On-line State and Parameter Estimation for Nonlinear Dynamic Systems: An NLP Framework

R. Lopez-Negrete and L. T. Biegler
Carnegie Mellon University
June, 2012

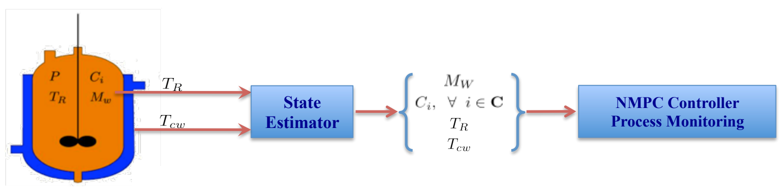
State Estimation Overview

- Introduction and Problem Setting
- Moving Horizon Estimation (MHE)
- Arrival Costs – Filtered
  - Bayesian properties
  - Sampling-based implementation
- Optimization Tools for MHE
  - IPOPT, sIPOPT
- Arrival Costs – Smoothed
  - Bayesian properties
  - Efficient covariance updates
- Dynamic Case Studies
  - Multi-reactor systems
  - Distillation
- Conclusions and Extensions

2

Motivation for State Estimation Strategy

- During process operation we need to know the state of the system
  - Process Monitoring
  - Process Control
- Nonlinear dynamic process models (DAEs) with measurement and process noise
- Use measurements, mathematical model to estimate full state set
- Obtained through time-critical on-line nonlinear programming (NLP) solutions



3

State Estimation Setting

- Dynamic nonlinear process represented by

States

$$z_{k+1} = f(z_k) + w_k$$

Measurements

$$y_k = h(z_k) + v_k$$

- where, w_k and v_k are Gaussian, zero mean, uncorrelated random variables
  - The process noise w_k may represent plant-model mismatch or unknown disturbances, assumed to be white with $\mathcal{N}(0, Q_k)$
  - Measurement noise, v_k comes from measurement equipment (sensors). Also white, but with covariance $\mathcal{N}(0, R_k)$.

4

State Estimation Setting

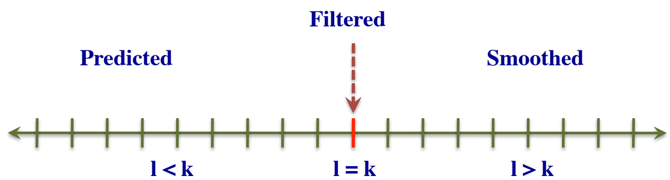
- State estimate pdf will be given by[1]

$$p\left(z_k|Y_l\right)$$

Probability of z_k given measurements up to l

$$Y_l = \{y_0, \ldots, y_l\}$$

  - If $l < k$ we get the **predicted state**
  - If $l = k$ we get the **filtered state**
  - If $l > k$ we get the **smoothed state**



1. Jazwinski, A. H. Stochastic Processes and Filtering Theory; Dover, 2007

5

State Estimation Setting

- State estimate pdf given by[1]

$$p\left(z_k|Y_l\right)$$

Probability of z_k given measurements up to l

$$Y_l = \{y_0, \ldots, y_l\}$$

- Assumptions
  - State sequence evolves as a first order Markov process
  - Measurement and system noise are independent, zero mean, Gaussian random variables
  - The initial prediction of the states (z_0) is a completely known random variable
  - We know the process and measurement models

1. Jazwinski, A. H. Stochastic Processes and Filtering Theory; Dover, 2007

6

State Estimation Background

- Linear Gaussian systems:
  - Simplest cases where we directly propagate and update the PDF:
  - Kalman Filter: Propagates mean and covariance (recursive solution)

- General nonlinear systems
  - Determining PDF structure/type and propagating usually not tractable
  - Extended Kalman Filter (EKF) – *state of practice*:[1]
    - Assume Gaussian states and use a linearized model
    - Loss of information from linearization
    - No bounds/constraints
  - Unscented KF, nonlinear evolution of states and covariance through pre-assigned σ points
  - Particle Filters, Ensemble KF, …, nonlinear evolution through sampling distributions

1. Haseltine, E. L. & Rawlings, J. B, Ind. Eng. Chem. Res., 2005, 44, 2451-2460

7

Moving Horizon Estimation

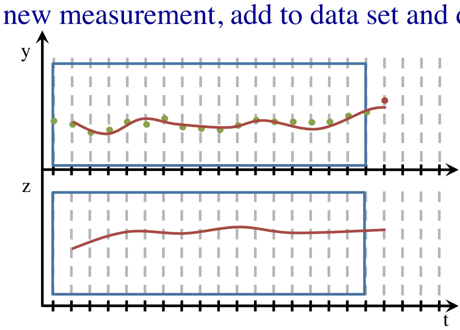
Full Information Estimation Problem

$$\{\hat{z}_{0|k}, \ldots, \hat{z}_{k|k}\} = \arg \max_{\{z_0, \ldots, z_k\}} p\left(Z_0^k | Y_0^k\right) =$$

$$= \arg \max_{\{z_0, \ldots, z_k\}} \log p\,(z_0) + \sum_{l=0}^{l=k} \log p\,(v_l) + \sum_{l=0}^{l=k-1} \log p\,(w_l)$$

Moving Horizon Estimation (MHE)[1]

- process only N measurements at each sampling time k
- with each new measurement, add to data set and drop oldest one



1. Rao, CV; Rawlings, JB & Lee, JH, Automatica, 2001, 37, 1619-1628

8

MHE Bayesian Derivation

- Find the conditional PDF of the trajectory of the states

$$p\left(Z_{k-N}^{k} \,\middle|\, Y_0^k\right) = \mathcal{K}\, p\left(z_0 \,\middle|\, Y_0^{k-N-1}\right) \prod_{l=k-N}^{k} p(y_l|z_l)p(z_l|z_{l-1})$$

- where $Z_a^b = \{z_a, z_{a+1}, \ldots, z_b\}$ $\quad Y_a^b = \{y_a, y_{a+1}, \ldots, y_b\}$
- Taking logarithms and maximizing the conditional PDF

Arrival Cost

$$\min_{z_{k-N},\ldots,z_k} \underbrace{\boxed{\|z_{k-N} - \bar{z}_{k-N|k-N-1}\|^2_{\Pi^{-1}_{k-N|k-N-1}}}}_{\text{Arrival Cost}} + \sum_{l=k-N}^{k-1} \|z_{l+1} - f(z_l, u_l)\|^2_{Q_l^{-1}} + \sum_{l=k-N}^{k} \|y_l - h(z_l)\|^2_{R_l^{-1}}$$

$$\text{s.t.} \quad z^{LB} \le z_l \le z^{UB}$$

  - where solution sequence contains smoothed and filtered state estimates:

$$\{\hat{z}_{k-N|k}, \ldots, \hat{z}_{k|k}\}$$

  - Can add bounds to states to include more process information

9

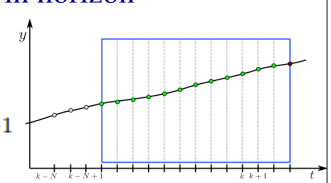
MHE Arrival Cost

- The **Arrival Cost**
  - Summarizes previous information <u>not included</u> in horizon window
  - As time moves forward, need to update its pdf parameters

$$\min_{z_{k-N},\ldots,z_k} \boxed{\mathcal{P}\left(z_{k-N}\right)} + \sum_{l=k-N}^{k-1} \|z_{l+1} - f(z_l, u_l)\|^2_{Q_l^{-1}} + \sum_{l=k-N}^{k} \|y_l - h(z_l)\|^2_{R_l^{-1}}$$

$$\text{s.t.} \quad z^{LB} \le z_l \le z^{UB}$$

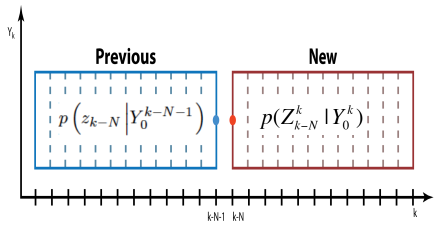
  - Penalizes the prediction of initial condition
  - Summarizes previous measurements not in horizon
  - AC changes as horizon moves forward
    - Update parameters for $\mathcal{P}(z_{k-N})$
      - Estimated initial condition, $\hat{z}_{k-N|k-N-1}$
      - Covariance associated with prediction $\Pi_{k-N|k-N-1}$



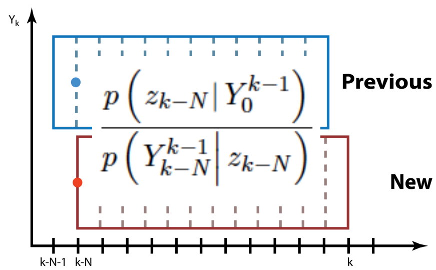
10

Updating the Arrival Cost

- The predicted initial condition can be updated in two ways



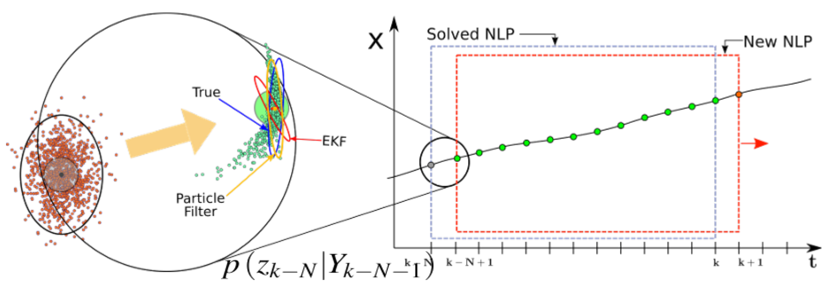
Filtered
- The filtered solution is propagated forward with the model
- The covariance is propagated with EKF



Smoothed
- Obtain smoothed state estimate directly from NLP solution
- The covariance is taken from EKF and Kalman Smoother

Sampled Approximation of Filtered Arrival Cost

- At each sampling time, the arrival cost parameters approximated with Sample based filter[1]



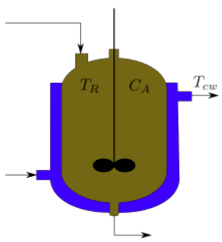
Approximated with SF
between sampling times

- AC update done in the background, avoiding online computations
  - Arrival cost is not Gaussian
    - AC Sampling through, UKF, PF/EnKF
    - MHE is constrained → Arrival Cost must be constrained
  - Horizon window length depends on accuracy of Arrival Cost

1. **Lopez-Negrete, R**, Patwardhan, SC & Biegler, LT, J. Process Control, doi:10.1016/j.jprocont.2011.03.004 (2011)

Filtered Arrival Cost - Results

- Example:
  - Non-isothermal CSTR with cooling water dynamics
  - The model consists of concentration and reactor and cooling water temperatures
  - Only measured variable is the cooling water temperature T_{cw}



$$\frac{dC_A}{dt} = \frac{F}{V}(C_{Ain} - C_A) - 2k(T_R)C_A^2$$

$$\frac{dT_R}{dt} = \frac{F}{V}(T_{in} - T) + 2\frac{(-\Delta H)_R}{\rho Cp}k(T_R)C_A^2 - \frac{UA}{V\rho Cp}(T_R - T_{cw})$$

$$\frac{dT_{cw}}{dt} = \frac{F_{cw}}{V_{cw}}(T_{cwin} - T_{cw}) + \frac{UA}{V_{cw}\rho_{cw}Cp_{cw}}(T_R - T_{cw})$$

$$k(T_R) = k_o \exp\left(\frac{-E}{RT_R}\right)$$

$$w_k \sim \mathcal{N}(0, Q_k), \quad Q_k = diag(10^{-8}, 0.25, 0.25)$$

$$v_k \sim \mathcal{N}(0, R_k), \quad R_k = 0.25$$

$$z_{0|0} \sim \mathcal{N}(0, P_0), \quad P_0 = diag(10^{-7}, 2.5, 2.5), \quad z_{0|0} = [0.0192, 384.0072, 371.2735]^T$$

13

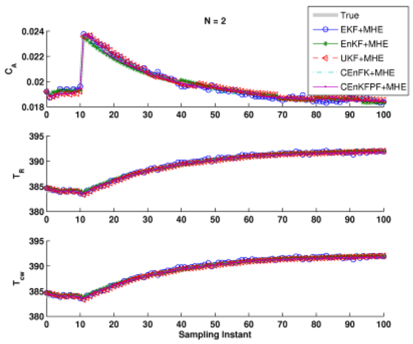
Comparison of Arrival Cost Updating Methods

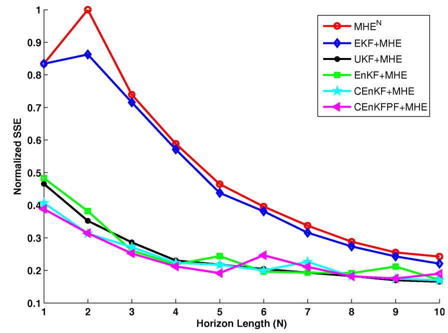
Filter	Sampling	Sample bounds	Update type	Update bounds
EKF	none	none	linear	unconstrained
UKF	sigma points	none	linear	unconstrained
EnKF	random	none	linear	unconstrained
EnKFPF*	random/random	none	linear	unconstrained
UKFPF*	sigma points/random	none	linear	unconstrained
ωEnKFPF*	random/random	none	linear	unconstrained
URNDDR	sigma points	clipping	nonlinear	QP
URNDDRPF*	sigma points/random	clipping	weights	QP/through density
CEnKF	random	clipping	weights	QP
CEnKFPF*	random/random	clipping	weights	QP/through density
ωCEnKFPF*	sigma points/random	clipping	weights	QP/through density

Unconstrained cases: sampling/importance distributions
Bound Constrained cases: truncated (clipped) state distributions

14

Filtered Arrival Cost - Results

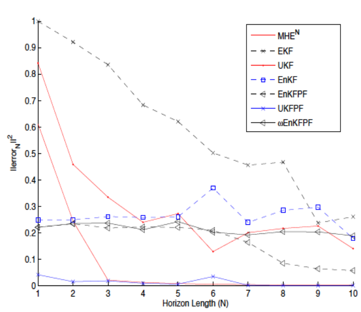


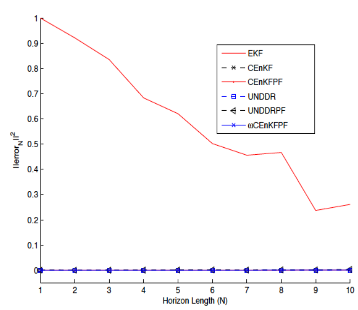


- Unconstrained Case Study
- Smaller horizon length with smaller estimation error - possible with better approximation of the arrival cost
- Number of required particles increases with number of states… Some times exponentially!

15

Filtered Arrival Cost - Results





- Bound Constrained Case Study
- Shorter horizon length with smaller estimation error - possible with better approximation of the arrival cost
- Smaller on-line NLP vs. greater off-line computation
- *Is sampling-based filtered AC worth the cost?*

16

Optimization Tools for Nonlinear Dynamic Optimization

$$\min \ \psi\left(z(t), y(t), u(t), p, t_f\right)$$

$$s.t. \quad \frac{dz(t)}{dt} = F\left(z(t), y(t), u(t), t, p\right)$$

$$G\left(z(t), y(t), u(t), t, p\right) = 0$$

$$z^o = z(0)$$

$$z^l \le z(t) \le z^u$$

$$y^l \le y(t) \le y^u$$

$$u^l \le u(t) \le u^u$$

$$p^l \le p \le p^u$$

t, time
z, differential variables
y, algebraic variables

t_f, final time
u, control variables
p, time independent parameters

Nonlinear Programming Formulation

Nonlinear Dynamic Optimization Problem

(Piecewise) Continuous profiles

Collocation on finite Elements

Nonlinear Programming Problem (NLP)

Discretized variables

Discretized DynamicNonlinear Program

$$\min \quad \psi\left(z_i, y_{i,q}, u_{i,q}, p, t_f\right)$$

s.t.

$$\left(\frac{dz}{dt}\right)_{i,j} = F\left(z_{i\text{-}1}, \frac{dz}{dt}_{i,j}, z_i, y_{i,j}, u_{i,j}, p\right)$$

$$G\left(z_{i\text{-}1}, \frac{dz}{dt}_{i,j}, z_i, y_{i,j}, u_{i,j}, p\right) = 0$$

$$z_i = f\left(\frac{dz}{dt}_{i-1,j}, z_{i-1}\right)_i$$

$$z_0^{\;o} = z(0)$$

$$z_i^{\;l} \le z_i \le z_i^{\;u}$$

$$y_{i,j}^{\;l} \le y_{\mathrm{i,j}} \le y_{i,j}^{\;u}$$

$$u_{i,j}^{\;l} \le u_{i,j} \le u_{i,j}^{\;u}$$

$$p^l \le p \le p^u$$

$$\min_{x \in \Re^n} f(x)$$

$$\text{s.t} \quad c(x) = 0$$

$$x^L \le x \le x^u$$

IPOPT– Interior Point Solver

- IPOPT[1]
  - Open source Interior Point NLP solver (see www.coin-or.org)
  - Solve large problems with ≤ 10^6 variables and constraints
  - Logarithmic barrier approach (bounds → penalty function)

$$\min \; f(x;p) - \mu \sum_{i=1}^{N_c} ln(x_i)$$

$$\text{s.t.} \; g(x;p) = 0$$

  - Solve for decreasing values of μ (μ → 0)
  - Apply Newton's method to optimality (KKT) conditions

$$\underbrace{\begin{bmatrix} H_k & A_k & -I \\ A_k^T & 0 & 0 \\ V_k & 0 & X_k \end{bmatrix}}_{\mathbf{K}} \begin{bmatrix} \Delta x_k \\ \Delta \lambda_k \\ \Delta \nu_k \end{bmatrix} = - \begin{bmatrix} \mathcal{L}(x_k;p) \\ g(x_k;p) \\ X_k V_k e - \mu e \end{bmatrix}$$

1. Wächter and L. T. Biegler, Mathematical Programming 106(1), pp. 25-57, 2006

20

IPOPT Factorization Byproduct

- Modify KKT (full space) matrix if nonsingular

$$\begin{bmatrix} H_k + \Sigma_k + \delta_1 I & A_k \\ A_k^T & -\delta_2 I \end{bmatrix}$$

  - δ_1 - Correct inertia to guarantee descent direction
  - δ_2 - Deal with rank deficient A_k
- KKT matrix factored by indefinite symmetric factorization
- Solution with δ_1=0 ➔ sufficient second order conditions
- Parameter Estimation Result – unique parameters
- Reduced Hessian available to approximate confidence regions

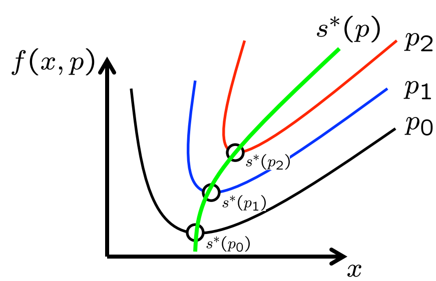
Sensitivity with IPOPT

- sIPOPT[1] – Open source library direct link with IPOPT
  - Barrier problem

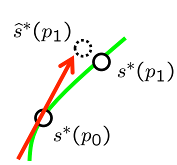
$$\min \; f(x;p) - \mu \sum^{N_c} ln(x_i) \quad \text{s.t.} \; g(x;p) = 0$$

$$\min \; f(x;p) \quad \text{s.t.} \; g(x;p) = 0, \; x \geq 0$$

$$s^\star(p)^T = [x^{\star T} \; \lambda^{\star T} \; \nu^{\star T}]$$



  - NLP sensitivity (Fiacco – 1983):
    - Regularity conditions must be met (SSOC, SC, LICQ)
    - Existence and differentiability of the path s*(p)
  - Taylor series expansion of solution (implicit function theorem)



$$\hat{s}^\star(p) = s^\star(p_0) + \boxed{\left.\frac{\partial s}{\partial p}^T\right|_{p_0}}(p - p_0) + o(\|p - p_0\|^2)$$

1. Pirnay, **Lopez-Negrete**, & Biegler, Optimal Sensitivity with IPOPT, Submitted to Math Prog Comp, 2011

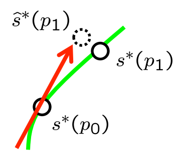
Sensitivity with IPOPT

- Sensitivity matrix can be obtained from the linearized optimality conditions

$$\mathbf{K}^* \begin{bmatrix} \Delta x_k \\ \Delta \lambda_k \\ \Delta \nu_k \end{bmatrix} = - \mathbf{M}$$

$$\left. \frac{\partial s}{\partial p}^T \right|_{p_0} = - \left[K^*(s(x, p_0)) \right]^{-1} M(s(x, p_0))$$

- At the solution we freeze K (already factorized!)
  – Change right hand sides and solve linear system
- Generate perturbed solution



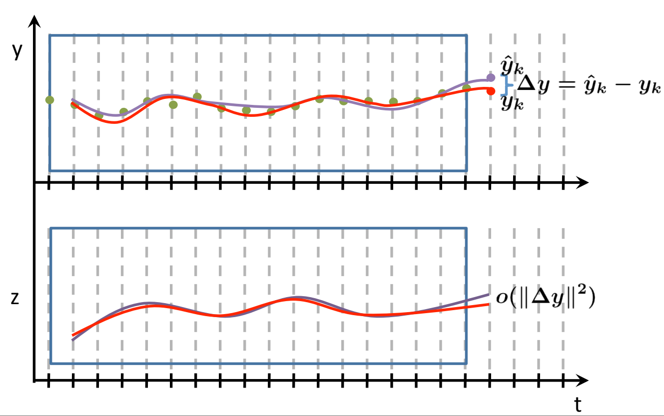
$$\hat{s}^\star(p) = s^\star(p_0) + \left. \frac{\partial s}{\partial p}^T \right|_{p_0} (p - p_0) + o(\|p - p_0\|^2)$$

- sIPOPT → Open source library freely available with IPOPT
  - **https://projects.coin-or.org/Ipopt/wiki/sIpopt**

23

Advanced Step MHE

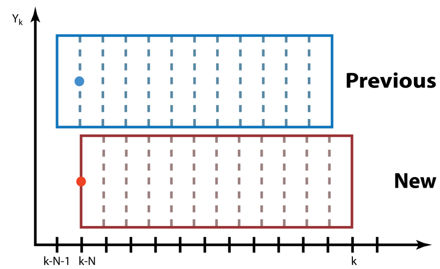
- Exploit NLP Sensitivity for approximate NLP solutions (Similar to asNMPC)
  – Predict new measurement
  – Solve NLP in background (between sampling times)
  – Correct solution with obtained measurement & NLP Sensitivity
  → On-line computation reduced ~ two orders of magnitude



24

Large Scale MHE

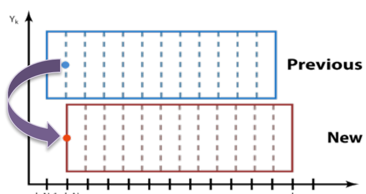
- Two possible AC formulations



- Filtered Arrival Cost applies to both formulations, but smoothed approach seems more efficient[1]
  - Avoids oscillations (specially when horizon is smaller)

1. Tenny, M. J. & Rawlings, J. B; American Control Conference, 2002, 4475-4480

Smoothed Arrival Cost



$$p\left(Z_{k-N}^{k}\middle|Y_0^k\right) = \mathcal{K}\,\frac{\overbrace{p\left(z_{k-N}\middle|Y_0^{k-1}\right)}^{\text{Smoothed PDF}}}{\underbrace{p\left(Y_{k-N}^{k-1}\middle|z_{k-N}\right)}_{\text{Correction term}}}\prod_{l=k-N}^{k} p(y_l|z_l)p(z_l|z_{l-1})$$

Assume Quadratic AC form

$$\min_{z_{k-N},\dots,z_k} \boxed{\left\|z_{k-N}-\bar{z}_{k-N|k-1}\right\|^2_{\Pi^{-1}_{k-N|k-1}} - \left\|\mathcal{Y}^{k-1}_{k-N} - \mathcal{O}z_{k-N}\right\|^2_{W^{-1}_{k-N}}} + \sum_{l=k-N}^{k-1}\left\|z_{l+1}-f(z_l,u_l)\right\|^2_{Q_l^{-1}} + \sum_{l=k-N}^{k}\left\|y_l-h(z_l)\right\|^2_{R_l^{-1}}$$

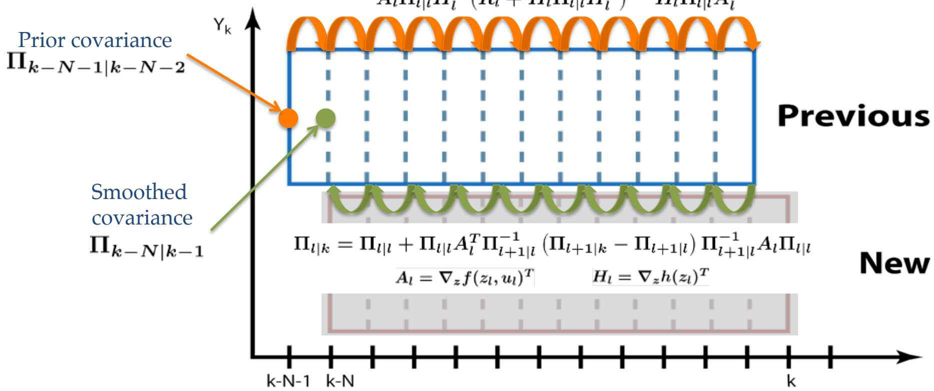
$$\text{s.t.} \quad z^{LB} \le z_l \le z^{UB}$$

- where:
  - $\mathcal{Y}$, $\mathcal{O}$, and W are known functions of previous data
  - Require smoothed estimate of initial condition
  - Covariance associated and updated with state estimate

Smoothed Arrival Cost

- Standard Approach: EKF propagates covariance forward, then Kalman smoother propagates backwards

$$\Pi_{l+1|l} = A_l \Pi_{l|l} A_l^T + \Gamma_l Q_{l|l} \Gamma_l^T - A_l \Pi_{l|l} H_l^T \left(R_l + H_l \Pi_{l|l} H_l^T\right)^{-1} H_l \Pi_{l|l} A_l^T$$



Recursive Solution to MHE

- MHE NLP is solved using Interior Point method
  - Newton type search direction found by linearizing the optimality conditions

$$\min_{z_{k-N},\dots,z_k} \left\| z_{k-N} - \bar{z}_{k-N|k-1} \right\|^2_{\Pi^{-1}_{k-N|k-1}} - \left\| \mathcal{Y}^{k-1}_{k-N} - \mathcal{O} z_{k-N} \right\|^2_{W^{-1}_{k-N}} + \sum_{l=k-N}^{k-1} \| z_{l+1} - f(z_l, u_l) \|^2_{Q_l^{-1}} + \sum_{l=k-N}^{k} \| y_l - h(z_l) \|^2_{R_l^{-1}}$$

Linearize optimality conditions

$$\left[\begin{array}{cccccccccc|cccc}
P_0 & -F_0 & & & & & & & & -A_0^T & & & \\
-F_0^T & W_0 & & & & & & & & -G_0 & & & \\
& & P_1 & -F_1 & & & & & & I & -A_1^T & & \\
& & -F_1 & W_1 & & & & & & & -G_1^T & & \\
& & & & \ddots & & & & & & & \ddots & \\
& & & & & P_{N-1} & -F_{N-1} & & & & & I & -A_{N-1}^T \\
& & & & & -F_{N-1}^T & W_{N-1} & & & & & & -G_{N-1}^T \\
& & & & & & & P_N & & & & & I \\
\hline
-A_0 & -G_0 & I & & & & & & & & & & \\
& & & & \ddots & & & & & & & & \\
& & & & & -A_{N-1} & -G_{N-1} & I & & & & &
\end{array}\right]
\left[\begin{array}{c}
\Delta_{z_0} \\ \Delta_{w_0} \\ \Delta_{z_1} \\ \Delta_{w_1} \\ \vdots \\ \Delta_{z_{N-1}} \\ \Delta_{w_{N-1}} \\ \Delta_{z_N} \\ \hline \Delta_{\lambda_1} \\ \vdots \\ \Delta_{\lambda_N}
\end{array}\right]
= -
\left[\begin{array}{c}
r_{z_0} \\ r_{w_0} \\ r_{z_1} \\ r_{w_1} \\ \vdots \\ r_{z_{N-1}} \\ r_{w_{N-1}} \\ r_{z_N} \\ \hline r_{\lambda_1} \\ \vdots \\ r_{\lambda_N}
\end{array}\right]$$

$A_l = \nabla_{z_l} f(\cdot)^T, G_l = \nabla_{w_l} f(\cdot)^T$

$P_l = \nabla^2_{z_l} \mathcal{L}, W_l = \nabla^2_{w_l} \mathcal{L}, F_l = \nabla_{z_l w_l} \mathcal{L}$

Recursive Solution of MHE

- Recursive solution to linear KKT system is

$$
\begin{aligned}
\Delta z_N &= -\Pi_N (M_N^{-1} r_{M_N} + r_{z_N}) \\
\Delta \lambda_l &= M_l^{-1} (\Delta z_l + r_{M_l})
\end{aligned}
$$

Propagation of the states

- where

$$
\begin{aligned}
\Pi_0 &= (\Pi_{k-N|k-1} - F_0 W_0^{-1} F_0^T)^{-1} \\
M_{l+1} &= (G_l W_l^{-1} F_l^T + A_l)\Pi_l (F_l W_l^{-1} G_l^T + A_l^T) + G_l W_l^{-1} G_l^T
\end{aligned}
$$

Propagation of Covariance

- and

$$
r_{M_1} = r_{\lambda_1} + G_0 W_0^{-1} r_{w_0} + (F_0 W_0^{-1} G_0^T + A_0)\Pi_0 (F_0 W_0^{-1} r_{w_0} + r_{z_0})
$$

Auxiliary equations

- The forward propagation of the covariance is done internally
- What about the smoothed covariance?

This is also generated as part of the optimality conditions of the NLP!

29

Reduced Hessian and Covariance

- We can show that the inverse of the Reduced Hessian is the smoothed covariance[1]

$$
\text{Cov}\left(\boldsymbol{Z}_{\boldsymbol{k-N}}^{\boldsymbol{k}}\right) = \left[\mathbb{Z}^{\boldsymbol{T}} \left(\boldsymbol{\nabla}^{\boldsymbol{2}} \boldsymbol{\mathcal{L}}\right) \mathbb{Z}\right]^{-\mathbf{1}}
$$

  – where $\mathbb{Z}$ is the null space basis of the constraint Jacobian

- Changing variables for simplicity

$$
\begin{aligned}
\min\ & \frac{1}{2}(y-\hat{y})^T V_y^{-1} (y-\hat{y}) + \frac{1}{2}\left(\theta-\hat{\theta}\right)^T V_\theta^{-1} \left(\theta-\hat{\theta}\right) \\
\text{s.t.}\ & c(y,\theta) = 0
\end{aligned}
$$

$$
y^T = \left[y_0^T, \ldots, y_N^T, w_0^T, \ldots, w_{N-1}^T\right] \qquad \theta^T = \left[z_0^T, \ldots, z_N^T\right]
$$

$$
c(y,\theta) = \begin{bmatrix} (y_0 - h(z_0)) \\ \vdots \\ (y_N - h(z_N)) \\ (z_1 - f(z_0, w_0)) \\ \vdots \\ (z_N - f(z_{N-1}, w_{N-1})) \end{bmatrix} \qquad
V_y = \begin{bmatrix} R_0 & & & & & \\ & \ddots & & & & \\ & & R_N & & & \\ & & & Q_0 & & \\ & & & & \ddots & \\ & & & & & Q_{N-1} \end{bmatrix} \qquad
V_\theta = \begin{bmatrix} \Pi_{k-N|k-1} & & \\ & \ddots & \\ & & 0 \end{bmatrix}
$$

1. Pirnay, **Lopez-Negrete**, & Biegler, Optimal Sensitivity with IPOPT, Math Prog Comp, 2012, to appear

30

Sensitivity of KKT Conditions

- Analyze sensitivity of estimates with changes in data

$$\min \frac{1}{2}(y-\hat{y})^T V_y^{-1}(y-\hat{y}) + \frac{1}{2}\left(\theta-\hat{\theta}\right)^T V_\theta^{-1}\left(\theta-\hat{\theta}\right)$$
$$\text{s.t.} \quad c(y,\theta)=0$$

- At the solution we have the linearized optimality conditions

$$\begin{bmatrix} \nabla_y^2 \mathcal{L} & 0 & \nabla_y c \\ 0 & \nabla_\theta^2 \mathcal{L} & \nabla_\theta c \\ \nabla_y c^T & \nabla_\theta c^T & \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta\theta \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} V_y^{-1}\delta\hat{y} \\ V_\theta^{-1}\delta\hat{\theta} \\ 0 \end{bmatrix}$$

- Introduce perturbations $y^\star + \delta\hat{y}$ and $\theta^\star + \delta\hat{\theta}$

$$\mathbb{E}\left\{\Delta\theta\Delta\theta^T\right\} = \mathrm{Cov}\left(Z_{k-N}^k\right) = \left[\mathbb{Z}^T\left(\nabla^2\mathcal{L}\right)\mathbb{Z}\right]^{-1}$$

- Note $\|\lambda\| = o\begin{pmatrix} y-\hat{y} \\ \theta-\hat{\theta} \end{pmatrix} \implies \mathbb{E}\{\lambda\}=0$

Extract Reduced Hessian from IPOPT

- If dynamic system is linear with Gaussian noise, this reduces to the Kalman Smoothing equations

- Interior point solvers do not form the Reduced Hessian, can be extracted from the optimality conditions[1]

KKT conditions at optimal solution

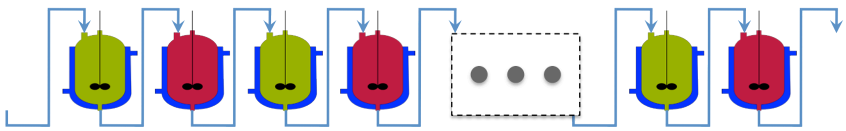
$$\begin{bmatrix} W & J \\ J^T & 0 \end{bmatrix} \Delta x = -\mathrm{rhs} = -\begin{bmatrix} 0 \\ 0 \\ I_z(:,j) \\ 0 \\ \vdots \end{bmatrix} \qquad \Delta x = \begin{bmatrix} \Delta z_{k-N} \\ \Delta w_{k-N} \\ \Delta z_{k-N+1} \\ \Delta w_{k-N+1} \\ \vdots \end{bmatrix}$$

- Δx_j is the j-th column of the inverted reduced Hessian
- In Ipopt KKT matrix is already factorized!
  - One back-solve per column of the covariance

1. Zavala, V. M.; Laird, C. D. & Biegler, L. T.; Journal of Process Control, 2008, 18, 876-884

Smoothed Arrival Cost - Results

- Simulation Example:
  - Network of CSTR's each has two states
  - We can increase the size of the network as needed



$$\frac{dC_i}{dt} = \frac{C_{i-1} - C_i}{\theta_i} - k_0 C_i e^{-\frac{R}{RT_i}}$$

State	Model variance	Measurement Variance
$C_{A,i}$	10^{-1}	–
T_i	10^{-1}	0.5

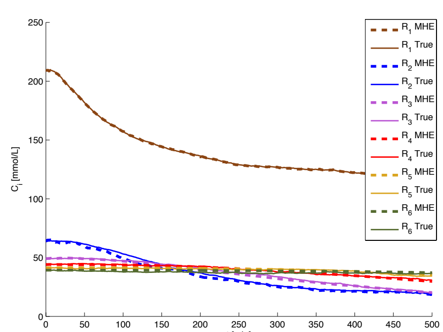
2, and

33

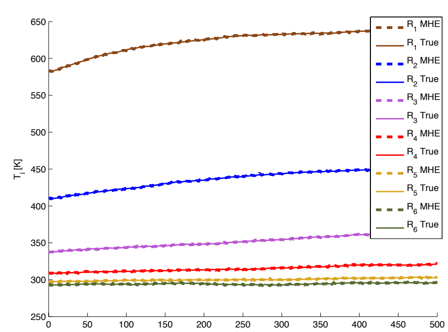
Smoothed Arrival Cost - Results

- 6 reactors in the network, and using 20 measurements in the horizon
  - The state estimates are very close to the true values obtained through simulation

Concentration



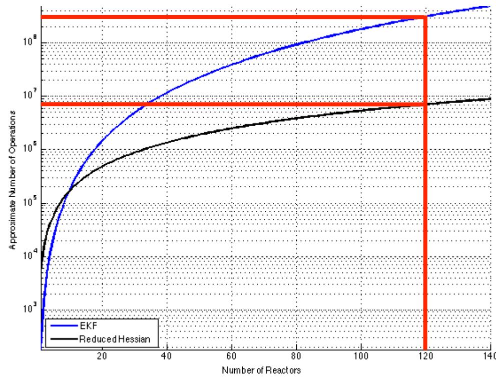
Temperature



34

Smoothed Arrival Cost - Results

- Extracting inverse Reduced Hessian from KKT conditions significantly reduces the number of operations required to obtain covariance

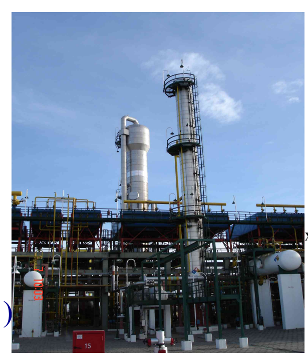


- Complexity of EKF/Smoother increases cubically with number of states, while essentially linear for the Reduced Hessian calculation

35

Large Scale Example – Distillation

- Binary distillation (methanol and n-propanol)[1]
- Column with 40 trays, and per tray we have
  - Detailed Dynamic MESH Model
- Index 1 DAE with 252 equations:
  - 84 Differential equations (methanol compositions of liquid phase and liquid molar holdup)
  - 168 Algebraic equations (T, L, V, y,…)



x_k, T_k, M_k

- Measurements are T and liquid volume holdup for each tray
- Sampling time is 60s
- Implementation of the asMHE with reduced Hessian approximation of smoothed covariance

1. Diehl, M., PhD Thesis, Heidelberg, Germany, 2001

36

Large Scale Example – Distillation

- Horizon length of 10 measurements
- Model after discretization (3 collocation points)
  - 19419 variables
  - 18579 equality constraints
  - 12180 between upper and lower bounds
- Average solution time
  - NLP: 42.38 CPU s (66 iterations)
  - Sensitivity: 0.529 CPU s
  - Reduced Hessian: 1.84 CPU s

Online Computation

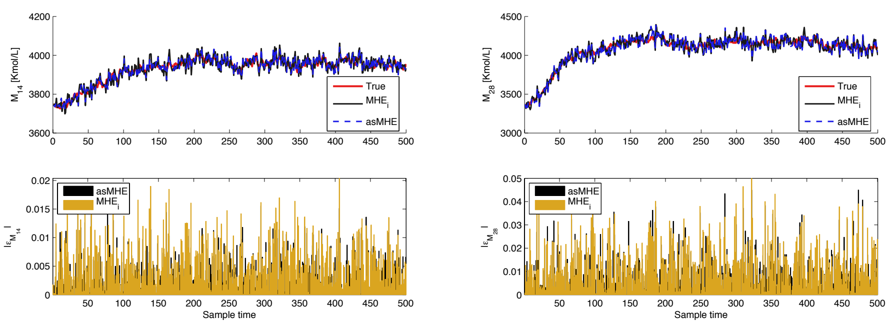
	Model noise variance	Measurement variance
T_i	–	6.25×10^{-2}
$V^m_{N_T+1}$	–	10^{-8}
x_i	10^{-5}	–
M_0	10	–
M_i	1	–
M_{N_T+1}	5	–

37

Large Scale Example – Distillation

Liquid Molar Holdup

Tray 14 Tray 28



38

Summary and Conclusions

- MHE: efficient, accurate framework for constrained state estimation
  - Applies directly to nonlinear dynamics with reasonable noise assumptions
  - Adding information about states through bounds improves quality of estimates
- Key aspect: treatment of arrival costs
  - Filtered AC: off-line sampling vs. shorter horizons
  - Smoothed AC: longer horizon problem solved more efficiently
- NLP (IPOPT) and Sensitivity (sIPOPT) provides fast online estimate of the states → asMHE
- Covariance information can be approximated from reduced Hessian and extracted very cheaply from IPOPT
  - Propagation of covariance is performed internally in the NLP (no need to do this twice)
- Large nonlinear MHE problems handled on-line (with hundreds of states)

39

Extensions and Future Work

- Multi-rate measurements, includes less frequent (possibly delayed) measurements
  - Some unobservable states can become observable
  - Straightforward to implement with MHE and smoothed AC
- Robust estimators to reduce the effects of measurement errors and outliers
  - Modify MHE with M-estimators (Hampl, Huber type)
  - Analyze observability of MHE with Robust M-Estimators
- Extend MHE formulation to include Fault detection/identification

Acknowledgements

- Prof. Sachin Patwardhan, IIT Bombay
- US National Science Foundation

40

Dynamic Distillation Model

•Mesh Equations for Distillation Column

Assumption:

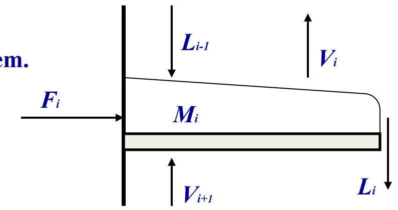
Vapor holdups are negligible. **Index 2 system.**

Ideal vapor phases.

Well mixed entering streams.

Constant pressure drop.

Equilibrium stage model.



Mass balance: $$\frac{dM_i}{dt} = L_{i-1} + V_{i+1} - L_i - V_i + F_i$$

Component balance: $$\frac{d(M_i x_{i,j})}{dt} = L_{i-1}x_{i-1,j} + V_{i+1}y_{i+1,j} - L_i x_{i,j} - V_i y_{i,j} + F_i x^f_{i,j}$$

Energy balance: $$\frac{d(M_i h^L_i)}{dt} = L_{i-1}h^L_{i-1} + V_{i+1}h^V_{i+1} - L_i h^L_i - V_i h^V_i + F_i h^f_i$$

Phase equilibrium: $$y_{i,j}p_i = \gamma_{i,j} x_{i,j} p^{sat}_{i,j}$$

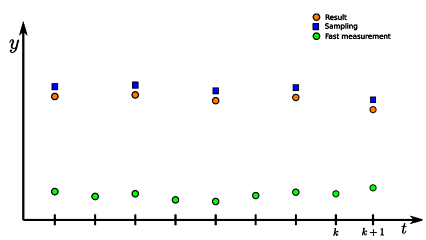
Summation: $$1 = \sum_{j \in \mathrm{COMP}} y_{i,j}$$

Hydrodynamics: $$L_i = k_d M_i$$

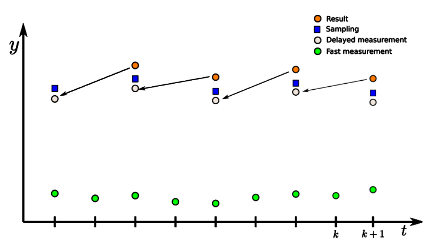
Reformulated index 1 system contains 320 ODEs, 1200 AEs.

Multi-Rate MHE

- No need to waste slow measurements
  - In some cases adding this information can help make unobservable states observable!



Different rates, no delays

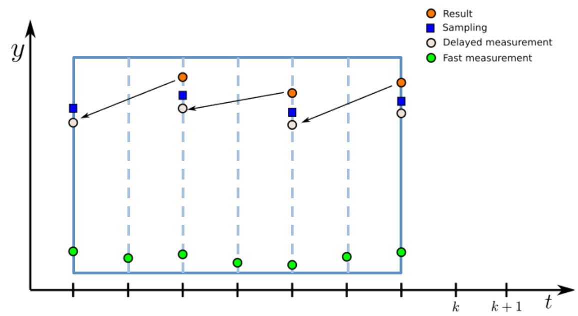


Different rates and delayed measurements

42

Multi-Rate MHE

- Just place the delayed measurements in their proper locations in the Horizon
  - Use the right measurement covariance matrix
- Arrival cost through reduced Hessian
  - Considers smoothing effects from all measurements automatically



43

Multi-Rate Example

- Example: Polymerization Reactor[1]
  - Styrene Polymerization
    - AsMHE not considered here, but application is straightforward
    - Fast measurements are:
      - Temperatures of reactor and cooling jacket
      - Concentration of monomer (viscosity - agitator)
    - Slow measurements are the molecular weight moments (GPC)
  - Assume that slow sample times are integer multiples of fast sample times
    - Fast sample rate 6 min
    - Slow sample rate 12 min (and delayed)

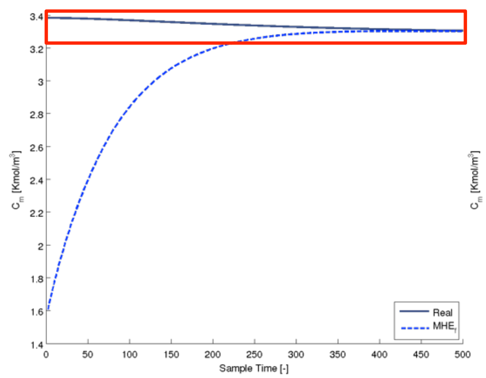
1. Tatiraju, S.; Soroush, M.; and Ogunnaike, B.A., AIChE Journal 45(4), 1999, pp. 769–780.

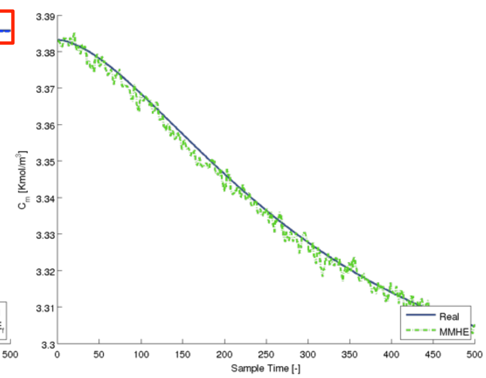
44

Multi-Rate MHE Example

Monomer Concentration

Only Fast Measurements | **Multi-Rated Measurements**

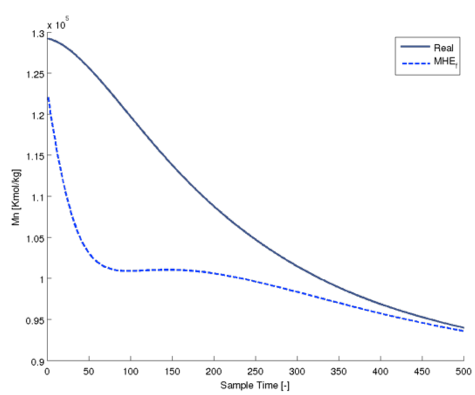


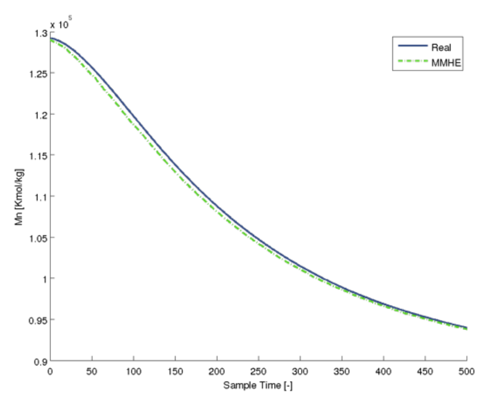


45

Weight Average Molecular Weight

Only Fast Measurements | **Multi-Rated Measurements**





46

Robust M-Estimators

- Outliers and gross errors are common during process operations
  - State/parameter estimates will be affected
  - Errors can propagate and destabilize the closed loop system
- Robust estimators give less weight to corrupted measurements

Fair Function[1]

$$\rho_j^F = C^2\left[\frac{|\epsilon_j|}{C} - \log\left(1 + \frac{|\epsilon_j|}{C}\right)\right]$$

Hampel's Re-descending Estimator

$$\rho_j^R = \begin{cases} \frac{1}{2}\epsilon_j^2, & 0 \le |\epsilon_j| \le a \\ a\,|\epsilon_j| - \frac{a^2}{2}, & a < |\epsilon_j| \le b \\ ab - \frac{a^2}{2} + \frac{a\,(c-b)}{2}\left[1 - \left(\frac{c - |\epsilon_j|}{c-b}\right)^2\right], & b < |\epsilon_j| \le c \\ ab - \frac{a^2}{2} + \frac{a\,(c-b)}{2}, & |\epsilon_j| > c \end{cases}$$

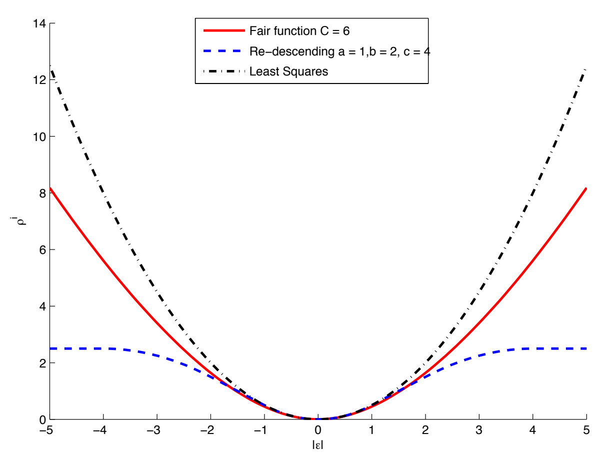
where ε_j are the studentized residuals

1. Arora, N & LT Biegler, Comput. Chem. Eng., 25, 1585—1599, 2001.

47

Robust M-Estimators

- Compare effects of the Fair Function, Hampel's Re-descending Estimator, and Least Squares.



48

Robust MHE

$$\min \quad \frac{1}{2}\left\|z_{k-N}-\hat{z}_{k-N|k-1}\right\|^2_{\hat{\Pi}^{-1}_{k-N|k-1}} - \frac{1}{2}\left\|\mathcal{Y}-\mathcal{O}z_{k-N}\right\|^2_{\mathcal{W}^{-1}} + \boxed{\frac{1}{2}\sum_{l=k-N}^{k}\rho_l^{ME}\left(v_l\right)} + \frac{1}{2}\sum_{l=k-N}^{k-1} w_l^T Q_l^{-1} w_l$$

$$\text{s.t.} \quad z_{l+1}-f\left(z_l\right)-w_l=0$$

$$y_l-h\left(z_l\right)-v_l=0$$

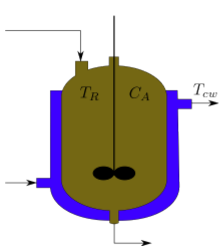
$$z^{LB}\leq z_l\leq z^{UB}$$

- Measurement error terms are substituted for M-Estimators to reduce effects of gross errors or outliers
  - When residuals are small we get least squares
  - Non-smooth terms in estimators are smoothed using an interior point smooth approximation

49

M-Estimator Results

- Example 1:
  - Non-isothermal CSTR with cooling water dynamics
  - The model consists of concentration and reactor and cooling water temperatures
  - Only measured variable are the reactor (T_R) and cooling water temperature (T_{cw})
  - A gross-error is introduced in the measured reactor temperature (e.g., the sensor drifts



$$\frac{dC_A}{dt}=\frac{F}{V}\left(C_{Ain}-C_A\right)-2k(T_R)C_A^2$$

$$\frac{dT_R}{dt}=\frac{F}{V}\left(T_{in}-T\right)+2\frac{(-\Delta H)_R}{\rho Cp}k(T_R)C_A^2-\frac{UA}{V\rho Cp}\left(T_R-T_{cw}\right)$$

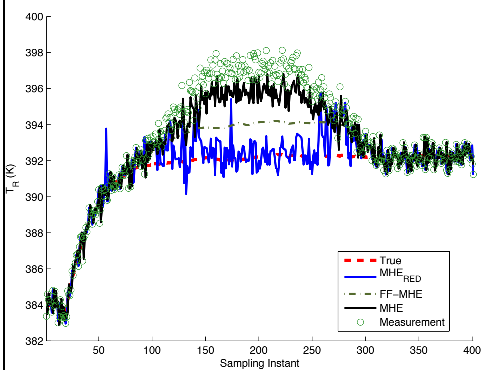
$$\frac{dT_{cw}}{dt}=\frac{F_{cw}}{V_{cw}}\left(T_{cwin}-T_{cw}\right)+\frac{UA}{V_{cw}\rho_{cw}Cp_{cw}}\left(T_R-T_{cw}\right)$$

$$k(T_R)=k_o\exp\left(\frac{-E}{RT_R}\right)$$

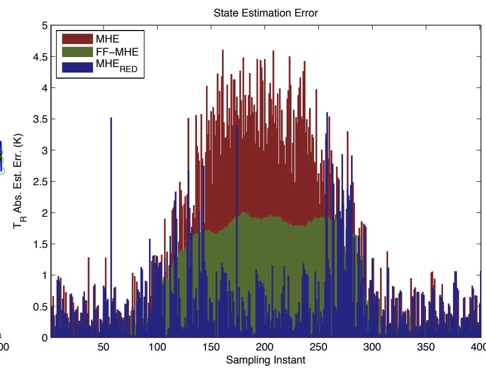
50

M-Estimator Results

Reactor Temperature



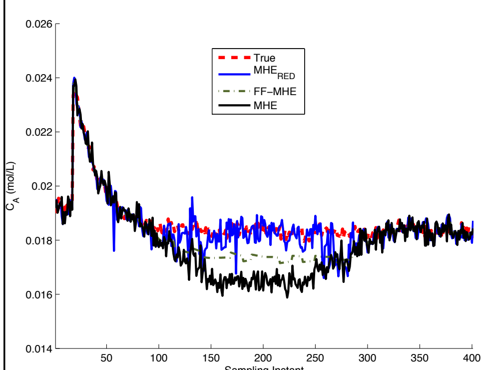
Estimation Error



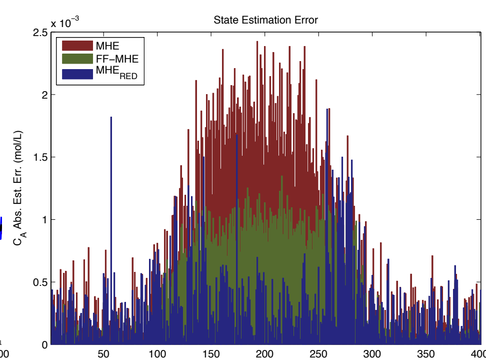
51

M-Estimator Results

Concentration

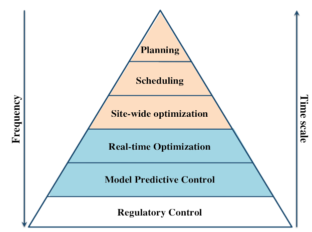


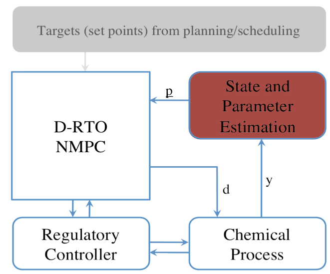
Estimation Error



52

Summary and Conclusions



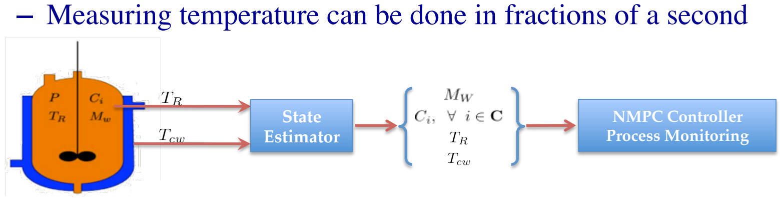


- Main contributions
  - sIPOPT - Sensitivity library for IPOPT
    - Fast approximation of neighboring problems
    - Reduced Hessian extraction
  - Real time nonlinear constrained estimation
    - AsMHE with smoothing arrival cost update
    - Covariance through reduced Hessian
    - Multi-rate asMHE extensions to MHE
    - Robust M-Estimators

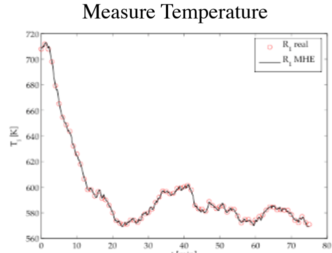
53

Motivational Example

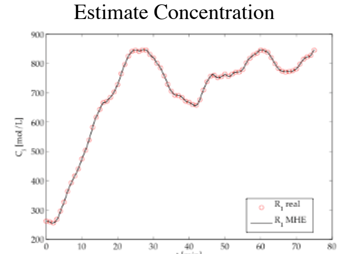
- Example Continuously Stirred Tank Reactor (CSTR)
  - Measuring concentration or molecular weights can take up to a few minutes
  - Measuring temperature can be done in fractions of a second



Measure Temperature



Estimate Concentration



54

Sample Based Approximation of Arrival Cost

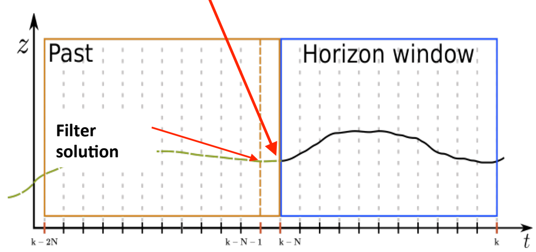
- Traditional update of arrival cost
  - Assume Gaussianity and use linearized model

$$\left\| z_{k-N} - \bar{z}_{k-N} \right\|^2_{\Pi^{-1}_{k-N}}$$

$\hat{z}_{k-N|k-N-1}$
Optimal prediction

$$\Pi_{k+1} = A_k \Pi_k A_k^T + \Gamma_k Q_k \Gamma_k^T - A_k \Pi_k H_k^T \left(R_k + H_k \Pi_k H_k^T \right)^{-1} H_k \Pi_k A_k^T$$

$$A_k = \nabla_{z_k} f(z_k, u_k)^T \Big|_{\hat{z}_{k-1|k-1}} \qquad H_k = \nabla_{z_k} h(z_k)^T \Big|_{\hat{z}_{k-1|k-1}}$$

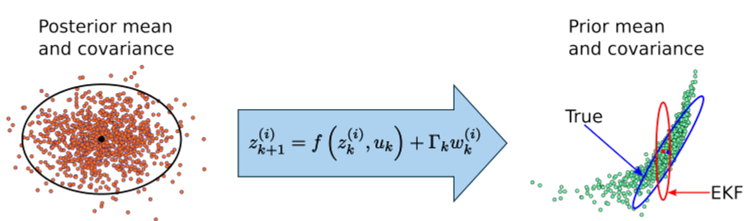


- Incorrect approximation:
  - Unconstrained method for propagating covariance
  - Using a linearized model for this transformation

55

Particle Fitlers

- Recognize that
  - Horizon window length is a function of Arrival Cost[1]
  - Arrival cost is not Gaussian
  - MHE is constrained → Arrival Cost must be constrained
- We propose to use Monte Carlo methods to approximate mean and covariance of arrival cost[2]



1. Rao, CV; Rawlings, JB & Lee, JH; Automatica, 2001, 37, 1619-1628
2. **Lopez-Negrete, R**, Patwardhan, SC & Biegler, LT, J. Process Control, doi:10.1016/j.jprocont.2011.03.004 (2011)

56