Using Moving Horizon Estimation to Overcome Extended Kalman Filtering Failure

Eric L. Haseltine and James B. Rawlings
Department of Chemical and Biological Engineering
University of Wisconsin, Madison, WI 53706

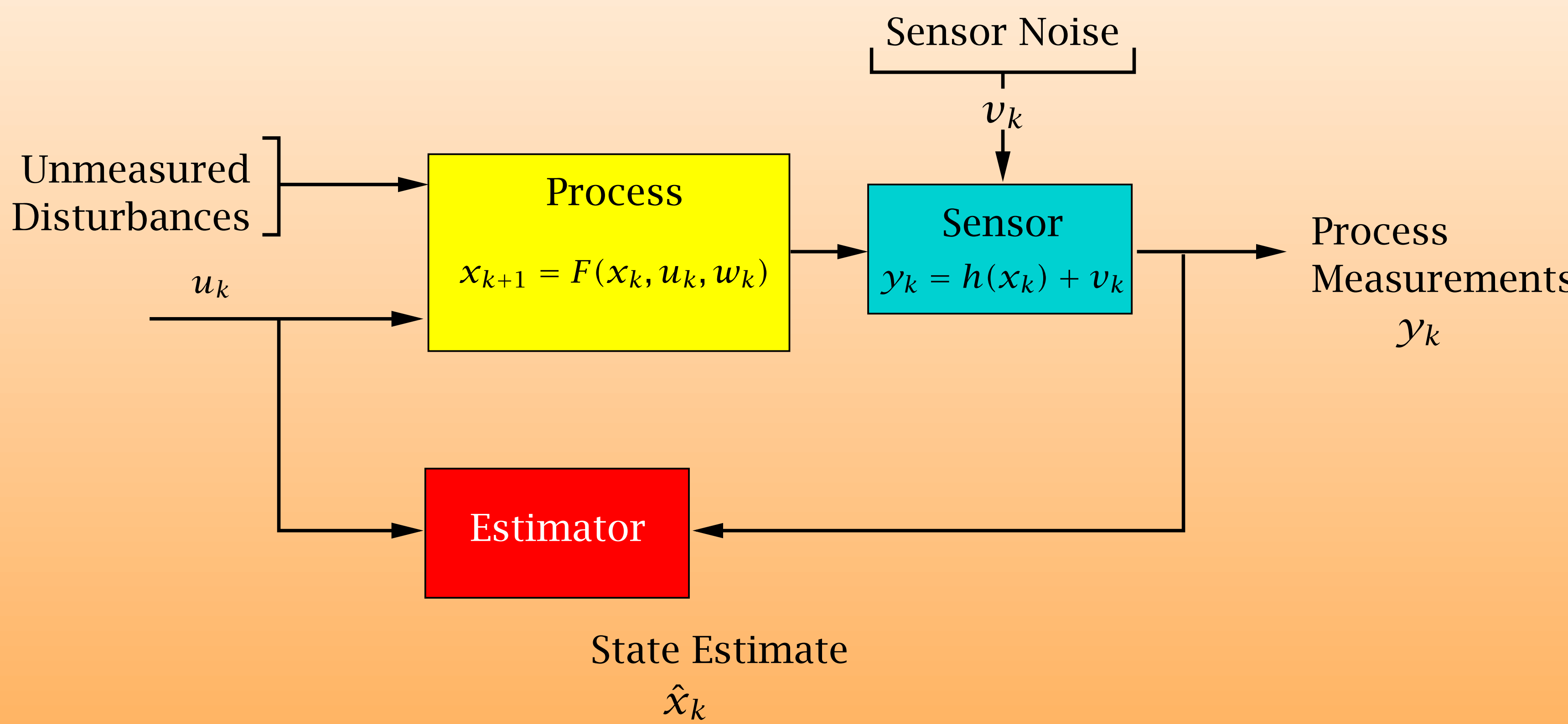
TWMCC

19 November 2003

Outline

- State estimation overview
  - problem formulation
  - extended Kalman filter
  - moving horizon estimation
- Effect of arrival cost
- Closed-loop control and plant-model mismatch
- Conclusions

State Estimation Overview



What estimate is desired?

- We will consider the model

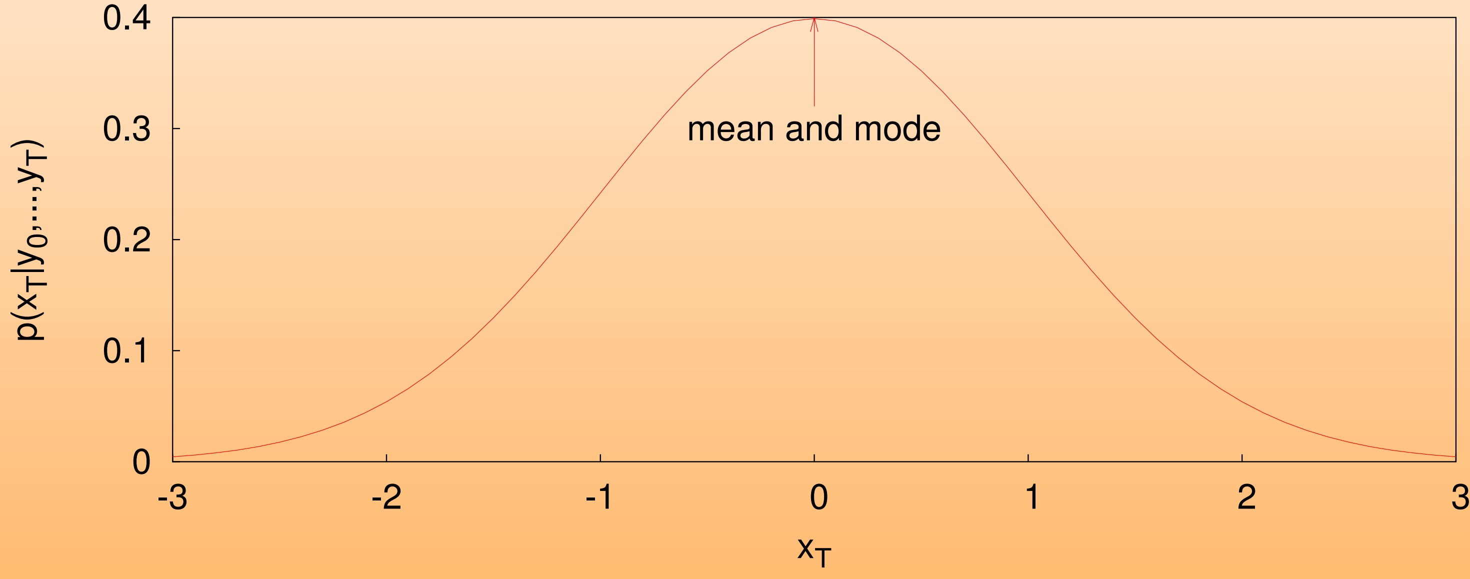
$$x_{k+1} = F(x_k, u_k) + Gw_k$$
$$y_k = h(x_k) + v_k$$

- If we knew the a posteriori distribution

$$p(x_T | y_0, \ldots, y_T)$$

what point estimate should we calculate?

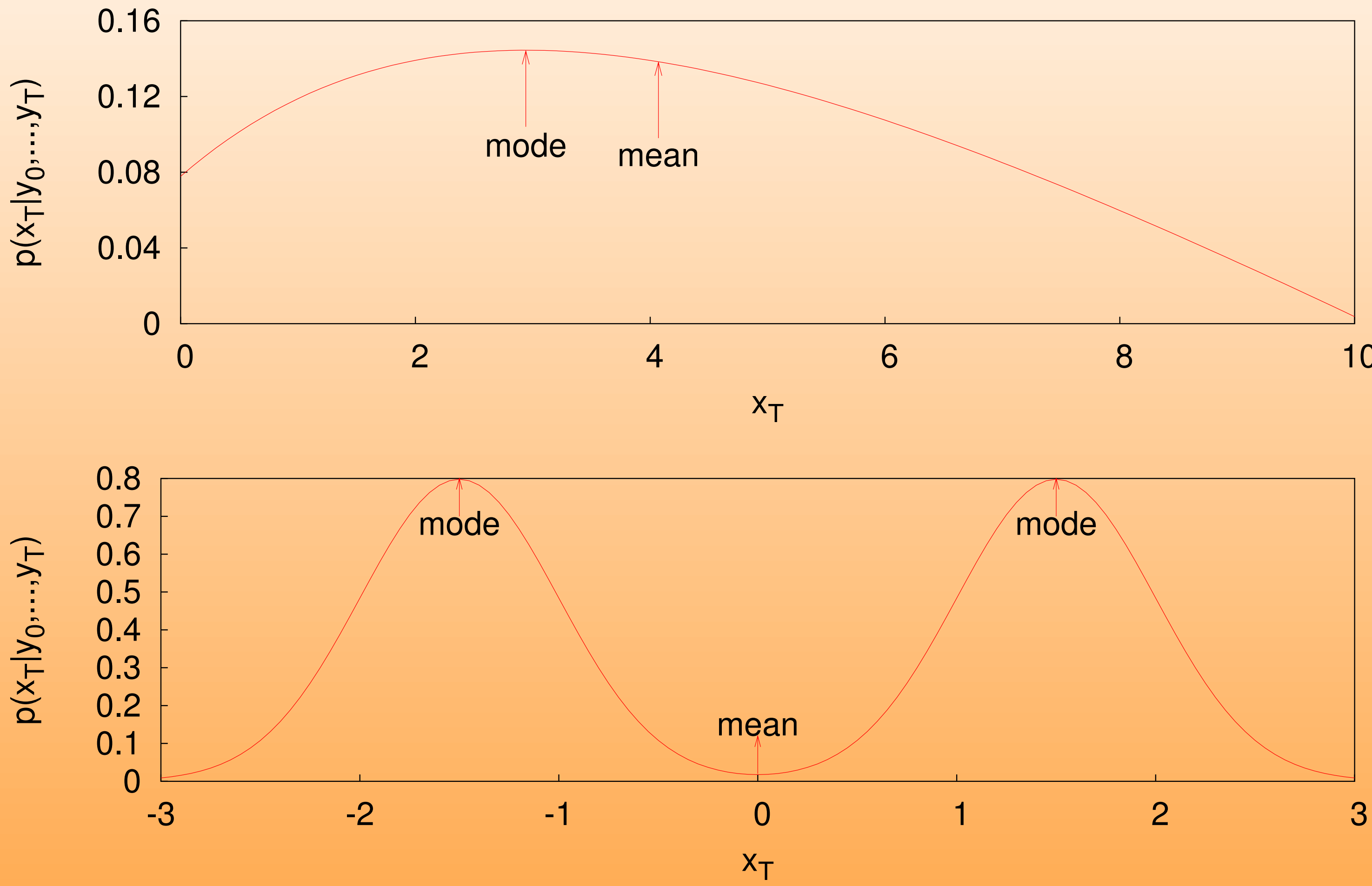
What estimate is desired?



For unconstrained linear estimation with Gaussian noise, the mean and mode of the probability distribution are the same.

Optimal estimator: Kalman filter (recursive).

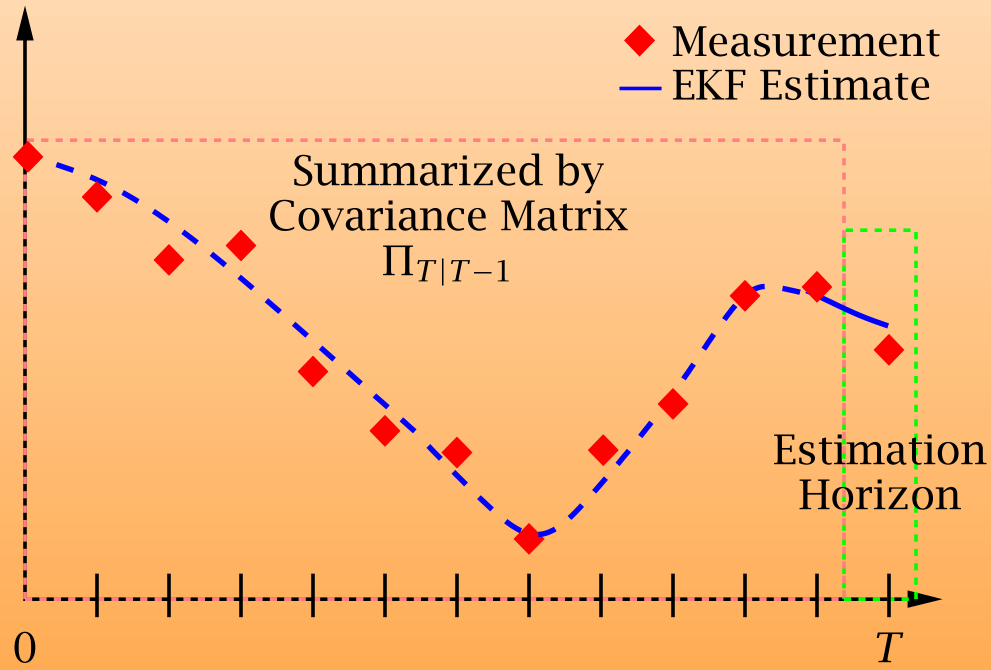
What estimate is desired?



The mean and mode of the probability distribution are generally different.

We would like to solve for the the maximum a posteriori estimate (MAP), i.e. the mode of this distribution.

Extended Kalman Filtering

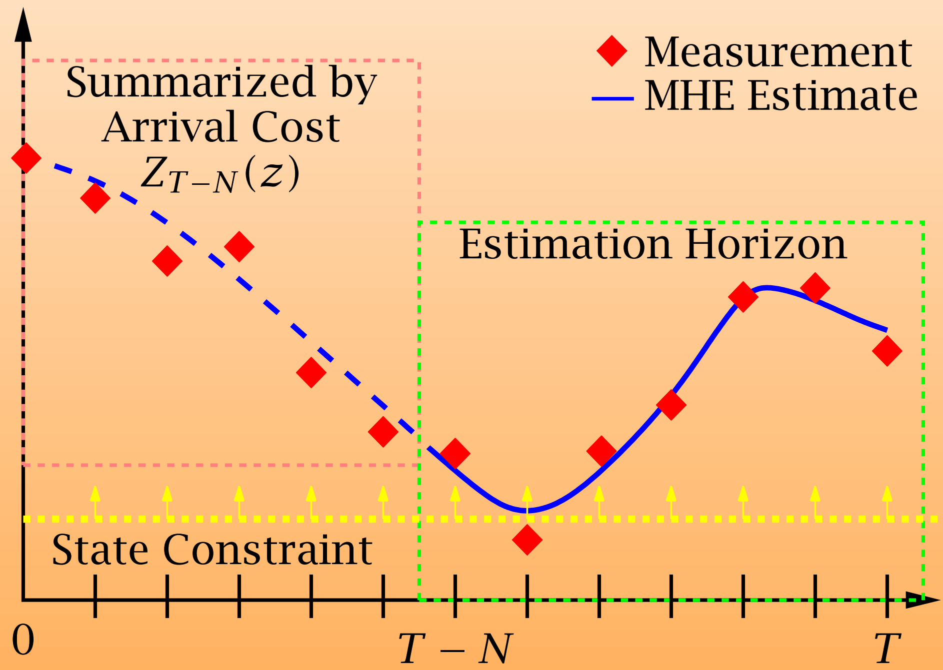


- Approximates

$$\hat{x}_{T|T} \approx \arg\max_{x_T} p(x_T | y_0, \ldots, y_T)$$

- Extension of the Kalman filter to nonlinear systems via linearization
- Summarizes past data with the covariance matrix
- Computationally trivial
- Most popular industrial method

Moving Horizon Estimation



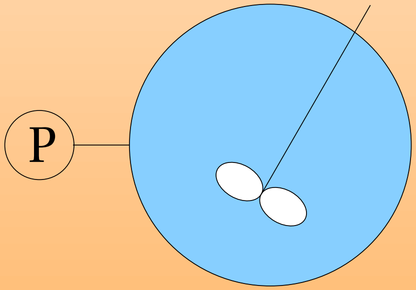
- Approximates

$$\{\hat{x}_{T-N|T}, \dots, \hat{x}_{T|T}\}$$

$$\approx \arg \max_{x_{T-N},\dots,x_T} p(x_{T-N}, \dots, x_T | y_0, \dots, y_T)$$

- Accurately employs the non-linear model
- Can incorporate constraints
- Requires on-line optimization
- Arrival cost?

Effect of Arrival Cost: An Illustrative Example



$$2\mathrm{A} \xrightarrow{k} \mathrm{B}$$

- Well-mixed, gas phase, batch reactor
- Estimate the partial pressures of A and B
- Model

$$\frac{dx}{dt} = \begin{bmatrix} -2 & 1 \end{bmatrix}^T k P_A^2$$

- Measure the total pressure

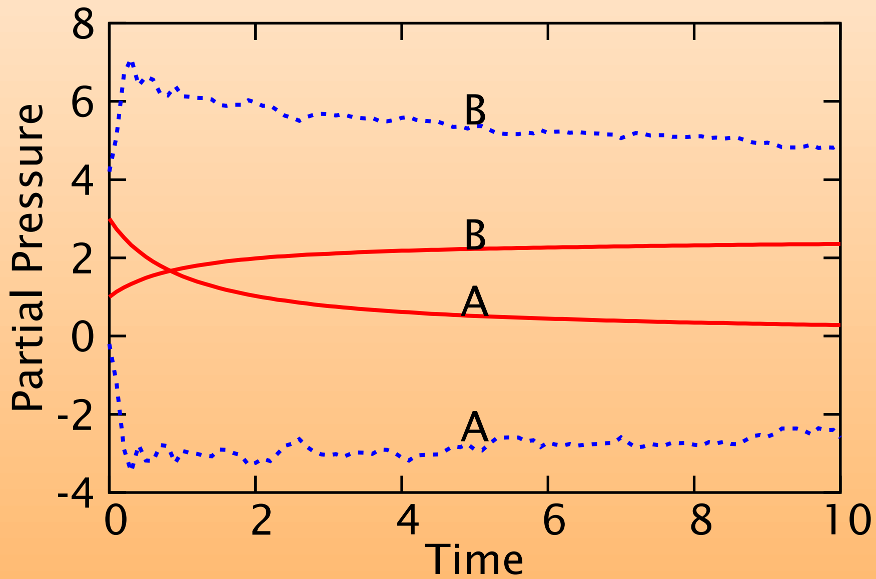
$$y = P_A + P_B$$

- Poor initial guess

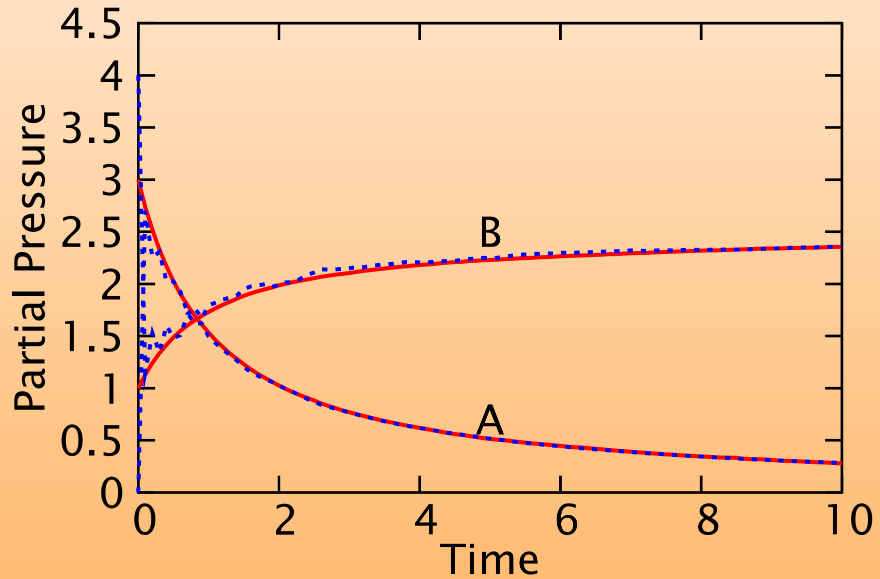
$$x_o = \begin{bmatrix} 3 & 1 \end{bmatrix}^T \text{ vs. } \bar{x}_o = \begin{bmatrix} 0.1 & 4.5 \end{bmatrix}^T$$

Estimator Comparisons

EKF



Constrained MHE



$P_A \geq 0$, $P_B \geq 0$

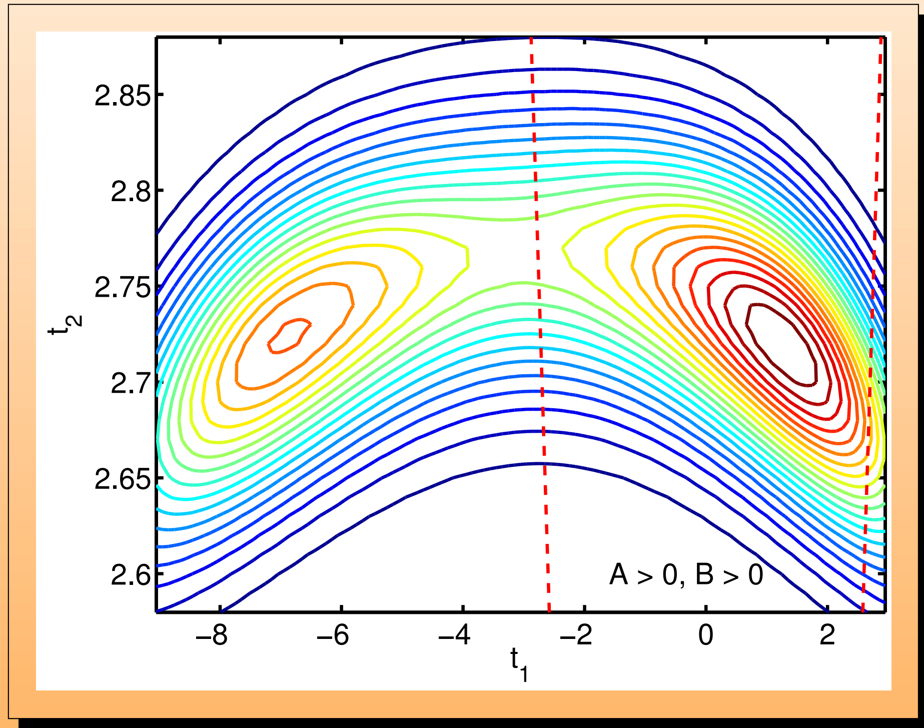
Horizon length of 2 minutes

Actual state (red)

Estimated state (blue)

Maximum a Posteriori Distribution

$$p(x_T | y_0, \ldots, y_T)$$

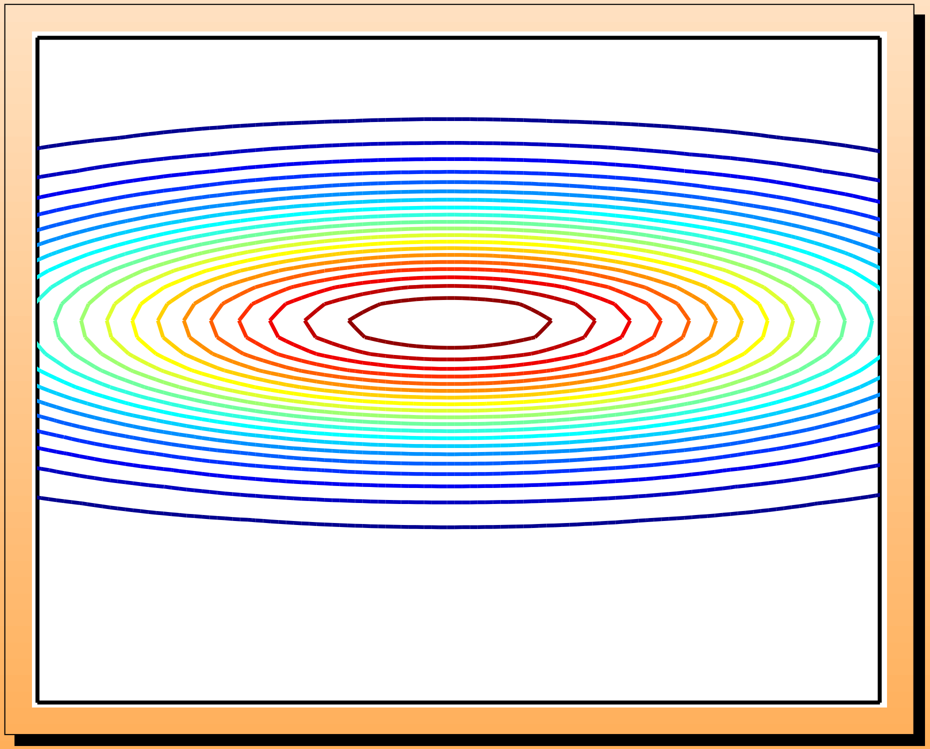


Maximum a posteriori distribution exhibits multiple optima.

Extended Kalman Filtering and Moving Horizon Estimation

Extended Kalman Filtering

$$\approx p(x_T | y_0, \ldots, y_T)$$



Moving Horizon Estimation

$$\approx \max_{x_{T-N},\ldots,x_{T-1}} p(x_{T-N}, \ldots, x_T | y_0, \ldots, y_T)$$

Depends on the arrival cost

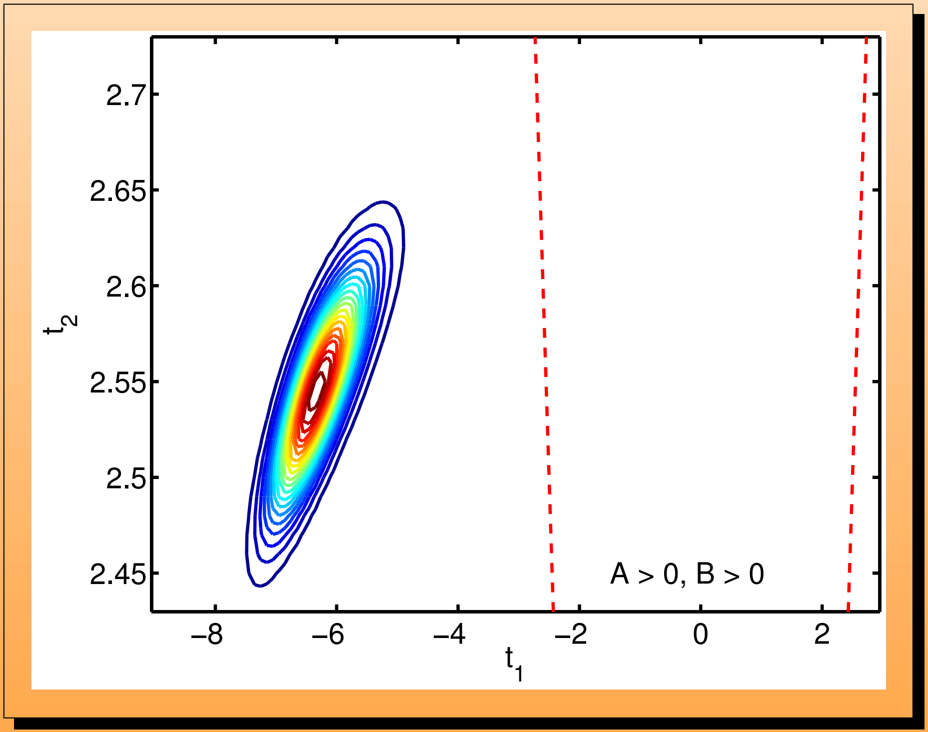
- approximate the process as a time-varying linear system?
- uniform prior?

Arrival Cost Strategies

Arrival cost approximations: smoothing update (assumes process is a time-varying linear system) or uniform prior

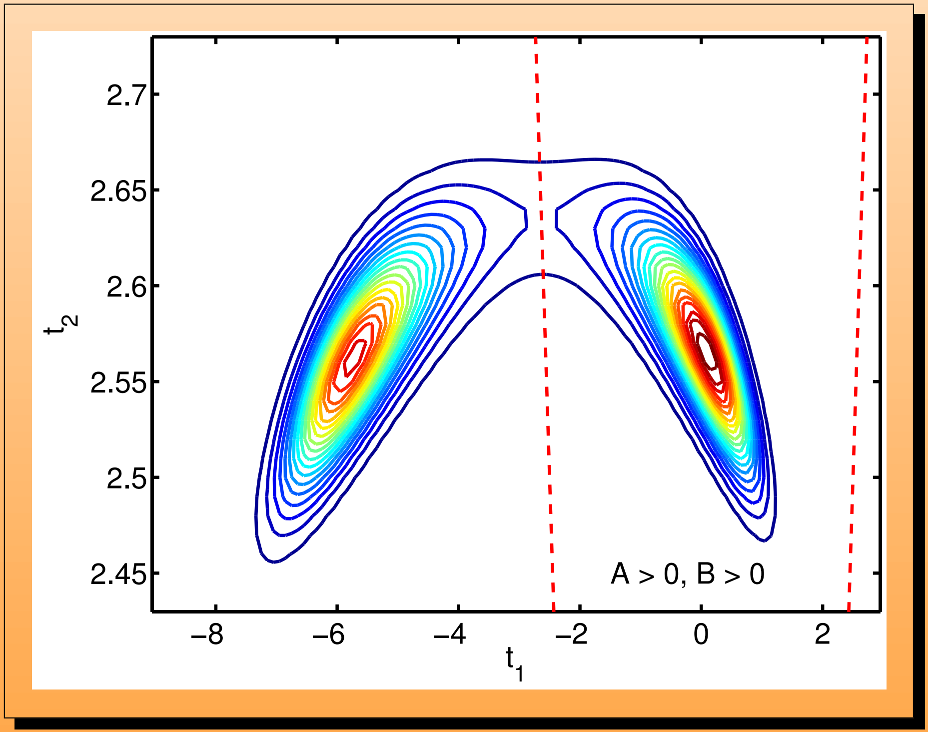
Smoothing Update

$$\max_{x_1,\ldots,x_3} p(x_1,\ldots,x_4|y_0,\ldots,y_4)$$



Uniform Prior

$$\max_{x_1,\ldots,x_3} p(x_1,\ldots,x_4|y_0,\ldots,y_4)$$



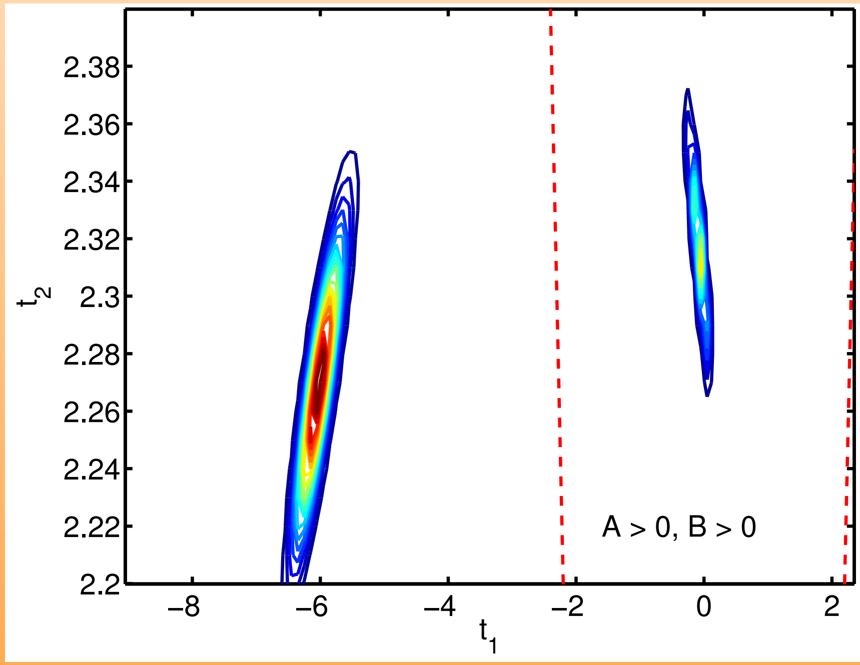
The prior (arrival cost) dominates and distorts the information contained in the data. Better to use a uniform prior if global optimization possible.

Arrival Cost Strategies: Effect of Horizon Length

Arrival cost approximations: smoothing update (assumes process is a time-varying linear system) or uniform prior

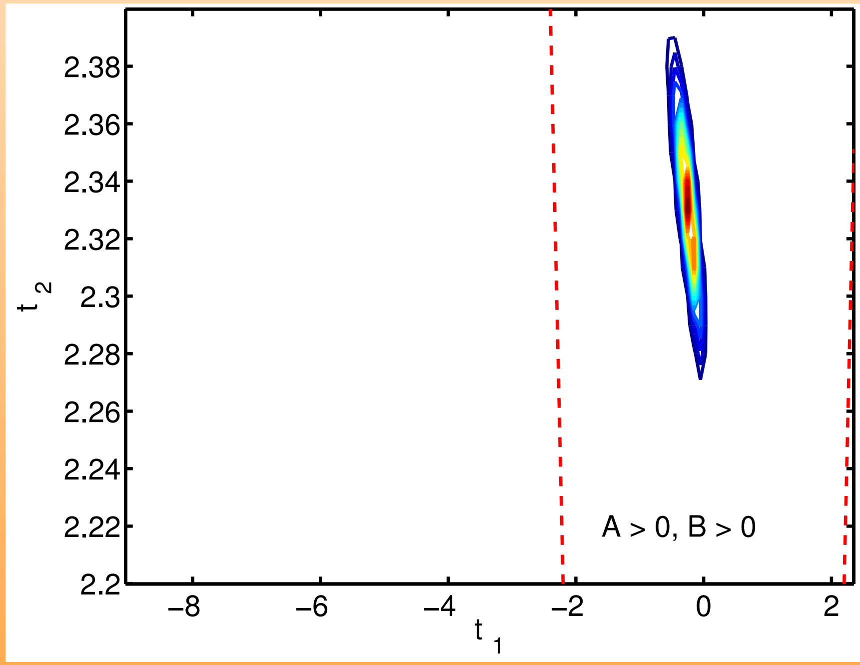
Smoothing Update

$$\max_{x_1,\ldots,x_9} p(x_1,\ldots,x_{10}|y_0,\ldots,y_{10})$$



Uniform Prior

$$\max_{x_1,\ldots,x_9} p(x_1,\ldots,x_{10}|y_0,\ldots,y_{10})$$

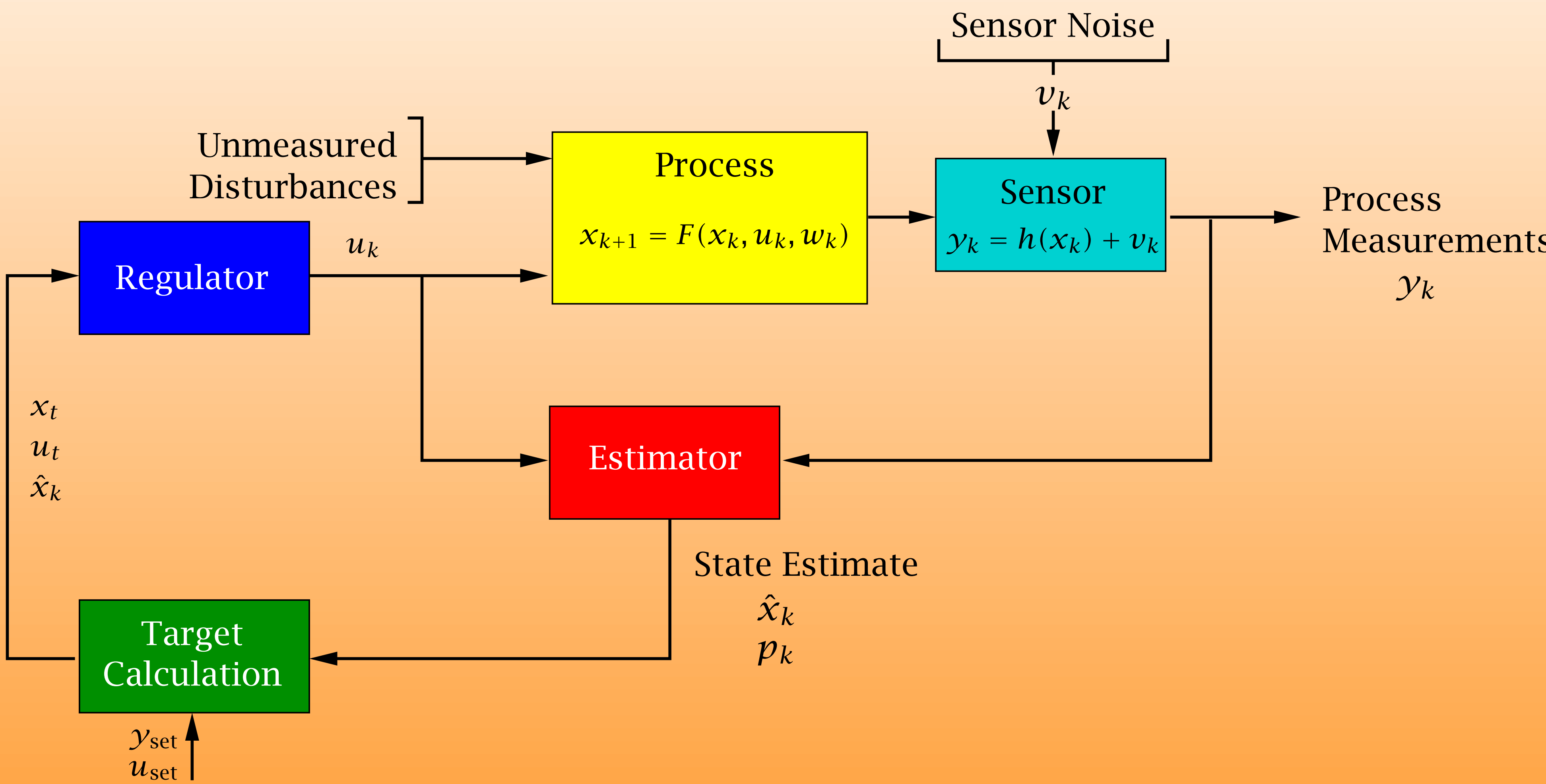


Arrival Cost Conclusions

- The EKF can predict only one optimum.
  Estimation behavior dictated by initial region of attraction.

- Behavior of MHE depends upon the arrival cost.
  - Approximating the past behavior as a time-varying, linear system introduces significant bias
  - Longer horizon can overcome effects of poor arrival cost
  - Best (current) option: uniform prior with global optimization

Closed-Loop Control



We will use the NMPC toolbox for the estimator, regulator, and target calculation (local optimization) [3].

Disturbance models for nonlinear models

- For offset free control, must account for discrepancies between the plant and the model.

- Augment the state with a disturbance model:

$$
\begin{aligned}
x_{k+1} &= F(x_k, u_k + X_u d_k) + G w_k \\
y_k &= h(x_k) + X_y d_k + v_k \\
d_{k+1} &= d_k + \xi_k \\
\xi_k &\sim \mathcal{N}(0, Q_d)
\end{aligned}
$$

Implies that d_k is stochastic!

Plant-model mismatch: exothermic CSTR example

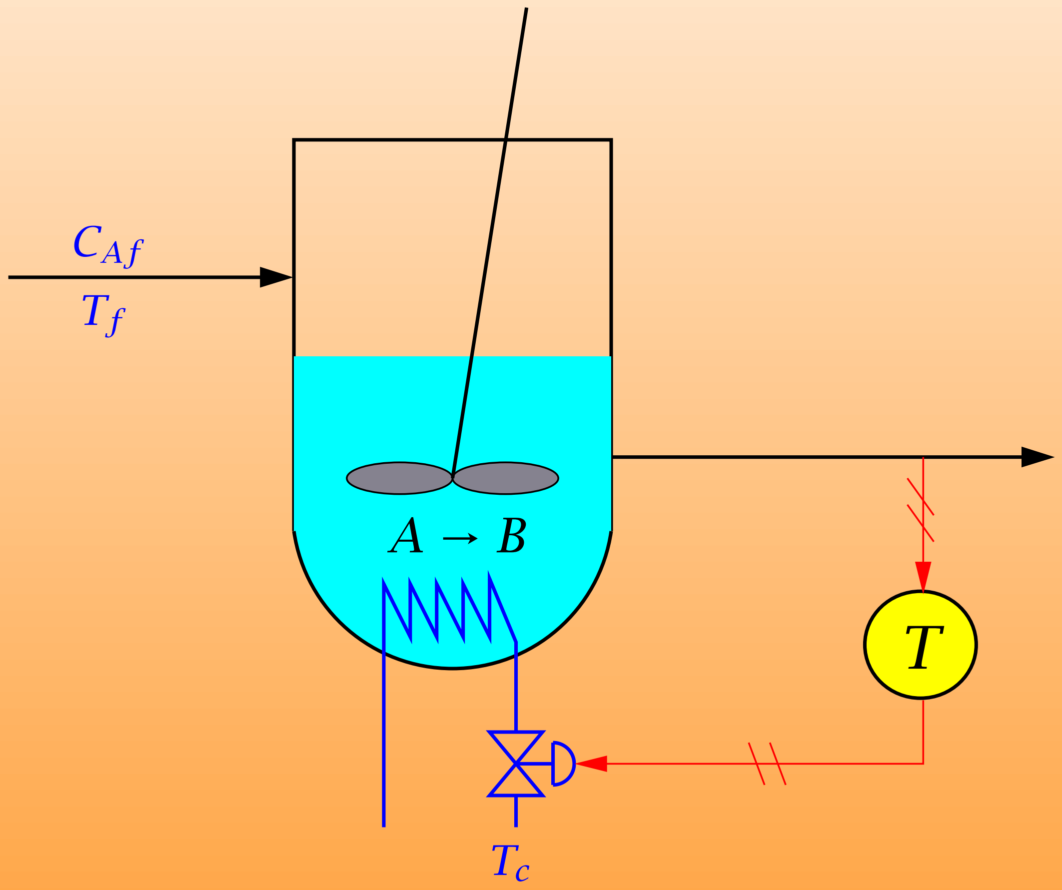
$$x = \begin{bmatrix} c_A \\ T \end{bmatrix}$$

$$u = T_c$$

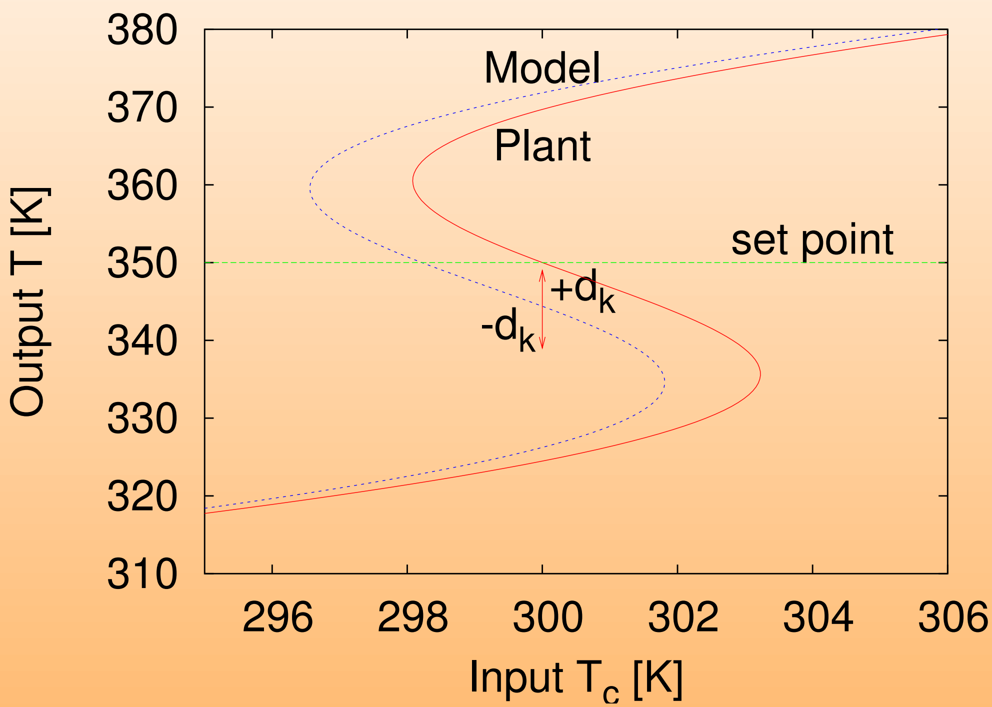
$$|\Delta u_k| \leq 15 \text{ K}$$

$$y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} x_k + d_k$$

Small mismatch in activation energy between the plant and the model.



Output disturbance model generates multiple steady states!



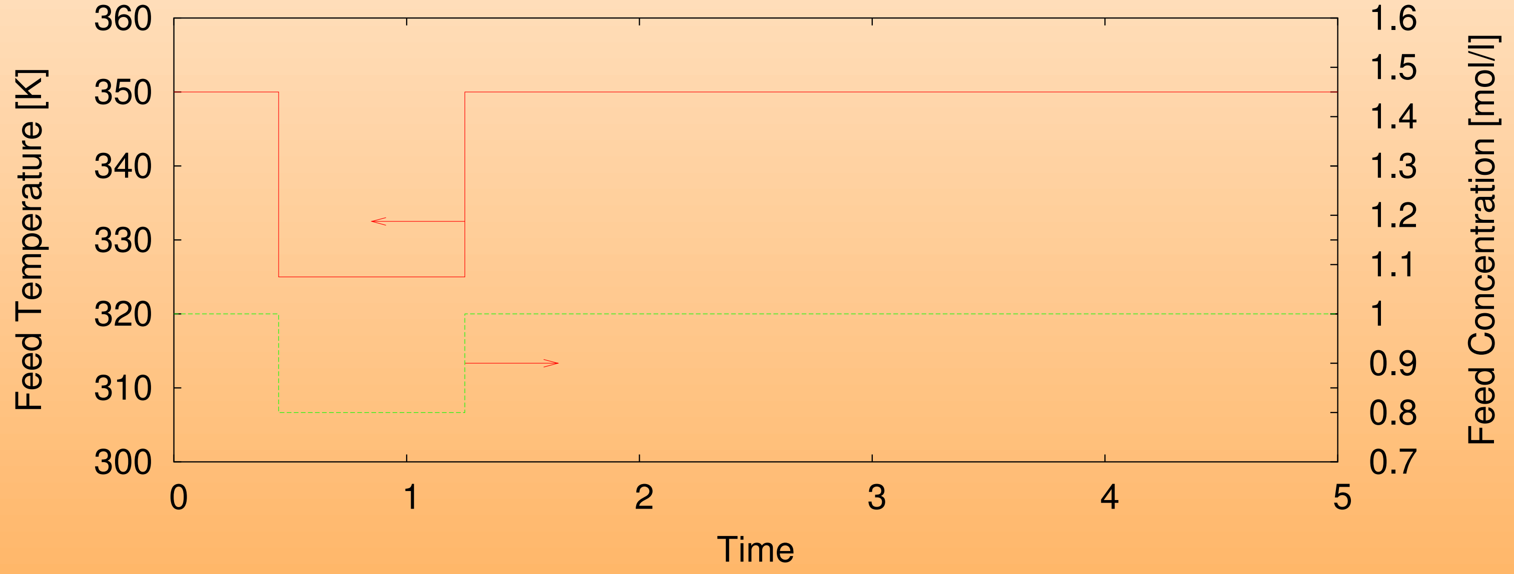
Multiple optima arise in the estimator. Do these optima affect control performance?

	c_A (mol/l)	Output T (K)	Disturbance d (K)
1	0.851	326.2	23.8
2	0.583	344.4	5.6
3	0.177	371.8	−21.8

Model Steady States for a Plant with $T_c = 300$ K, $T = 350$ K

Example: Disturbance rejection

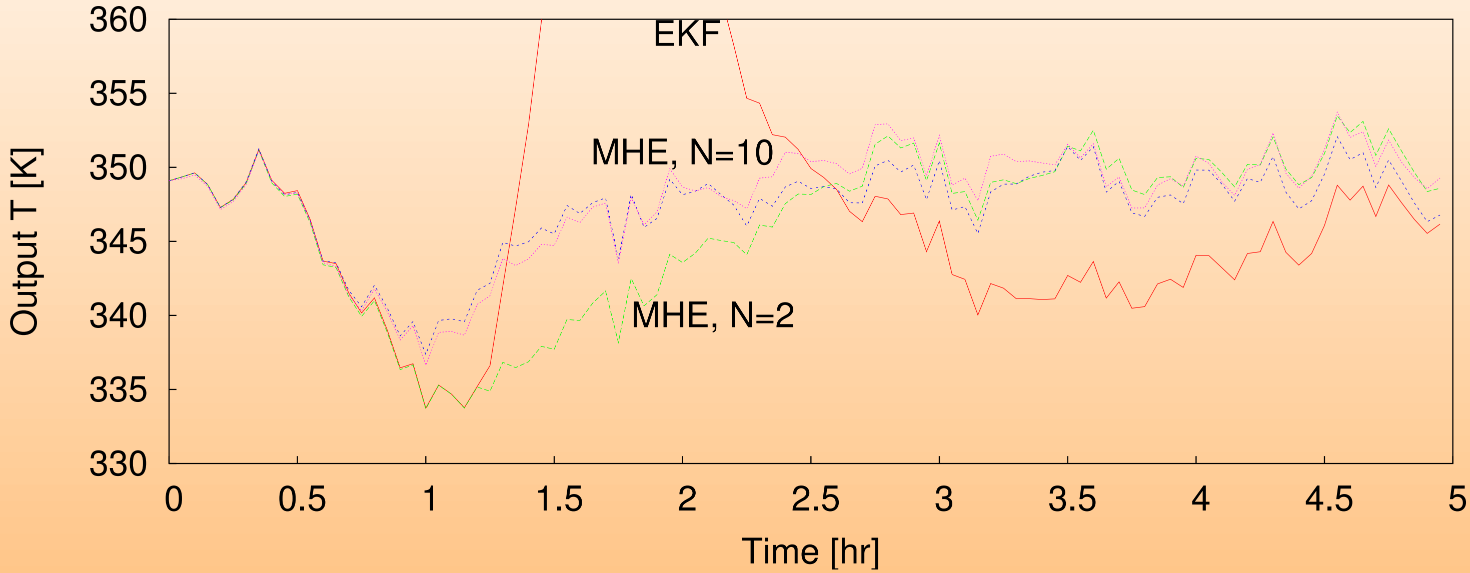
- Consider a disturbance in the feed:

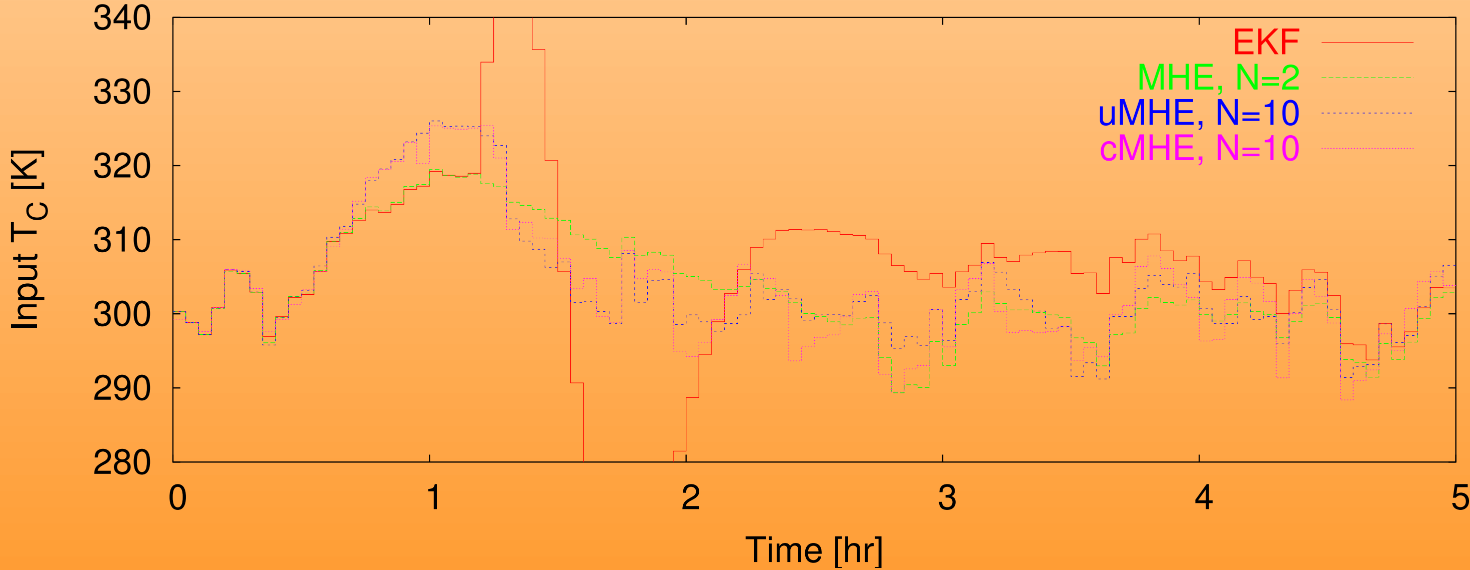


- Estimators:

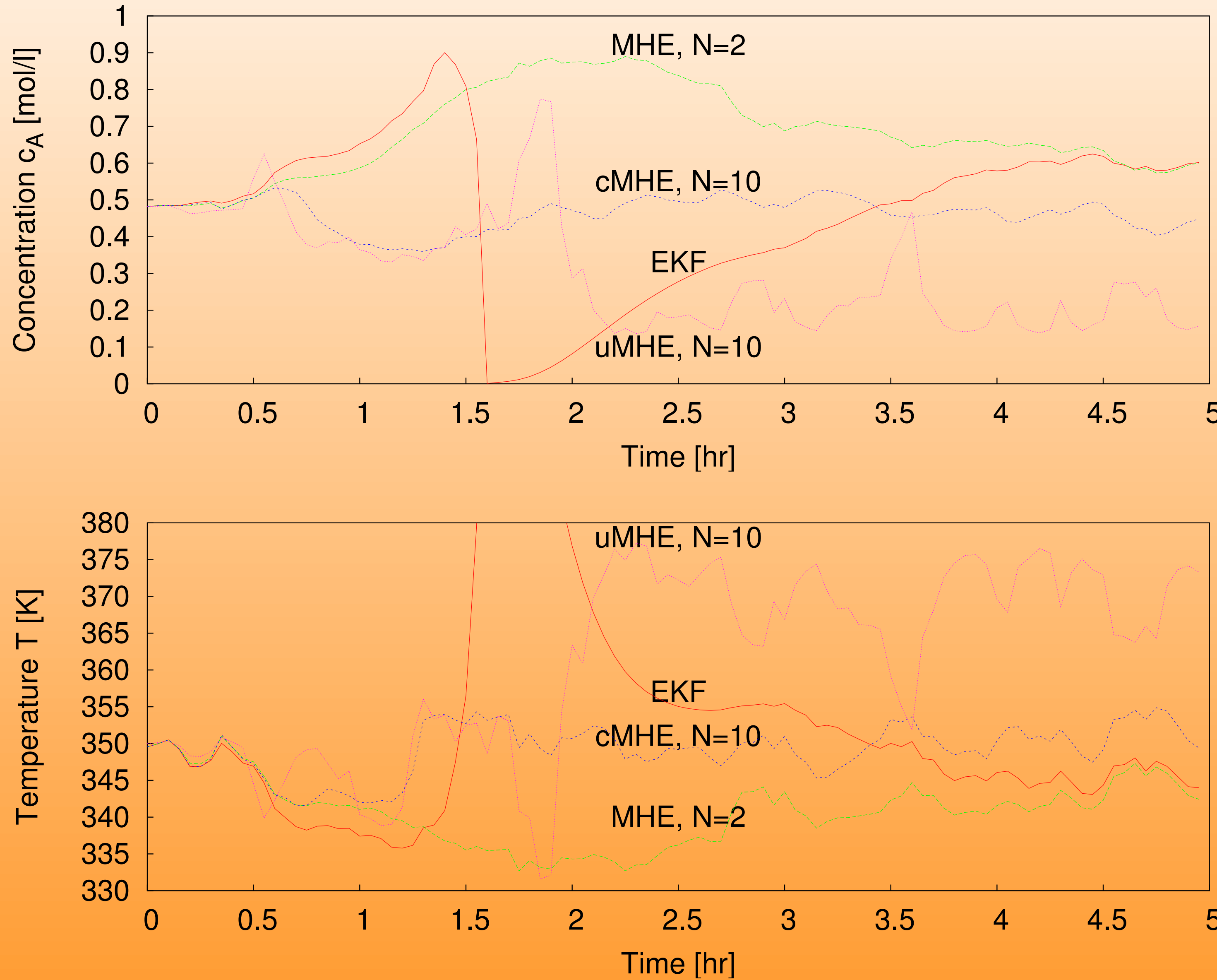
1. EKF
2. MHE with $N = 2$, smoothing update
3. MHE with $N = 10$, no initial penalty
4. MHE with $N = 10$, constant initial penalty

Exothermic CSTR Results

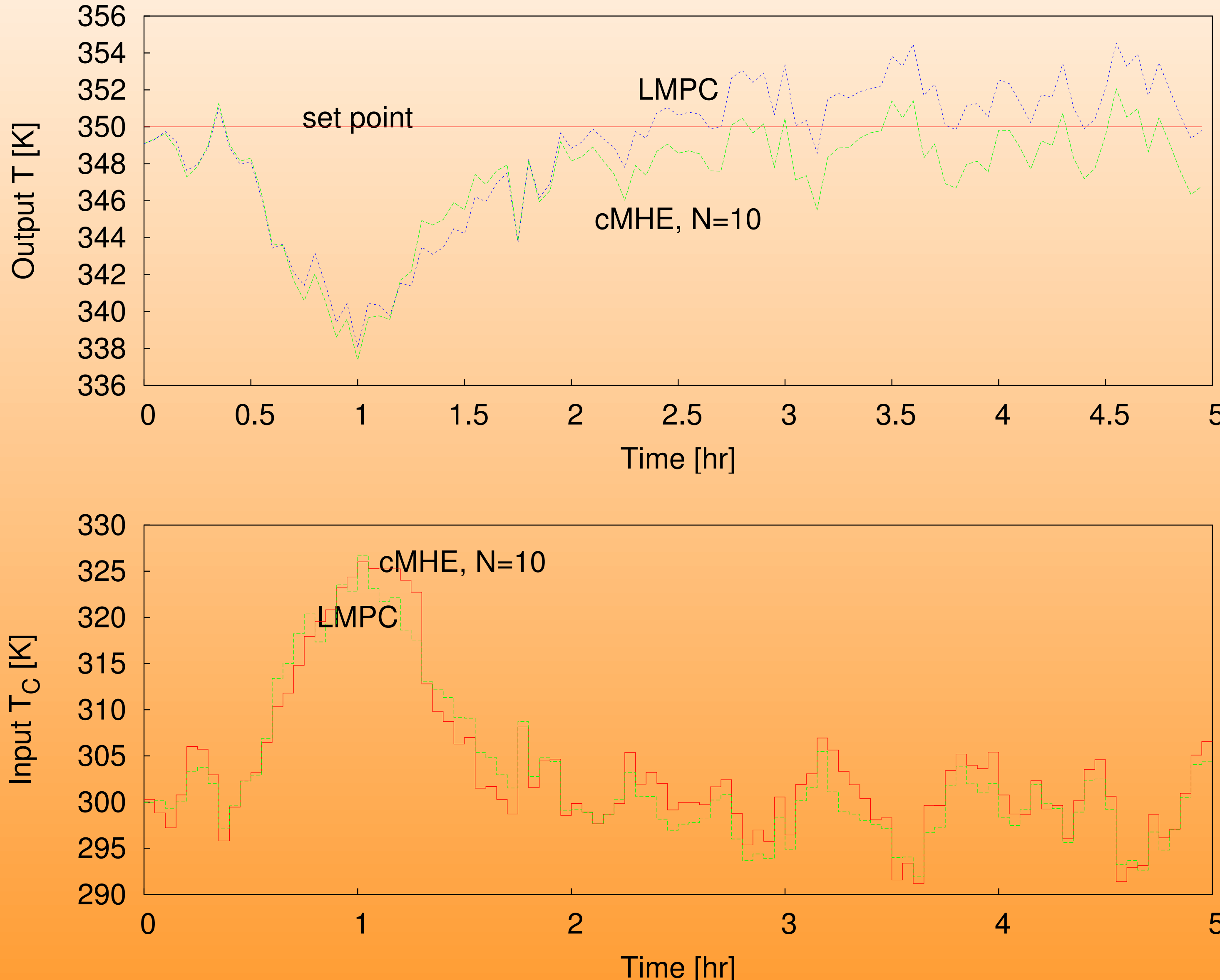




Exothermic CSTR Results



Comparison to Linear MPC Results



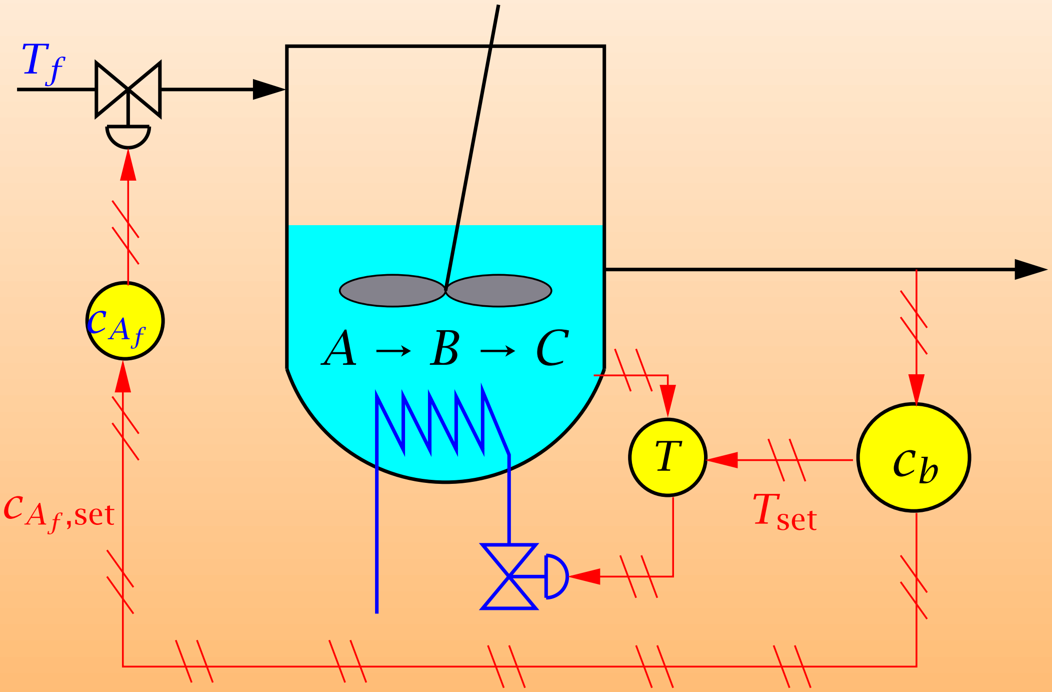
Plant-model mismatch: maximum yield example

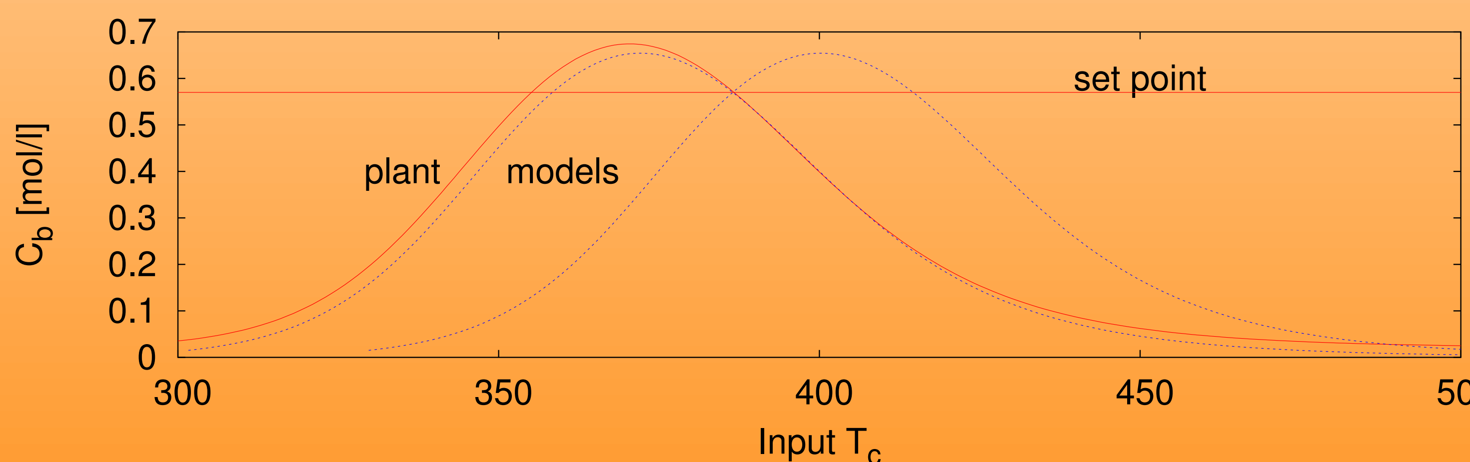
$$x = \begin{bmatrix} c_A & c_B \end{bmatrix}^T$$

$$u = \begin{bmatrix} T_c & c_{Af} \end{bmatrix}^T$$

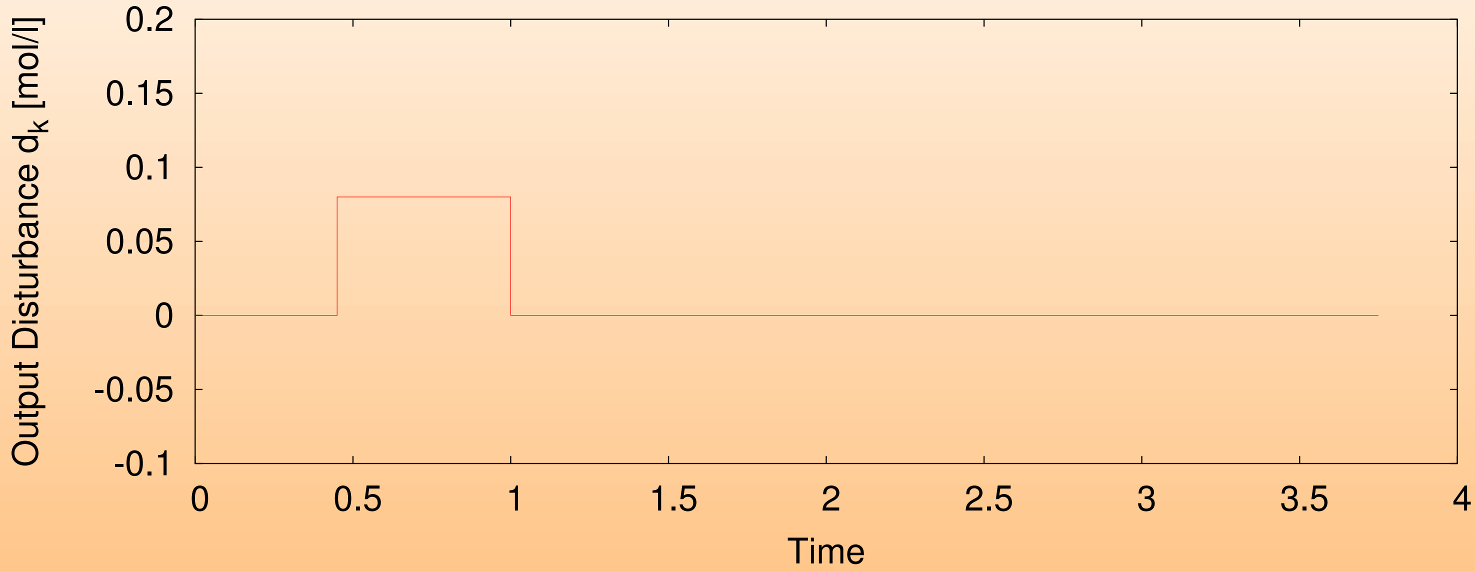
$$u_k = \begin{bmatrix} T_c \\ c_{Af} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} d_k$$

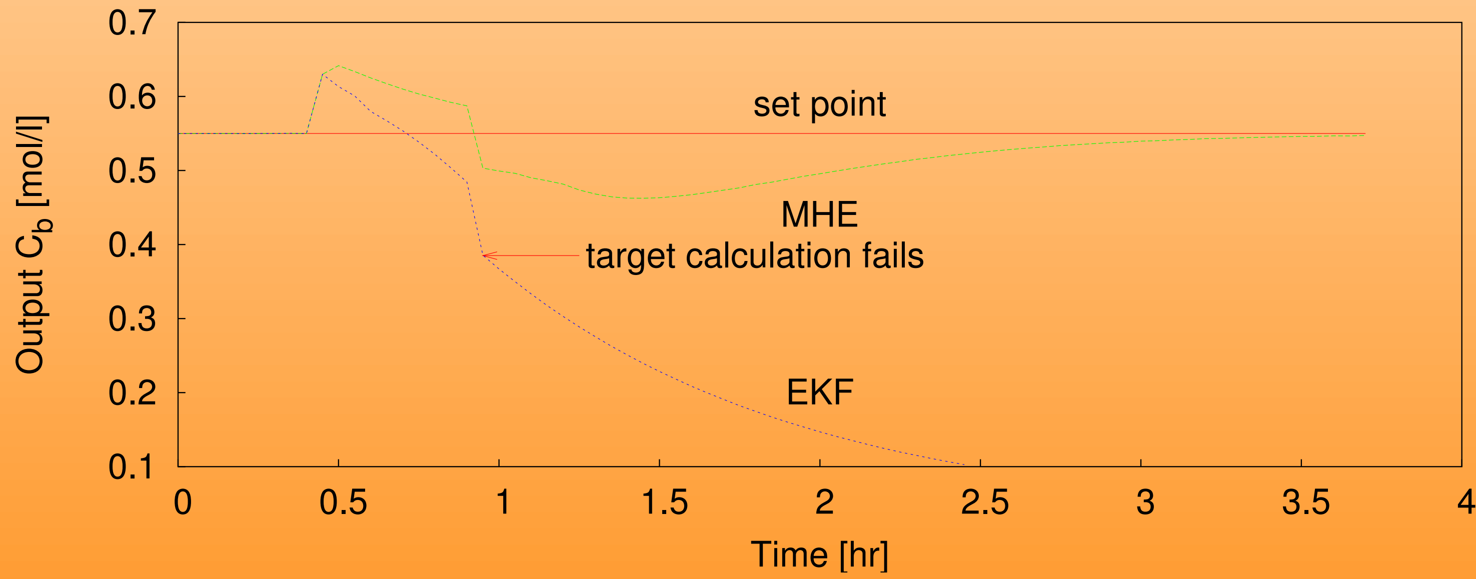
$$y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} x_k$$



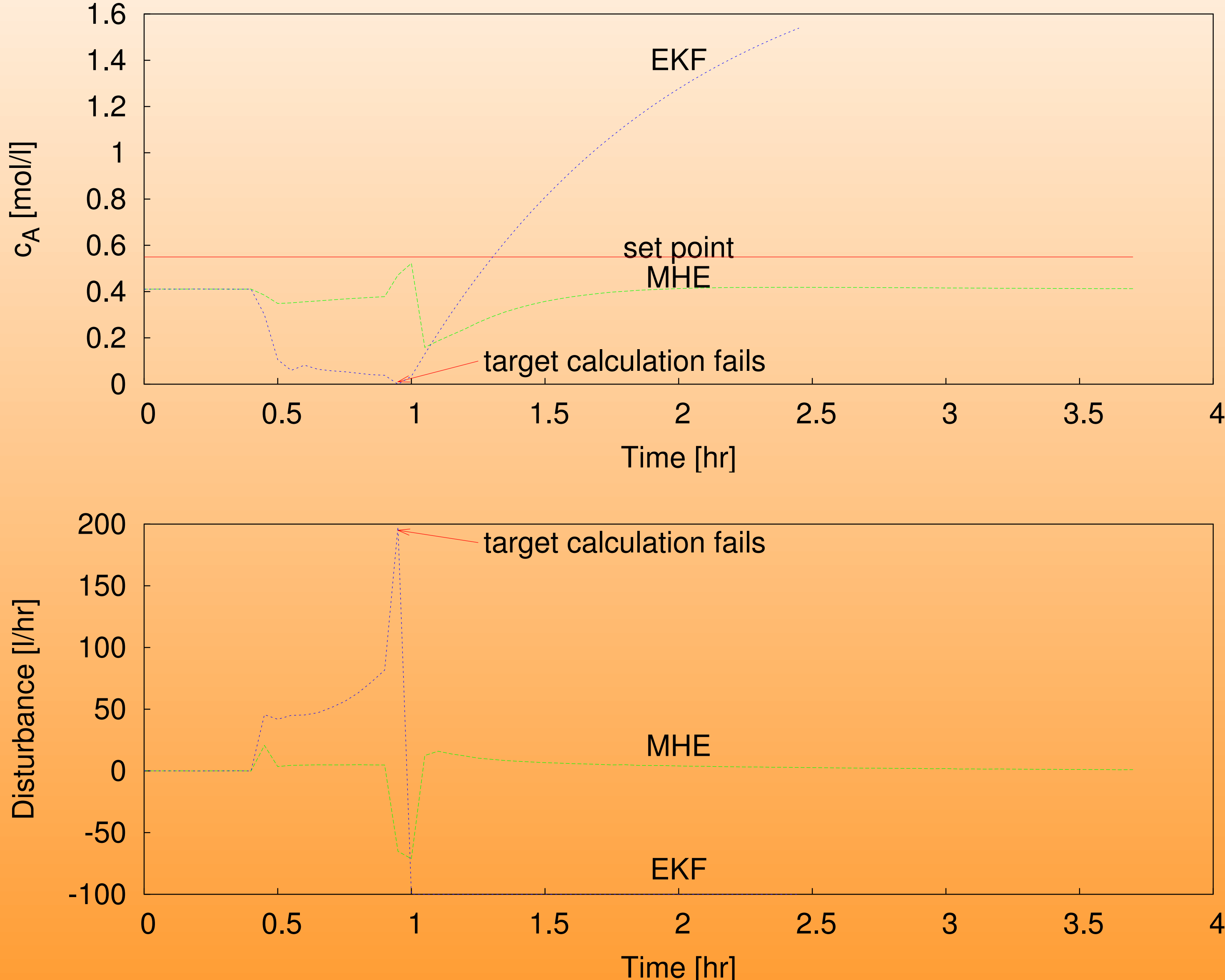


Maximum Yield CSTR Results





Maximum Yield CSTR Results



Conclusions

- Integrated disturbance models can induce multiple optima with nonlinear models.
- EKF: poor state tracking results in poor control.
- MHE
  1. better control than the EKF
  2. increased horizon length results in improved performance
- No significant improvement in nonlinear over linear control for disturbance rejection.
- If you want better disturbance rejection, you need a better disturbance model.

Acknowledgements

- Organizers of the Gordon Research Conference for Statistics in Chemistry & Chemical Engineering
- Wen-shiang Chen and Prof. Bhavik Bakshi (Ohio State University)

Questions?

Improving MHE

- Primary concern: longer horizon length = greater computational expense
- Basis of MHE:

$$\max_{x_{T-N+1},\ldots,x_T} p(x_{T-N+1},\ldots,x_T|y_0,\ldots,y_T)$$

$$= \max_{x_{T-N+1}\ldots,x_T} \underbrace{p(x_{T-N+1}|y_0,\ldots,y_T)}_{\text{Arrival Cost}} \underbrace{\left(\prod_{k=T-N+1}^{T-1} p(x_{k+1}|x_k)\right)\left(\prod_{k=T-N+1}^{T} p(y_k|x_k)\right)}_{\text{Estimation Horizon}}$$

- If we had a better estimate for $p(x_{T-N+1}|y_0,\ldots,y_T)$, we could shorten the estimation horizon
- Can we use Monte Carlo filters to estimate this density?

State Estimation via Monte Carlo Filters

- Basic idea: reconstruct state estimates from simulations of the stochastic process

$$\int h(x)P(x)dx = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} h(x^i)$$

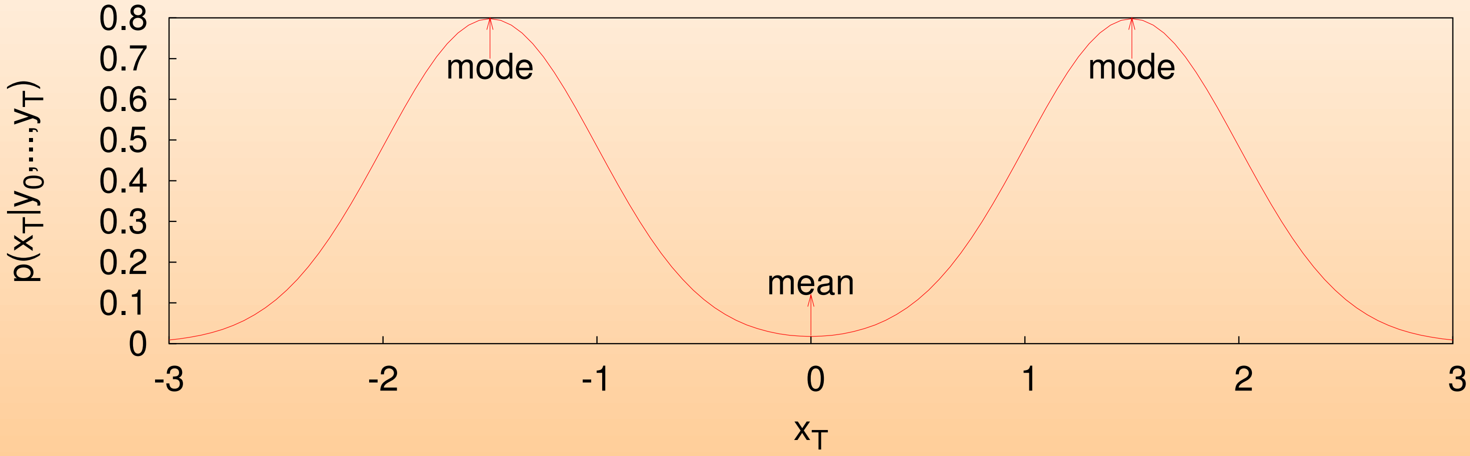
- Most MC filters propose estimation of the mean

$$E[x] = \int xP(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} x^i$$

- Permits use of any combination of model, random noise

- We will consider rejection sampling [1]

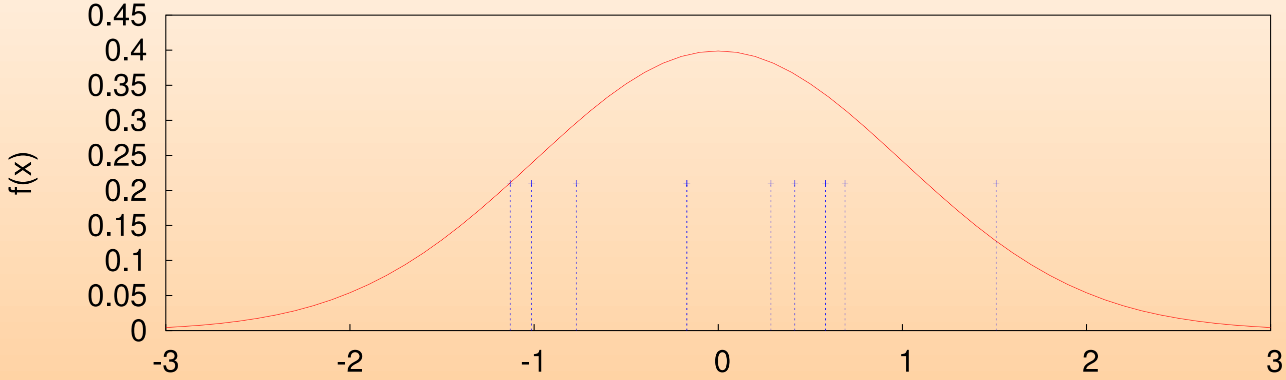
What estimate is desired?



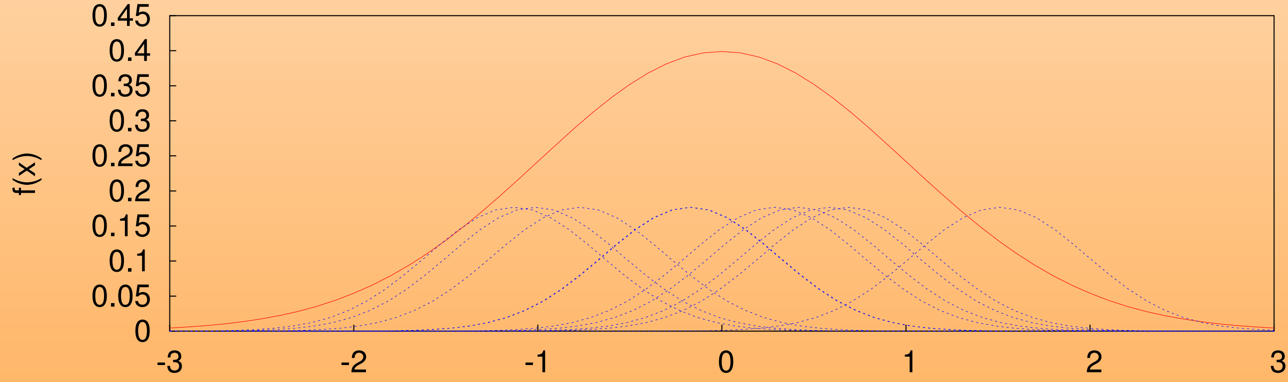
- MHE estimates the mode
- Monte Carlo filters must estimate the entire probability density to calculate the mode

1. density estimation
2. optimization

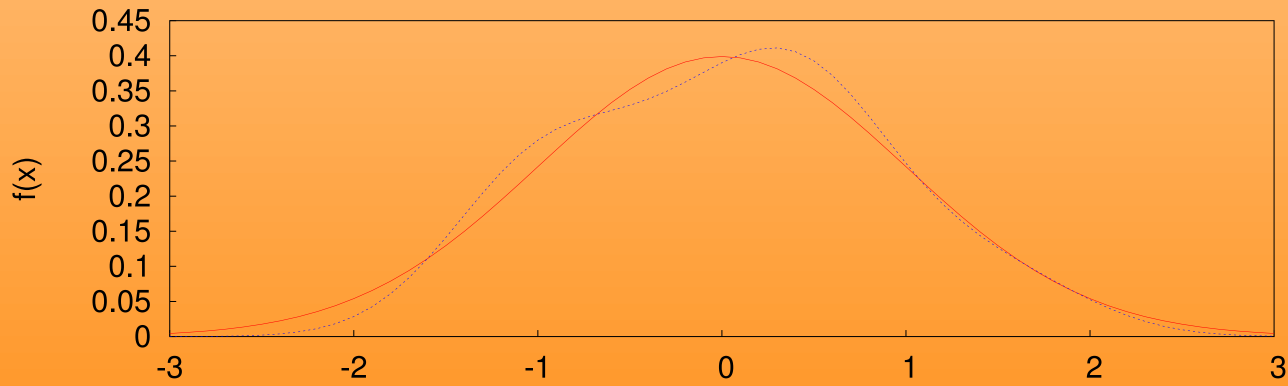
Example: density estimation of a normal distribution



Draw samples from the underlying distribution.

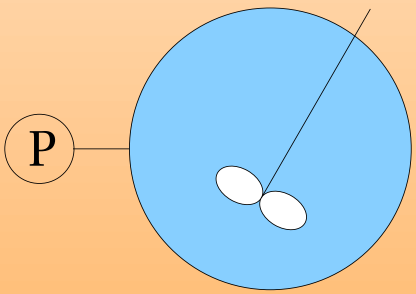


Apply a symmetric “kernel” density at each sample.



Sum the kernel densities to approximate the underlying distribution.

A Simple Example



$$2A \xrightarrow{k} B$$

- Well-mixed, gas phase, batch reactor
- Estimate the partial pressures of A and B
- Model

$$\frac{dx}{dt} = \begin{bmatrix} -2 & 1 \end{bmatrix}^T k P_A^2$$

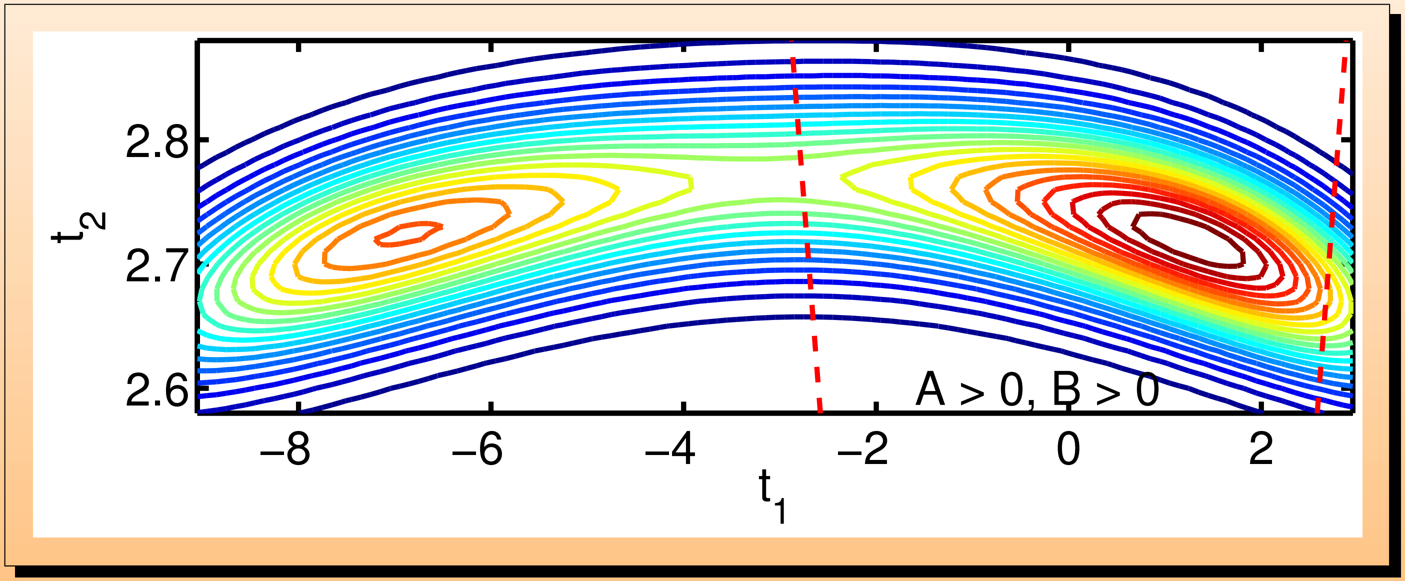
- Measure the total pressure

$$y = P_A + P_B$$

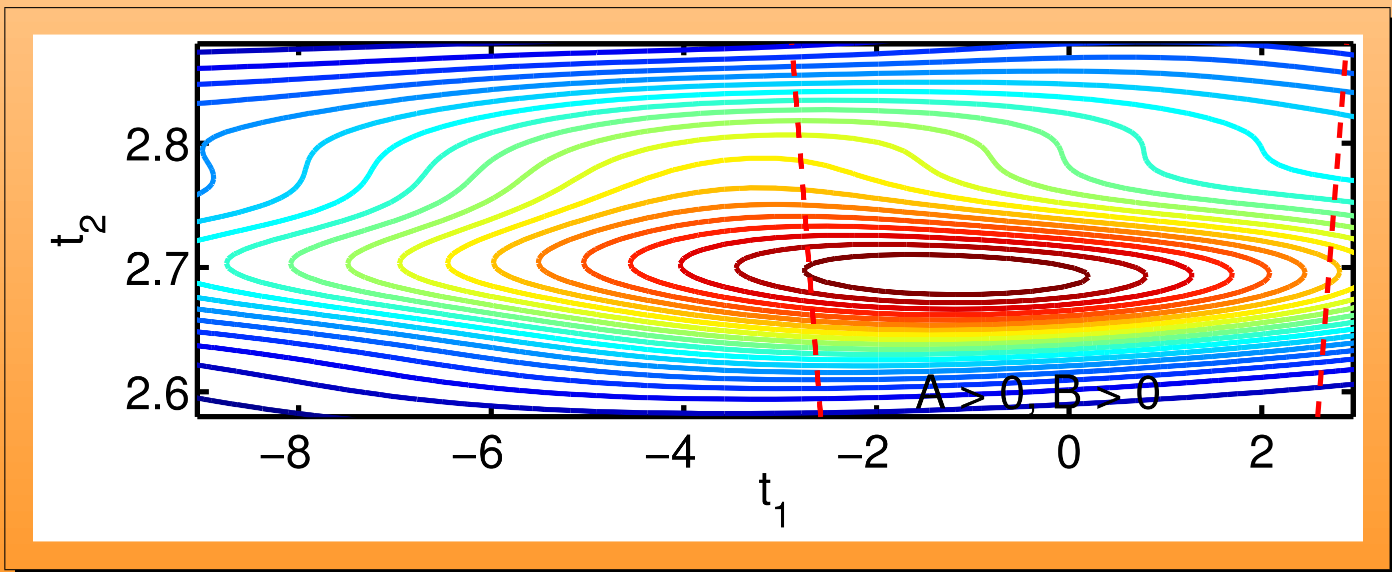
- Poor initial guess

$$x_o = \begin{bmatrix} 3 & 1 \end{bmatrix}^T \text{ vs. } \bar{x}_o = \begin{bmatrix} 0.1 & 4.5 \end{bmatrix}^T$$

A Posteriori Comparison: Actual vs. Monte Carlo $p(x_1|y_0, y_1)$



Actual



Monte Carlo Reconstruction (100 Accepted Samples)

Comments on Monte Carlo Estimation

- Not very accurate estimation of the mode.

  Sources of error:

  1. finite number of samples
  2. density estimation approximation

- Simple to code

- May provide a useful estimate of the arrival cost if computationally inexpensive

Density estimation: the curse of dimensionality [2]

Dimensionality	Required Sample Size
1	4
2	19
3	67
4	223
5	768
6	2790
7	10700
8	43700
9	187000
10	842000

Sample size required to ensure that the relative mean square error at zero (a single point) is less than 0.1. The underlying distribution is a standard multivariate normal density.

References

[1] E. Bølviken, P. J. Acklam, N. Christopherson, and J.-M. Størdal. Monte Carlo filters for non-linear state estimation. *Automatica*, 37(2):177–183, February 2001.

[2] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.

[3] M. Tenny. *Computational Strategies for Nonlinear Model Predictive Control*. PhD thesis, University of Wisconsin–Madison, 2002.