A hybrid approach for state estimation: combining moving horizon estimation and particle filtering

James B. Rawlings and Murali R. Rajamani

Department of Chemical and Biological Engineering
University of Wisconsin–Madison

Sandia CSRI Workshop
Large-Scale Inverse Problems and Quantification of Uncertainty
Santa Fe, New Mexico
September 10–12, 2007

Outline

1. State Estimation of Linear Systems
   - Limitations
2. Moving Horizon Estimation (MHE)
3. Particle Filtering
4. Combining Particle Filtering and MHE
5. Conclusions

The conditional density function

For the linear, time invariant model with Gaussian noise,

$$x(k+1) = Ax + Bu + Gw$$
$$y = Cx + v$$

$$w \sim N(0, Q) \qquad v \sim N(0, R) \qquad x(0) \sim N(\overline{x}_0, Q_0)$$

The conditional density function

For the linear, time invariant model with Gaussian noise,

$$x(k+1) = Ax + Bu + Gw$$
$$y = Cx + v$$

$$w \sim N(0, Q) \qquad v \sim N(0, R) \qquad x(0) \sim N(\overline{x}_0, Q_0)$$

We can compute the conditional density function exactly

$$p_{x|Y}(x|Y(k-1)) = N(\widehat{x}^-, P^-) \qquad \text{(before } y(k)\text{)}$$
$$p_{x|Y}(x|Y(k)) = N(\widehat{x}, P) \qquad \text{(after } y(k)\text{)}$$

Mean and covariance before and after measurement

Forecast

$$\widehat{x}^-(k+1) = A\widehat{x} + Bu \qquad \text{(estimate)}$$

$$P^-(k+1) = APA' + GQG' \qquad \text{(covariance)}$$

$$\widehat{x}^-(0) = \overline{x}_0 \qquad P^-(0) = Q_0 \qquad \text{(initial condition)}$$

Mean and covariance before and after measurement

Forecast

$$\widehat{x}^-(k+1) = A\widehat{x} + Bu \qquad \text{(estimate)}$$

$$P^-(k+1) = APA' + GQG' \qquad \text{(covariance)}$$

$$\widehat{x}^-(0) = \overline{x}_0 \qquad P^-(0) = Q_0 \qquad \text{(initial condition)}$$

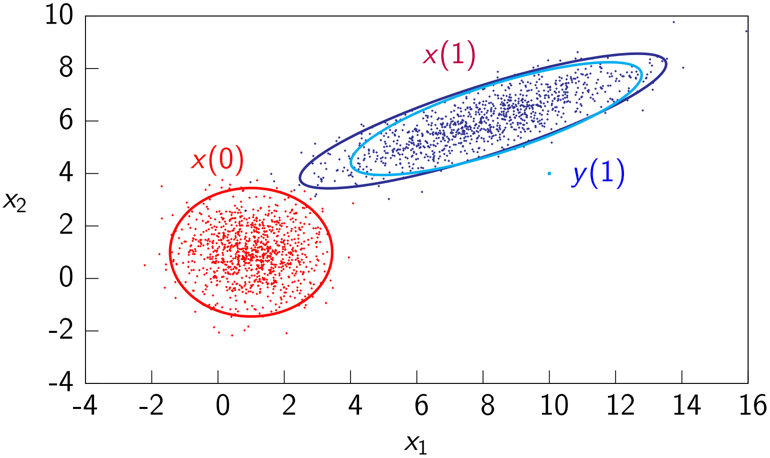
Correction

$$\widehat{x} = \widehat{x}^- + L(y - C\widehat{x}^-) \qquad \text{(estimate)}$$

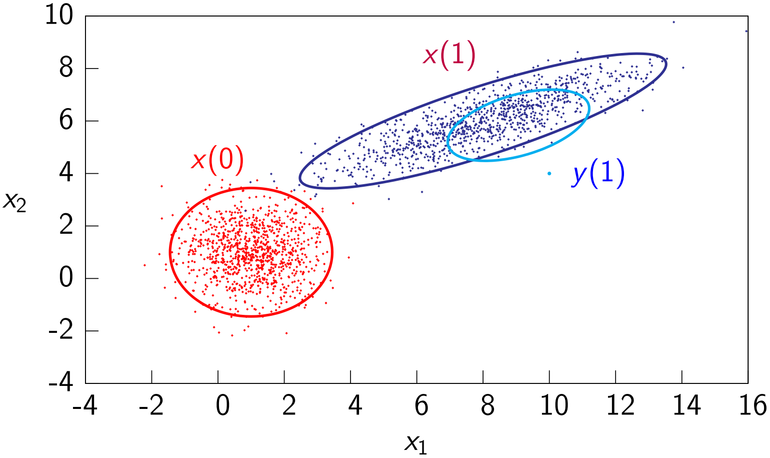
$$L = P^- C'(R + CP^- C')^{-1} \qquad \text{(gain)}$$

$$P = P^- - LCP^- \qquad \text{(covariance)}$$

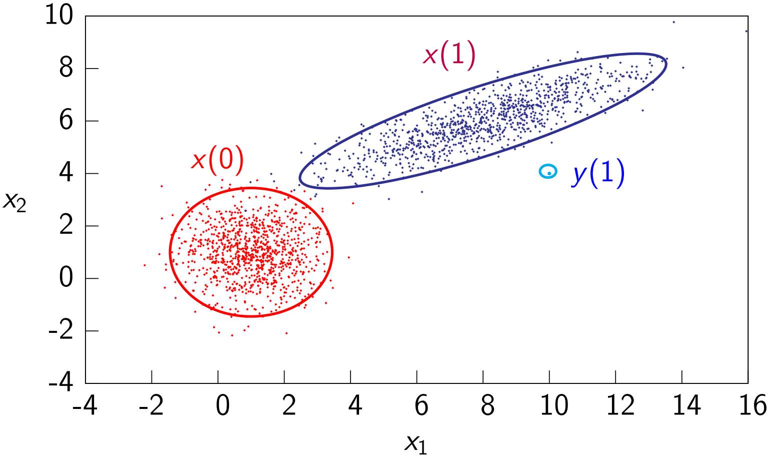
Large R, ignore the measurement, trust the forecast



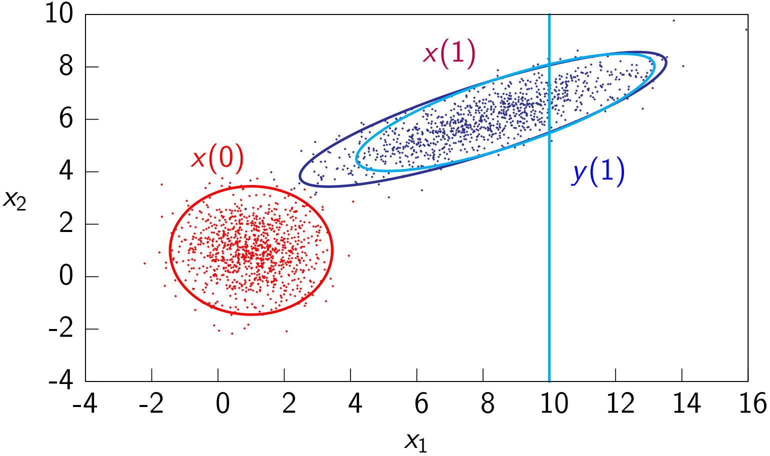
Medium R, blend the measurement and the forecast



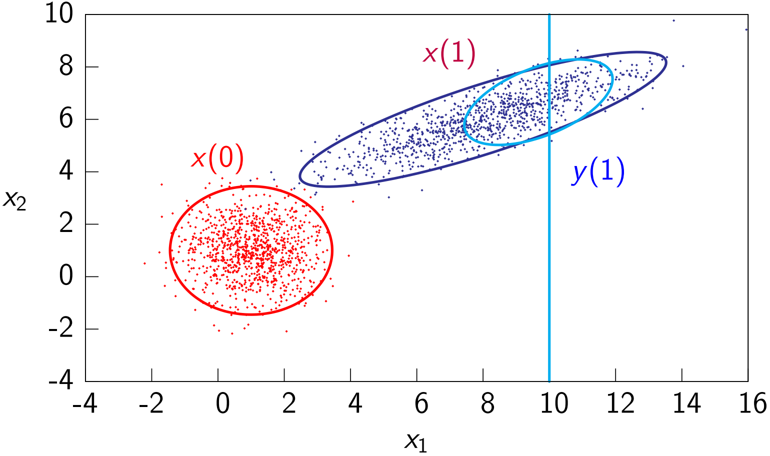
Small R, trust the measurement, override the forecast



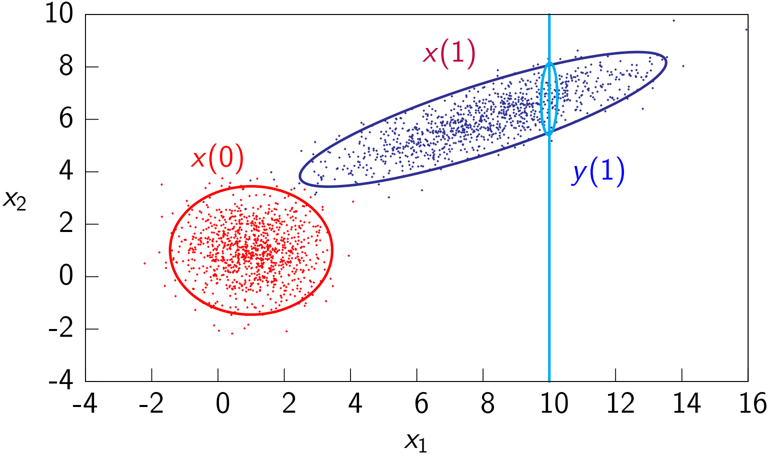
Large R, y measures x_1 only



Medium R, y measures x_1 only



Small R, y measures x_1 only



Limitations of the Kalman filter

- What about constraints?

Limitations of the Kalman filter

- What about constraints?
  - Concentrations, particle size distributions, pressures, temperatures are positive.
  - Using this extra information provides more accurate estimates.
  - Projecting the unconstrained KF estimates to the feasible region is an ad hoc solution that does not satisfy the model.

Limitations of the Kalman filter

- What about constraints?
  - Concentrations, particle size distributions, pressures, temperatures are positive.
  - Using this extra information provides more accurate estimates.
  - Projecting the unconstrained KF estimates to the feasible region is an ad hoc solution that does not satisfy the model.
- What about nonlinear models?

Limitations of the Kalman filter

- What about constraints?
  - Concentrations, particle size distributions, pressures, temperatures are positive.
  - Using this extra information provides more accurate estimates.
  - Projecting the unconstrained KF estimates to the feasible region is an ad hoc solution that does not satisfy the model.
- What about nonlinear models?
  - Almost all physical models in chemical and biological applications are nonlinear differential equations or nonlinear Markov processes.
  - Linearizing the nonlinear model and using the standard update formulas (extended Kalman filter) is the standard industrial approach.

Extended Kalman filter — assessment

The extended Kalman filter is probably the most widely used estimation algorithm for nonlinear systems.

Julier and Uhlmann (2004).

Extended Kalman filter — assessment

The extended Kalman filter is probably the most widely used estimation algorithm for nonlinear systems.

However, more than 35 years of experience in the estimation community has shown that it is difficult to implement, difficult to tune, and only reliable for systems that are almost linear on the time scale of the updates.

Julier and Uhlmann (2004).

Extended Kalman filter — assessment

The extended Kalman filter is probably the most widely used estimation algorithm for nonlinear systems.

However, more than 35 years of experience in the estimation community has shown that it is difficult to implement, difficult to tune, and only reliable for systems that are almost linear on the time scale of the updates.

Many of these difficulties arise from its use of linearization.

Julier and Uhlmann (2004).

Alternative approaches

Options for handling constraints and nonlinearity in state estimation

1. Optimization (moving horizon estimation (MHE))
2. Sampling (particle filtering)

Full information estimation

Nonlinear model, Gaussian noise,

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

Full information estimation

Nonlinear model, Gaussian noise,

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

The trajectory of states

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

Full information estimation

Nonlinear model, Gaussian noise,

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

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

$$\min_{X(T)} V_0(x_0) + \sum_{j=1}^{T-1} L_w(w_j) + \sum_{j=0}^{T} L_v(y_j - h(x_j))$$

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

$$V_0(x) := -\log(p_{x_0}(x))$$

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

Arrival cost and moving horizon estimation

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

Arrival cost and moving horizon estimation

Most recent N states $X(T-N:T) := \{x_{T-N} \ldots x_T\}$

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

Arrival cost approximation

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 approximation

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

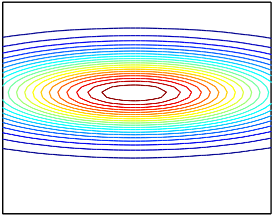
$$V_{T-N}(x) = -\log p_{x_{T-N}|Y}(x|Y(T-N-1))$$

Arrival cost approximations (Rao et al., 2003)

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

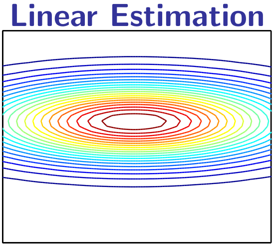
The challenge of nonlinear estimation

Linear Estimation



Estimation Possibilities:

The challenge of nonlinear estimation

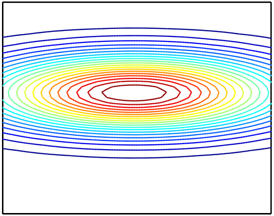


Estimation Possibilities:

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

The challenge of nonlinear estimation

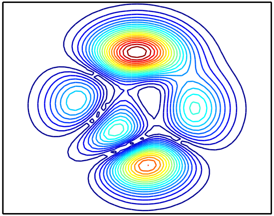
Linear Estimation



Estimation Possibilities:

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

Nonlinear Estimation

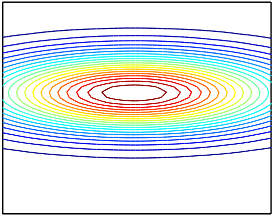


Estimation Possibilities:

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

The challenge of nonlinear estimation

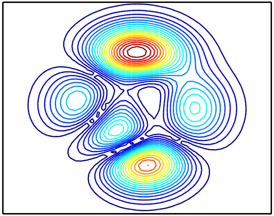
Linear Estimation



Estimation Possibilities:

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

Nonlinear Estimation

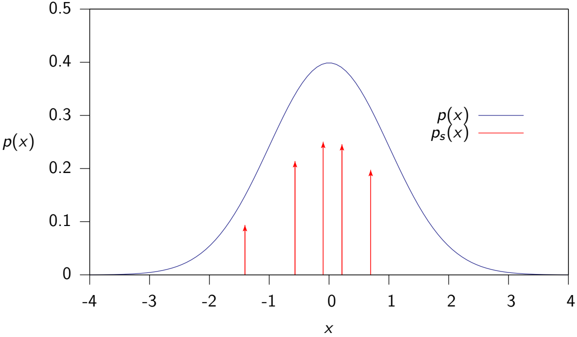


Estimation Possibilities:

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

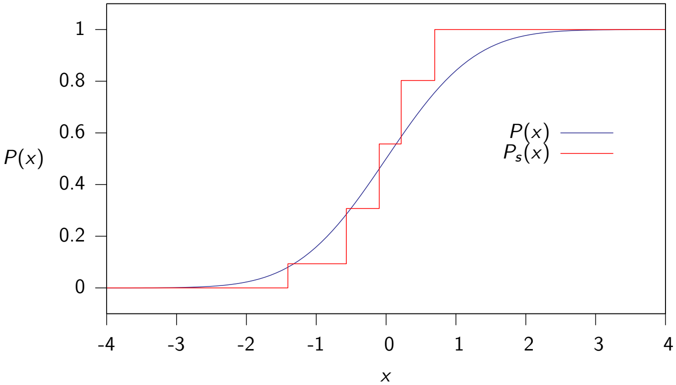
Particle filtering — sampled densities

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



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*.

Importance sampling

In state estimation, p of interest is easy to *evaluate* but difficult to *sample*. We choose an *importance function*, q, instead.

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

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

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

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

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).

Importance sampled particle filter (Arulampalam et al., 2002)

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

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 $q(x(k+1)|x_i(k), y(k+1))$

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 $q(x(k+1)|x_i(k), y(k+1))$

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

Importance sampled particle filter (Arulampalam et al., 2002)

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

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

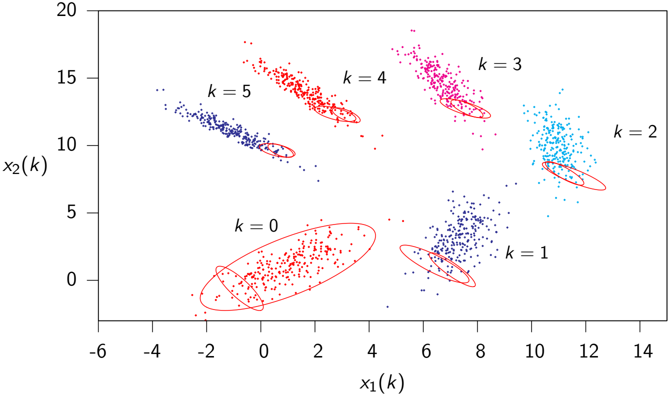
$$w_i(k+1) = w_i(k)\,\frac{p(y(k+1)|x_i(k+1))p(x_i(k+1)|x_i(k))}{q(x_i(k+1)|x_i(k), y(k+1))}$$

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

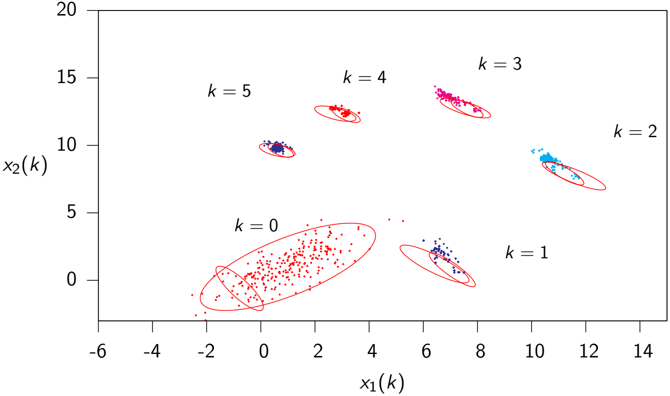
Optimal importance function



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



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

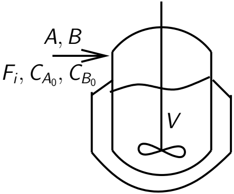
The MHE and particle filtering hybrid approach

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

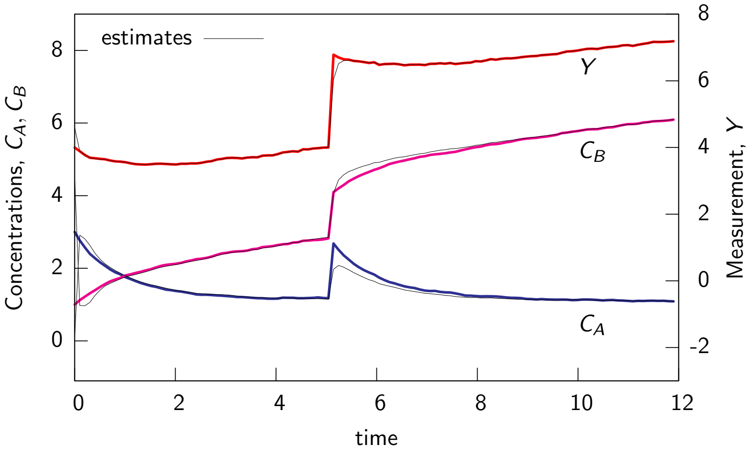


$$\frac{dC_A}{dt} = -2kC_A^2 + \frac{F_i}{V} C_{A_0} \qquad \Delta t = 0.1$$
$$\frac{dC_B}{dt} = kC_A^2 + \frac{F_i}{V} C_{B_0}$$

- Noise covariances $Q_w = \text{diag}\left(0.01^2, 0.01^2\right)$ 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 $t_k = 5$

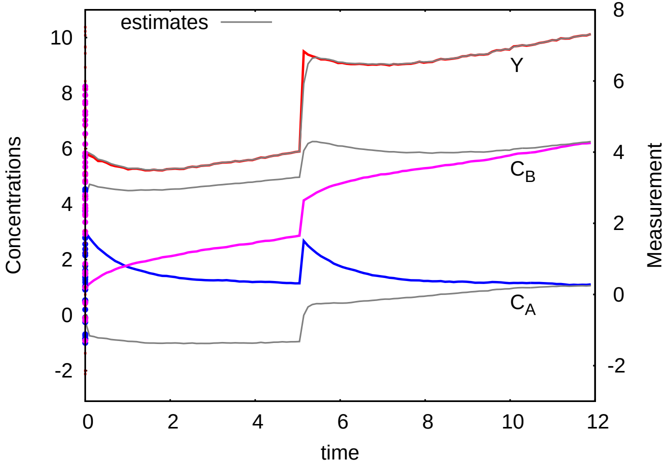
Using only MHE

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



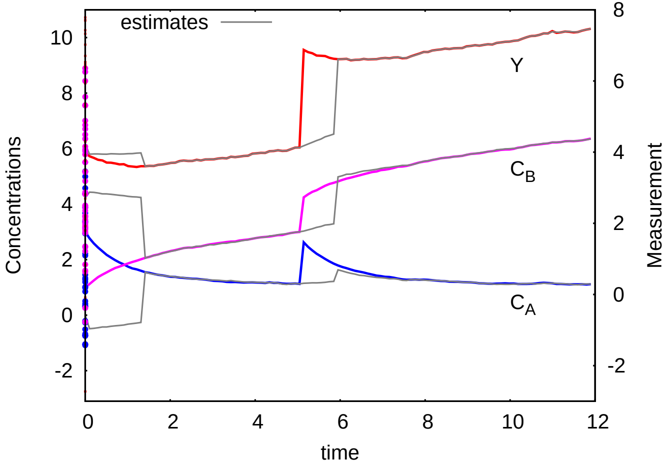
Using only particle filter

- 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 bad $\bar{x}_0$ and is unreliable



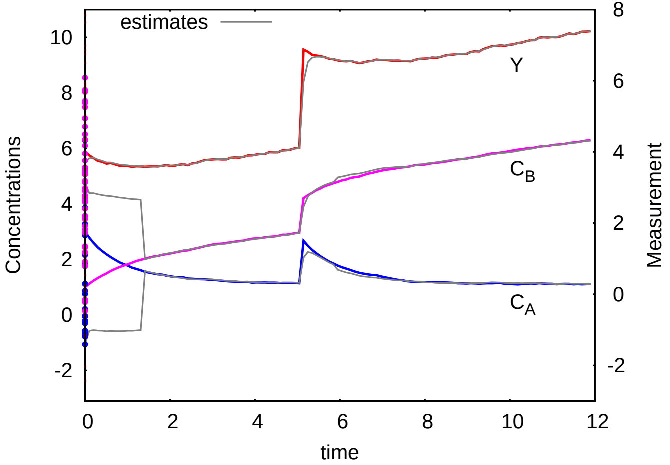
MHE/PF hybrid with a simple importance function

- Importance function for PF: $p(x_k|x_{k-1})$, 50 samples
- The PF samples recover from a bad $\bar{x}_0$ 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 bad $\bar{x}_0$, 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.

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.

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.

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.
- MHE and particle filtering are higher-quality solutions for nonlinear models.

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.
- MHE and particle filtering are higher-quality solutions for nonlinear models. MHE is robust to modeling errors but requires an online optimization.

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.
- 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.

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.
- 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.
- 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.

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.

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

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

Further Reading I

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.

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

S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proc. IEEE*, 92(3):401–422, March 2004.

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.

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