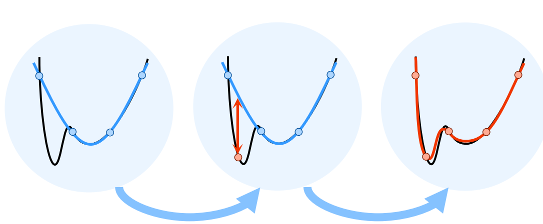
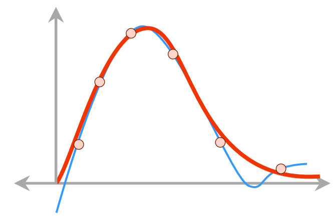




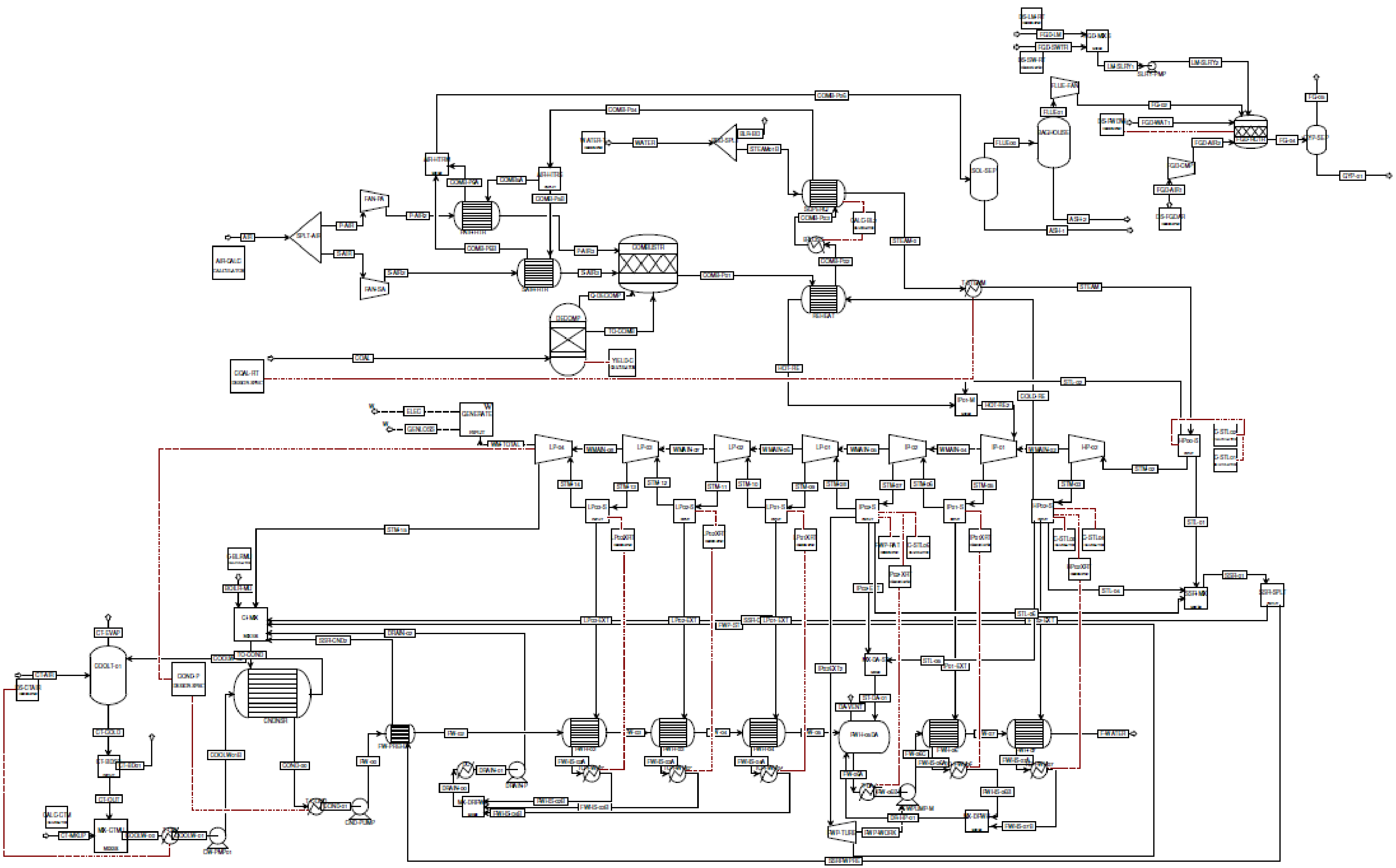
Extending the Scope of Algebraic MINLP Solvers to Black- and Grey-box Optimization

Nick Sahinidis

Acknowledgments:

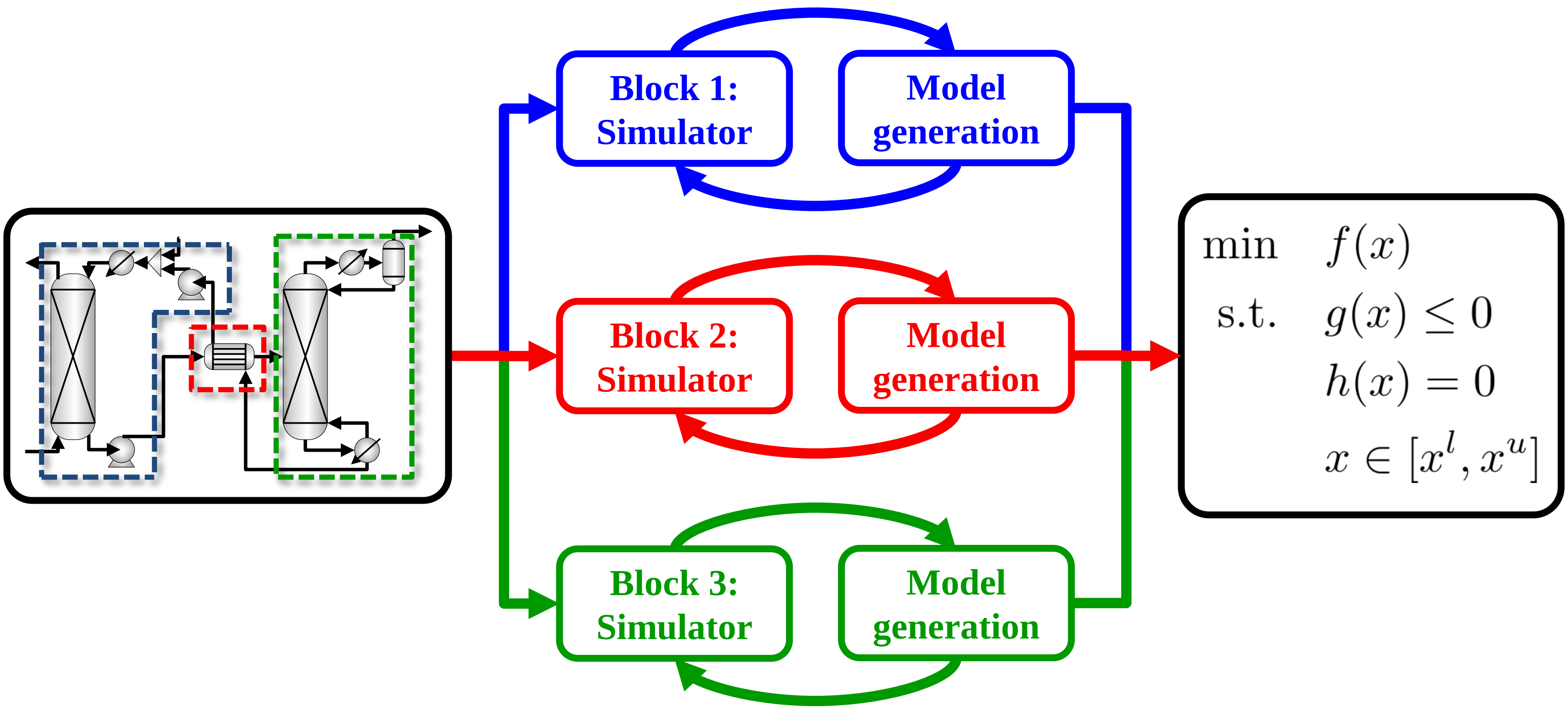
Alison Cozad, David Miller, Zach Wilson

SIMULATION OPTIMIZATION



Pulverized coal plant Aspen Plus® simulation provided by the National Energy Technology Laboratory

PROCESS DISAGGREGATION



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

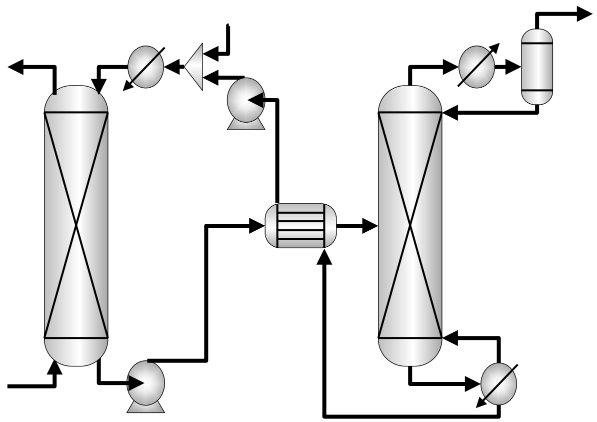
Add algebraic constraints design specs, heat/mass balances, and logic constraints

LEARNING PROBLEM

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

$$x \in \mathbb{R}^k \qquad x^l \leq x \leq x^u \qquad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_k \end{pmatrix} \Rightarrow$$

Process simulation or Experiment



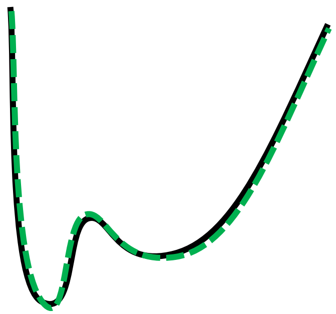
$$\Rightarrow \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_l \\ \vdots \\ z_m \end{pmatrix} \qquad z \in \mathbb{R}^m \qquad z = f(x)$$

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

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

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*



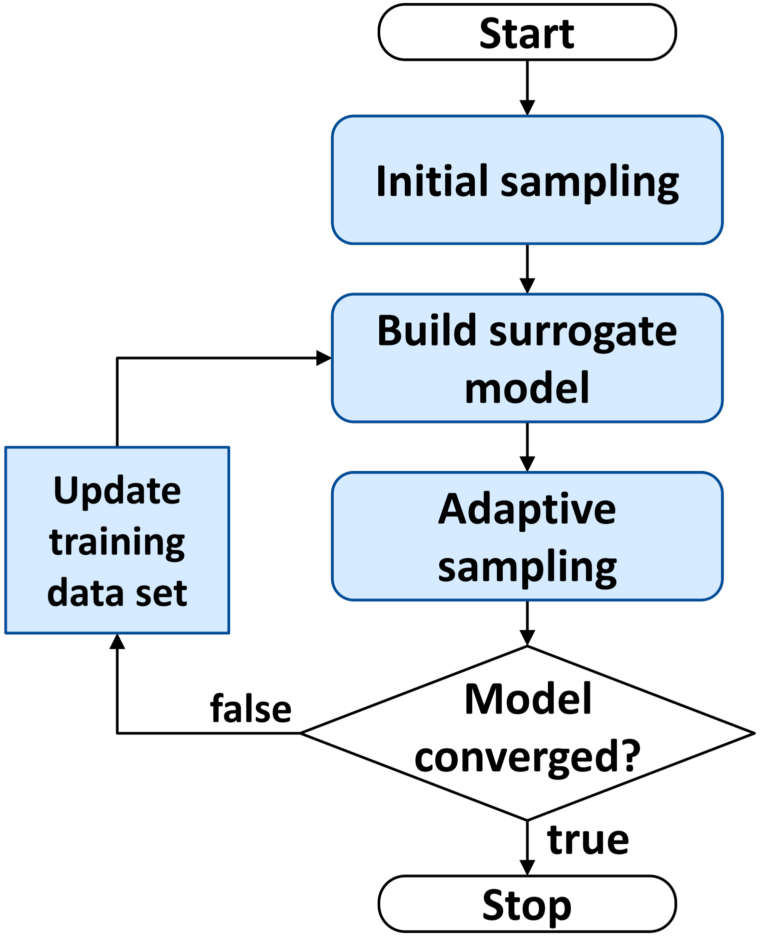
$$\hat{f}(x) = \sum_{i=1}^{n} \gamma_i \exp\left(\frac{\|x\|}{\sigma^2}\right) + \beta_0 + \beta_1 x + \ldots$$

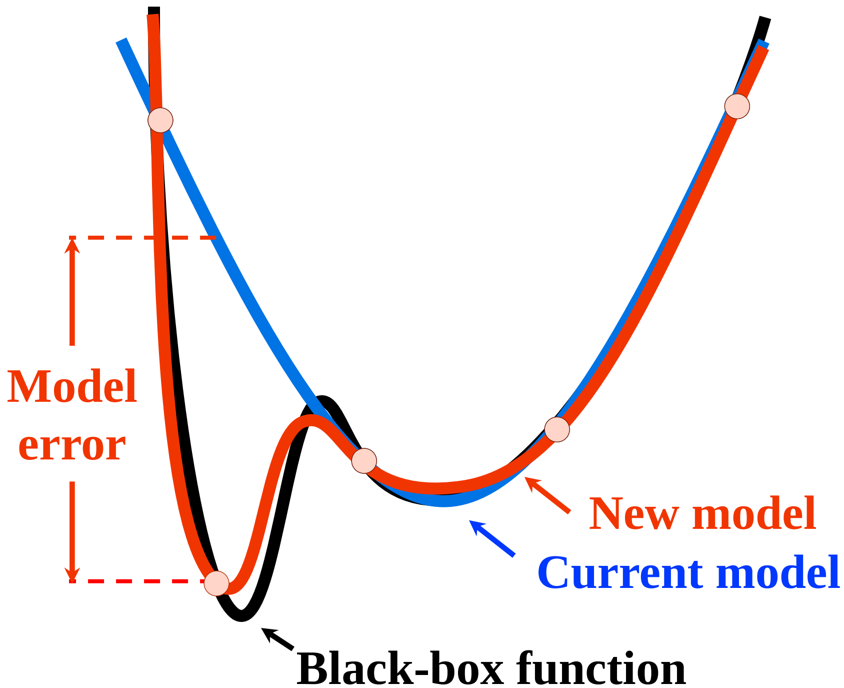
$$\hat{f}(x) = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 e^x$$

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

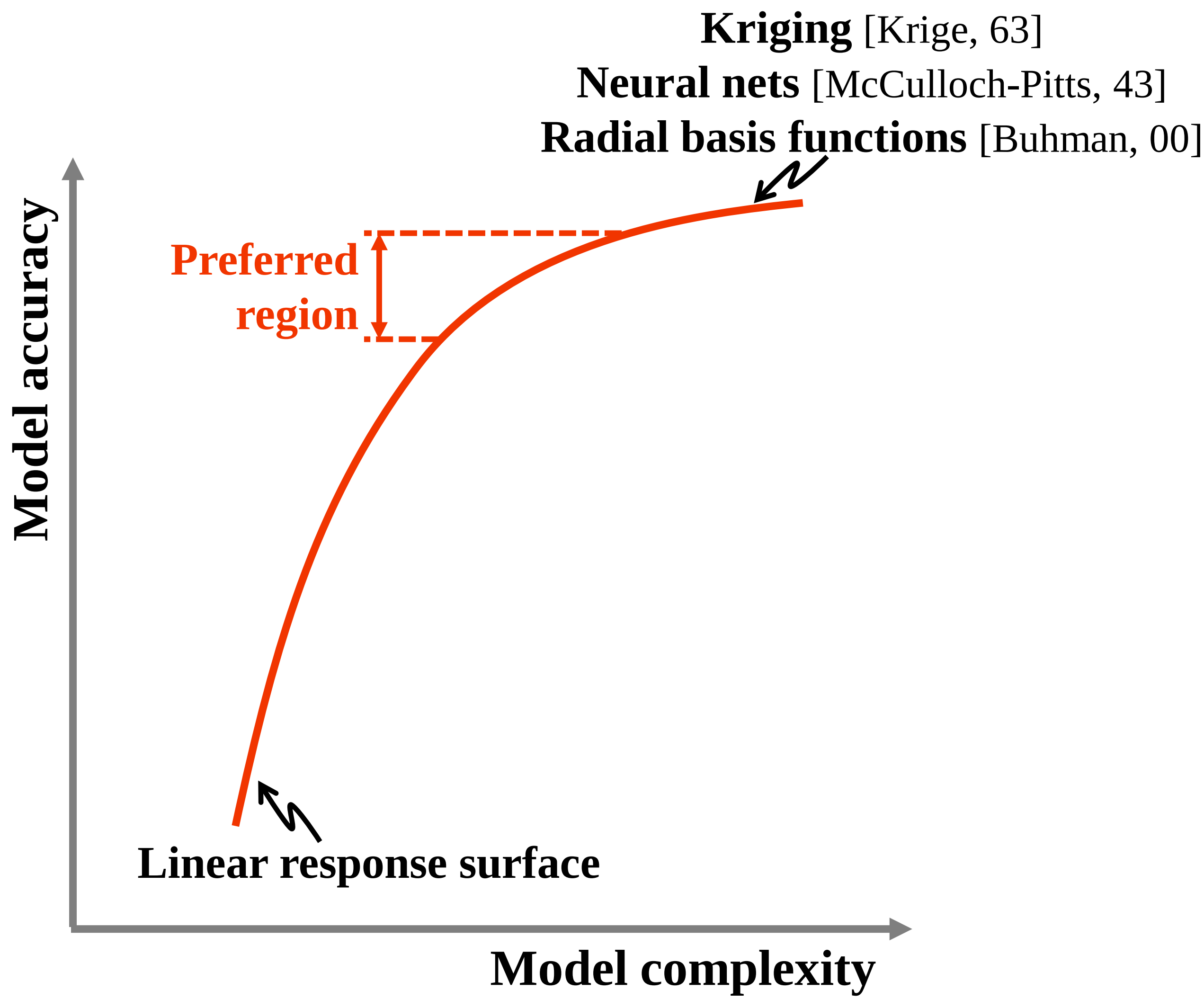
ALAMO

Automated Learning of Algebraic Models for Optimization





MODEL COMPLEXITY TRADEOFF



MODEL IDENTIFICATION

- **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

Category		$X_j(x)$
I.	Polynomial	$(x_d)^\alpha$
II.	Multinomial	$\prod_{d \in \mathcal{D}' \subseteq \mathcal{D}} (x_d)^{\alpha_d}$
III.	Exponential and logarithmic	$\exp\left(\frac{x_d}{\gamma}\right)^\alpha$, $\log\left(\frac{x_d}{\gamma}\right)^\alpha$
IV.	Expected bases	From experience, simple inspection, physical phenomena, etc.

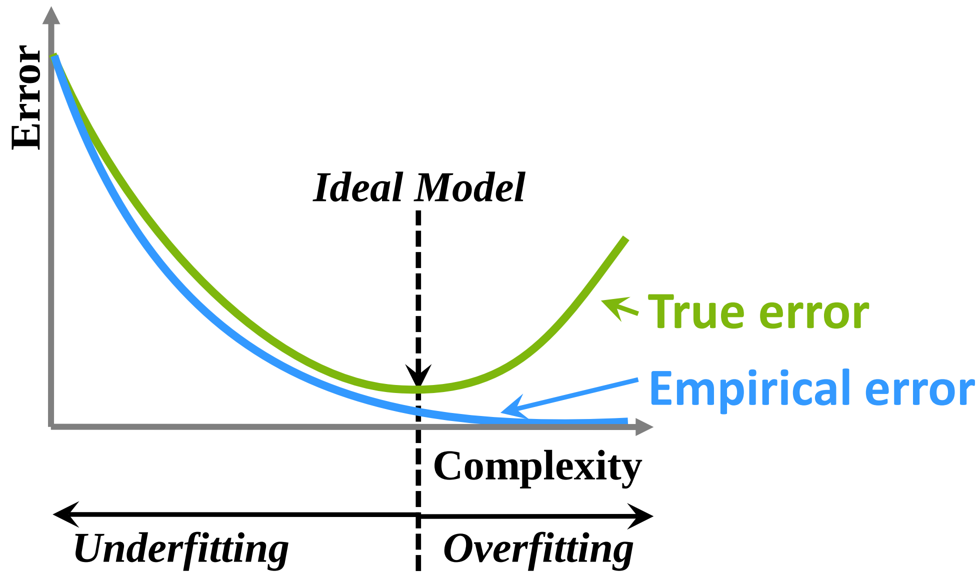
OVERFITTING AND TRUE ERROR

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

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

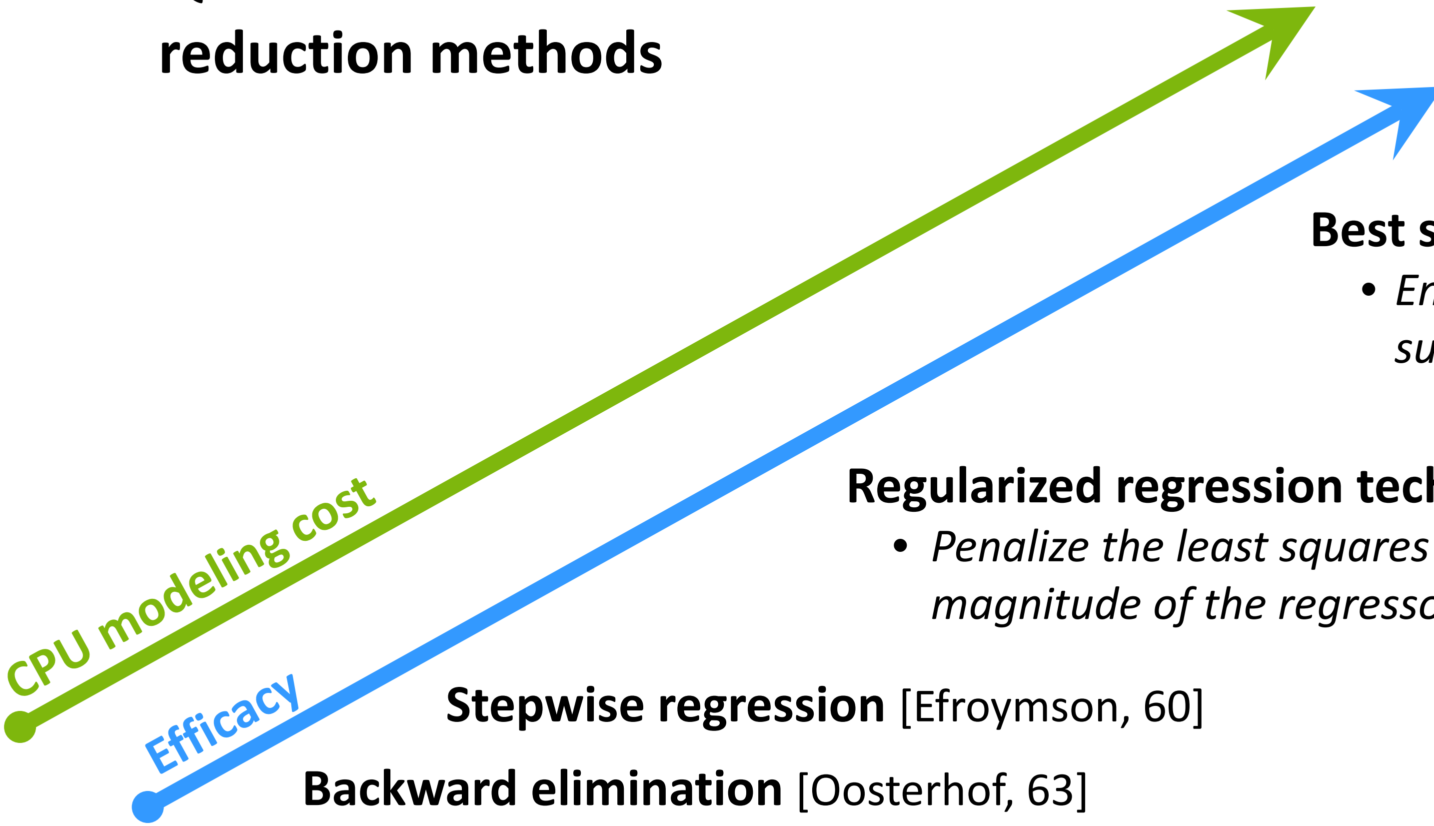
- **Step 2: Model reduction**

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



MODEL REDUCTION TECHNIQUES

- **Qualitative tradeoffs of model reduction methods**



Best subset methods

- *Enumerate all possible subsets*

Regularized regression techniques

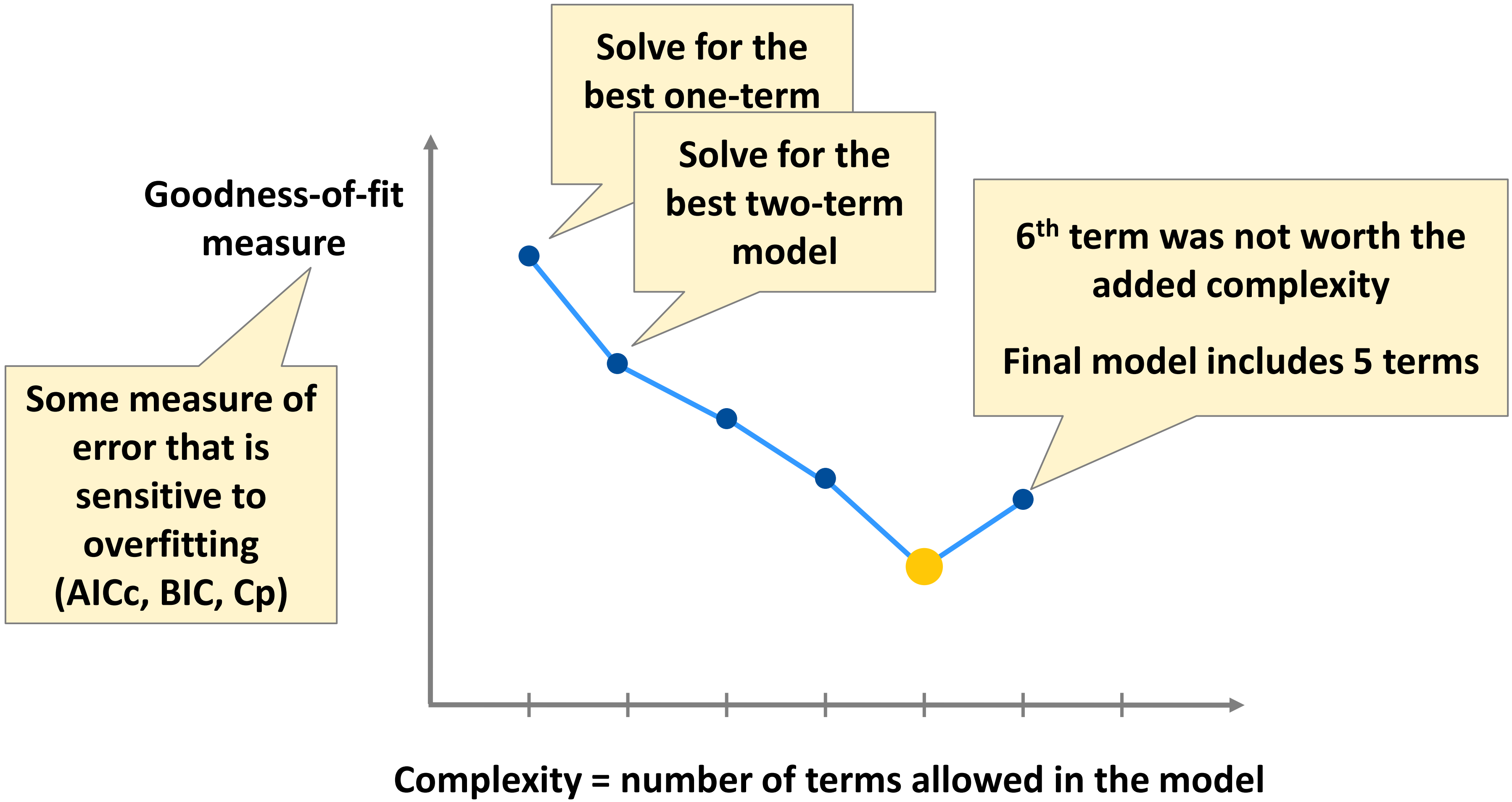
- *Penalize the least squares objective using the magnitude of the regressors* [Tibshirani, 95]

Stepwise regression [Efroymson, 60]

Backward elimination [Oosterhof, 63]

Forward selection [Hamaker, 62]

MODEL SIZING



BASIS FUNCTION SELECTION

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

$$\text{s.t.} \quad \sum_{j \in \mathcal{B}} y_j = T$$

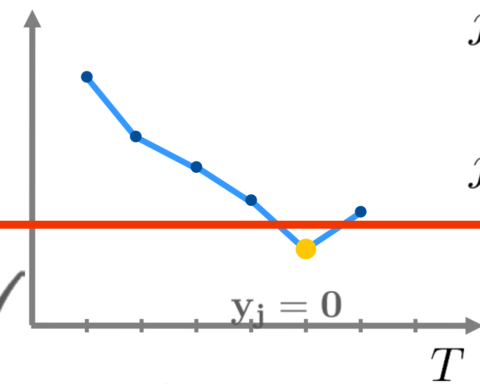
$$-U(1-y_j) \le \sum_{i=1}^{N} X_{ij} \left(z^i - \sum_{j \in \mathcal{B}} \beta_j X_{ij} \right) \le U(1-y_j) \qquad j \in \mathcal{B}$$

$$\beta^l y_j \le \beta_j \le \beta^u y_j \qquad j \in \mathcal{B}$$

$$y_j = \{0,1\} \qquad j \in \mathcal{B}$$

Find the model with the least error

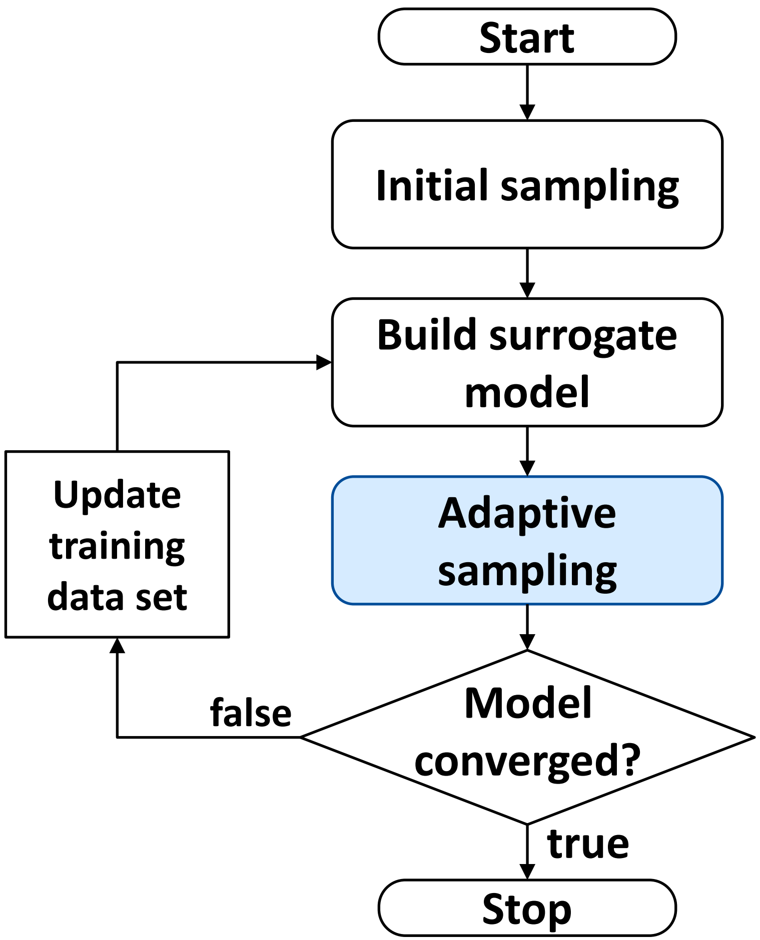
$$\left(\begin{array}{c} \mathrm{y_j} = 1 \\ \text{Basis function used in the model} \\ \beta_j \text{ is chosen to satisfy a least squares regression} \\ (\textit{assumes loose bounds on } \beta_j) \end{array} \right) \vee \left(\begin{array}{c} \mathrm{y_j} = 0 \\ \text{Basis function NOT used in the model} \\ \beta_j = 0 \end{array} \right)$$

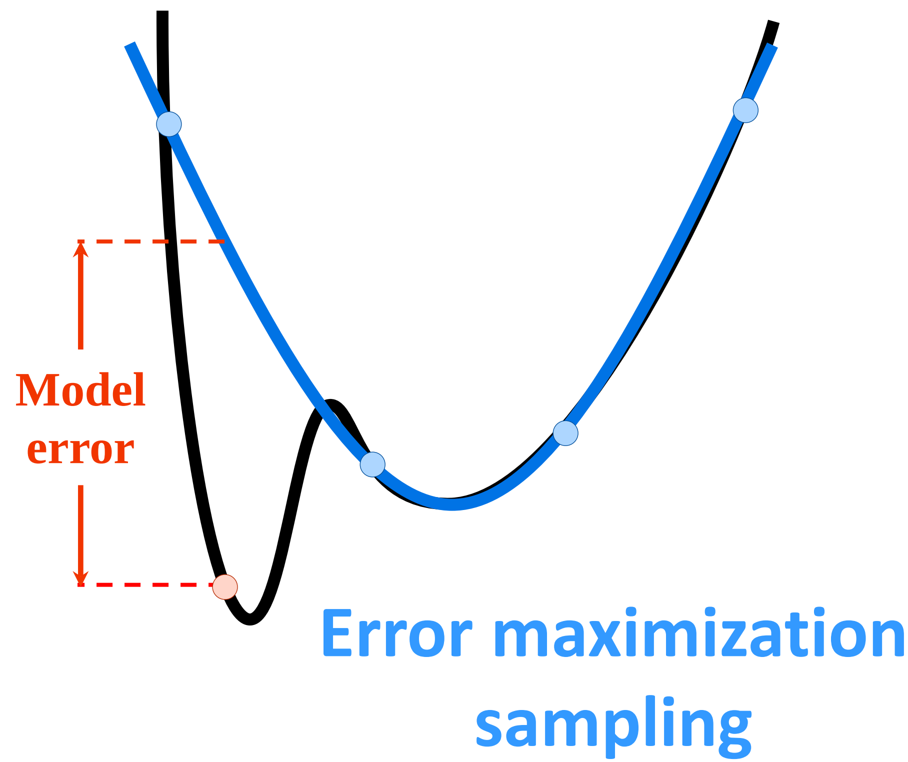


We will solve this model for increasing T until we determine a model size

ALAMO

Automated Learning of Algebraic Models for Optimization





ERROR MAXIMIZATION SAMPLING

- **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**

Surrogate model

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

  - **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

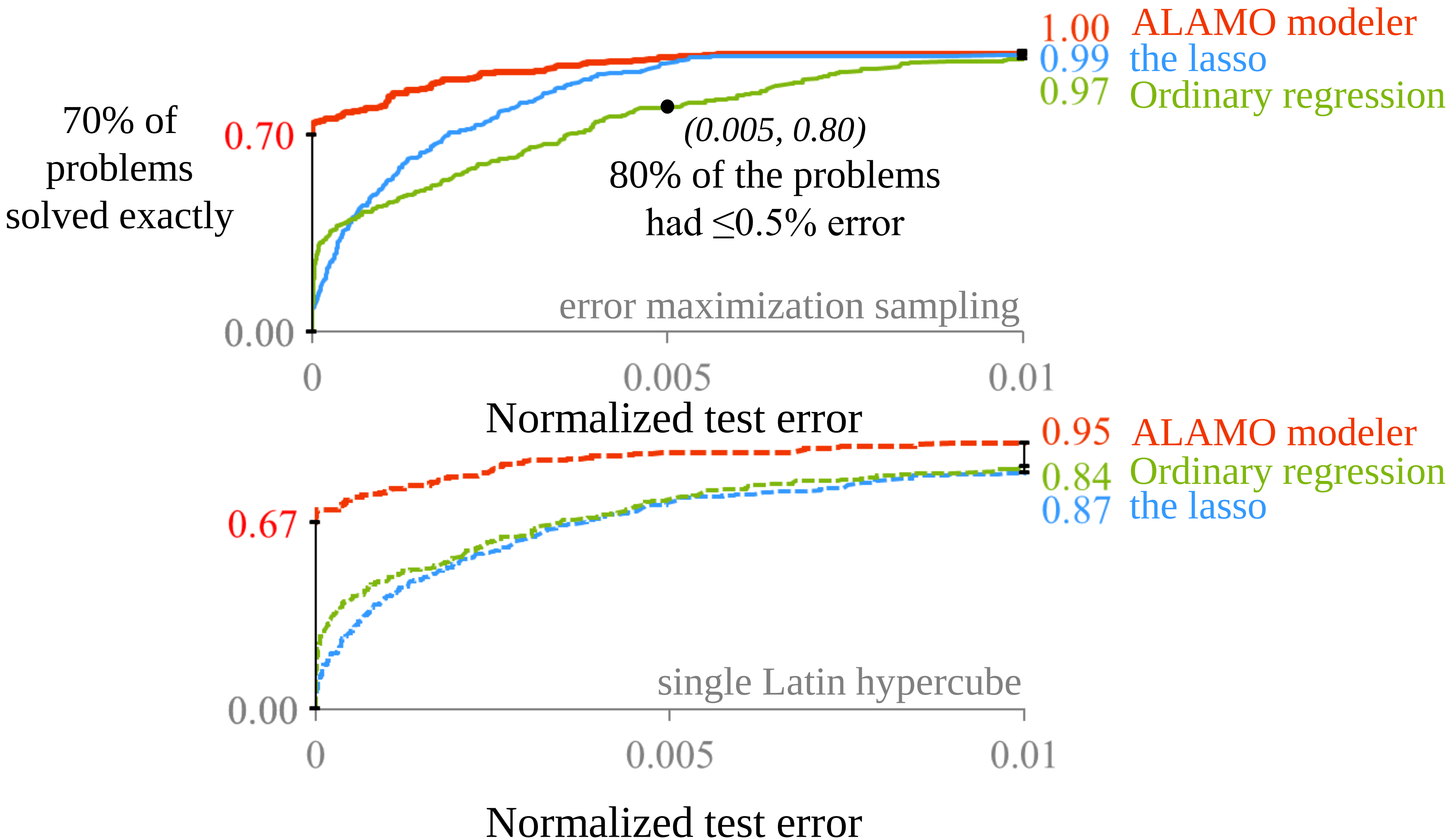
- **Goal – Compare methods on three target metrics**

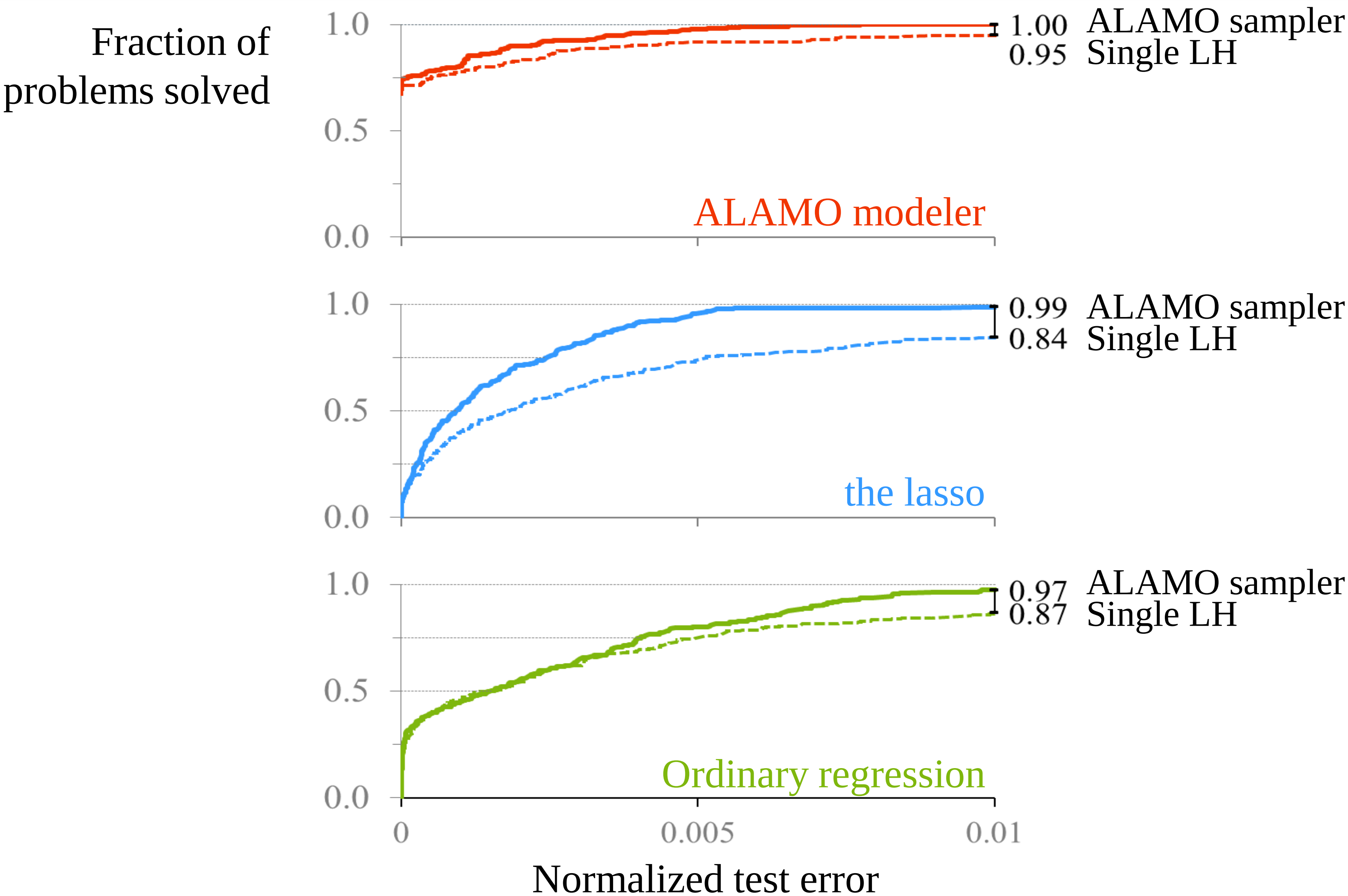
1 Model accuracy | **2 Data efficiency** | **3 Model simplicity**

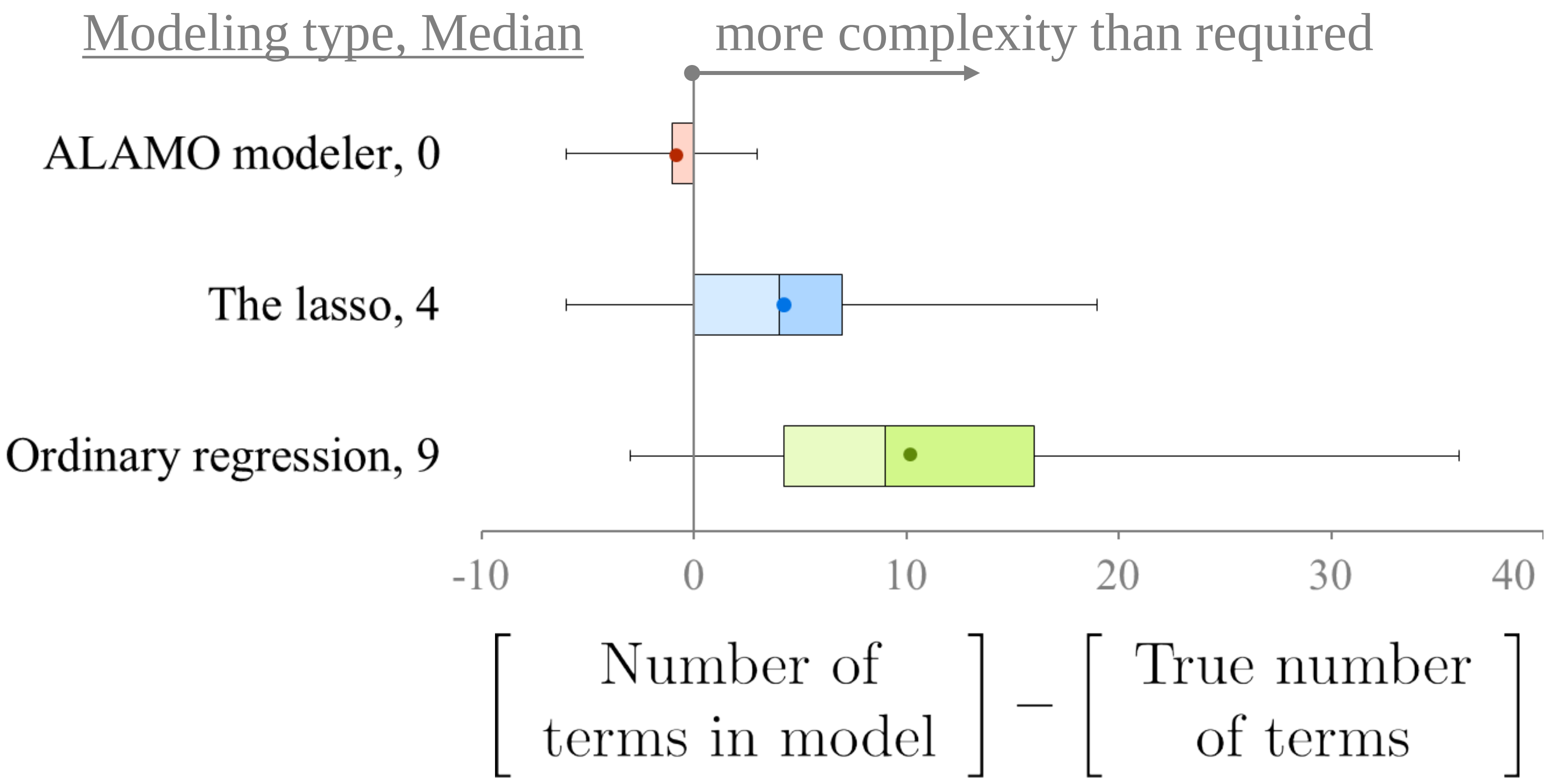
- **Modeling methods compared**
  - **ALAMO modeler – Proposed methodology**
  - **The LASSO – The lasso regularization**
  - **Ordinary regression – Ordinary least-squares regression**

- **Sampling methods compared (over the same data set size)**
  - **ALAMO sampler – Proposed error maximization technique**
  - **Single LH – Single Latin hypercube (no feedback)**

Fraction of problems solved




Fraction of problems solved
1.0
0.5
0.0
1.00 ALAMO sampler
0.95 Single LH
ALAMO modeler
0.99 ALAMO sampler
0.84 Single LH
the lasso
0.97 ALAMO sampler
0.87 Single LH
Ordinary regression
0
0.005
0.01
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

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

Mallows' Cp

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

Hannan-Quinn Information Criterion

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

Bayes Information Criterion

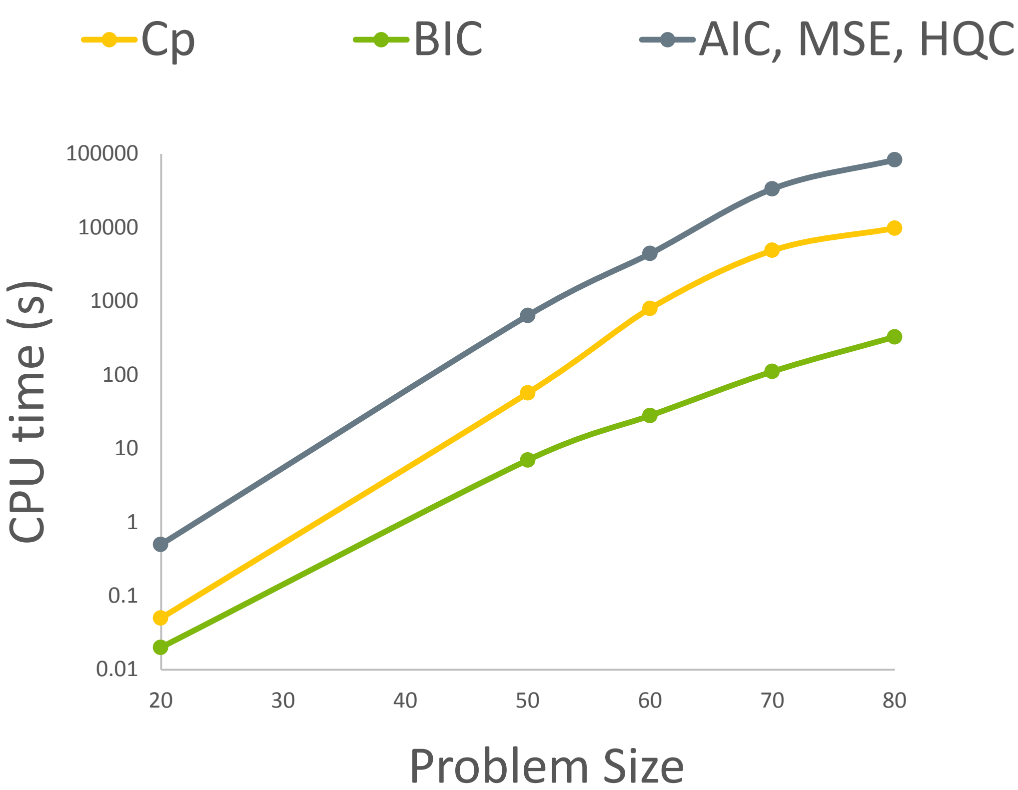
$$BIC = \frac{\sum_{i=1}^{N}(z_i - X_i\beta)^2}{\widehat{\sigma^2}} + \boldsymbol{p}\log(N)$$

Mean Squared Error

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

CPU TIME COMPARISON

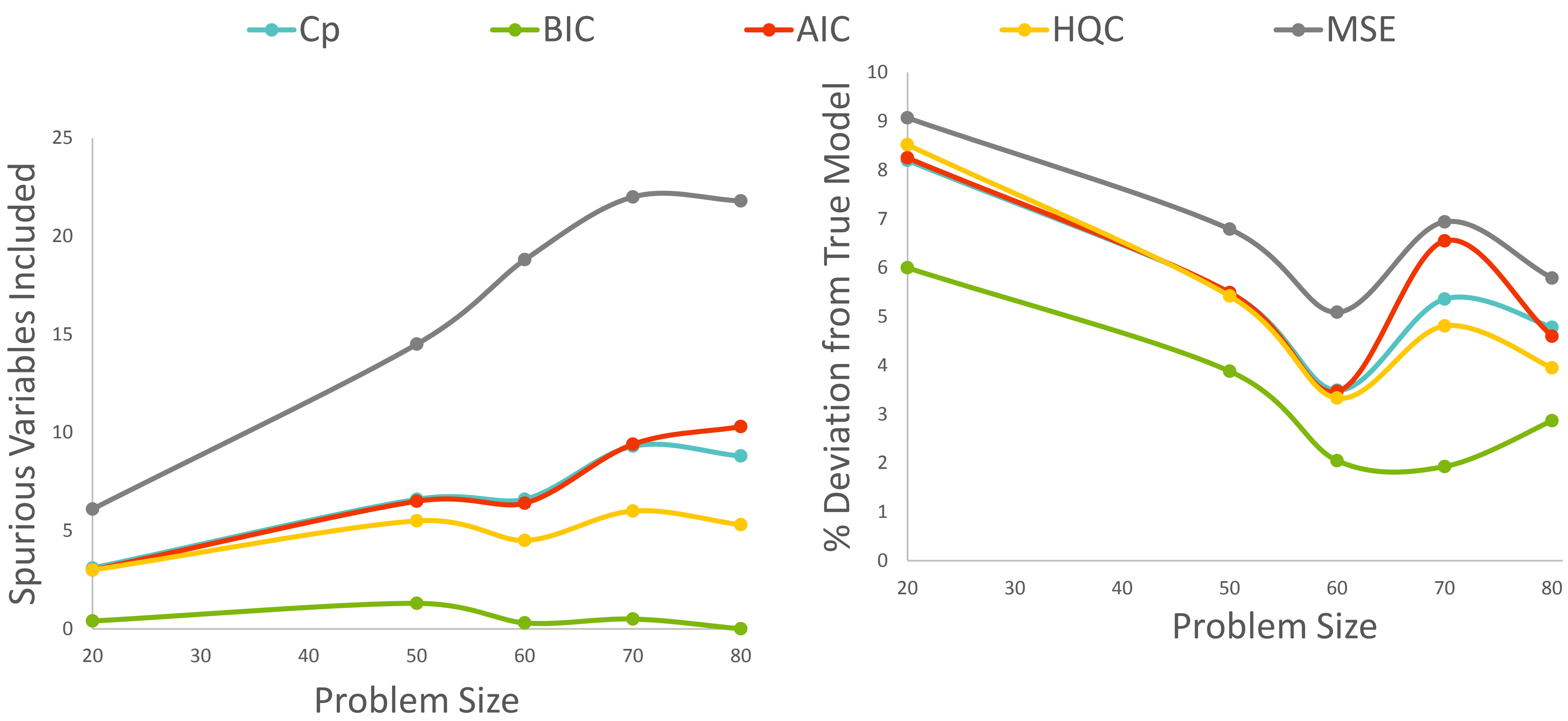
- **Eight benchmarks from the UCI and CMU data sets**
- **Seventy noisy data sets were generated with multicolinearity and increasing problem size (number of bases)**



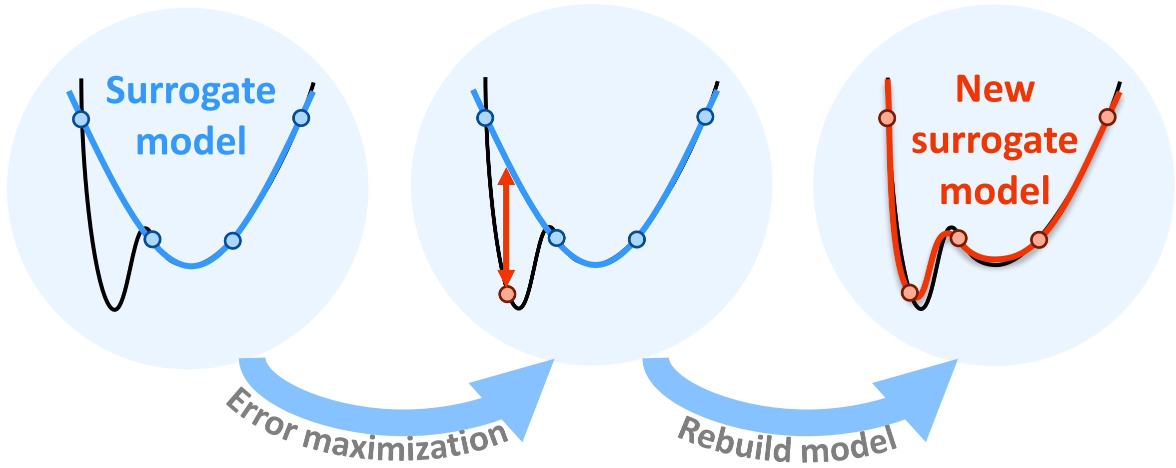
- **BIC solves more than two orders of magnitude faster than AIC, MSE and HQC**
  - **Optimized directly via a single mixed-integer convex quadratic model**

MODEL QUALITY COMPARISON

- **BIC leads to smaller, more accurate models**
  - Larger penalty for model complexity

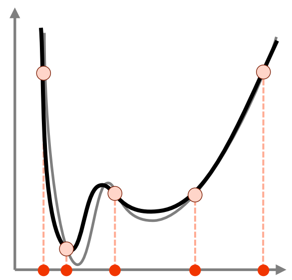


ALAMO REMARKS

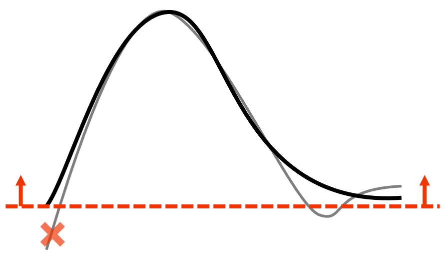


- **Expanding the scope of algebraic optimization**
  - Using low-complexity surrogate models to strike a balance between optimal decision-making and model fidelity
- **Surrogate model identification**
  - Simple, accurate model identification – Integer optimization
- **Error maximization sampling**
  - More information found per simulated data point

THEORY UTILIZATION



empirical data



non-empirical information

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

- **Non-empirical restrictions can be applied to general regression problems**

CONSTRAINED REGRESSION

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

Standard regression

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

Surrogate model

easy

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{4} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$
$$\text{s.t.} \quad \beta_1 \geq \beta_2$$

tough

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{4} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$
$$\text{s.t.} \quad \hat{z}(x_i;\beta_1,\beta_2) \geq 0 \quad \forall x$$

IMPLIED 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

$$\min_{\beta_1,\beta_2} \quad \sum_{i=1}^{4} \left(z_i - \left[\beta_1\, x + \beta_2\, x^3\right]\right)^2$$

$$\text{s.t.} \quad \beta_1\, x + \beta_2\, x^3 \geq 0 \quad x \in [0, 1]$$

1 parametric constraint

4 β-constraints

Step 2: Identify a sufficient set of β-space constraints

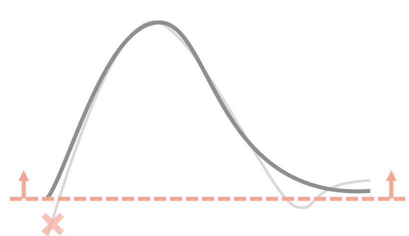
$$\min_{\beta_1,\beta_2} \quad \sum_{i=1}^{4} \left(z_i - \left[\beta_1\, x + \beta_2\, x^3\right]\right)^2$$

$$\text{s.t.} \quad \left\{ \begin{array}{l} 0.240\,\beta_1 + 0.0138\,\beta_2 \geq 0 \\ 0.281\,\beta_1 + 0.0223\,\beta_2 \geq 0 \\ 0.120\,\beta_1 + 0.00173\,\beta_2 \geq 0 \\ 0.138\,\beta_1 + 0.00263\,\beta_2 \geq 0 \end{array} \right.$$

TYPES OF RESTRICTIONS

Response bounds

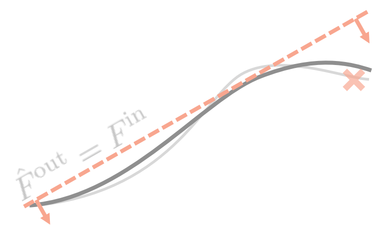
$$[\hat{A}]_t \geq 0$$



pressure, temperature, compositions

Individual responses

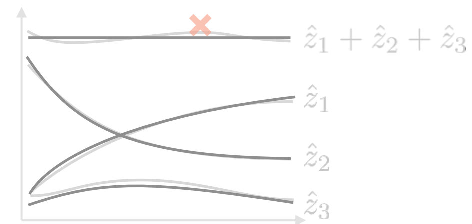
$$\hat{F}^{\text{out}}(x) \leq F^{\text{in}}$$



mass and energy balances, physical limitations

Multiple responses

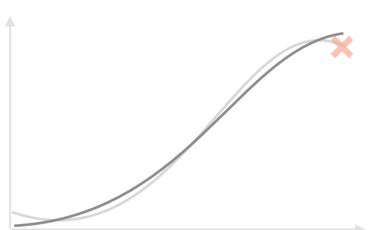
$$\hat{z}_1 + \hat{z}_2 + \hat{z}_3 = 1$$



mass balances, sum-to-one, state variables

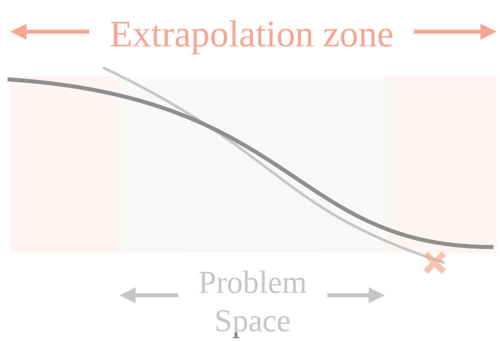
Response derivatives

$$\frac{dT}{dx} \geq 0$$



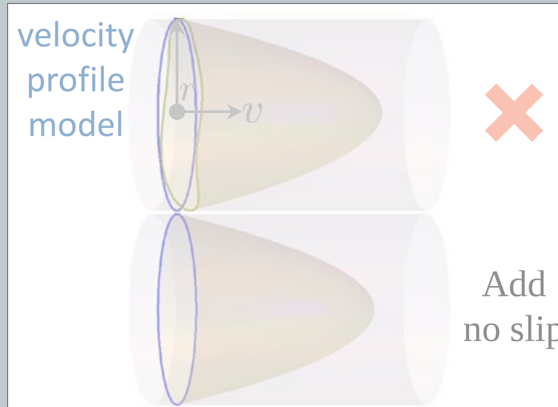
monotonicity, numerical properties, convexity

Alternative domains



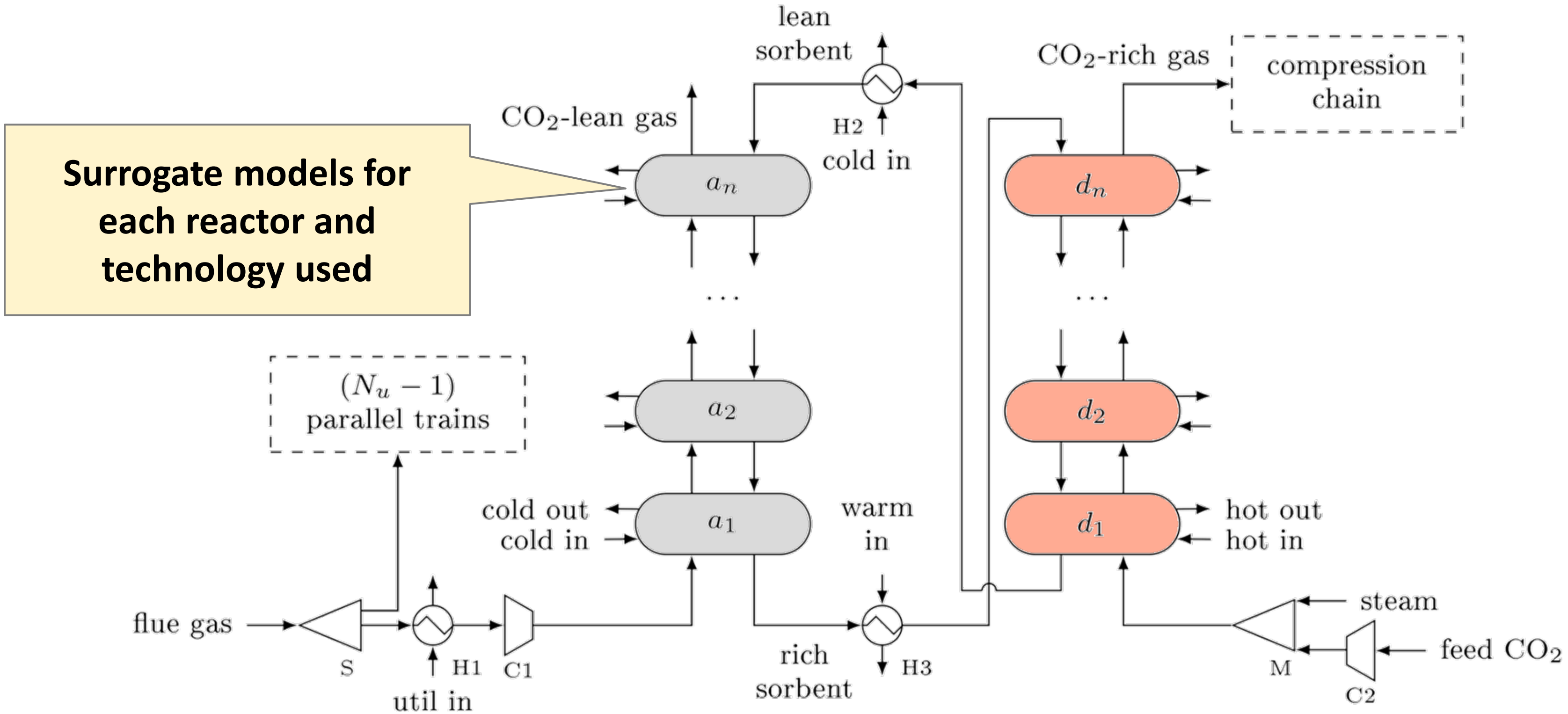
safe extrapolation, boundary conditions

Boundary conditions



$$\hat{v}(R, \theta) = 0 \qquad \forall \theta$$

CARBON CAPTURE SYSTEM DESIGN

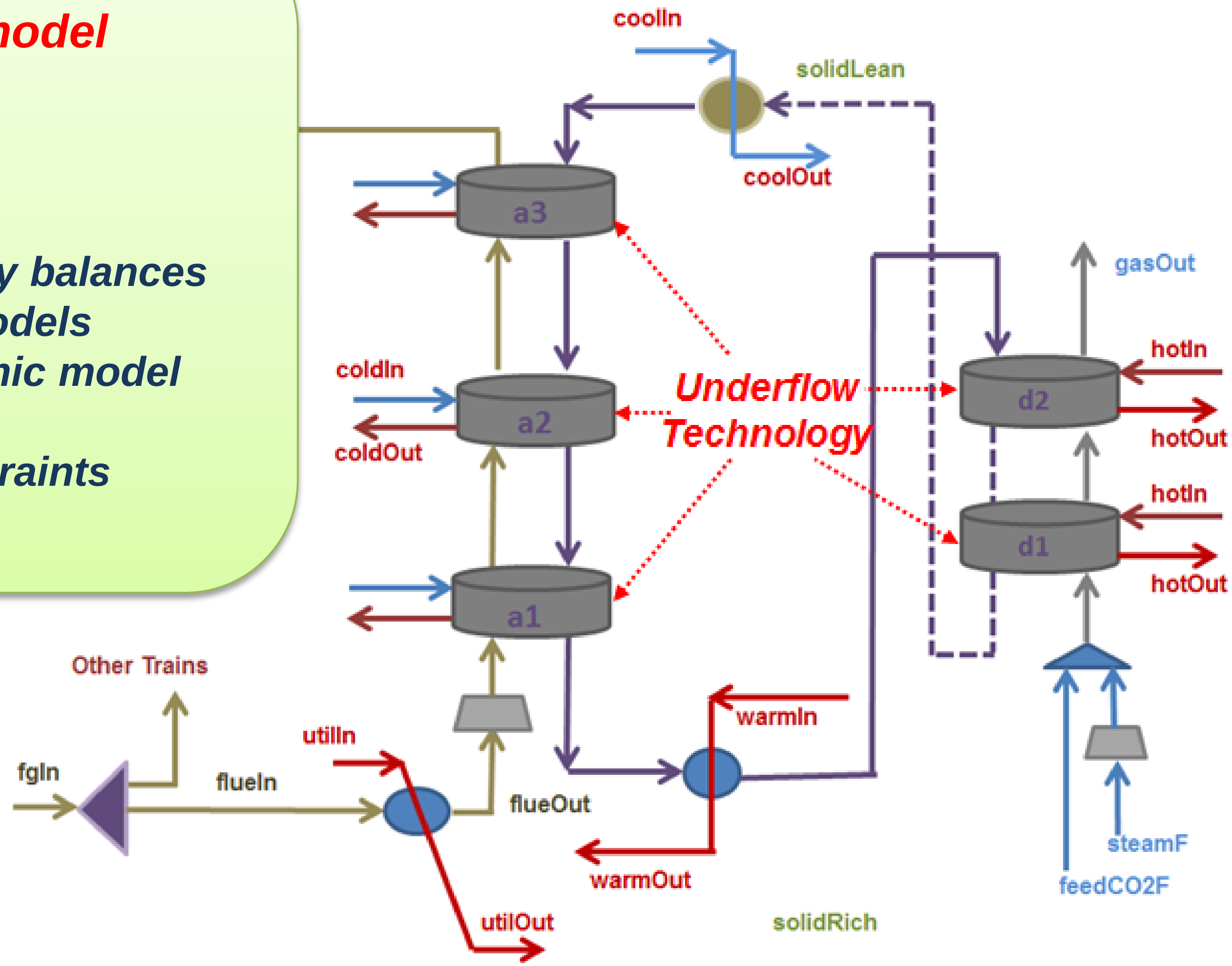


- **Discrete decisions:** How many units? Parallel trains? What technology used for each reactor?
- **Continuous decisions:** Unit geometries
- **Operating conditions:** Vessel temperature and pressure, flow rates, compositions

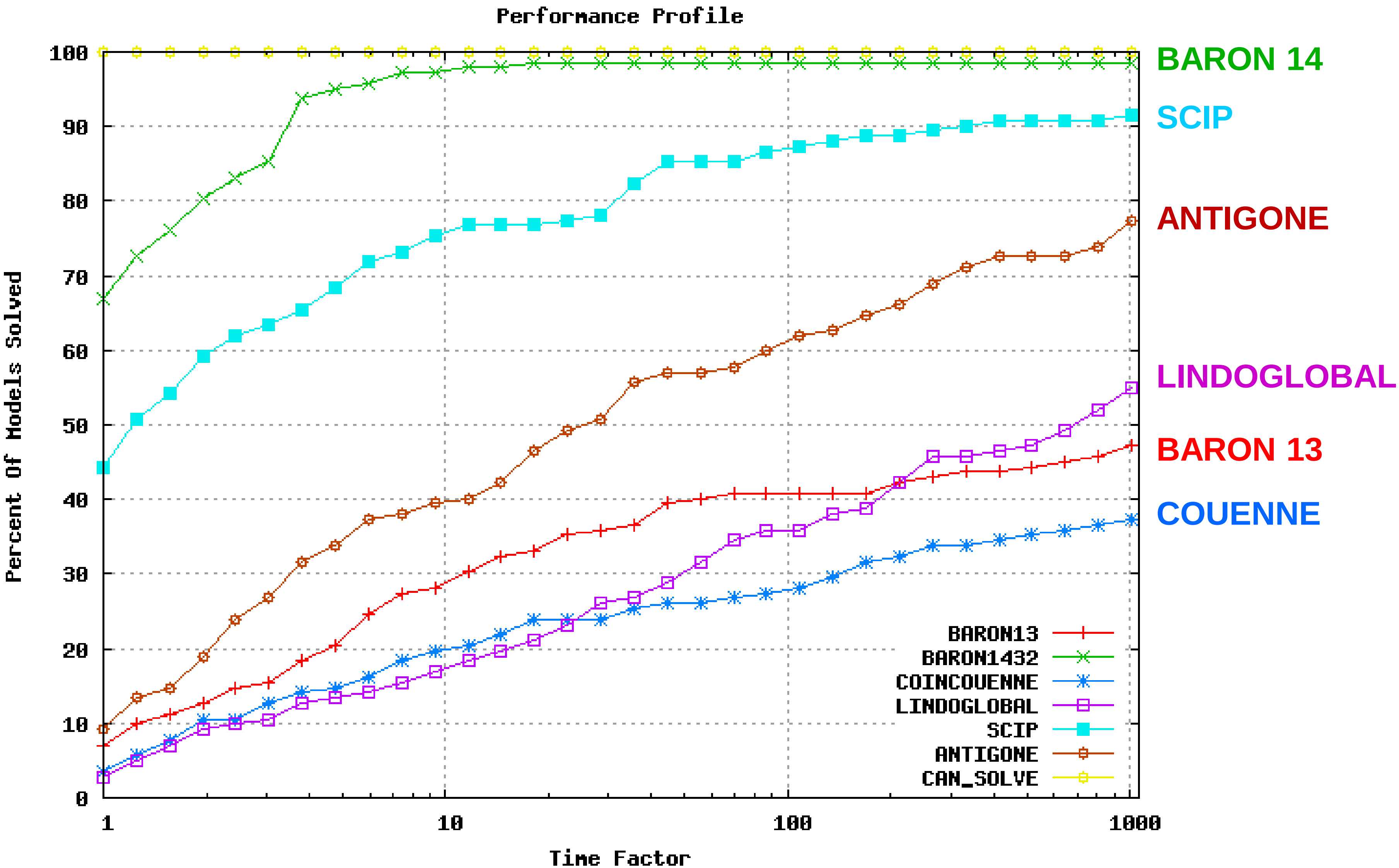
SUPERSTRUCTURE OPTIMIZATION

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