State Estimation of Linear and Nonlinear Dynamic Systems
Part IV: Nonlinear Systems: Moving Horizon Estimation (MHE) and Particle Filtering (PF)

James B. Rawlings and Fernando V. Lima
Department of Chemical and Biological Engineering
University of Wisconsin–Madison
AICES Regional School
RWTH Aachen
March 17, 2008

Outline

1. The Challenge of Nonlinear Estimation
2. Moving Horizon Estimation (MHE)
3. Particle Filtering (PF)
4. Combining PF and MHE
5. Conclusions
6. Further Reading
The Challenge of Nonlinear Estimation

**Linear Estimation**

Estimation Possibilities:
- **one** state is the optimal estimate
- **infinitely many** states are optimal estimates (unobservable)

**Nonlinear Estimation**

Estimation Possibilities:
- **one** state is the optimal estimate
- **infinitely many** states are optimal estimates (unobservable)
- **finitely many** states are locally optimal estimates

---

Full Information Estimation

Nonlinear model, Gaussian noise,

\[
\begin{align*}
x(k + 1) &= F(x, u) + G(x, u)w \\
y(k) &= h(x) + v
\end{align*}
\]

The trajectory of states

\[
X(T) := \{x(0), \ldots, x(T)\}
\]

Maximizing the conditional density function

\[
\max_{X(T)} p_{X|Y}(X(T)|Y(T))
\]
Equivalent Optimization Problem

Using the model and taking logarithms

\[
\min_{X(T)} V_0(x(0)) + \sum_{j=0}^{T-1} L_w(w(j)) + \sum_{j=0}^{T} L_v(y(j) - h(x(j)))
\]

subject to \( x(j + 1) = F(x, u) + w \quad (G(x, u) = I) \)

\[
V_0(x) := -\log(p_{x(0)}(x))
\]
\[
L_w(w) := -\log(p_{w}(w))
\]
\[
L_v(v) := -\log(p_{v}(v))
\]

What do we do when we have a new measurement?

- Resolve optimization problem with \( T + 1 \) stages.
  \( \implies \) Size of optimization increases with time.
- Employ moving horizon approximation.
  \( \implies \) Bound size of optimization with approximate estimator.
Adding new Observations to the Estimation Problem

It occasionally happens that after we have completed all parts of an extended calculation on a sequence of observations, we learn of a new observation that we would like to include. In many cases we will not want to have to redo the entire elimination but instead to find the modifications due to the new observation in the most reliable values of the unknowns and in their weights.

C.F. Gauss, 1823

Moving Horizon Estimation

- In the Moving Horizon Estimation (MHE) strategy
  - The most recent $N$ states are considered
Most recent $N$ states $X(T - N : T) := \{x(T - N) \ldots x(T)\}$

Optimization problem

$$\min_{X(T-N:T)} \mathbb{V}_{T-N}(x(T-N)) + \sum_{j=T-N}^{T-1} L_w(w(j)) + \sum_{j=T-N}^{T} L_v(y(j) - h(x(j)))$$

subject to $x(j + 1) = F(x, u) + w$.

What does moving horizon estimation have to offer?

linear model
Gaussian noise
stability \quad \Rightarrow \quad \text{Kalman Filter}

linear model
general noise
inequality constraints
stability \quad \Rightarrow \quad \text{MHE}

nonlinear model
Gaussian noise \quad \Rightarrow \quad \text{Extended Kalman Filter}

nonlinear model
general noise
inequality constraints
stability \quad \Rightarrow \quad \text{MHE}
Literature Summary

- **Data Reconciliation/Moving Horizon Estimation**

- **Moving Horizon Observers**

- **Constrained Moving Horizon Estimation**
  - Meadows et al. (1993): Linear constrained estimation
  - Muske and Rawlings (1995): Linear and nonlinear MHE
  - Robertson and Lee (Robertson et al., 1996; Robertson and Lee, 2002): Linear and nonlinear MHE, constraints, truncated distributions
  - Tyler and Morari (1996): Linear MHE, constraints
  - Findeisen (1997): Linear MHE, constraints
  - Rao, Rawlings, Mayne, Lee (Rao et al., 2003, 2001; Rao and Rawlings, 2002; Michalska and Mayne, 1995): Linear and nonlinear MHE, constraints

---

Moving Horizon Approximation

What are the consequences of neglecting old data?

- Sensitivity to noise, high gain estimator.
- Divergence.
Potential pitfalls of neglecting past data

By neglecting or too weakly weighting the past data, the estimator may be sensitive to outliers or noise.
⇒ Account for past data using approximate statistic.

Potential pitfalls of improperly approximating past data

By weighting the past data or the prior information too strongly, the estimator may be unable to keep up with data. Estimator divergence may result.
⇒ We require some forgetting to improve the robustness.
How are full information and moving horizon estimation related?

⇒ Forward Dynamic Programming

\[
\Phi_T = \Gamma(x_0 - \bar{x}) + \sum_{k=0}^{T-1} L(w_k, v_k)
\]

\[
= \Gamma(x_0 - \bar{x}) + \sum_{k=0}^{T-N-1} L(w_k, v_k) + \sum_{k=T-N}^{T-1} L(w_k, v_k)
\]

Cost associated with arriving at \(x_{T-N}\)

Arrival Cost

\(\Xi_{T-N}(x_{T-N})\)

Uniquely determined by \(x_{T-N}\) and \(\{w_k\}_{k=T-N}^{T-1}\)

Fixed Horizon Estimation Problem

Arrival Cost

\[
\Xi_T
\]

\[
\bar{x}
\]

\[
T
\]
Moving Horizon Estimation — Optimization Problem

\[
\min_{\{x_k\}} \Theta_T = \sum_{k=T-N}^{T-1} L(w_k, v_k) + \Gamma_{T-N}(x_{T-N} - \hat{x}_{T-N|T-N-1})
\]

- Initial penalty \(\Gamma_{T-N}\) summarizes past data by penalizing deviation away from past estimate.
- If the initial penalty is equal to the arrival cost, then the full information and moving horizon estimation problems are equivalent.
For unconstrained linear systems with quadratic objectives, we can calculate the arrival cost with the Kalman filter covariance. Moving horizon estimation reduces to Kalman filtering.

For constrained linear systems with quadratic objectives, we can globally lower bound the arrival cost with the Kalman filter covariance.

When the system is nonlinear, we cannot in general calculate a globally lower bound to the arrival cost with the exception of the trivial choice: $\Gamma_T = 0$.

One solution: Generate lower bound online (Rao, 2000).

One year from now I try to reconstruct my travel to Aachen

- Flew from Madison to a larger airport.
  Data: “I don’t remember very well, but it was a short flight...”
- Flew from that larger airport to Frankfurt.
  Data: “I remember this flight very well. It took forever; it was about 7 hours...”
Arrival Cost Approximation — Current Research in MHE

The statistically correct choice for the arrival cost is the conditional density of $x(T - N)|Y(T - N - 1)$

$$V_{T-N}(x) = -\log p_{x(T-N)|Y}(x|Y(T - N - 1))$$

Arrival cost approximations (Rao et al., 2003)
- uniform prior (and large $N$)
- EKF covariance formula
- MHE smoothing

Particle filtering — sampled densities

$$p_s(x) = \sum_{i=1}^{s} w_i \delta(x - x_i) \quad x_i \text{ samples (particles)} \quad w_i \text{ weights}$$

Exact density $p(x)$ and a sampled density $p_s(x)$ with five samples for $\xi \sim N(0,1)$
Convergence — cumulative distributions

![Cumulative distribution graph]

Corresponding exact $P(x)$ and sampled $P_s(x)$ cumulative distributions

Importance sampling

In state estimation, $p$ of interest is easy to evaluate but difficult to sample. We choose an importance function, $q$, instead. When we can sample $p$, the sampled density is

$$p_s = \left\{ x_i, \quad w_i = \frac{1}{s} \right\} \quad p_{sa}(x_i) = p(x_i)$$

When we cannot sample $p$, the importance sampled density $\overline{p}_s(x)$ is

$$\overline{p}_s = \left\{ x_i, \quad w_i = \frac{1}{s} \frac{p(x_i)}{q(x_i)} \right\} \quad p_{is}(x_i) = q(x_i)$$

Both $\overline{p}_s(x)$ and $p_s(x)$ are unbiased and converge to $p(x)$ as sample size increases (Smith and Gelfand, 1992).
Density to be sampled $p(x)$.

Importance function that can be sampled $q(x)$.

Importance function $q(x)$ and its histogram based on 5000 samples.

Exact density $p(x)$ and its histogram based on 5000 importance samples.
Importance sampled particle filter (Arulampalam et al., 2002)

\[ p(x(k+1)|Y(k+1)) = \{x_i(k+1), w_i(k+1)\} \]

\( x_i(k+1) \) is a sample of \( q(x(k+1)|x_i(k), y(k+1)) \)

\[ w_i(k+1) = w_i(k) \frac{p(y(k+1)|x_i(k+1))p(x_i(k+1)|x_i(k))}{q(x_i(k+1)|x_i(k), y(k+1))} \]

The importance sampled particle filter converges to the conditional density with increasing sample size. It is biased for finite sample size.

Research challenge — placing the particles

- Optimal importance function (Doucet et al., 2000). Restricted to linear measurement \( y = Cx + v \).
- Resampling
- Curse of dimensionality
Optimal importance function

Particles’ locations versus time using the optimal importance function; 250 particles. Ellipses show the 95% contour of the true conditional densities before and after measurement.

Resampling

How to resample without bias

- Partition $[0, 1]$ with original sample weights, $w_i$.
- Arrows depict the outcome of drawing three uniformly distributed random numbers.
- Sample $x_2$ is discarded and sample $x_3$ is repeated twice in the resample.
- The new sample’s weights are $\tilde{w}^1 = \tilde{w}^2 = \tilde{w}^3 = 1/3$. 
Resampling

<table>
<thead>
<tr>
<th>Original sample</th>
<th>Resample</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>state</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$\bar{x}_1$</td>
</tr>
<tr>
<td>$w_1 = \frac{3}{10}$</td>
<td>$\bar{x}_2 = x_3$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$\bar{x}_2 = x_3$</td>
</tr>
<tr>
<td>$w_2 = \frac{1}{10}$</td>
<td>$\bar{x}_3 = x_3$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$\bar{x}_3 = x_3$</td>
</tr>
<tr>
<td>$w_3 = \frac{6}{10}$</td>
<td>$\bar{x}_3 = x_3$</td>
</tr>
</tbody>
</table>

The properties of the resamples are summarized by

$$p_{rc}(\bar{x}_i) = \begin{cases} w_j, & \bar{x}_i = x_j \\ 0, & \bar{x}_i \neq x_j \end{cases} \quad \bar{w}_i = \frac{1}{s}, \quad \text{all } i$$

Particles’ locations versus time using the optimal importance function with resampling; 250 particles.
The MHE and particle filtering hybrid approach

Hybrid implementation
- Use the MHE optimization to locate/relocate the samples
- Use the PF to obtain fast state estimates between MHE optimizations

Application: Semi-Batch Reactor
- Reaction: \( 2A \rightarrow B \)
- \( k = 0.16 \)
- Measurement is \( C_A + C_B \)
- \( x_0 = [3 \ 1]^T \)

\[
\frac{dC_A}{dt} = -2kC_A^2 + \frac{F_i}{V} C_A0 \\
\frac{dC_B}{dt} = kC_A^2 + \frac{F_i}{V} C_B0
\]

- Noise covariances \( Q_w = \text{diag}(0.01^2, 0.01^2) \) and \( R_v = 0.01^2 \)
- **Bad Prior**: \( \bar{x}_0 = [0.1 \ 4.5]^T \) with a large \( P_0 \)
- **Unmodelled Disturbance**: \( C_{A0}, C_{B0} \) is pulsed at \( t_k = 5 \)
Using only MHE

- MHE implemented with $N = 15 (t = 1.5)$ and a smoothed prior
- MHE recovers robustly from poor priors and unmodelled disturbances

![Graph showing concentration and measurement estimates over time](image)

Using only particle filter

- Particle filter implemented with the Optimal importance function: 
  $p(x_k|x_{k-1}, y_k)$, 50 samples, Resampling
- The PF samples never recover from a poor $\bar{x}_0$. Not robust

![Graph showing concentration and measurement estimates over time](image)
MHE/PF hybrid with a simple importance function

- Importance function for PF: \( p(x_k|x_{k-1}) \), 50 samples
- The PF samples recover from a poor \( \bar{x}_0 \) and the unmodelled disturbance only after the MHE relocates the samples

MHE/PF hybrid with an optimal importance function

- The optimal importance function: \( p(x_k|x_{k-1}, y_k) \), 50 samples
- MHE relocates the samples after a poor \( \bar{x}_0 \), but samples recover from the unmodelled disturbance without needing the MHE
Conclusions

- Optimal state estimation of the linear dynamic system is the gold standard of state estimation.
- MHE is a good option for linear, constrained systems.
- The classic solution for nonlinear systems, the EKF, has been superseded. The UKF, for example, is an easily implemented alternative worth evaluating.
- MHE and particle filtering are higher-quality solutions for nonlinear models. MHE is robust to modeling errors but requires an online optimization. PF is simple to program and fast to execute but may be sensitive to model errors. They require more user experience to set up properly and more computational resources to execute. The payoff can be substantial, however.
- Hybrid MHE/PF methods can combine these complementary strengths.

Future challenges

- Process systems are typically unobservable or ill-conditioned, i.e. nearby measurements do not imply nearby states. We must decide on the subset of states to reconstruct from the data – an additional part to the modeling question.
- Nonlinear systems produce multi-modal densities. We need better solutions for handling these multi-modal densities in real time.
Acknowledgments

- Murali Rajamani of BP
- Professor Bhavik Bakshi of OSU for helpful discussion.
- NSF grant #CNS–0540147
- PRF grant #43321–AC9

Further Reading I


Further Reading II


---

Further Reading III


