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

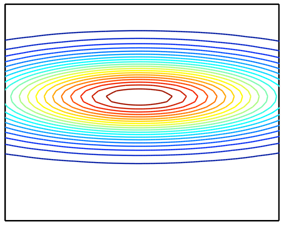
Department of Chemical and Biological Engineering
University of Wisconsin–Madison

AICES Regional School
RWTH Aachen
March 17, 2008

Outline

The Challenge of Nonlinear Estimation

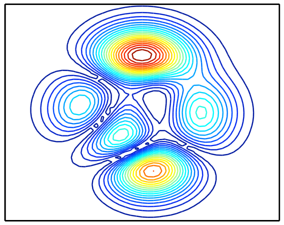
Linear Estimation



Estimation Possibilities:

1. *one* state is the optimal estimate
2. *infinitely many* states are optimal estimates (unobservable)

Nonlinear Estimation



Estimation Possibilities:

1. *one* state is the optimal estimate
2. *infinitely many* states are optimal estimates (unobservable)
3. *finitely many* states are locally optimal estimates

Full Information Estimation

Nonlinear model, Gaussian noise,

$$x(k+1) = F(x,u) + G(x,u)w$$
$$y(k) = h(x) + v$$

The trajectory of states

$$X(T) := \{x(0), \dots x(T)\}$$

Maximizing the conditional density function

$$\max_{X(T)} p_{X|Y}(X(T)|Y(T))$$

Equivalent Optimization Problem

Using the model and taking logarithms

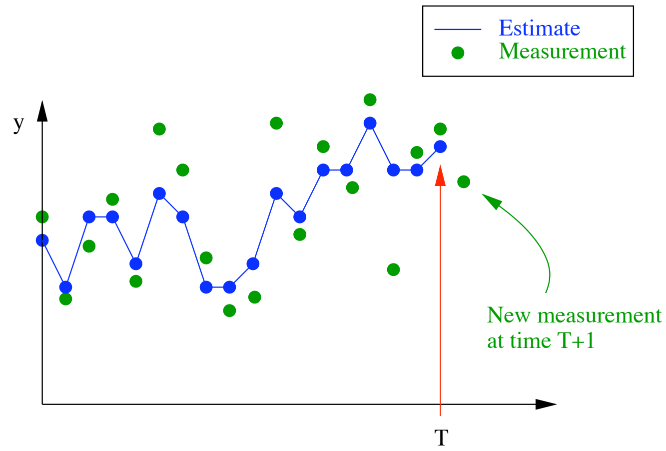
$$\min_{X(T)} V_0(x(0)) + \sum_{j=0}^{T-1} L_w(w(j)) + \sum_{j=0}^{T} L_v(y(j) - h(x(j)))$$

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

$$V_0(x) := -\log(p_{x(0)}(x))$$

$$L_w(w) := -\log(p_w(w)) \qquad L_v(v) := -\log(p_v(v))$$

What do we do when we have a new measurement?



- Resolve optimization problem with $T+1$ stages.
  $\Longrightarrow$ Size of optimization increases with time.
- Employ moving horizon approximation.
  $\Longrightarrow$ Bound size of optimization with approximate estimator.

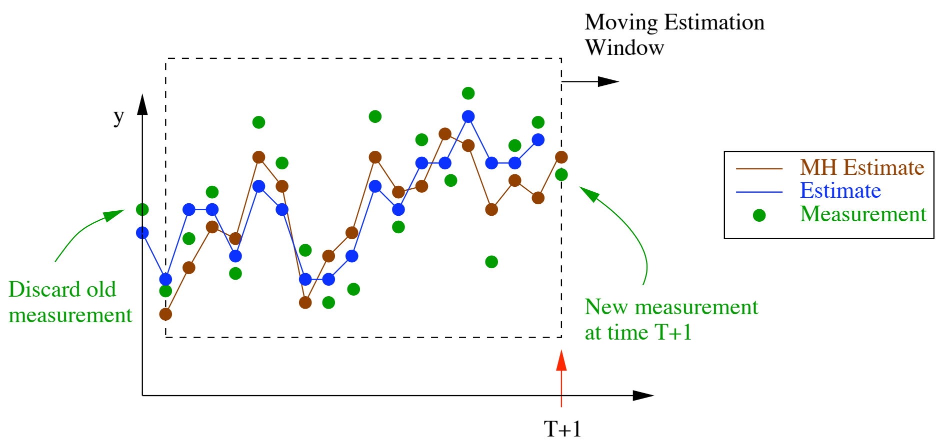
Adding new Observations to the Estimation Problem

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
G.W. Stewart Translation, 1995, p. 191.

Moving Horizon Estimation



- In the Moving Horizon Estimation(MHE) strategy
  - The most recent N states are considered

Arrival Cost and Moving Horizon Estimation

Most recent N states $X(T-N:T) := \{x(T-N)\dots x(T)\}$

Optimization problem

$$\min_{X(T-N:T)} \underbrace{V_{T-N}(x(T-N))}_{\text{arrival cost}} + \sum_{j=T-N}^{T-1} L_w(w(j)) + \sum_{j=T-N}^{T} L_v(y(j)-h(x(j)))$$

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

What does moving horizon estimation have to offer?

linear model, Gaussian noise, stability $\Longrightarrow$ Kalman Filter

linear model, general noise, inequality constraints, stability $\Longrightarrow$ MHE

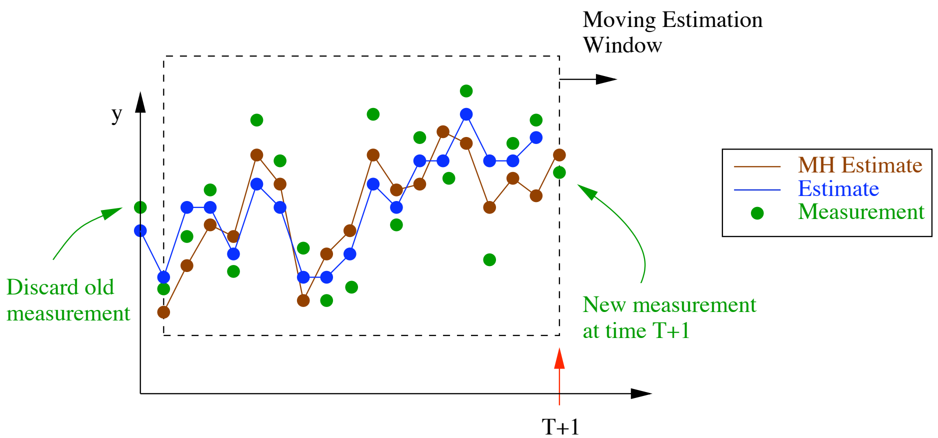
nonlinear model, Gaussian noise $\Longrightarrow$ Extended Kalman Filter

nonlinear model, general noise, inequality constraints, stability $\Longrightarrow$ MHE

Literature Summary

- Data Reconciliation/Moving Horizon Estimation
  Liebman et al. (1992), Kim et al. (1991), Bequette (1991), Ramamurthi et al. (1993), Tjoa and Biegler (1991), Albuquerque and Biegler (1996), Marquardt et al. (M'hamdi et al., 1996; Binder et al., 2002) . . .
- Moving Horizon Observers
  Jang et al. (1986), Zimmer (1994), Michalska and Mayne (1995), Moraal and Grizzle (1995)
- 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

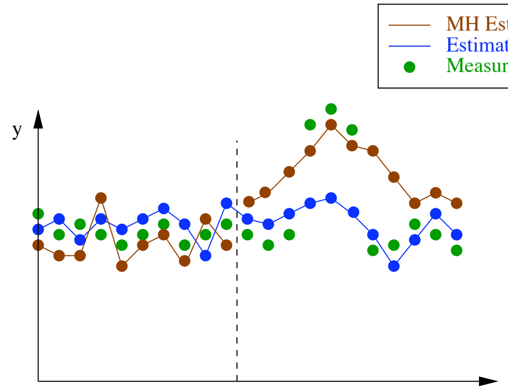
Moving Horizon Approximation



What are the consequences of neglecting old data?

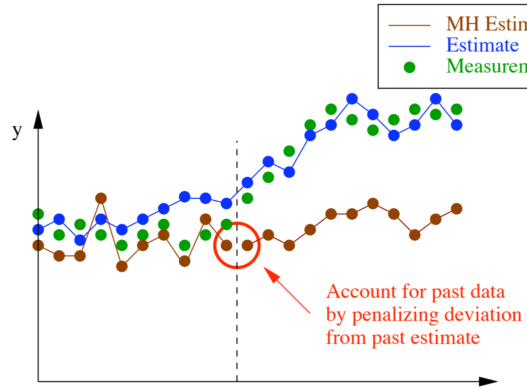
- Sensitivity to noise, high gain estimator.
- Divergence.

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.

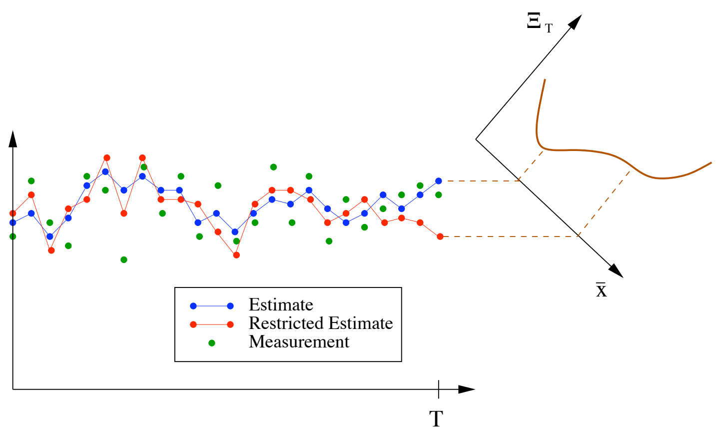
How are full information and moving horizon estimation related?
$\Longrightarrow$ Forward Dynamic Programming

$$\begin{aligned}\Phi_T &= \Gamma(x_0 - \bar{x}) + \sum_{k=0}^{T-1} L(w_k, v_k) \\ &= \underbrace{\Gamma(x_0 - \bar{x}) + \sum_{k=0}^{T-N-1} L(w_k, v_k)}_{\substack{\text{Cost associated with} \\ \text{arriving at } x_{T-N}}} + \underbrace{\sum_{T-N}^{T-1} L(w_k, v_k)}_{\substack{\text{Uniquely determined by} \\ x_{T-N} \text{ and } \{w_k\}_{k=T-N}^{T-1}}}\end{aligned}$$

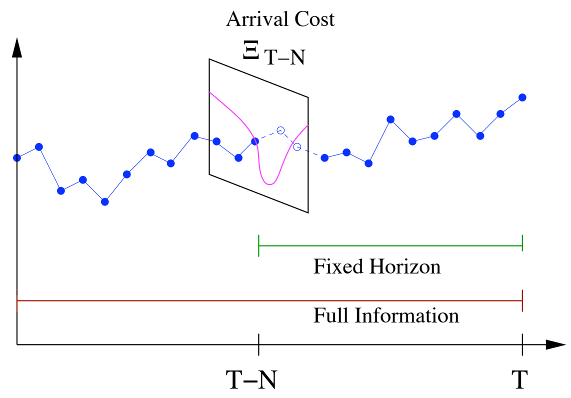
Arrival Cost
$\Xi_{T-N}(x_{T-N})$

Fixed Horizon
Estimation Problem

Arrival Cost



Forward Dynamic Programming Structure



Moving Horizon Estimation — Optimization Problem

$$\min_{\{x_k\}} \Theta_T = \sum_{k=T-N}^{T-1} L(w_k, v_k) + \underbrace{\Gamma_{T-N}(x_{T-N} - \widehat{x}_{T-N|T-N-1})}_{\substack{\text{prior information} \\ \text{arrival cost}}}$$

- Initial penalty Γ_{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.

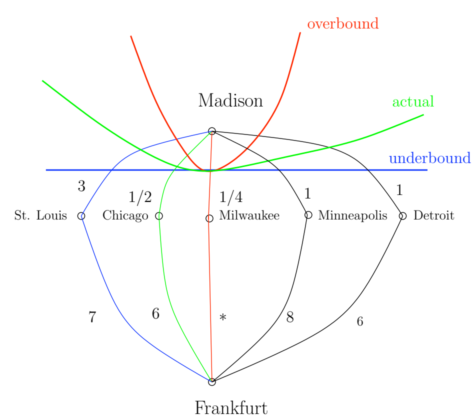
Comments

- 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 $x(T-N)|Y(T-N-1)$

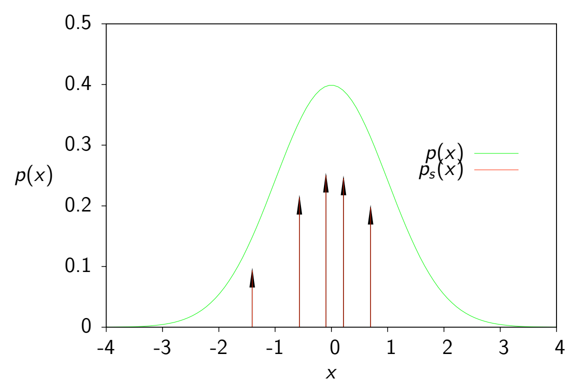
$$V_{T-N}(x) = -\log p_{x(T-N)|Y}(x|Y(T-N-1))$$

Arrival cost approximations (Rao et al., 2003)

- uniform prior (and large N)
- EKF covariance formula
- MHE smoothing

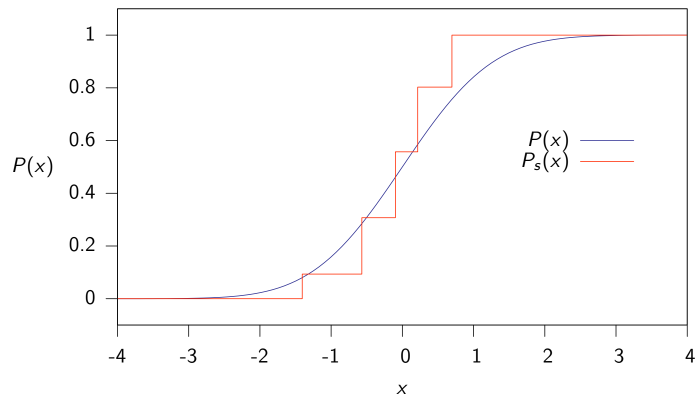
Particle filtering — sampled densities

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



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

Convergence — cumulative distributions



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

Importance sampling

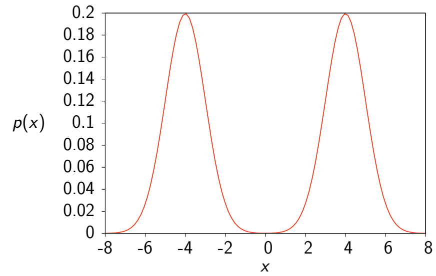
In state estimation, p of interest is easy to *evaluate* but difficult to *sample*.
We choose an *importance function*, q, instead.
When we can sample p, the sampled density is

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

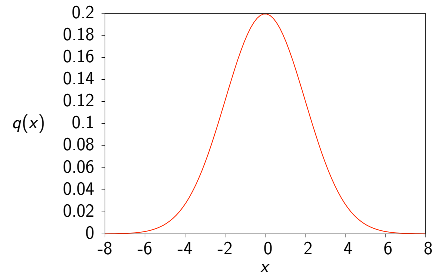
When we cannot sample p, the importance sampled density $\overline{p}_s(x)$ is

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

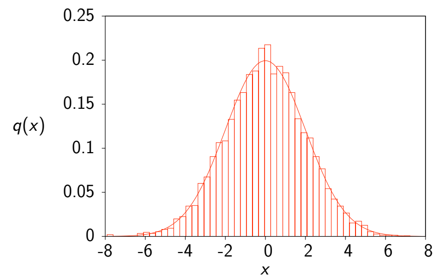
Both $\overline{p}_s(x)$ and $p_s(x)$ are *unbiased* and *converge* to $p(x)$ as sample size increases (Smith and Gelfand, 1992).



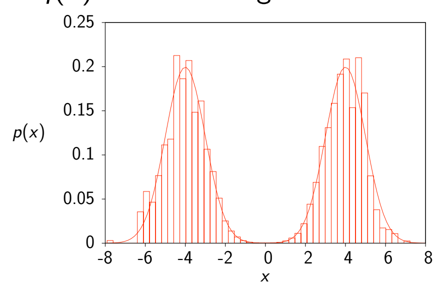
Density to be sampled $p(x)$.



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)

$$p(x(k+1)|Y(k+1)) = \{x_i(k+1), \overline{w}_i(k+1)\}$$

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

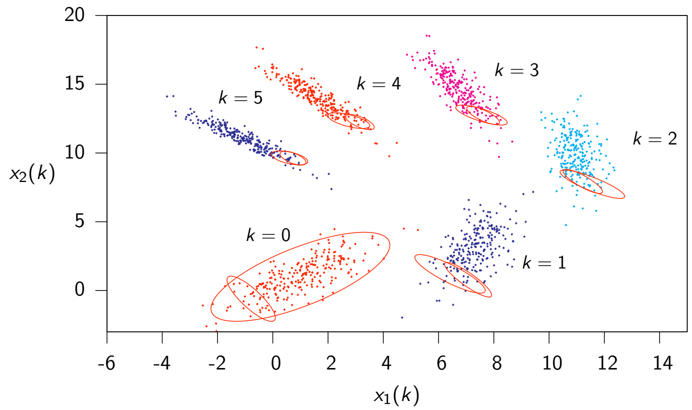
$$w_i(k+1) = w_i(k)\,\frac{p(y(k+1)|x_i(k+1))p(x_i(k+1)|x_i(k))}{q(x_i(k+1)|x_i(k), y(k+1))}$$

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

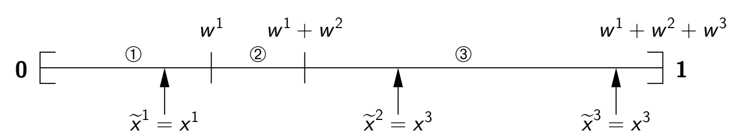
- Optimal importance function (Doucet et al., 2000). Restricted to linear measurement $y = Cx + v$.
- Resampling
- Curse of dimesionality

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

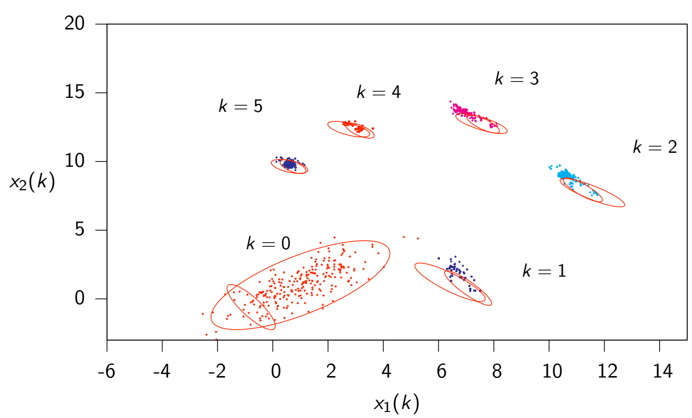
Original sample		Resample	
state	weight	state	weight
x_1	$w_1 = \frac{3}{10}$	$\widetilde{x}_1 = x_1$	$\widetilde{w}_1 = \frac{1}{3}$
x_2	$w_2 = \frac{1}{10}$	$\widetilde{x}_2 = x_3$	$\widetilde{w}_2 = \frac{1}{3}$
x_3	$w_3 = \frac{6}{10}$	$\widetilde{x}_3 = x_3$	$\widetilde{w}_3 = \frac{1}{3}$

The properties of the resamples are summarized by

$$p_{\rm re}(\widetilde{x}_i) = \begin{cases} w_j, & \widetilde{x}_i = x_j \\ 0, & \widetilde{x}_i \neq x_j \end{cases}$$

$$\widetilde{w}_i = 1/s, \quad \text{all } i$$

Resampling



Particles' locations versus time using the optimal importance function with resampling; 250 particles.

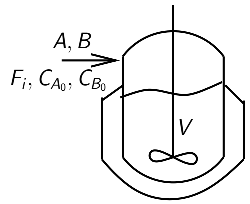
The MHE and particle filtering hybrid approach

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 $C_A + C_B$
- $x_0 = \begin{bmatrix} 3 & 1 \end{bmatrix}^T$

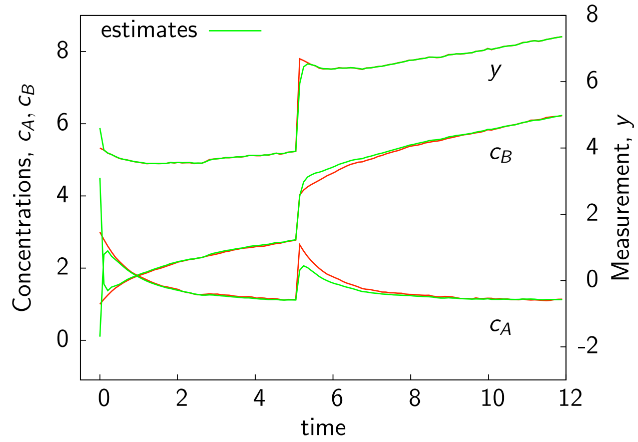


$$\frac{dC_A}{dt} = -2kC_A^2 + \frac{F_i}{V} C_{A_0} \qquad \Delta t = 0.1$$
$$\frac{dC_B}{dt} = kC_A^2 + \frac{F_i}{V} C_{B_0}$$

- Noise covariances $Q_w = \text{diag}\,(0.01^2, 0.01^2)$ and $R_v = 0.01^2$
- **Bad Prior:** $\bar{x}_0 = \begin{bmatrix} 0.1 & 4.5 \end{bmatrix}^T$ with a large P_0
- **Unmodelled Disturbance:** C_{A_0}, C_{B_0} is pulsed at $t_k = 5$

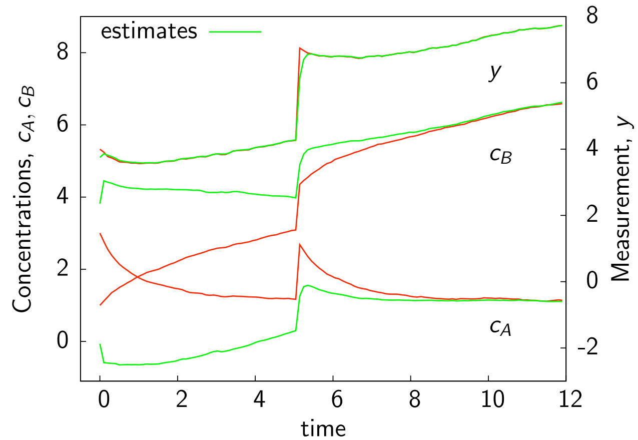
Using only MHE

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



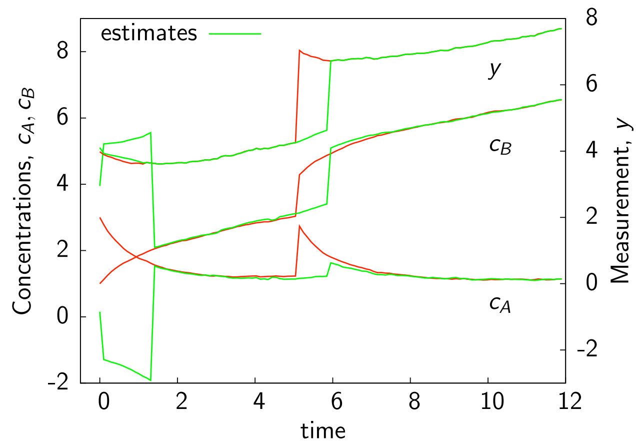
Using only particle filter

- 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



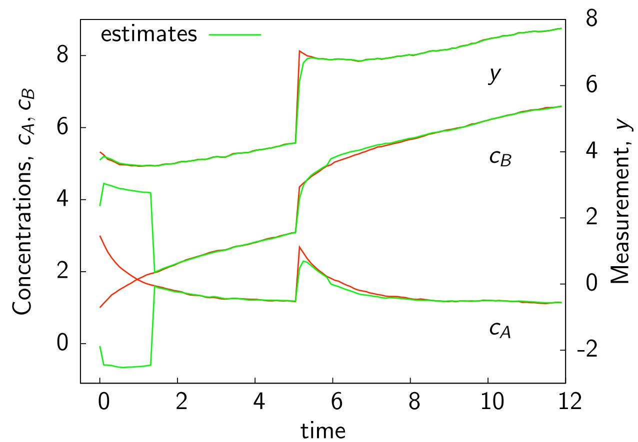
MHE/PF hybrid with a simple importance function

- 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

J. Albuquerque and L. T. Biegler. Data reconciliation and gross-error detection for dynamic systems. *AIChE J.*, 42(10):2841–2856, 1996.

M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50 (2):174–188, February 2002.

B. W. Bequette. Nonlinear predictive control using multi-rate sampling. *Can. J. Chem. Eng.*, 69:136–143, February 1991.

T. Binder, L. Blank, W. Dahmen, and W. Marquardt. On the regularization of dynamic data reconciliation problems. *J. Proc. Cont.*, 12(4):557–567, 2002.

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. and Comput.*, 10:197–208, 2000.

P. Findeisen. Moving horizon state estimation of discrete time systems. Master's thesis, University of Wisconsin–Madison, 1997.

S.-S. Jang, B. Joseph, and H. Mukai. Comparison of two approaches to on-line parameter and state estimation of nonlinear systems. *Ind. Eng. Chem. Proc. Des. Dev.*, 25:809–814, 1986.

Further Reading II

I. Kim, M. Liebman, and T. Edgar. A sequential error-in-variables method for nonlinear dynamic systems. *Comput. Chem. Eng.*, 15(9):663–670, 1991.

M. Liebman, T. Edgar, and L. Lasdon. Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Comput. Chem. Eng.*, 16 (10/11):963–986, 1992.

E. S. Meadows, K. R. Muske, and J. B. Rawlings. Constrained state estimation and discontinuous feedback in model predictive control. In *Proceedings of the 1993 European Control Conference*, pages 2308–2312. European Automatic Control Council, June 1993.

A. M'hamdi, A. Helbig, O. Abel, and W. Marquardt. Newton-type receding horizon control and state estimation. In *Proceedings of the 1996 IFAC World Congress*, pages 121–126, San Francisco, California, 1996.

H. Michalska and D. Q. Mayne. Moving horizon observers and observer-based control. *IEEE Trans. Auto. Cont.*, 40(6):995–1006, 1995.

P. E. Moraal and J. W. Grizzle. Observer design for nonlinear systems with discrete-time measurements. *IEEE Trans. Auto. Cont.*, 40(3):395–404, 1995.

Further Reading III

K. R. Muske and J. B. Rawlings. Nonlinear moving horizon state estimation. In R. Berber, editor, *Methods of Model Based Process Control*, NATO Advanced Study Institute series: E Applied Sciences 293, pages 349–365, Dordrecht, The Netherlands, 1995. Kluwer.

Y. Ramamurthi, P. Sistu, and B. Bequette. Control-relevant dynamic data reconciliation and parameter estimation. *Comput. Chem. Eng.*, 17(1):41–59, 1993.

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 and J. B. Rawlings. Constrained process monitoring: moving-horizon approach. *AIChE J.*, 48(1):97–109, January 2002.

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

C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Trans. Auto. Cont.*, 48(2):246–258, February 2003.

D. G. Robertson and J. H. Lee. On the use of constraints in least squares estimation and control. *Automatica*, 38(7):1113–1124, 2002.

Further Reading IV

D. G. Robertson, J. H. Lee, and J. B. Rawlings. A moving horizon-based approach for least-squares state estimation. *AIChE J.*, 42(8):2209–2224, August 1996.

A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *Amer. Statist.*, 46(2):84–88, 1992.

I. B. Tjoa and L. T. Biegler. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Comput. Chem. Eng.*, 15(10):679–690, 1991.

M. L. Tyler and M. Morari. Stability of constrained moving horizon estimation schemes. *Preprint AUT96–18*, Automatic Control Laboratory, Swiss Federal Institute of Technology, 1996.

G. Zimmer. State observation by on-line minimization. *Int. J. Control*, 60(4):595–606, 1994.