, Chapter 4)
Given all this extra generality, what do we lose?

- More restrictive assumptions about $w, v$.
  Bounded disturbances $\rightarrow$ *convergent* disturbances
Given all this extra generality, what do we lose?

- More restrictive assumptions about $w, v$.
  Bounded disturbances $\rightarrow$ convergent disturbances
- Lose the strict equivalence between the the full information problem and its moving horizon approximation.
  But expect similar behavior for large horizon, $N$
Given all this extra generality, what do we lose?

- More restrictive assumptions about $w, v$.
  Bounded disturbances $\rightarrow$ *convergent* disturbances

- Lose the strict equivalence between the the full information problem and its moving horizon approximation.
  But expect similar behavior for large horizon, $N$

- Lose the analytical, recursive solution.
  Least squares problem $\rightarrow$ nonconvex optimization problem
Given all this extra generality, what do we lose?

- More restrictive assumptions about $w, \nu$.
  Bounded disturbances $\rightarrow$ convergent disturbances
- Lose the strict equivalence between the full information problem and its moving horizon approximation. But expect similar behavior for large horizon, $N$
- Lose the analytical, recursive solution.
  Least squares problem $\rightarrow$ nonconvex optimization problem
- Because of this online computational complexity, the big divide in state estimation is between linear and nonlinear models
What are some alternatives to optimization?

Two general alternatives to optimization: *sampling* and *linearization*.
What are some alternatives to optimization?

Two general alternatives to optimization: sampling and linearization

- **Sampling**: AKA particle filtering. Generate samples of \( p(x(k) \mid y(k)) \). Many options have been explored: importance sampling, resampling, etc.
What are some alternatives to optimization?

Two general alternatives to optimization: *sampling* and *linearization*

- **Sampling**: AKA particle filtering. Generate samples of $p(x(k) \mid y(k))$. Many options have been explored: importance sampling, resampling, etc.

- An attractive feature is that the sampled density converges to the true conditional density as sample number goes to infinity. However, the required number of samples is excessive for even modest state dimension.
What are some alternatives to optimization?

Two general alternatives to optimization: *sampling* and *linearization*

- **Sampling**: AKA particle filtering. Generate samples of $p(x(k) \mid y(k))$. Many options have been explored: importance sampling, resampling, etc.

- An attractive feature is that the sampled density converges to the true conditional density as sample number goes to infinity. However, the required number of samples is excessive for even modest state dimension.

- Statistical sampling is not taught to systems engineers. For a brief tutorial that doesn’t leave out any steps, see (Rawlings and Mayne, 2009, pp. 304–348).
Sampling applied to a linear system — 250 particles

Using optimal importance function
Sampling applied to a linear system — 250 particles

Using optimal importance function

... plus resampling
Two general alternatives to optimization: *sampling* and *linearization*. 

**Linearization**

AKA the extended Kalman filter. Linearize the nonlinear model at the current estimate. Apply the linear system update formulas.

\[
A_k = \frac{\partial f(x, w)}{\partial x} \bigg|_{(\hat{x}_k, 0)}
\]

\[
G_k = \frac{\partial f(x, w)}{\partial w} \bigg|_{(\hat{x}_k, 0)}
\]

\[
C_k = \frac{\partial h(x)}{\partial x} \bigg|_{\hat{x}_k - k}
\]

**Propagation**

\[
\hat{x}_k - k + 1 = f(\hat{x}_k, 0)
\]

\[
P_{k - 1} = A_k P_k A_k' + G_k Q_k G_k'
\]

**Measurement Update**

\[
L_k = P_{k - 1} C_k' \left[ C_k P_{k - 1} C_k' + R_k \right]^{-1}
\]

\[
\hat{x}_k = \hat{x}_k - k + L_k (y_k - h(\hat{x}_k - k))
\]

\[
P_k = P_{k - 1} - L_k C_k P_{k - 1}
\]

The extended Kalman filter remains popular. So let’s look further.
What are some alternatives to optimization?

Two general alternatives to optimization: *sampling* and *linearization*

- **Linearization**: AKA the extended Kalman filter. Linearize the nonlinear model at the current estimate. Apply the linear system update formulas.
What are some alternatives to optimization?

Two general alternatives to optimization: *sampling* and *linearization*

- **Linearization**: AKA the extended Kalman filter. Linearize the nonlinear model at the current estimate. Apply the linear system update formulas.

\[
A_k = \left. \frac{\partial f(x, w)}{\partial x} \right|_{(\hat{x}_k, 0)} \\
G_k = \left. \frac{\partial f(x, w)}{\partial w} \right|_{(\hat{x}_k, 0)} \\
C_k = \left. \frac{\partial h(x)}{\partial x} \right|_{\hat{x}_k}
\]

**Propagation**

\[
\hat{x}_{k+1}^- = f(\hat{x}_k, 0) \\
P_{k+1}^- = A_k P_k A_k' + G_k Q_k G_k
\]

**Measurement Update**

\[
L_k = P_k^- C_k' [C_k P_k^- C_k + R_k]^{-1} \\
\hat{x}_k = \hat{x}_k^- + L_k (y_k - h(\hat{x}_k^-)) \\
P_k = P_k^- - L_k C_k P_k^-
\]

The extended Kalman filter remains popular. So let's look further.
Two general alternatives to optimization: *sampling* and *linearization*

- **Linearization**: AKA the extended Kalman filter. Linearize the nonlinear model at the current estimate. Apply the linear system update formulas.

\[
A_k = \left. \frac{\partial f(x, w)}{\partial x} \right|_{(\hat{x}_k, 0)} \\
G_k = \left. \frac{\partial f(x, w)}{\partial w} \right|_{(\hat{x}_k, 0)} \\
C_k = \left. \frac{\partial h(x)}{\partial x} \right|_{\hat{x}_k^–}
\]

**Propagation**

\[
\hat{x}_{k+1}^– = f(\hat{x}_k, 0) \\
P_{k+1}^– = A_k P_k A_k' + G_k Q_k G_k
\]

**Measurement Update**

\[
L_k = P_k^- C_k' [C_k P_k^- C_k' + R_k]^{-1} \\
\hat{x}_k = \hat{x}_{k}^- + L_k (y_k - h(\hat{x}_k^-)) \\
P_k = P_k^- - L_k C_k P_k^–
\]

- The extended Kalman filter remains popular. So let’s look further.
EKF Summary

- Limited supporting theory.
Limited supporting theory.
Convergence only if start close enough to the true state
deterministic: (Reif and Unbehauen, 1999; Song and Grizzle, 1995)
stochastic: (Reif et al., 1999, 2000).
EKF Summary

- Limited supporting theory.
  Convergence only if start close enough to the true state
deterministic: (Reif and Unbehauen, 1999; Song and Grizzle, 1995)
stochastic: (Reif et al., 1999, 2000).
- Optimal choice of $Q_w$, $R_v$ and $P_0$ known only for linear case (KF).
EKF Summary

- Limited supporting theory. Convergence only if start close enough to the true state
deterministic: (Reif and Unbehauen, 1999; Song and Grizzle, 1995)
stochastic: (Reif et al., 1999, 2000).

- Optimal choice of $Q_w$, $R_v$ and $P_0$ known only for linear case (KF).
Performance and convergence properties of the EKF depend on the
parameter choices. (Reif et al., 1996; Boutayeb and Aubry, 1999).
EKF Summary

- Limited supporting theory. Convergence only if start close enough to the true state.
  Deterministic: (Reif and Unbehauen, 1999; Song and Grizzle, 1995)
  Stochastic: (Reif et al., 1999, 2000).

- Optimal choice of \( Q_w, R_v \) and \( P_0 \) known only for linear case (KF).
  Performance and convergence properties of the EKF depend on the parameter choices.
  (Reif et al., 1996; Boutayeb and Aubry, 1999).

- Hard constraints are difficult to handle.
Limited supporting theory. Convergence only if start close enough to the true state
deterministic: (Reif and Unbehauen, 1999; Song and Grizzle, 1995)
stochastic: (Reif et al., 1999, 2000).

Optimal choice of $Q_w$, $R_v$ and $P_0$ known only for linear case (KF).
Performance and convergence properties of the EKF depend on the
parameter choices. (Reif et al., 1996; Boutayeb and Aubry, 1999).

Hard constraints are difficult to handle. Many modifications proposed to include constraints, but none have proven generally effective (Simon, 2010).
Let’s break the EKF. What are the ingredients?

1. Nonlinearity and a bad prior, i.e., a disturbance
Let’s break the EKF. What are the ingredients?

1. Nonlinearity and a bad prior, i.e., a disturbance
2. Processing one measurement at a time is inadequate
Let’s break the EKF. What are the ingredients?

1. Nonlinearity and a bad prior, i.e., a disturbance
2. Processing one measurement at a time is inadequate
3. Multiple local optima (multimodal probability density)
If the world were linear...we probably wouldn’t be here!

Figure: Density Propagation in Linear Systems.
Nonlinear System: Multimodal Density

Bimodal Conditional Density

\[ p(x_0|y_0) \]

\[ p(x_1|y_1, y_0) \]

\[ p(x_2|y_2, y_1) \]

\[ p(x_3|y_3, y_2) \]
Estimate the concentrations of A, B, and C
Estimate the concentrations of A, B, and C
Model

\[
\frac{d}{dt} \begin{bmatrix} c_A \\ c_B \\ c_C \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} k_1 c_A - k_{-1} c_B c_C \\ k_2 c_B^2 - k_{-2} c_C \end{bmatrix}
\]

\[
A \xrightleftharpoons[k_{-1}]^{k_1} B + C
\]

\[
2B \xrightleftharpoons[k_{-2}]^{k_2} C
\]

Measure the total pressure

\[
y = RT (c_A + c_B + c_C)
\]

Poor initial guess

\[
x(0) = \begin{bmatrix} 0.5 \\ 0.05 \\ 0 \end{bmatrix}
\]

\[
'x(0) = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}
\]
Estimate the concentrations of A, B, and C

Model

\[
\frac{d}{dt} \begin{bmatrix}
c_A \\
c_B \\
c_C 
\end{bmatrix} = \begin{bmatrix}
-1 & 0 \\
1 & -2 \\
1 & 1 
\end{bmatrix} \begin{bmatrix}
k_1 c_A - k_{-1} c_B c_C \\
k_2 c_B^2 - k_{-2} c_C 
\end{bmatrix}
\]

Measure the total pressure

\[
y = RT \left( c_A + c_B + c_C \right)
\]

A  \xrightleftharpoons[k_{-1}]{k_1}  B + C

2B  \xrightleftharpoons[k_{-2}]{k_2}  C
Estimate the concentrations of A, B, and C

Model

\[
\frac{d}{dt} \begin{bmatrix} c_A \\ c_B \\ c_C \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} k_1 c_A - k_{-1} c_B c_C \\ k_2 c_B^2 - k_{-2} c_C \end{bmatrix}
\]

Measure the total pressure

\[
y = RT \left( c_A + c_B + c_C \right)
\]

Poor initial guess

\[
x(0) = \begin{bmatrix} 0.5 \\ 0.05 \\ 0 \end{bmatrix} \quad \text{vs.} \quad \bar{x}(0) = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}
\]
Performance of EKF and Nonlinear MHE

Both estimators give good output estimation (noise filtering)
NMHE converges to the true states. The EKF does not.

- The tradeoff of NMHE window size: estimation accuracy (large $N$) versus online efficiency (small $N$)
A large disturbance plus a small initial variance \((P_0)\) spells trouble

Could we weaken the prior (increase \(P_0\)) when a disturbance is detected?

Let’s see how the EKF behaves given different \(P_0\) values
The EKF converges in the blue region and does not converge in the red region.

Given the complexity of this map, it’s difficult to choose tuning parameters systematically.
Improve EKF by tuning the initial variance?

- The EKF converges in the **blue** region and does not converge in the **red** region.
- Given the complexity of this map, it’s difficult to choose tuning parameters systematically.
Simplify NMHE – just use LMHE in the nonlinear system?

- Simplify the algorithm from NMHE to LMHE while keeping the optimization basis
- Separate linearization from optimization
- For a window with size $N$, use either the old LMHE results (from time $T - N + 1$ to $T - 1$) or the current model prediction (time $T$) as the basis of the linearization
- Still keep the same constraint setting
Simplify NMHE – just use LMHE in the nonlinear system?

- LMHE's convergence is similar to NMHE
- The computation time is reduced
Another way to simplify NMHE computation

- If we really want an optimization-based estimator but as fast as EKF
- Try such a simplified version of MHE:
  - Use a horizon of **one**
  - Keep constraints the same

\[
\begin{align*}
\min_{x_T} \quad & V = |x_T - \bar{x}_T|^2_P + |v_T|^2_R \\
\text{s.t.} \quad & v_T = y_T - h(x_T) \\
& x_k \geq 0
\end{align*}
\]

- It considers only **one** measurement at one time like EKF
- Information of \( w \) is just used to determine the prior weight \( (P) \), similar with EKF
- The speed is very close to EKF
Another way to simplify NMHE computation

With the positivity constraint:

The same approach without constraints does not work – presence of constraints can improve the performance significantly after a large disturbance.

EKF cannot implement hard constraints as well as MHE. A simple clipping approach does not work.
Conclusions

- The theory perspective

- We have reasonable theory for full information estimation.
- By extension, we can expect reasonable theoretical properties for MHE.
- The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

- Try the EKF! It usually (sometimes, never) works!
- If that fails, try MHE, it usually (always, sometimes) works! But it’s a slog to compute.
- If that fails, try full information, it always (usually, sometimes) works! Good luck managing the computation.
Conclusions

The theory perspective

▶ We have reasonable theory for full information estimation.
Conclusions

- The theory perspective
  - We have reasonable theory for full information estimation.
  - By *extension*, we can expect reasonable theoretical properties for MHE.
Conclusions

The theory perspective

- We have reasonable theory for full information estimation.
- By *extension*, we can expect reasonable theoretical properties for MHE.
- The “theory” for EKF remains: if the system is (almost) linear, the EKF works.
Conclusions

The theory perspective

- We have reasonable theory for full information estimation.
- By *extension*, we can expect reasonable theoretical properties for MHE.
- The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.
Conclusions

- **The theory perspective**
  - We have reasonable theory for full information estimation.
  - By *extension*, we can expect reasonable theoretical properties for MHE.
  - The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

- **The practice perspective**
Conclusions

- The theory perspective
  - We have reasonable theory for full information estimation.
  - By *extension*, we can expect reasonable theoretical properties for MHE.
  - The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

- The practice perspective
  - Try the EKF! It usually (sometimes, never) works!
Conclusions

- The theory perspective
  - We have reasonable theory for full information estimation.
  - By extension, we can expect reasonable theoretical properties for MHE.
  - The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

- The practice perspective
  - Try the EKF! It usually (sometimes, never) works!
  - If that fails, try MHE, it usually (always, sometimes) works!
Conclusions

The theory perspective

- We have reasonable theory for full information estimation.
- By *extension*, we can expect reasonable theoretical properties for MHE.
- The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

The practice perspective

- Try the EKF! It usually (sometimes, never) works!
- If that fails, try MHE, it usually (always, sometimes) works! But it’s a slog to compute.
Conclusions

- The theory perspective
  - We have reasonable theory for full information estimation.
  - By extension, we can expect reasonable theoretical properties for MHE.
  - The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

- The practice perspective
  - Try the EKF! It usually (sometimes, never) works!
  - If that fails, try MHE, it usually (always, sometimes) works!
    But it’s a slog to compute.
  - If that fails, try full information, it always (usually, sometimes) works!
Conclusions

- The theory perspective
  - We have reasonable theory for full information estimation.
  - By *extension*, we can expect reasonable theoretical properties for MHE.
  - The “theory” for EKF remains: if the system is (almost) linear, the EKF works. But the practice is more encouraging.

- The practice perspective
  - Try the EKF! It usually (sometimes, never) works!
  - If that fails, try MHE, it usually (always, sometimes) works!
    But it’s a slog to compute.
  - If that fails, try full information, it always (usually, sometimes) works!
    Good luck managing the computation.
Where next?

- Open theory problems.

- Tighten link between full information's and MHE's theoretical properties.

- Suboptimal MHE to reduce computational requirements.

- Treat bounded disturbances rather than converging disturbances.

Open computational problems.

- Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.

- Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

Open organizational problems.

- Software engineering. How does the community want to develop, extend, maintain, and share software?

- What is the role of proprietary, commercial software?

- What is the role of sponsored university software?

- What is the role (if any) of free software?
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.

- Open computational problems.
  - It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous.
    - Scalability of optimization software is mandatory.

- Open organizational problems.
  - Software engineering.
    - How does the community want to develop, extend, maintain, and share software?
  - What is the role of proprietary, commercial software?
  - What is the role of sponsored university software?
  - What is the role (if any) of free software?
Where next?

* Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency!
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
Where next?

- **Open theory problems.**
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- **Open computational problems.**
  - Computational efficiency! It would be a lot easier to do research, *even on theory*, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

- Open organizational problems.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, *even on theory*, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

- Open organizational problems.
  - Software engineering.
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

- Open organizational problems.
  - Software engineering. How does the community want to develop, extend, maintain, and share software?
Where next?

- **Open theory problems.**
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- **Open computational problems.**
  - Computational efficiency! It would be a lot easier to do research, *even on theory*, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

- **Open organizational problems.**
  - Software engineering. How does the community want to develop, extend, maintain, and share software?
  - What is the role of proprietary, commercial software?
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

- Open organizational problems.
  - Software engineering. How does the community want to develop, extend, maintain, and share software?
  - What is the role of proprietary, commercial software?
  - What is the role of sponsored university software?
Where next?

- Open theory problems.
  - Tighten link between full information’s and MHE’s theoretical properties.
  - Suboptimal MHE to reduce computational requirements.
  - Treat bounded disturbances rather than converging disturbances.

- Open computational problems.
  - Computational efficiency! It would be a lot easier to do research, even on theory, if MHE computations were faster and more reliable.
  - Multiprocessor are ubiquitous. Scalability of optimization software is mandatory.

- Open organizational problems.
  - Software engineering. How does the community want to develop, extend, maintain, and share software?
  - What is the role of proprietary, commercial software?
  - What is the role of sponsored university software?
  - What is the role (if any) of free software?
Further reading


