Extending the Scope of Algebraic MINLP Solvers to Black- and Grey-box Optimization

Nick Sahinidis

Acknowledgments:
Alison Cozad, David Miller, Zach Wilson
Pulverized coal plant Aspen Plus® simulation provided by the National Energy Technology Laboratory
**Process Disaggregation**

1. **Block 1: Simulator**
   - Model generation

2. **Block 2: Simulator**
   - Model generation

3. **Block 3: Simulator**
   - Model generation

---

**Process Simulation**
- Disaggregate process into process blocks

**Surrogate Models**
- Build simple and accurate models with a functional form tailored for an optimization framework

**Optimization Model**
- Add algebraic constraints, design specs, heat/mass balances, and logic constraints

Mathematical formulation:

\[
\begin{align*}
\text{min} & \quad f(x) \\
\text{s.t.} & \quad g(x) \leq 0 \\
& \quad h(x) = 0 \\
& \quad x \in [x^l, x^u]
\end{align*}
\]
LEARNING PROBLEM

Build a model of output variables $z$ as a function of input variables $x$ over a specified interval

$x \in \mathbb{R}^k$
$x^l \leq x \leq x^u$

$\begin{pmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_j \\
  \vdots \\
  x_k
\end{pmatrix}

\rightarrow

\begin{pmatrix}
  Z_1 \\
  Z_2 \\
  \vdots \\
  Z_l \\
  \vdots \\
  Z_m
\end{pmatrix}

z \in \mathbb{R}^m$

$z = f(x)$

Independent variables:
Operating conditions, inlet flow properties, unit geometry

Dependent variables:
Efficiency, outlet flow conditions, conversions, heat flow, etc.

Process simulation or Experiment
HOW TO BUILD THE SURROGATES

• We aim to build surrogate models that are
  – Accurate
    • *We want to reflect the true nature of the simulation*
  – Simple
    • *Tailored for algebraic optimization*

\[
\hat{f}(x) = \sum_{i=1}^{n} \gamma_i \exp \left( \frac{\|x\|}{\sigma^2} \right) + \beta_0 + \beta_1 x + \ldots
\]

\[
\hat{f}(x) = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 e^x
\]

• Generated from a minimal data set
  • *Reduce experimental and simulation requirements*
ALAMO

Automated Learning of Algebraic Models for Optimization

Start

Initial sampling

Build surrogate model

Adaptive sampling

Update training data set

false

Model converged?

true

Stop

Model error

New model

Current model

Black-box function
MODEL COMPLEXITY TRADEOFF

- **Kriging** [Krige, 63]
- **Neural nets** [McCulloch-Pitts, 43]
- **Radial basis functions** [Buhman, 00]

Model accuracy vs. Model complexity diagram:

- Preferred region
- Linear response surface
• **Goal:** Identify the **functional form and complexity** of the surrogate models

\[ z = f(x) \]

• **Functional form:**
  - General functional form is unknown: Our method will identify models with combinations of **simple basis functions**

<table>
<thead>
<tr>
<th>Category</th>
<th>( X_j(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Polynomial</td>
<td>( (x_d)^\alpha )</td>
</tr>
<tr>
<td>II. Multinomial</td>
<td>( \prod_{d \in D' \subseteq D} (x_d)^{\alpha_d} )</td>
</tr>
<tr>
<td>III. Exponential and logarithmic</td>
<td>( \exp\left(\frac{x_d}{\gamma}\right)^\alpha, \log\left(\frac{x_d}{\gamma}\right)^\alpha )</td>
</tr>
<tr>
<td>IV. Expected bases</td>
<td>From experience, simple inspection, physical phenomena, etc.</td>
</tr>
</tbody>
</table>
OVERFITTING AND TRUE ERROR

- **Step 1:** Define a large set of potential basis functions

  \[
  \hat{z}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 e^{x_1} + \beta_5 e^{x_2} + \ldots
  \]

- **Step 2:** Model reduction

  \[
  \hat{z}(x) = 2 + x_2 + 5 e^{x_1}
  \]

**Graph:**
- Ideal Model
- True error
- Empirical error
- Underfitting
- Overfitting
MODEL REDUCTION TECHNIQUES

• Qualitative tradeoffs of model reduction methods

Regularized regression techniques
• Penalize the least squares objective using the magnitude of the regressors [Tibshirani, 95]

Best subset methods
• Enumerate all possible subsets

Stepwise regression [Efroymson, 60]
Backward elimination [Oosterhof, 63]
Forward selection [Hamaker, 62]
MODEL SIZING

Complexity = number of terms allowed in the model

Goodness-of-fit measure

- Some measure of error that is sensitive to overfitting (AICc, BIC, Cp)
- Solve for the best one-term model
- Solve for the best two-term model
- 6th term was not worth the added complexity
- Final model includes 5 terms

Carnegie Mellon University
BASIS FUNCTION SELECTION

Find the model with the least error

\[
\min \quad SE = \sum_{i=1}^{N} \left| z_i - \sum_{j \in B} \beta_j X_{ij} \right|
\]

s.t.

\[
\sum_{j \in B} y_j = T
\]

\[-U(1 - y_j) \leq \sum_{i=1}^{N} X_{ij} \left( z_i - \sum_{j \in B} \beta_j X_{ij} \right) \leq U(1 - y_j) \quad j \in B\]

\[
\beta_l y_j \leq \beta_j \leq \beta_u y_j
\]

\[y_j = \{0, 1\}\]

We will solve this model for increasing \(T\) until we determine a model size

Basis function used in the model
\(\beta_j\) is chosen to satisfy a least squares regression (assumes loose bounds on \(\beta_j\))

Basis function NOT used in the model
\(\beta_j = 0\)
**ALAMO**

Automated Learning of Algebraic Models for Optimization

- **Start**
- **Initial sampling**
- **Build surrogate model**
- **Adaptive sampling**
- **Update training data set**
- **Model converged?**
  - true → **Stop**
  - false → **Adaptive sampling**
- **Error maximization sampling**
- **Model error**
• Search the problem space for areas of model inconsistency or model mismatch

• Find points that maximize the model error with respect to the independent variables

\[
\max_x \left( \frac{z(x) - \hat{z}(x)}{z(x)} \right)^2
\]

– Derivative-free solvers work well in low-dimensional spaces [Rios and Sahinidis, 12]

– Optimized using a black-box or derivative-free solver (SNOBFIT) [Huyer and Neumaier, 08]
COMPUTATIONAL RESULTS

• Goal – Compare methods on three target metrics
  
<table>
<thead>
<tr>
<th></th>
<th>Model accuracy</th>
<th>Data efficiency</th>
<th>Model simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Model accuracy</td>
<td>Data efficiency</td>
<td>Model simplicity</td>
</tr>
</tbody>
</table>

• Modeling methods compared
  – **ALAMO modeler** – Proposed methodology
  – **The LASSO** – The lasso regularization
  – **Ordinary regression** – Ordinary least-squares regression

• Sampling methods compared (over the same data set size)
  – **ALAMO sampler** – Proposed error maximization technique
  – **Single LH** – Single Latin hypercube (no feedback)
70% of problems solved exactly

80% of the problems had ≤0.5% error

Normalized test error
Results over a test set of 45 known functions treated as black boxes with bases that are available to all modeling methods.
MODEL SELECTION CRITERIA

• Balance fit (sum of square errors) with model complexity (number of terms in the model; denoted by $p$)

Corrected Akaike Information Criterion

$$AIC_c = N \log \left( \frac{1}{N} \sum_{i=1}^{N} (z_i - X_i \beta)^2 \right) + 2p + \frac{2p(p + 1)}{N - p - 1}$$

Mallows’ $C_p$

$$C_p = \frac{\sum_{i=1}^{N} (z_i - X_i \beta)^2}{\sigma^2} + 2p - N$$

Hannan-Quinn Information Criterion

$$HQC = N \log \left( \frac{1}{N} \sum_{i=1}^{N} (z_i - X_i \beta)^2 \right) + 2p \log(\log(N))$$

Bayes Information Criterion

$$BIC = \frac{\sum_{i=1}^{N} (z_i - X_i \beta)^2}{\sigma^2} + p \log(N)$$

Mean Squared Error

$$MSE = \frac{\sum_{i=1}^{N} (z_i - X_i \beta)^2}{N - p - 1}$$
CPU TIME COMPARISON

- Eight benchmarks from the UCI and CMU data sets
- Seventy noisy data sets were generated with multicollinearity and increasing problem size (number of bases)

![](chart.png)

- BIC solves more than two orders of magnitude faster than AIC, MSE and HQC
  - Optimized directly via a single mixed-integer convex quadratic model
MODEL QUALITY COMPARISON

- BIC leads to smaller, more accurate models
  - Larger penalty for model complexity
• **Expanding the scope of algebraic optimization**
  – Using low-complexity surrogate models to strike a balance between optimal decision-making and model fidelity

• **Surrogate model identification**
  – Simple, accurate model identification – Integer optimization

• **Error maximization sampling**
  – More information found per simulated data point
THEORY UTILIZATION

- Use **freely available** system knowledge to strengthen model
  - Physical limits
  - First-principles knowledge
  - Intuition

- Non-empirical restrictions can be applied to general regression problems
CONSTRAI NED REGRESSION

• Challenging due to the semi-infinite nature of the regression constraints

Standard regression

\[
\min_{\beta_1, \beta_2} \sum_{i=1}^{4} (z_i - \hat{z}(x_i; \beta_1, \beta_2))^2
\]

Surrogate model

\[
\min_{\beta_1, \beta_2} \sum_{i=1}^{4} (z_i - \hat{z}(x_i; \beta_1, \beta_2))^2 \\
\text{s.t. } \hat{z}(x_i; \beta_1, \beta_2) \geq 0 \quad \forall x
\]
IMPLIRED PARAMETER RESTRICTIONS

Find a model \( \hat{z} \) such that \( \hat{z}(x) \geq 0 \) with a fixed model form:

\[
\hat{z}(x) = \beta_1 x + \beta_2 x^3
\]

Step 1: Formulate constraint in z- and x-space

\[
\begin{align*}
\min_{\beta_1, \beta_2} & \quad \sum_{i=1}^{4} (z_i - [\beta_1 x + \beta_2 x^3])^2 \\
\text{s.t.} & \quad \beta_1 x + \beta_2 x^3 \geq 0 \quad x \in [0, 1]
\end{align*}
\]

Step 2: Identify a sufficient set of \( \beta \)-space constraints

\[
\begin{align*}
\min_{\beta_1, \beta_2} & \quad \sum_{i=1}^{4} (z_i - [\beta_1 x + \beta_2 x^3])^2 \\
\text{s.t.} & \quad 0.240 \beta_1 + 0.0138 \beta_2 \geq 0 \\
& \quad 0.281 \beta_1 + 0.0223 \beta_2 \geq 0 \\
& \quad 0.120 \beta_1 + 0.00173 \beta_2 \geq 0 \\
& \quad 0.138 \beta_1 + 0.00263 \beta_2 \geq 0
\end{align*}
\]

1 parametric constraint
4 \( \beta \)-constraints
### TYPES OF RESTRICTIONS

<table>
<thead>
<tr>
<th>Response bounds</th>
<th>Individual responses</th>
<th>Multiple responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\hat{A}]_t \geq 0$</td>
<td>$\hat{F}^{\text{out}}(x) \leq F^{\text{in}}$</td>
<td>$\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 = 1$</td>
</tr>
<tr>
<td>pressure, temperature, compositions</td>
<td>mass and energy balances, physical limitations</td>
<td>mass balances, sum-to-one, state variables</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response derivatives</th>
<th>Alternative domains</th>
<th>Boundary conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{dT}{dx} \geq 0$</td>
<td>Extrapolation zone</td>
<td>velocity profile model</td>
</tr>
<tr>
<td>monotonicity, numerical properties, convexity</td>
<td>Problem Space</td>
<td>Add no slip</td>
</tr>
<tr>
<td>safe extrapolation, boundary conditions</td>
<td>$\hat{v}(R, \theta) = 0 \quad \forall \theta$</td>
<td></td>
</tr>
</tbody>
</table>
CARBON CAPTURE SYSTEM DESIGN

- **Discrete decisions:** How many units? Parallel trains? What technology used for each reactor?
- **Continuous decisions:** Unit geometries
- **Operating conditions:** Vessel temperature and pressure, flow rates, compositions
SUPERSTRUCTURE OPTIMIZATION

Mixed-integer nonlinear programming model

- Economic model
- Process model
- Material balances
- Hydrodynamic/Energy balances
- Reactor surrogate models
- Link between economic model and process model
- Binary variable constraints
- Bounds for variables
GLOBAL MINLP SOLVERS ON CMU/IBMLIB
CONCLUSIONS

• **ALAMO provides algebraic models that are**
  - Accurate and simple
  - Generated from a minimal number of function evaluations

• **ALAMO’s constrained regression facility allows modeling of**
  - Bounds on response variables
  - Convexity/monotonicity of response variables

• **On-going efforts**
  - Uncertainty quantification
  - Symbolic regression

• **ALAMO site:** [archimedes.cheme.cmu.edu/?q=alamo](archimedes.cheme.cmu.edu/?q=alamo)