Large-Scale Nonlinear Parameter Estimation with IPOPT

Victor M. Zavala & Lorenz T. Biegler

Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

INFORMS Annual Meeting
November 5th 2006

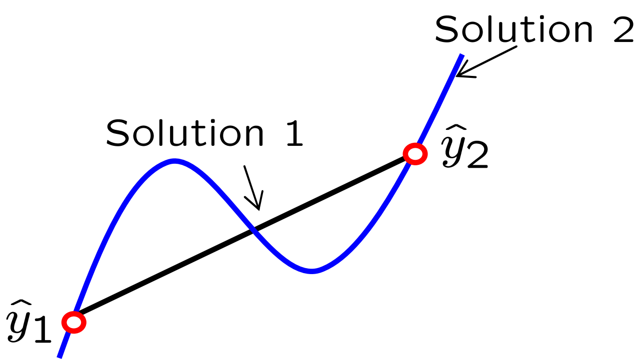
Background

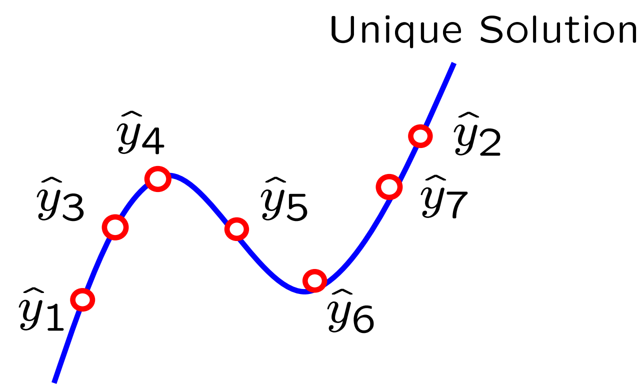
Parameter Estimation

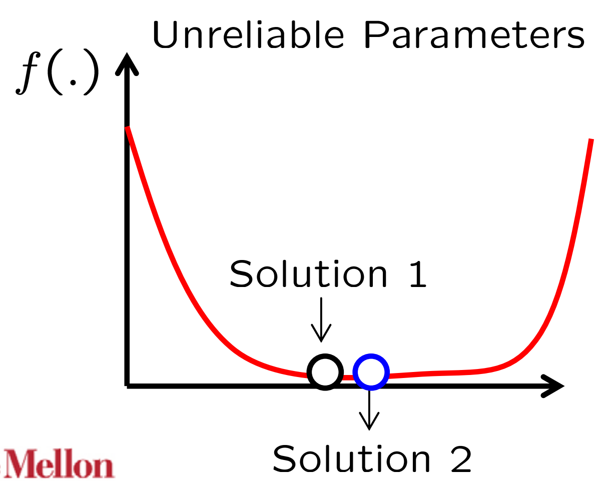
$$\min \quad f(.) = \sum_{k=1}^{NS} (y_k - y_k^M)^2$$

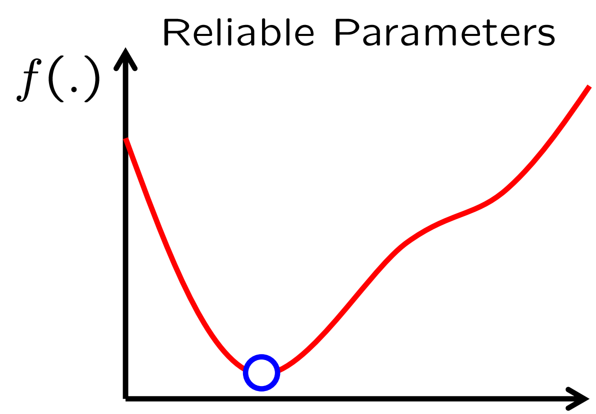
$$y_k = \pi_1 + \pi_2\, x_k + \pi_3\, x_k^2 + \pi_4\, x_k^3$$

$$k = 1, ..., NS$$

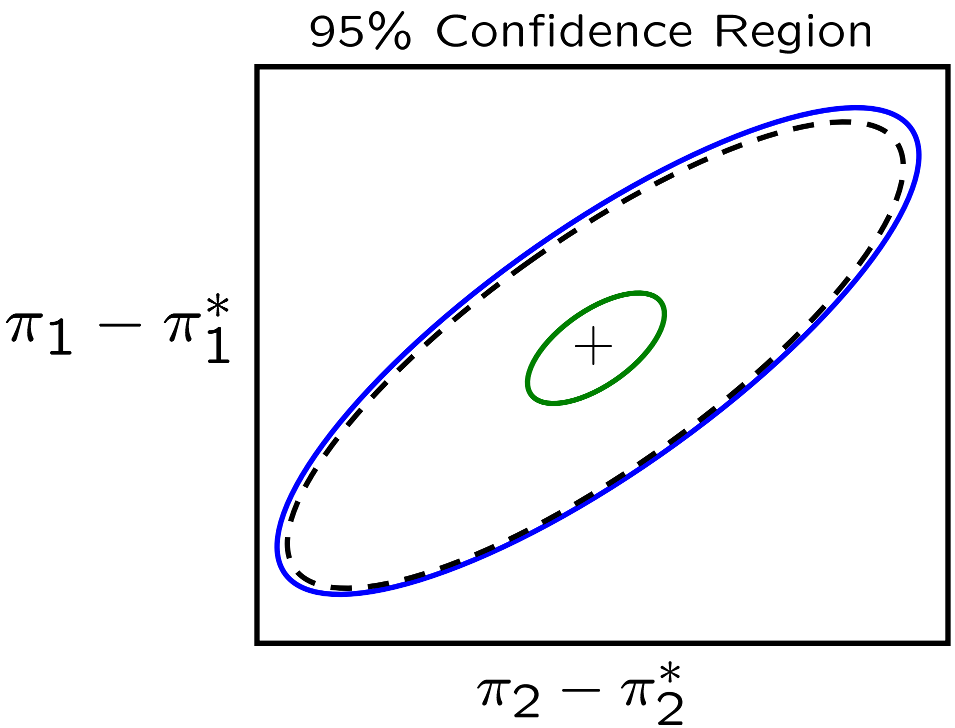


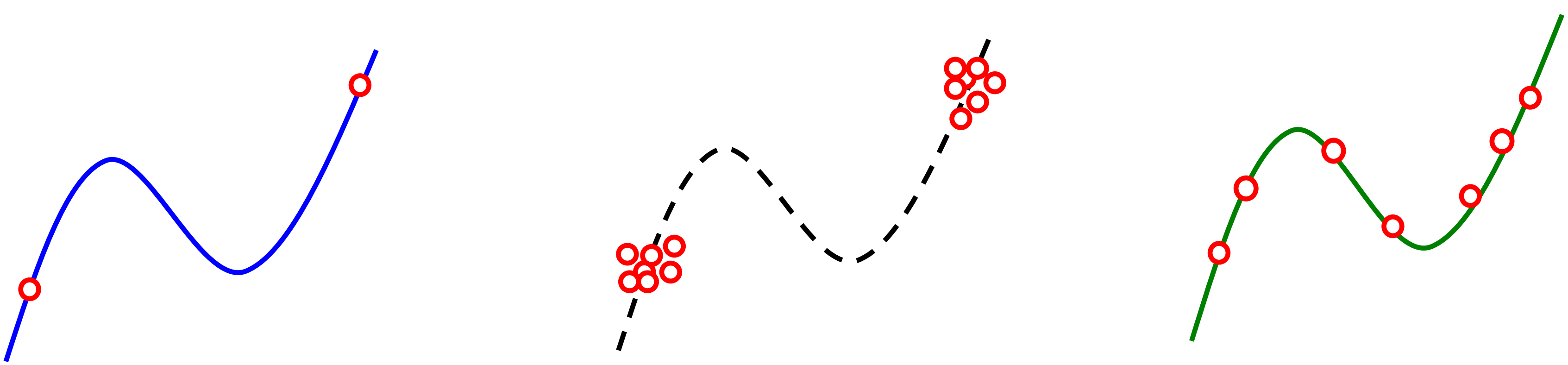






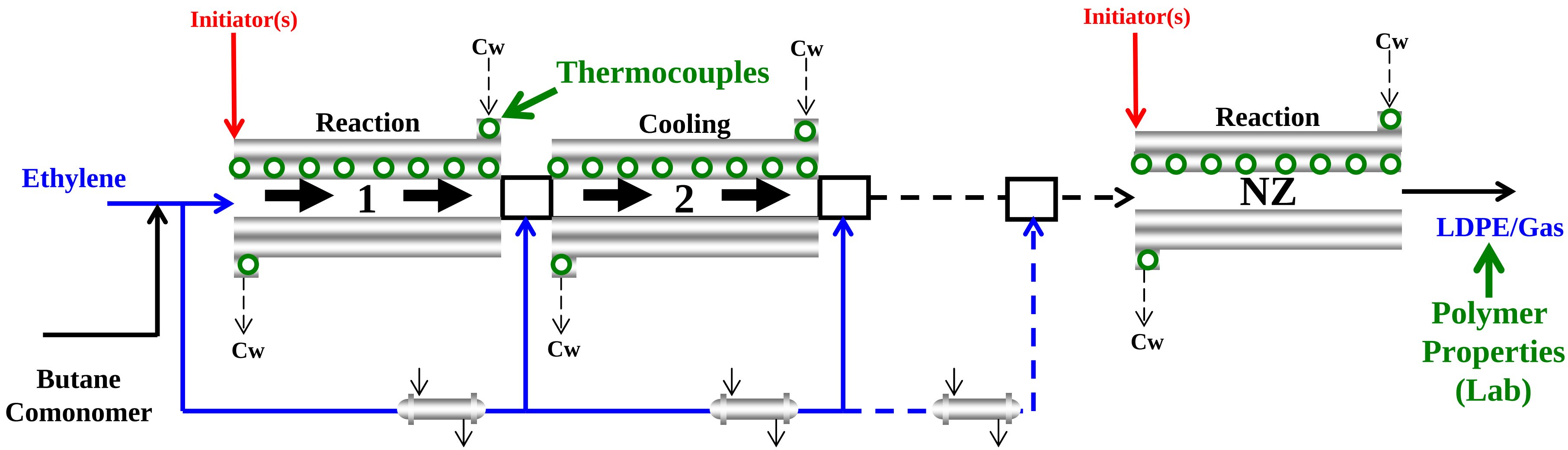
Confidence Regions





Rigorous Models in Process Engineering

High-Pressure LDPE Tubular Reactors



On-line Parameters   Kinetic Parameters

Material & Energy

$$\mathbf{F}_j\left[\frac{dy_j(z)}{dz}, y_j(z), w_j(z), z, \pi_j, \Pi\right] = 0$$

Physical Properties

$$\mathbf{G}_j\left[y_j(z), w_j(z), z, \pi_j, \Pi\right] = 0$$

Zone Transitions

$$y_j(0) = \phi(y_{j-1}(z_{L_{j-1}}), F_{f_j})$$

$$j \in \{1..NZ\}$$

500 ODEs

1000 AEs

× **Stiffness + Highly Nonlinear + Parametric Sensitivity + Algebraic Coupling**

High-Pressure LDPE Tubular Reactors

Complex Kinetic Mechanisms

~ 35 Elementary Reactions
~100 Kinetic Parameters

$$k_i = k_i^0 \exp\left[-\frac{\Delta E_{a_i} + P\Delta E_{v_i}}{RT}\right]$$

Initiator decomposition

$$I_i \xrightarrow{k_{d_i}} 2R \quad i = 1, N_I$$

Chain Initiation

$$R^{\cdot} + M_1 \xrightarrow{k_{I_1}} P_{1,0}$$
$$R^{\cdot} + M_2 \xrightarrow{k_{I_2}} Q_{0,1}$$

Chain Propagation

$$P_{r,s} + M_1 \xrightarrow{k_{p11}} P_{r+1,s}$$
$$P_{r,s} + M_2 \xrightarrow{k_{p12}} Q_{r,s+1}$$
$$Q_{r,s} + M_1 \xrightarrow{k_{p21}} P_{r+1,s}$$
$$Q_{r,s} + M_2 \xrightarrow{k_{p22}} Q_{r,s+1}$$

Chain Transfer to Monomer

$$P_{r,s} + M_1 \xrightarrow{k_{fm11}} P_{1,0} + M_{r,s}$$
$$P_{r,s} + M_2 \xrightarrow{k_{fm12}} Q_{0,1} + M_{r,s}$$
$$Q_{r,s} + M_1 \xrightarrow{k_{fm21}} P_{1,0} + M_{r,s}$$
$$Q_{r,s} + M_2 \xrightarrow{k_{fm22}} Q_{0,1} + M_{r,s}$$

Chain Transfer to Solvent

$$P_{r,s} + S_i \xrightarrow{k_{si1}} P_{1,0} + M_{r,s}$$
$$Q_{r,s} + S_i \xrightarrow{k_{si2}} Q_{0,1} + M_{r,s}$$

Chain Transfer to Polymer

$$P_{r,s} + M_{x,y} \xrightarrow{k_{fp11}} P_{x,y} + M_{r,s}$$
$$P_{r,s} + M_{x,y} \xrightarrow{k_{fp12}} Q_{x,y} + M_{r,s}$$
$$Q_{r,s} + M_{x,y} \xrightarrow{k_{fp21}} P_{x,y} + M_{r,s}$$
$$Q_{r,s} + M_{x,y} \xrightarrow{k_{fp22}} Q_{x,y} + M_{r,s}$$

Termination by Combination

$$P_{r,s} + P_{x,y} \xrightarrow{k_{tc11}} M_{r+x,s+y}$$
$$P_{r,s} + Q_{x,y} \xrightarrow{k_{tc12}} M_{r+x,s+y}$$
$$Q_{r,s} + Q_{x,y} \xrightarrow{k_{tc22}} M_{r+x,s+y}$$

Termination by Disproportionation

$$P_{r,s} + P_{x,y} \xrightarrow{k_{td11}} M_{r,s} + M_{x,y}$$
$$P_{r,s} + Q_{x,y} \xrightarrow{k_{td12}} M_{r,s} + M_{x,y}$$
$$Q_{r,s} + Q_{x,y} \xrightarrow{k_{td22}} M_{r,s} + M_{x,y}$$

Backbitting

$$P_{r,s} \xrightarrow{k_{b1}} P_{r,s} \ or \ Q_{r,s}$$
$$P_{r,s} \xrightarrow{k_{b2}} Q_{r,s} \ or \ P_{r,s}$$

β-scission

$$P_{r,s} \xrightarrow{k_{\beta 1}} M_{r,s}^{=} + P_{1,0}$$
$$P_{r,s} \xrightarrow{k_{\beta 2}} M_{r,s}^{=} + Q_{0,1}$$

Large-Scale Parameter Estimation

Parameter Estimation Problem

- Maximum Likelihood Estimation

$$\min_{\Pi, \pi_{k,j}} \sum_{k=1}^{NS} \sum_{j=1}^{NZ} \sum_{i=1}^{NM(j)} \left(y_{k,j}(z_i) - y_{k,j,i}^{M}\right)^T \mathbf{V}_{\mathbf{y}}^{-1} \left(y_{k,j}(z_i) - y_{k,j,i}^{M}\right) + \sum_{k=1}^{NS} \left(w_{k,NZ} - w_{k,NZ}^{M}\right)^T \mathbf{V}_{\mathbf{w}}^{-1} \left(w_{k,NZ} - w_{k,NZ}^{M}\right)$$

Standard Least-Squares

$$s.t.$$

$$\mathbf{F}_{k,j}\left[\frac{dy_{k,j}(z)}{dz}, y_{k,j}(z), w_{k,j}(z), z, \pi_{k,j}, \Pi\right] = 0$$

$$\mathbf{G}_{k,j}\left[y_{k,j}(z), w_{k,j}(z), z, \pi_{k,j}, \Pi\right] \; 0$$

$$y_{k,j}(0) = \phi(y_{k,j-1}(z_{L_{k,j-1}}), F_{f_{k,j}})$$

$$j \in \{1..NZ\}, \quad k \in \{1..NS\}$$

Rigorous DAE Model

1 data set		5 data sets
500 ODEs	x 5 →	2500 ODEs
1000 AEs		5000 AEs

- Previously Intractable Problems

  - Full Discretization Approach → Large-Scale NLPs

Solution of Large-Scale NLP

- *Robust* General Algorithms -IPOPT- vs. *Ad-Hoc* Estimation Algorithms

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0 \\ & x \geq 0 \end{aligned} \quad \text{NLP} \quad \longrightarrow \quad \begin{aligned} \min \quad & \varphi_{\mu_\ell}(x) = f(x) - \mu_\ell \sum_{i=1}^{nx} \ln\left(x^{(i)}\right) \\ \text{s.t.} \quad & c(x) = 0 \end{aligned} \quad \text{BP}$$

Main Idea: Solve Sequence of BPs with $\mu_\ell \to 0$

Factorization of KKT Matrix -MA27+MC19-

$$\begin{aligned} \nabla_x f(x) + \nabla_x c(x)\,\lambda - \nu &= 0 \\ c(x) &= 0 \\ XVe - \mu_\ell\, e &= 0 \end{aligned} \quad \xrightarrow[\text{Exact Derivatives}]{\text{Newton's Method}} \quad \begin{bmatrix} H_k & A_k & -I \\ A_k^T & 0 & 0 \\ V_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) + A_k \lambda_k - \nu_k \\ c(x_k) \\ X_k V_k e - \mu_\ell e \end{bmatrix}$$

- Non-Convex, Ill-Posed – Negative Curvature- Filter Line-Search

– Jacobian Rank-Deficient

Inertia Correction
$$\begin{bmatrix} H_k + \delta_1 I & A_k & -I \\ A_k^T & -\delta_2 I & 0 \\ V_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) + A_k \lambda_k - \nu_k \\ c(x_k) \\ X_k V_k e - \mu_\ell e \end{bmatrix}$$

- IF $\delta_1, \delta_2 = 0$ at Solution – Parameters Uniquely Determined

Results – Standard Least-Squares

Sparsity Structure

Fill-In

- Single Data Set (On-line Parameters)

Grade	Constraints	Parameters	LB	UB	Iterations	CPUs	NZJ	NZH
A	11955	32	374	361	11	17.03	166425	87954
B	11283	32	374	361	8	10.06	138666	76890

On-Line Estimation

- Multiple Data Sets (On-line Parameters + Kinetics)

Data Sets	Constraints	DOF	LB	UB	Iterations	CPUs	NZJ	NZH
3	33900	121	1246	1207	68	451.51	520275	552738
6	68421	217	2467	2389	58	900.21	1058412	1119258

Bottleneck -Memory Requirements-
Factorization KKT Matrix

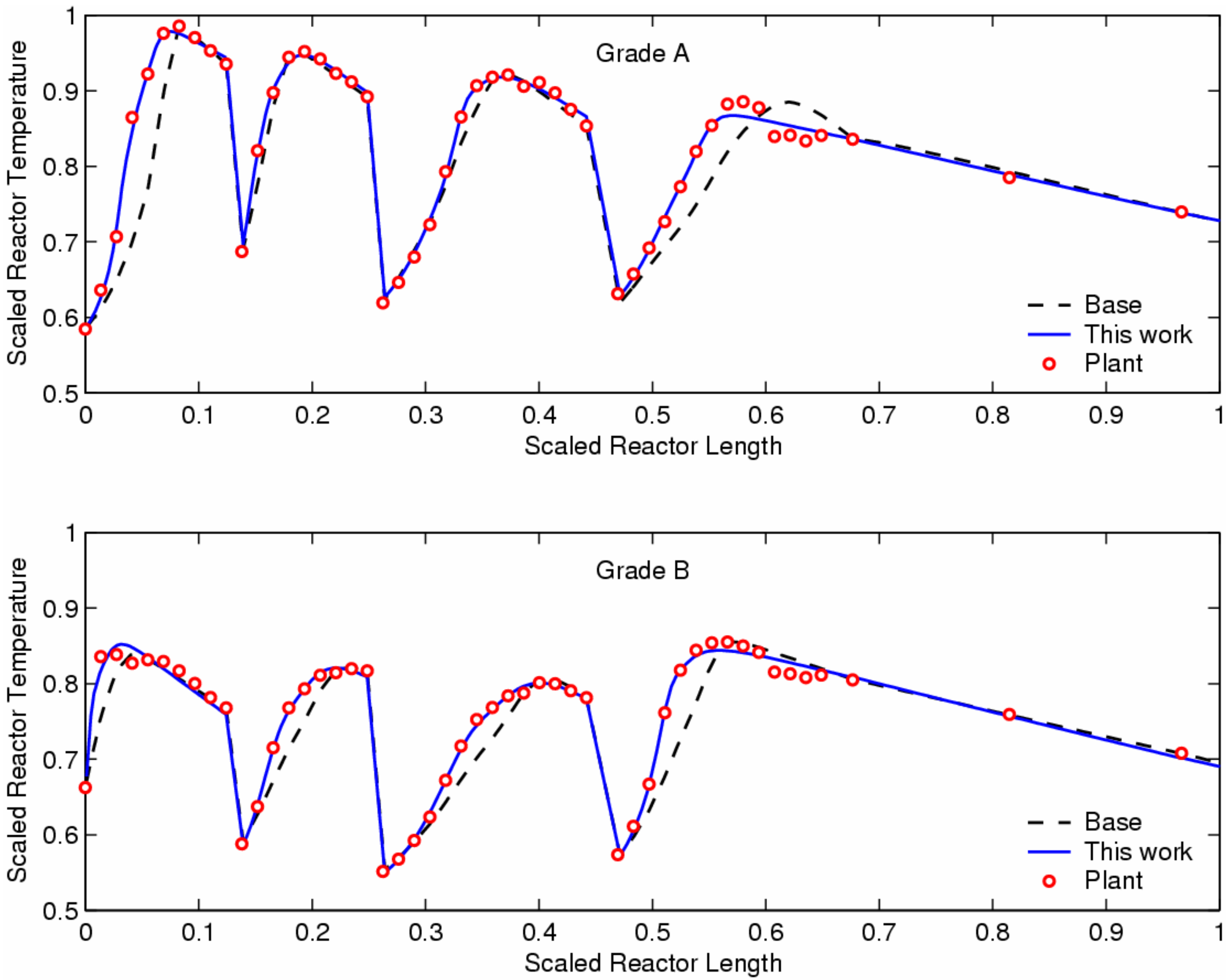
$$\underbrace{\begin{bmatrix} H_k+\delta_1 I & A_k & -I \\ A_k^T & -\delta_2 I & 0 \\ V_k & 0 & X_k \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k)+A_k\lambda_k-\nu_k \\ c(x_k) \\ X_k V_k e-\mu_\ell e \end{bmatrix}}$$

Tailored Linear Algebra

$$\begin{bmatrix} \mathbf{K}_1 & & & Q_1 \\ & \mathbf{K}_2 & & Q_2 \\ & & \ddots & \vdots \\ Q_1^T & Q_2^T & \dots & D_\Pi \end{bmatrix}\begin{bmatrix} \Delta s_1 \\ \Delta s_2 \\ \vdots \\ \Delta\Pi \end{bmatrix} = -\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_\Pi \end{bmatrix}$$

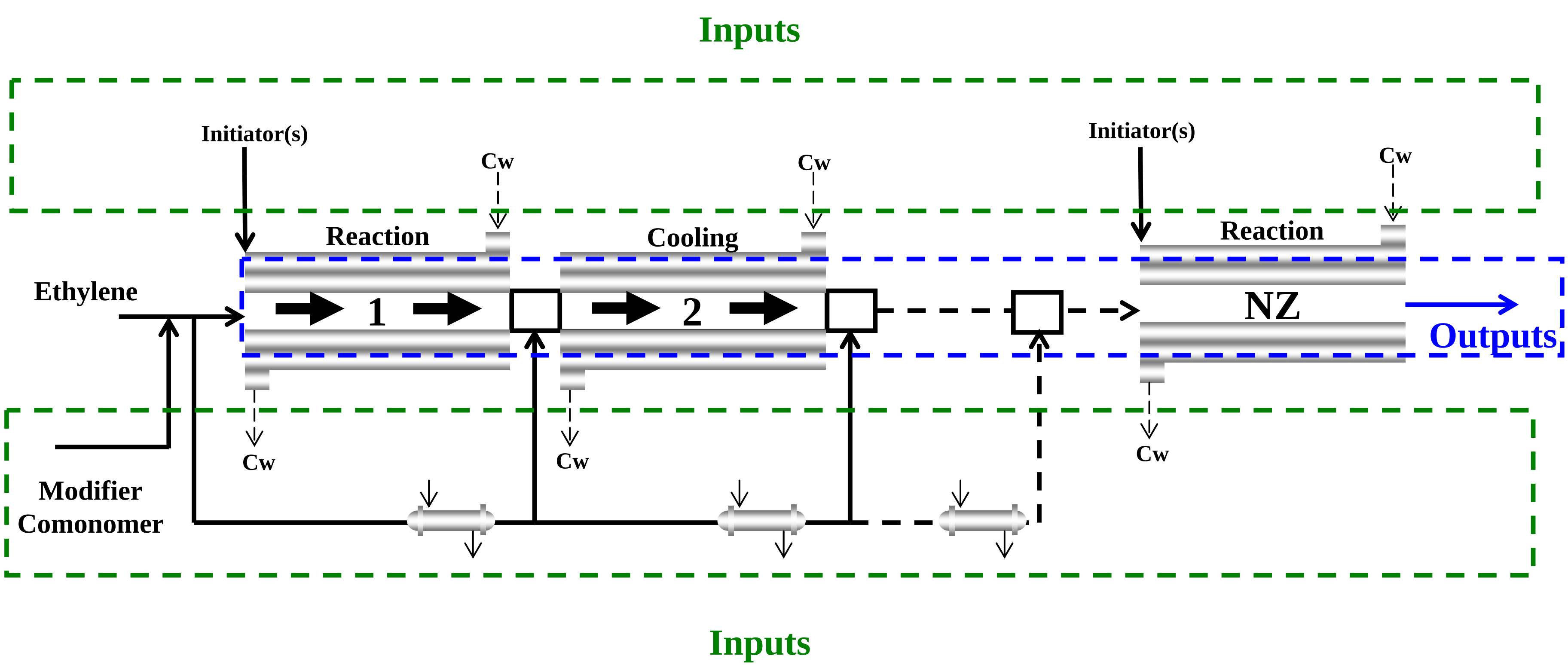
Results – Standard Least-Squares

Improved Match of Reactor Temperature Profile → On-line Estimation



Errors-In-Variables (EVM)

- Standard Least Squares - Errors in Output Variables - *Biased* Parameters
- EVM - Errors in Output *AND* Input Variables - *Unbiased* Parameters



Errors-In-Variables (EVM)

EVM Drawback - Degrees of Freedom

$$DOF = \Pi + \sum_{k-1}^{NS}\sum_{j=1}^{NZ}\pi_{k,j} + \sum_{k=1}^{NS}\sum_{j=1}^{NZ}\sum_{i=1}^{NM_u(j)} u_{k,j,i}$$

Formulation is Straightforward

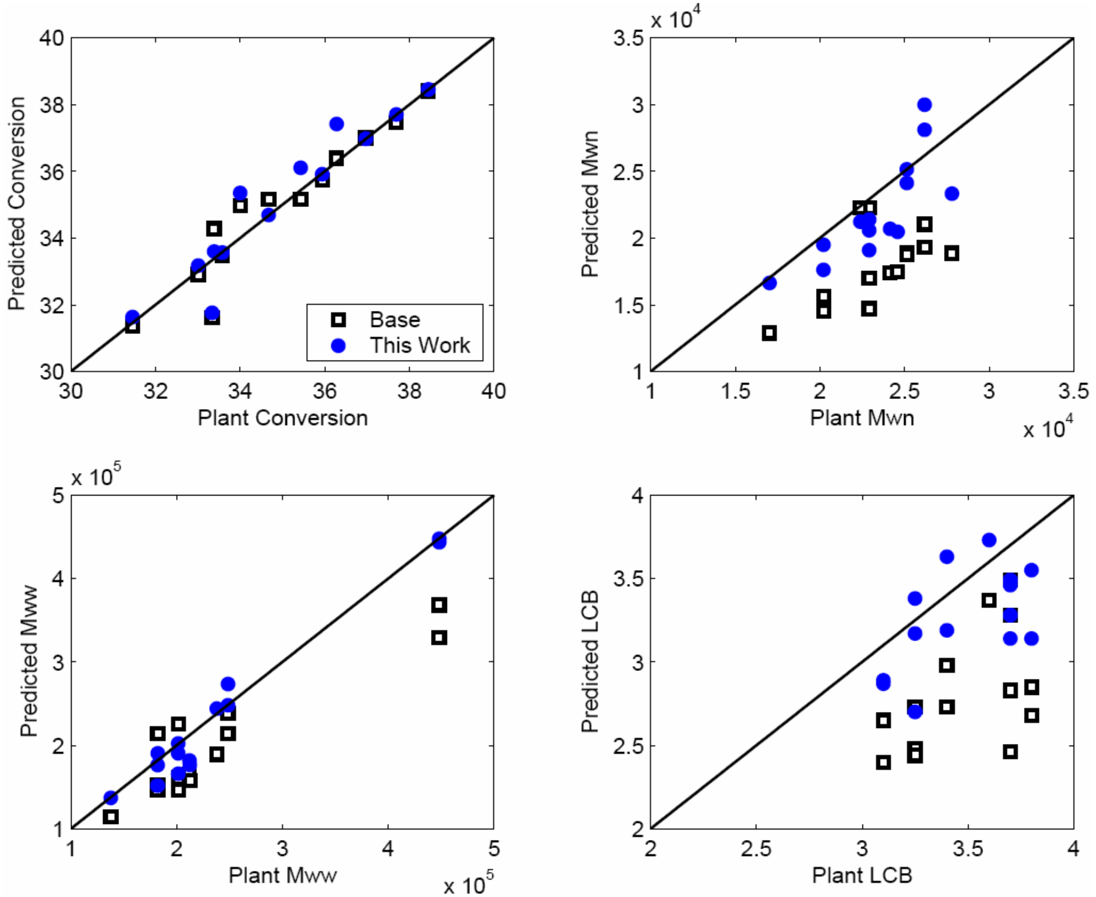
$$\min_{\Pi,\,\pi_{k,j},\,u_{k,j}} \sum_{k=1}^{NS}\sum_{j=1}^{NZ}\sum_{i=1}^{NM(j)} \left(y_{k,j}(z_i) - y^M_{k,j,i}\right)^T \mathbf{V}_\mathbf{y}^{-1} \left(y_{k,j}(z_i) - y^M_{k,j,i}\right)$$
$$+ \sum_{k=1}^{NS} \left(w_{k,NZ} - w^M_{k,NZ}\right)^T \mathbf{V}_\mathbf{w}^{-1} \left(w_{k,NZ} - w^M_{k,NZ}\right)$$
$$+ \sum_{k=1}^{NS}\sum_{j=1}^{NZ} \left(u_{k,j} - u^M_{k,j}\right)^T \mathbf{V}_\mathbf{u}^{-1} \left(u_{k,j} - u^M_{k,j}\right)$$

EVM vs. Standard Least Squares

Data Sets	Constraints	DOF	LB	UB	Iterations	CPUs	NZJ	NZH
6 (EVM)	68627	529	2653	2575	71	1010.74	1059512	1119780
6 (SLS)	68421	217	2467	2389	58	900.21	1058412	1119258

Results – Advanced Regression Methods

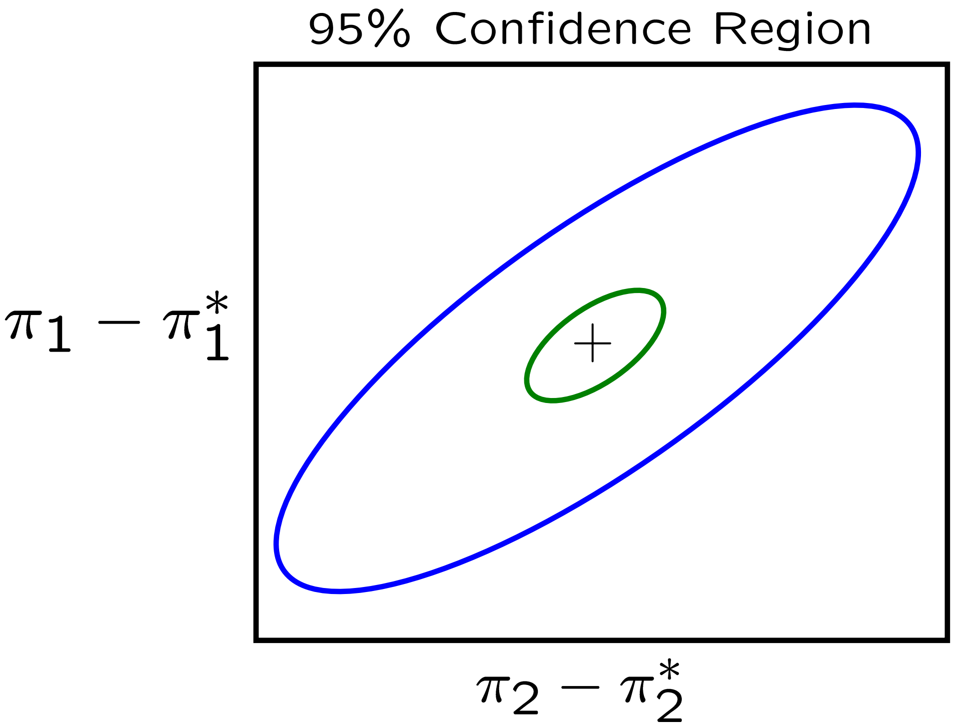
Improved Prediction of Polymer Molecular Properties



Average Deviation - 14 Different Data Sets

	Conversion(%)	Mwn(%)	Mww(%)	LCB(%)	Density(%)
Industrial Model	1.49	23.24	18.58	19.20	0.0965
This Work	0.12	6.20	3.31	6.27	0.0875

Post-Optimal Analysis

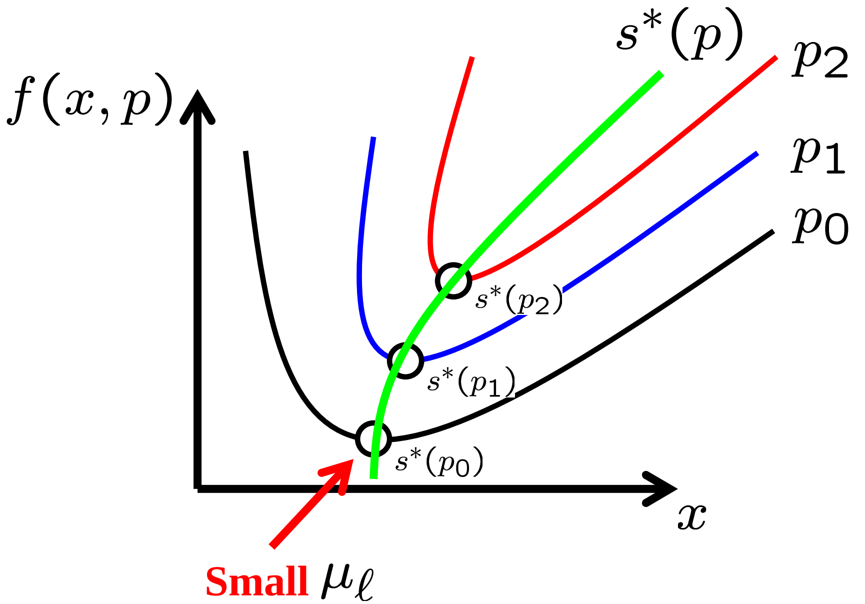


$$(\pi - \pi^*)^T V_\pi^{-1} (\pi - \pi^*) \leq \chi^2(\alpha, n_\pi)$$

Sensitivity Information

$$V_\pi^{-1} \approx \frac{\partial \pi}{\partial y^M}^T V_y^{-1} \frac{\partial \pi}{\partial y^M}$$

$$\left.\begin{array}{ll} \text{min} & f(x,p) \\ \text{s.t.} & c(x,p) = 0 \\ & x \geq 0 \end{array}\right\} \mathrm{P}(p)$$



Solution Triplet

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

$$\downarrow \qquad \downarrow$$

$$\hat{y}^M \qquad \pi_j,\ \Pi$$

Optimality Conditions of $\mathrm{P}(p)$

$$\begin{aligned} \nabla_x f(x,p) + \nabla_x c(x,p)\,\lambda - \nu &= 0 \\ c(x,p) &= 0 \\ XVe &= 0 \end{aligned}$$

NLP Sensitivity - Existence and Differentiability of Path $s^*(p)$ - *Fiacco, 1982*

- $\left.\frac{\partial s}{\partial p}\right|_{p_0}$ **Exists and is Unique**

Post-Optimal Analysis - NLP Sensitivity

Obtaining $\left.\dfrac{\partial s}{\partial p}\right|_{p_0}$

Optimality Conditions of $\mathrm{P}(p)$

$$\left.\begin{aligned} \nabla_x \mathcal{L} = \nabla_x f(x,p) + \nabla_x c(x,p)\,\lambda - \nu &= 0 \\ c(x,p) &= 0 \\ XVe &= 0 \end{aligned}\right\} \quad \mathbf{F}(s,p) = 0$$

Apply Implicit Function Theorem to $\mathbf{F}(s,p) = 0$ **around** $(p_0, s^*(p_0))$

$$\frac{\partial \mathbf{F}(s^*(p_0),p_0)}{\partial s} \left.\frac{\partial s}{\partial p}\right|_{p_0} + \frac{\partial \mathbf{F}(s^*(p_0),p_0)}{\partial p} = 0$$

$$\begin{bmatrix} W(s^*(p_0)) & A(x^*(p_0)) & -I \\ A(x^*(p_0))^T & 0 & 0 \\ V^*(p_0) & 0 & X^*(p_0) \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial p} \\ \frac{\partial \lambda}{\partial p} \\ \frac{\partial \nu}{\partial p} \end{bmatrix} + \begin{bmatrix} \nabla_{x,p} \mathcal{L}(s^*(p_0)) \\ \nabla_p\, c(x^*(p_0)) \\ 0 \end{bmatrix} = 0$$

KKT Matrix IPOPT

$$\begin{bmatrix} W(x_k, \lambda_k) & A(x_k) & -I \\ A(x_k)^T & 0 & 0 \\ V_k & 0 & X_k \end{bmatrix}$$

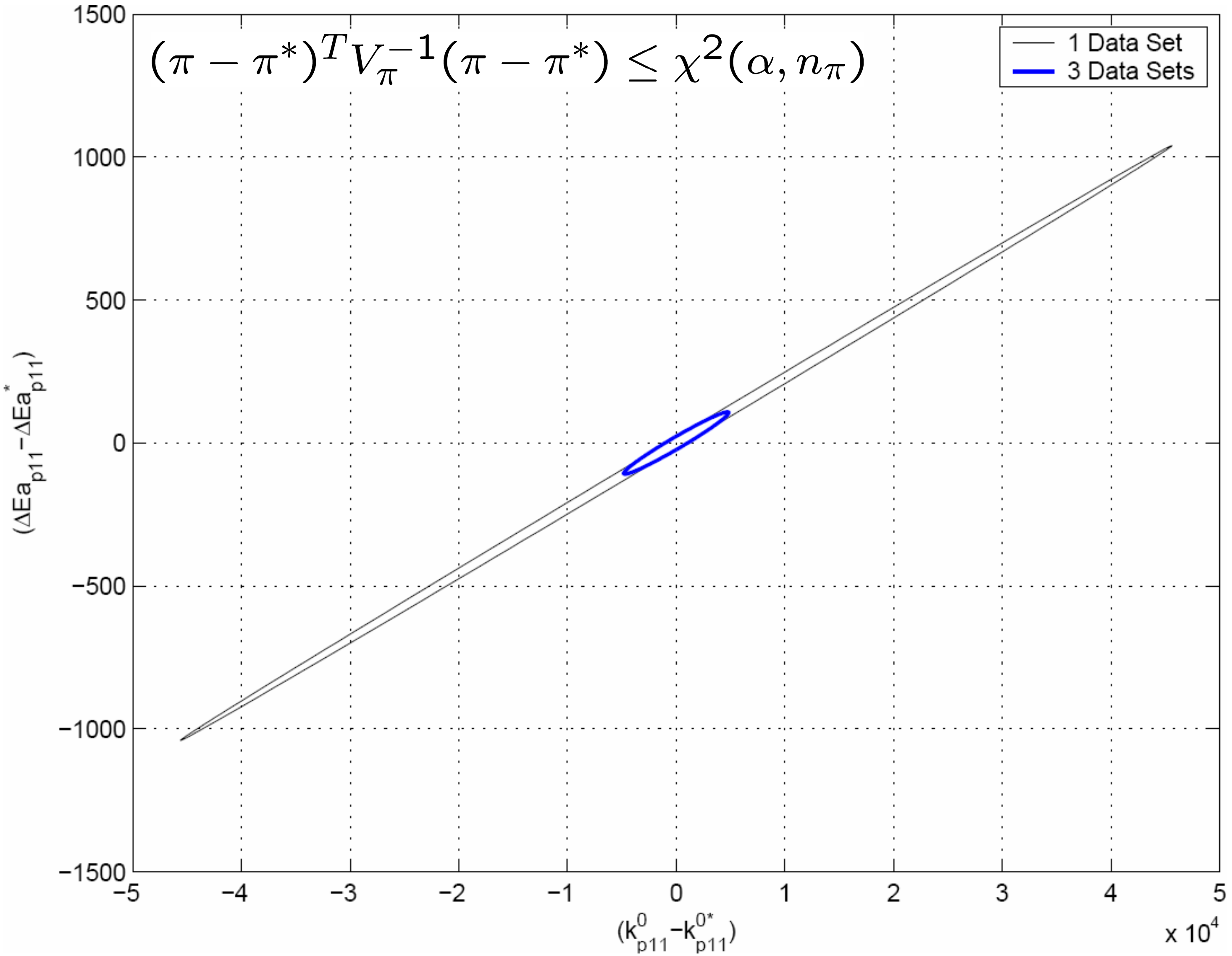
- Already Factored at Solution

- Cheap Sensitivity Calculations

- Confidence Regions $V_\pi^{-1} \approx \dfrac{\partial \pi}{\partial y^M}^T V_y^{-1} \dfrac{\partial \pi}{\partial y^M}$

Post-Optimal Analysis

Confidence Regions - Reliability of Parameters



Summary and Conclusions

Summary and Conclusions

Parameter Estimation Fundamental in Science and Engineering

- **Large and Complicated Optimization Problems**

Robust and Comprehensive Optimization Algorithms -IPOPT-

- **Cope with General Complexity**
- **Adapt to Problem Specifics**

Solve Previously Intractable Estimation Problems

- **Highly Nonlinear, Many Degrees of Freedom**
- **Advanced Regression Methods**

Nonlinear Estimation Problems - Globalization Strategies -

- **Exploit Directions of Negative Curvature *Naturally* - Trust-Region Methods**

Questions