Computational Strategies for Large-Scale Estimation in Dynamic Systems

Victor M. Zavala
Postdoctoral Researcher
Mathematics and Computer Science Division
Argonne National Laboratory
vzavala@mcs.anl.gov

Joint Work with: Larry Biegler (Carnegie Mellon), Carl Laird (Texas A&M)

Eindhoven, Netherlands
June 9th, 2009
Motivation

**Degrees of Freedom**

\[
\begin{align*}
\min & \quad u_k(t), P(t), z_k^0 \\
\text{s.t.} & \quad \sum_{k=1}^{N_s} \int_{t_0}^{t_f} \varphi_k(x_k(t), u_k(t), P(t)) \, dt \\
\frac{dz_k}{dt} & = f_k(z_k(t), y_k(t), u_k(t), P(t)) \\
0 & = g_k(z_k(t), y_k(t), u_k(t), P(t)) \\
\chi_k(t) & = \phi_k(z_k(t), y_k(t), u_k(t), P(t)) \\
0 & \geq h_k(z_k(t), y_k(t), u_k(t), P(t)) \\
z_k(0) & = z_k^0, \quad k = 1, \ldots, N_s
\end{align*}
\]

- **Maximum Likelihood** (Multi-Set)
- **DAEs** (Physical Model)
- **Output Mapping** (Measured)
- **Inequalities** (Physical Limits)
- **Initial Conditions**

**Objective:** - Detailed Models (>1,000 DAEs), Many Experimental Data Sets
- Errors-in-Variables, Inference Analysis
- Fast Solutions (Feedback Control, Profiling Methods)

**Challenge:** - Degrees of Freedom, Computational Expensive DAEs
- Nonlinearity, Ill-Posedness (Observability)
**Outline**

**Message:** Current Optimization Tools Suitable for *Large-Scale* Estimation Tasks

1. **Nonlinear Programming (NLP) Framework**
   - Discretization and Interior Point Solvers
   - Internal Regularization and Observability

2. **Inference Analysis and Parallel Decomposition**
   - Exploiting Linear Algebra

3. **Industrial Case Study**
   - Polyethylene Reactor

4. **Conclusions and Future Work**
1. Nonlinear Programming Framework
NLP Framework

\[
\begin{align*}
\min_{u_k(t), P(t), z_k^0} & \quad \sum_{k=1}^{N_s} \int_{t_0}^{t_f} \varphi_k(x_k(t), u_k(t), P(t)) dt \\
\frac{dz_k}{dt} & = f_k(z_k(t), y_k(t), u_k(t), P(t)) \\
0 & = g_k(z_k(t), y_k(t), u_k(t), P(t)) \\
\chi_k(t) & = \phi_k(z_k(t), y_k(t), u_k(t), P(t)) \\
0 & \geq h_k(z_k(t), y_k(t), u_k(t), P(t)) \\
z_k(0) & = z_k^0, \quad k = 1, ..., N_s
\end{align*}
\]

Traditional Estimation Tools: Simulation-Based, Gauss-Newton (Avoid Second Derivatives)

Limitations: Degrees of Freedom, Boundary Conditions, Unstable Systems

Full Discretization + Sparse NLP Solvers

All-At-Once Solution (Collocation, Splines, Galerkin)

Cheap Exact First and Second Derivatives - Automatic Differentiation (AMPL, ADOL-C)
**NLP Framework**

**IPOPT** Wächter, Biegler 2006

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad c(x) = 0 \\
& \quad x \geq 0 \quad \text{NLP}
\end{align*}
\]

\[
\begin{align*}
\min & \quad f(x) - \mu \sum_{i=1}^{nx} \ln(x^{(i)}) \\
\text{s.t.} & \quad c(x) = 0 \quad \text{Barrier Problem}
\end{align*}
\]

Solve Sequence of BPs with \(\mu \to 0\)

\[
\nabla_x f(x) + \nabla_x c(x) \lambda - \nu = 0 \quad \text{Newton’s Method}
\]

\[
\begin{bmatrix}
H_x & A_j & -I \\
A_j^T & 0 & 0 \\
V_j & 0 & X_j
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \lambda \\
\Delta \nu
\end{bmatrix} = -
\begin{bmatrix}
\nabla f(x_j) + A_j \lambda_j - \nu_j \\
\c(x_j) \\
X_j V_j e - \mu e
\end{bmatrix}
\]

**Advantages of General NLP Solvers**

- **Favorable Complexity** \(O(N(n_z + n_y + n_P))^{1,2} \) vs. \(O((N(n_z + n_y))^2 + (n_P)^3)\)

- **Inequalities, Second Derivatives** (Number of Iterations Not Affected by Degrees of Freedom)

- **Crucial to Exploit Sparsity** (Reduce Time and Memory Requirements)
NLP Framework

KKT Matrix

\[
\begin{bmatrix}
H & A & -I \\
A^T & 0 & 0 \\
V & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \lambda \\
\Delta \nu
\end{bmatrix}
= -
\begin{bmatrix}
\nabla f(x) + A\lambda - \nu \\
c(x) \\
XVe - \mu e
\end{bmatrix}
\]

- General Sparse Linear Solvers - MA57 from Harwell
  - Phase I: Reordering to Preserve Sparsity

Original

Reordered

( Default) Approximate Minimum-Degree Reordering Duff, 1993
  - Local Analysis – Fails at Identifying High-Level Structures

Nested Dissection Reordering Gould, 2004, Karypis, 1999
  - Global Analysis - Identifies High-Level Structures
  - Efficient for DAEs and PDEs Discretizations

Elimination Tree Davis, 2006
2. Inference Analysis and Parallel Decomposition
Inference Analysis

Limitation of General NLP Solvers

Solution Analysis Difficult: Are Parameters Observable? How to Get Posterior Covariance?

Observability Basu & Bresler, 2000

- Parameters Observable IF Can be Inferred Uniquely from Measurements
- Parameters Observable IF NLP Has a Unique Local Minimizer

Context of NLP Solver

- Internal Regularization of Hessian in Presence of Nonconvexity (as Levenberg-Marquardt)

\[
\begin{bmatrix}
H_j + \delta_R & A_j & -I \\
A_j^T & 0 & 0 \\
V_j & 0 & X_j
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \lambda \\
\Delta \nu
\end{bmatrix} =
\begin{bmatrix}
\nabla f(x_j) + A_j \lambda_j - \nu_j \\
0 \\
X_j V_j e - \mu e
\end{bmatrix}
\]

Strong Second-Order Conditions Nocedal & Wright, 1999

- \( \delta_R = 0 \) at Solution Implies Local Isolated Minimizer (Needs Second Derivatives!)
Inference Analysis

Consider Formulation

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad c(x) = 0 \\
& \quad x \geq 0
\end{align*}
\]

\[
\begin{align*}
\min & \quad f(\bar{x}, d) \\
\text{s.t.} & \quad c(\bar{x}, d) = 0 \\
& \quad \bar{x}, d \geq 0
\end{align*}
\]

**Reduced Hessian** \( Z^T H Z \) **Hessian Projected onto Null Space of Constraints** \( Z = \begin{bmatrix} -A_x^{-1} A_d \\ I_d \end{bmatrix} \)

**Approximation of Parameters Posterior Covariance** \( V_d \) **Bard, 1974, Generalized**

\[
\begin{bmatrix}
A_x \\
A_d \\
A_d \\
I_{n_d}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\bar{X} \\
D \\
\Lambda
\end{bmatrix} = -\begin{bmatrix} 0 \\ I_{n_d} \\ 0 \end{bmatrix}
\]

**Theorem:** \( D = (Z^T H Z)^{-1} \) **Zavala & Biegler, 2007**

Large-Scale Inference Analysis Possible using KKT Matrix (Available AMPL and C++ Interface)
Toy Problem

Fit Third Order Polynomial to Increasing Number of Measurements
- Unique solution only for $N_s \geq 3$, Obtain Confidence Regions from IPOPT

\[
\begin{align*}
\min \quad & \sum_{k=1}^{N_s} (y_k - \hat{y}_k)^2 \\
y_k &= P_1 x_k + P_2 x_k^2 + P_3 x_k^3 \\
k &= 1, \ldots, N_s
\end{align*}
\]

AMPL Model

```
option solver ipopt;
option ipopt_options "compute_reduced_hessian yes";
suffix ipopt_idx_red_hessian, IN, integer;
param nsets:=2;
var P1,P2,P3;
let P1.ipopt_idx_red_hessian :=1;
let P2.ipopt_idx_red_hessian :=2;
let P3.ipopt_idx_red_hessian :=3;

minimize obj: sum{k in 1..nsets}(y[k]-yhat[k])^2;
s.t. cons{k in 1..nsets}: y[k] = P1*x[k] + P2*x[k]^2 + P3*x[k]^3;
solve;
```
**Toy Problem**

**IPOPT Output** \( N_s = 2 \)

| Number of nonzero in equality constraint Jacobian... | 5 |
| Number of nonzero in inequality constraint Jacobian | 0 |
| Number of nonzero in Lagrangian Hessian...        | 2 |
| Total number of variables...                       | 5 |
| variables with only lower bounds                  | 0 |
| variables with only upper bounds                  | 0 |
| Total number of equality constraints...           | 2 |
| Total number of inequality constraints...         | 0 |
| inequality constraints with lower and upper bounds| 0 |
| inequality constraints with only lower bounds     | 0 |
| inequality constraints with only upper bounds     | 0 |

<table>
<thead>
<tr>
<th>iter</th>
<th>objective</th>
<th>inf_pr</th>
<th>inf_du</th>
<th>lg(mu)</th>
<th>lssrg</th>
<th>alpha_du</th>
<th>alpha_pr</th>
<th>ls_c</th>
<th>ls_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.8272221e-002</td>
<td>1.0000000</td>
<td>8.68e+000</td>
<td>-1.0</td>
<td>0.00e+40</td>
<td>0.0000000</td>
<td>0.00e+000</td>
<td>0.0000000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.3340000e-004</td>
<td>8.38e-017</td>
<td>1.12e-005</td>
<td>-1.0</td>
<td>1.00e+000</td>
<td>-1.00e+000</td>
<td>1.00e+000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.1340000e-004</td>
<td>7.81e-019</td>
<td>3.41e-007</td>
<td>-5.7</td>
<td>1.00e-014</td>
<td>-4.9e+000</td>
<td>1.00e+000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of iterations...: 2

\( \delta_R > 0 \)

Confidence Regions \( N_s \geq 3 \)

- **Unobservable Parameters!**
- **Solution 1**
- **Solution 2**

\[ P_2 - P_2^* \]

\[ P_3 - P_3^* \]
Parallel Decomposition

Multi-Set Estimation Problem

\[ \min_{u_k(t), P(t), z_k^0} \sum_{k=1}^{N_s} \int_{t_0}^{t_f} \varphi_k(x_k(t), u_k(t), P(t)) dt \]

\[ \frac{dz_k}{dt} = f_k(z_k(t), y_k(t), u_k(t), P(t)) \]

\[ 0 = g_k(z_k(t), y_k(t), u_k(t), P(t)) \]

\[ x_k(t) = \phi_k(z_k(t), y_k(t), u_k(t), P(t)) \]

\[ 0 \geq h_k(z_k(t), y_k(t), u_k(t), P(t)) \]

\[ z_k(0) = z_k^0, \quad k = 1, \ldots, N_s \]

\[ \min f(x) \]

s.t. \[ c(x) = 0 \]

\[ x \geq 0 \]

General Optimization

Multi-Scenario Optimization

Direct Factorization
Reaches Memory Bottleneck
Factorization Time Increases (Linearly)

Tailored Factorization, Laird & Biegler 2006
Distribute Over Parallel Processors
Factorization Time Constant
3. Industrial Case Study
Low-Density Polyethylene

- **Polymerization at High Pressures** (2000 atm, Complex Transport and Kinetic Effects)
- **Persistent Fouling** (Spatio-Temporal Disturbance)
- **Wide Operating Regions** (> 20 Grades)
Low-Density Polyethylene

\[
\begin{cases}
\frac{\partial z}{\partial t} + \nu(x,t) \frac{\partial z}{\partial x} = f(z(x,t), w(x,t), u(x,t), P(x,t)) & \text{Conservation Mass, Energy, Momentum} \\
0 = g(z(x,t), w(x,t), u(x,t), P(x,t)) & \text{Thermodynamic & Transport Properties} \\
\chi(x,t) = \phi(z(x,t), w(x,t), u(x,t), P(x,t)) & \text{Output Mapping} \\
z(x,0) = z_0(x) & \text{Initial Conditions} \\
0 = B \left( z(t,0), z(t,x^L), \frac{\partial z}{\partial x}(t,0), \frac{\partial z}{\partial x}(t,x^L), w(t,0), u(x,t) \right) & \text{Boundary Conditions}
\end{cases}
\]

Large Set of Partial Differential and Algebraic Equations (PDAEs)
Discretize in Space - 10,000 DAEs in Time
**Polymerization Kinetic Mechanism**

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

\(~ 35 \text{ Elementary Reactions}~

**Kinetic Parameters Under Controlled Laboratory Conditions**

- **Initiator decomposition**
  \[ I_i \xrightarrow{k_{di}} 2R \quad i = 1, N_I \]

- **Chain Initiation**
  \[ R^* + M_1 \xrightarrow{k_{i1}} P_{1,0} \]
  \[ R^* + M_2 \xrightarrow{k_{i2}} Q_{0,1} \]

- **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 Polymer**
  \[ P_{r,s} + M_{x,y} \xrightarrow{k_{tp11}} P_{x,y} + M_{r,s} \]
  \[ P_{r,s} + Q_{x,y} \xrightarrow{k_{tp12}} Q_{x,y} + M_{r,s} \]
  \[ Q_{r,s} + M_{x,y} \xrightarrow{k_{tp21}} P_{x,y} + M_{r,s} \]
  \[ Q_{r,s} + Q_{x,y} \xrightarrow{k_{tp22}} Q_{x,y} + M_{r,s} \]

- **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**
  \[ P_{r,s} + P_{x,y} \xrightarrow{k_{td11}} M_{r+s} + M_{x,y} \]
  \[ P_{r,s} + Q_{x,y} \xrightarrow{k_{td12}} M_{r+s} + M_{x,y} \]
  \[ Q_{r,s} + Q_{x,y} \xrightarrow{k_{td22}} M_{r+s} + M_{x,y} \]

- **Chain Transfer to Monomer**
  \[ P_{r,s} \xrightarrow{k_{tm11}} P_{1,0} + M_{r,s} \]
  \[ P_{r,s} \xrightarrow{k_{tm12}} Q_{0,1} + M_{r,s} \]
  \[ Q_{r,s} \xrightarrow{k_{tm21}} P_{1,0} + M_{r,s} \]
  \[ Q_{r,s} \xrightarrow{k_{tm22}} Q_{0,1} + M_{r,s} \]

- **\(\beta\)-scission**
  \[ P_{r,s} \xrightarrow{k_{b1}} P_{r,s} \quad \text{or} \quad Q_{r,s} \]
  \[ P_{r,s} \xrightarrow{k_{b2}} Q_{r,s} \quad \text{or} \quad P_{r,s} \]

- **\(\beta\)-scission**
  \[ P_{r,s} \xrightarrow{k_{g1}} M_{r,s} = P_{1,0} \]
  \[ P_{r,s} \xrightarrow{k_{g2}} M_{r,s} + Q_{0,1} \]
Low-Density Polyethylene

Estimation with Industrial Process Data

From Industrial Collaborator (Steady-State Data Sets)

**Output Measurements** (Temperatures at Discrete Positions, Laboratory Polymer Properties)

**Input Measurements** (Inlet Temperatures, Flow rates, etc.)

**Off-Line Estimation Objective**

Refit **Kinetic** Parameters and Disturbances (e.g. Heat-Transfer Coefficients)
Low-Density Polyethylene

Off-Line Estimation Problem

\[
\begin{align*}
\min_{u_k(x), P(x)} & \quad \sum_{k=1}^{N_s} \int_{x_0}^{x_L} \varphi_k(z_k(x), y_k(x), u_k(x), P(x)) \, dx \\
\frac{dz_k}{dx} &= f_k(z_k(x), y_k(x), u_k(x), P(x)) \\
0 &= g_k(z_k(x), y_k(x), u_k(x), P(x)) \\
\chi_k(x) &= \phi_k(z_k(x), y_k(x), u_k(x), P(x)) \\
0 &\geq h_k(z_k(x), y_k(x), u_k(x), P(x)) \\
z_k(0) &= z_k^0, \quad k = 1, \ldots, N_s
\end{align*}
\]

Least-Squares Objective

Steady-State Reactor Model

1) Standard Least-Squares (SLS) - Parameters (Do not Correct Noisy Inputs) -Biased-

2) Errors-in-Variables Measured (EVM) - Parameters (Correct Noisy Inputs) –UnBiased-

**NLP Size** with Number of Data Sets

1 Data Set 10 Data Sets

10,000 Constraints \(\times 10\) 100,000 Constraints

Large Number of Degrees of Freedom

Order of 100 (SLS) to 1,000 (EVM)
Low-Density Polyethylene

Fit of Axial Temperature Profile

![Graphs showing fit of axial temperature profile for Grade A and Grade B. The graphs compare the base case, this work, and plant data.]
Low-Density Polyethylene

Fit of Polymer Properties  Zavala & Biegler, 2006

Is Model Structure Correct?

Average Deviation  14 Different Grades -6 Estimation, 8 Trial-

<table>
<thead>
<tr>
<th></th>
<th>Conversion(%)</th>
<th>Mwn(%)</th>
<th>Mww(%)</th>
<th>LCB(%)</th>
<th>Density(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial Model</td>
<td>1.49</td>
<td>23.24</td>
<td>18.58</td>
<td>19.20</td>
<td>0.0965</td>
</tr>
<tr>
<td>This Work</td>
<td>0.12</td>
<td>6.20</td>
<td>3.31</td>
<td>6.27</td>
<td>0.0875</td>
</tr>
</tbody>
</table>
Most Parameters Estimated Reliably With Few Sets (Except Pressure-Dependent)
Low-Density Polyethylene

Computational Results – IPOPT & Default Linear Algebra (Pentium IV PC)

1) One Data Set (SLS)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Equations</th>
<th>DOF</th>
<th>Iterations</th>
<th>CPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11955</td>
<td>32</td>
<td>11</td>
<td>17.03</td>
</tr>
</tbody>
</table>

2) Multiple Data Sets (SLS)

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Equations</th>
<th>DOF</th>
<th>Iterations</th>
<th>CPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>33900</td>
<td>121</td>
<td>68</td>
<td>451.51</td>
</tr>
<tr>
<td>6</td>
<td>68421</td>
<td>217</td>
<td>58</td>
<td>900.21</td>
</tr>
</tbody>
</table>

3) Multiple Data Sets (EVM)

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Equations</th>
<th>DOF</th>
<th>Iterations</th>
<th>CPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 (EVM)</td>
<td>68627</td>
<td>529</td>
<td>71</td>
<td>1010.74</td>
</tr>
<tr>
<td>6 (SLS)</td>
<td>68421</td>
<td>217</td>
<td>58</td>
<td>900.21</td>
</tr>
</tbody>
</table>

Solution Strategy is Efficient
But, Memory Requirements Can Become Bottleneck in Very Large Problems
Low-Density Polyethylene

IPOPT with Parallel Linear Algebra  Zavala, Laird & Biegler 2007

Beowulf Cluster, Pentium IV, 3.0 GHz Processors

Parallelization Avoids Memory Limitations
Low-Density Polyethylene

On-Line Estimation Problem (MHE)

Least-Squares Objective

\[ \min_{u_k(x, t), P(x, t), z^0(x)} \int_{x_0}^{x_L} \int_{t_0}^{t_0+T} \varphi(z(x, t), y(x, t), u(x, t), P(x, t)) \, dx \, dt \]

Dynamic Reactor Model

\[ \frac{\partial z}{\partial t} + \nu(x, t) \frac{\partial z}{\partial x} = f(z(x, t), y(x, t), u(x, t), P(x, t)) \]

\[ 0 = g(z(x, t), y(x, t), u(x, t), P(x, t)) \]

\[ \chi(x, t) = \phi(z(x, t), y(x, t), u(x, t), P(x, t)) \]

\[ 0 \geq h(z(x, t), y(x, t), u(x, t), P(x, t)) \]

\[ z(x, 0) = z^0(x) \]

Closed-Loop Inference of Model States and Parameters using Moving Measurements Window

Inputs and Temperature Measurements Available On-Line (No Laboratory Data)

Solution Time Critical (Difficult Parallelization), Uncertainty Required
Low-Density Polyethylene

Fouling-Defouling with Synthetic Data

Measured Core Temperature

Measured Jacket Temperature

Time
Low-Density Polyethylene

Reconstruction of Wall Profile

Covariance Matrix of Axial Wall Profile

Steady-State Covariance Matrix Reached After Few Time Steps
Low-Density Polyethylene

Computational Results (IPOPT with Nested Dissection on Pentium IV PC)

NLP ~ 80,000 Constraints, 370 Degrees of Freedom

Sampling Time = 2 min

Effect of KKT Matrix Reordering on Solution Time

Dissection ~ 350,000 Constraints and 1,000 Degrees of Freedom ~ 2 Minutes (Warm-Start)
4. Conclusions and Future Work
Conclusions and Future Work

Current Optimization Technology Suitable for Large-Scale Estimation
Automatic Differentiation, Sparse Linear Algebra

General Modeling and Optimization Tools - Try!
- **AMPL** for Modeling and Derivatives (Commercial but Cheap)
- **ADOL-C** for Exact Derivatives (Open-Source)
- **IPOPT** for Solution and Inference (Open-Source)

Future Work
- Strategies for 4-D PDE Systems
- Lithium-Ion Batteries (Complex Aging Phenomena, Systematic Estimation Needed)
Computational Strategies for Large-Scale Estimation in Dynamic Systems

Victor M. Zavala
Postdoctoral Researcher
Mathematics and Computer Science Division
Argonne National Laboratory
vzavala@mcs.anl.gov

Joint Work with: Larry Biegler (Carnegie Mellon), Carl Laird (Texas A&M)

Eindhoven, Netherlands
June 9th, 2009