Data-driven Optimization with Truth Models

J. P. Eason and L. T. Biegler
Chemical Engineering Dept.
Carnegie Mellon University
Pittsburgh, PA 15213
June, 2018

Overview

Introduction
- Why optimize?
- Equation Oriented Optimization

How RM Optimizers Cheat
- Truth vs. Reduced Models in Optimization
- Toy Problem → Fundamental RM Concerns

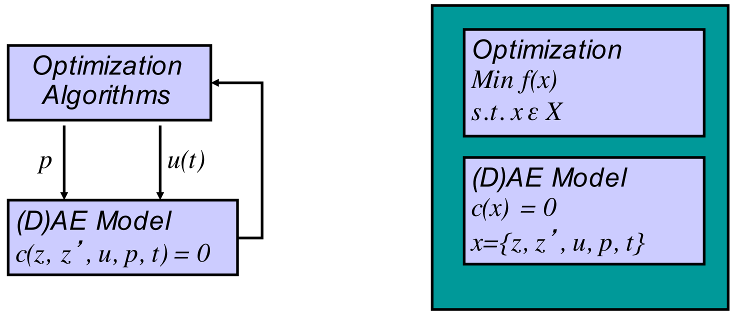
Trust Region Framework
- Algorithm and Convergence Properties
- Glass Box Optimization with Embedded Black Box Models

Process Case Studies
- Ammonia Process
- Oxycombustion Power Plant

Summary and Conclusions

Why Process Optimization?

- Equipment and Flowsheet Design
- Process Operations, Transients and Upsets
- Parameter Estimation and Model Discrimination



- Optimization Gives Better Results than with "Experience"
- Consistent Results among all Practitioners
- Reduce Solution Time by Orders of Magnitude
- Support and Enhance Process Understanding

Evolution of NLP Algorithms & Tasks

'80s: Flowsheet optimization ~ 100 variables and constraints — SQP

'90s: Static real-time optimization (RTO) over 100 000 variables & constraints — rSQP

'00s: Simultaneous dynamic optimization over 1 000 000 variables and constraints — IPOPT

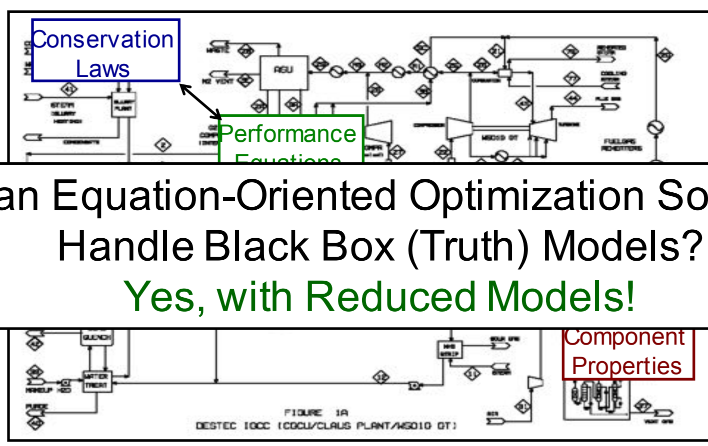
'10s: Sensitivity-based dynamic on-line optimization for large NLPs: < 1 CPUs — sIPOPT

Equation-Oriented Utopia for Process Optimization

- Exact Jacobians/Hessians and sparse equation structure
- NLP Reformulation for MPECs (for nonsmooth models, bi-level formulations, phase changes,…)
- Fast Newton-based NLP solvers
- NLP sensitivity (post-optimality and interpretation, multi-level opt., …)
- EO-Modeling Enables:
  – Efficient MINLP Strategies
  – Deterministic Global Optimization
  – Robust and Stochastic Optimization for Uncertainty

DON'T SEARCH…SOLVE!

From Nested to Equation-Oriented Optimization?



Can Equation-Oriented Optimization Solvers Handle Black Box (Truth) Models?
Yes, with Reduced Models!

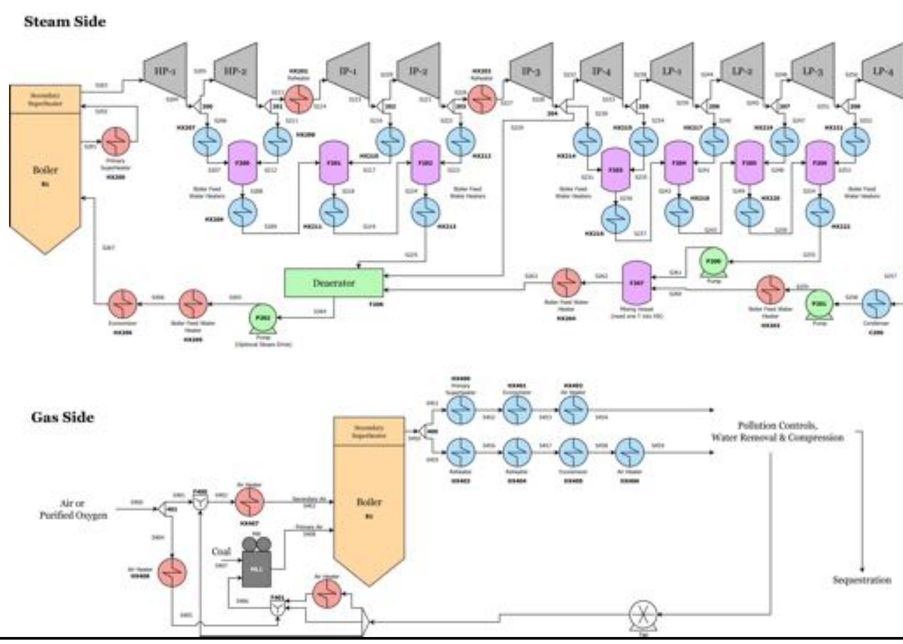
Model Hierarchies

Conservation Laws:	Often linear, straightforward to satisfy	
Physical properties:	Ideal	→ Specialized Nonideal
Separation Models:	Shortcut	→ MESH, mass transfer
Reaction Elements:	Stoichiometric	→ CFD, Multiphase

Reduced models

- In the engineering literature, common approach is to build a reduced model (RM) for the truth model (TM)
  - Also known as surrogate model, meta-model
  - Physics-based reduction: Shortcut models, POD
  - Data-driven reduction: Polynomial regression, Kriging, neural networks etc., Machine Learning
  - Provide a Glass Box RM for the Black Box TM
- Accuracy of the reduced model is a challenge
- Common practice:
  - Update reduced model during optimization process (Machine Learning)

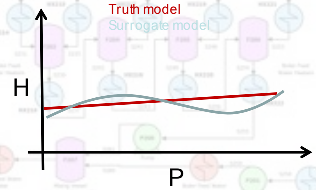
Reduced models in process optimization?



Reduced models in process optimization

Optimization Solvers Cheat!

Surrogate models may not lead to accurate optimization



Optimizer can exploit small errors

Example: Enthalpy of vapor stream

Propose surrogate $r(T,P) \approx H(T,P)$

$\frac{\partial H}{\partial P} \approx 0$ at high T and low P (equality holds for ideal gas)

However, supposed $\frac{\partial r}{\partial P} = -\epsilon$ at some T, P

This means that an arbitrarily accurate surrogate can give us ***compressors that create work!***

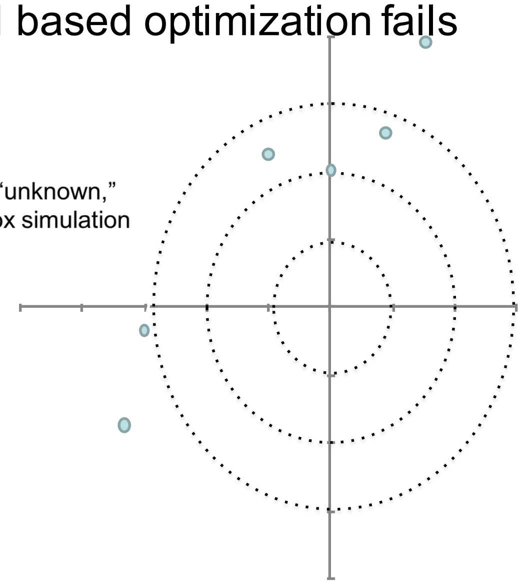
Toy example – RM based optimization fails

Truth model: $d(w)$

$\min \;\; y^2 + w^2$

$s.t. \;\; y - d(w) = 0$

$d(w)$ is "unknown," black-box simulation



Biegler et al., Computers and Chem Eng, 9, 2, (1985)

Toy example – RM based optimization fails

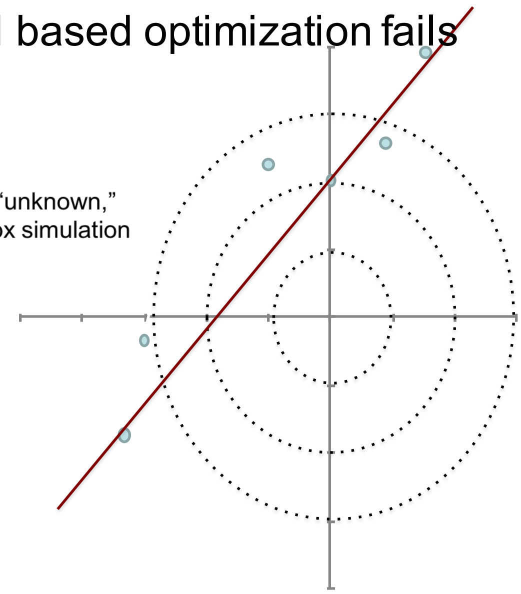
Truth model: $d(w)$

$$\min \; y^2 + w^2$$
$$s.t. \quad y - d(w) = 0$$

$d(w)$ is "unknown," black-box simulation

Reduced model: $r(w)$

$$\min \; y^2 + w^2$$
$$s.t. \quad y - r(w) = 0$$
$$r(x) = -(w + \boldsymbol{b})$$



Biegler et al., Computers and Chem Eng, 9, 2, (1985)

Toy example – RM based optimization fails

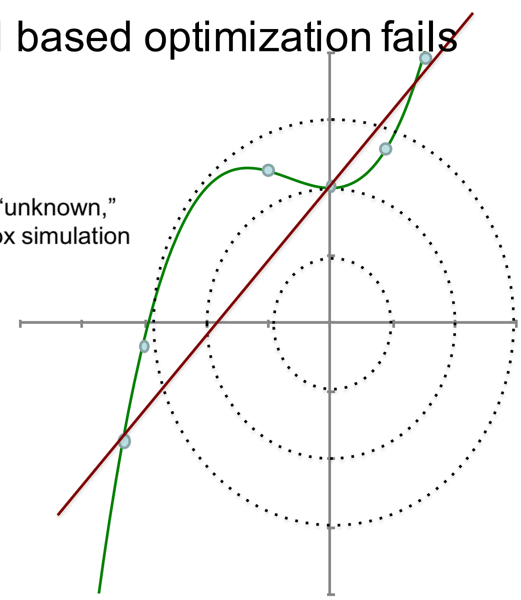
Truth model: $d(w)$

$$\min \; y^2 + w^2$$
$$s.t. \quad y - d(w) = 0$$
$$d(w) = w^3 + w^2 + 1$$

$d(w)$ is "unknown," black-box simulation

Reduced model: $r(w)$

$$\min \; y^2 + w^2$$
$$s.t. \quad y - r(w) = 0$$
$$r(x) = -(w + b)$$



Biegler et al., Computers and Chem Eng, 9, 2, (1985)

Toy example – RM based optimization fails

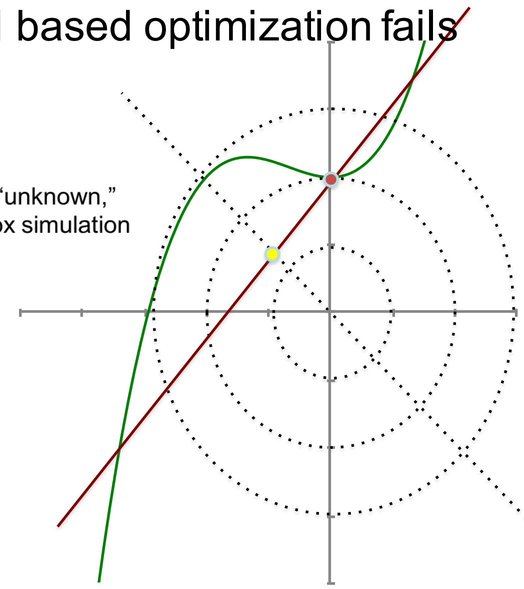
True model: $d(w)$

$$\min \; y^2 + w^2$$
$$s.t. \;\; y - d(w) = 0$$
$$d(w) = w^3 + w^2 + 1$$

$d(w)$ is "unknown," black-box simulation

Reduced model: $r(w)$

$$\min \; y^2 + w^2$$
$$s.t. \;\; y - r(w) = 0$$
$$r(x) = -(w + b)$$



Biegler et al., Computers and Chem Eng, 9, 2, (1985)

Toy example – RM based optimization fails

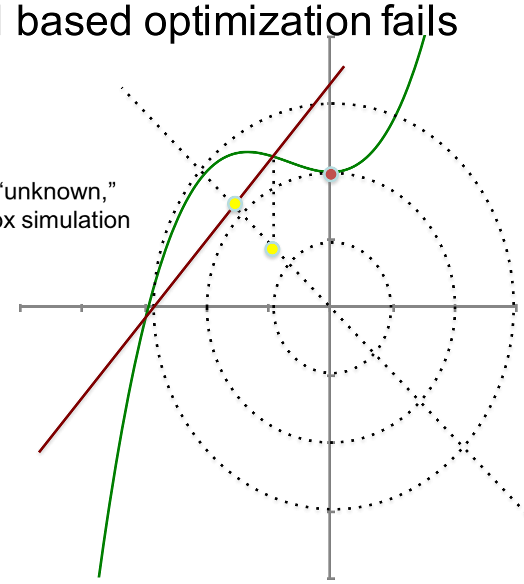
Truth model: $d(w)$

$$\min \; y^2 + w^2$$
$$s.t. \;\; y - d(w) = 0$$
$$d(w) = w^3 + w^2 + 1$$

$d(w)$ is "unknown," black-box simulation

Reduced model: $r(w)$

$$\min \; y^2 + w^2$$
$$s.t. \;\; y - r(w) = 0$$
$$r(x) = -(w + b)$$



Biegler et al., Computers and Chem Eng, 9, 2, (1985)

Toy example – RM based optimization fails

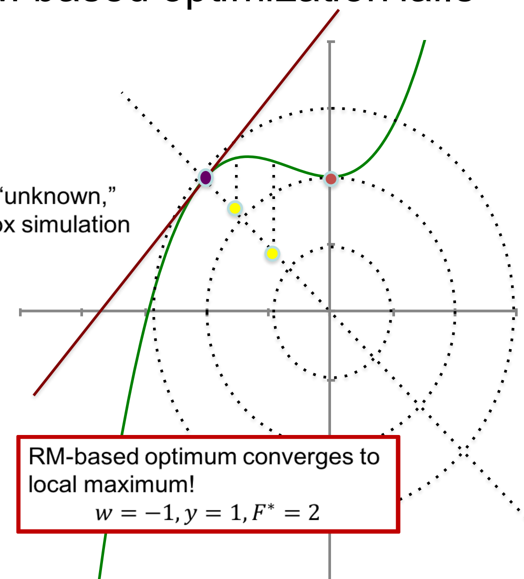
Truth model: $d(w)$

$$\min \; y^2 + w^2$$
$$s.t. \quad y - d(w) = 0$$
$$d(w) = w^3 + w^2 + 1$$

$d(w)$ is "unknown," black-box simulation

Reduced model: $r(w)$

$$\min \; y^2 + w^2$$
$$s.t. \quad y - r(w) = 0$$
$$r(x) = -(w + b)$$

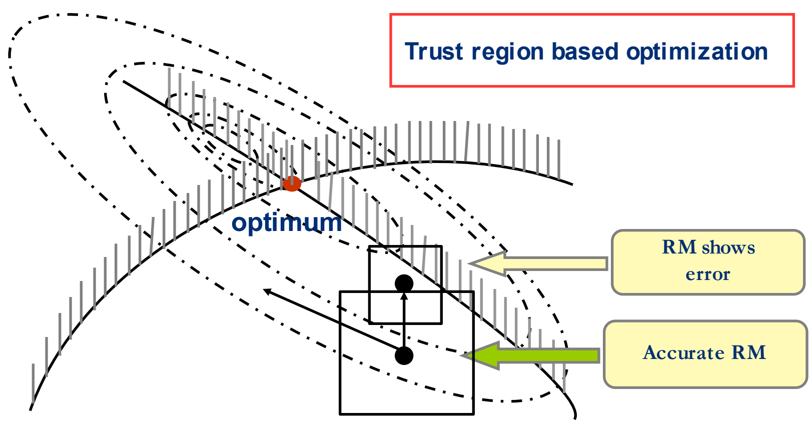


Biegler et al., Computers and Chem Eng, 9, 2, (1985)

Validate TM Optimum with RMs?

RM depends on TM information at current parameter values
TM gradients - often not available

Trust region based optimization



Validating the TM Optimum: RM Properties
(Conn, Scheinberg, Vicente, 2009)

- The key to convergence is the **κ-fully linear property**: there exist finite κ_f and κ_g such that for all iterations k,
$$\|d(w) - r_k(w)\| \le \kappa_f \Delta_k^2$$
$$\|\nabla d(w) - \nabla r_k(w)\| \le \kappa_g \Delta_k$$
- As trust region vanishes, function values and gradients approach original model
- Any type of RM may be used satisfying this property
  - Kriging, Neural Nets, Machine Learning …
  - Polynomial models applied in our applications!
- $\Delta_k \rightarrow 0$ required if TM derivatives unavailable

17

Elements of TR Subproblem
(from Fletcher, et al. 2002)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \le 0 \\ & y = r_k(w) \\ & \|x - x_k\| \le \Delta_k. \end{aligned}$$

Trust Region Subproblem

$$\begin{aligned} \min_x \quad & \|y - r_k(w)\| \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \le 0 \\ & \|x - x_k\| \le \kappa_\Delta \Delta_k \min[1, \kappa_\mu \Delta_k^\mu] \end{aligned}$$

Compatibility Check
(analogous to 'normal problem')

$$\begin{aligned} \chi(x) \quad = \quad & \left| \min_v \quad \nabla f(x)^T v \right| \\ \text{s.t.} \quad & \nabla h(x)^T v = 0 \\ & g(x) + \nabla g(x)^T v \le 0 \\ & v_y - \nabla r_k(w)^T v_w = 0 \\ & \|v\| \le 1 \end{aligned}$$

Criticality Measure

$$\theta(x) = \|y - d(w)\|.$$

$$\rho_k = \frac{ared_k}{pred_k}$$
$$= \frac{\theta(x_k) - \theta(x_k + s_k)}{\theta^r(x_k) - \theta^r(x_k + s_k)} = 1 - \frac{\theta(x_k + s_k)}{\theta^r(x_k)}$$

Trust Region Management

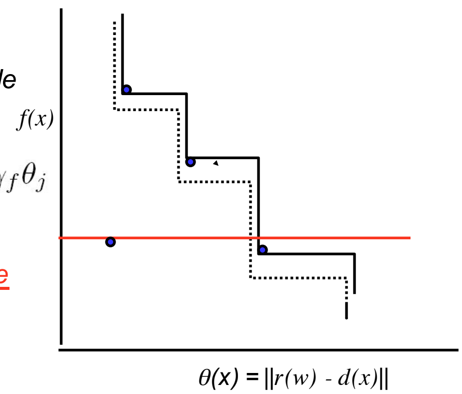
Trust Region Filter Method (Fletcher, et al., 2002)

- *Store (f_j, θ_j) at allowed iterates ➔ $\mathcal{F}$*
- *Allow TR step if trial point is acceptable to filter with θ_j margin*

$$\theta(x) \leq (1-\gamma_\theta)\theta_j \ \text{ or } \ f(x) \leq f_j - \gamma_f \theta_j$$

$$\forall(\theta_j, f_j) \in \mathcal{F}$$

- *If switching condition is satisfied (f-type step), only f(x) should be sufficiently reduced in TR.*



- Exact derivatives from TM (Agarwal, B., 2012)
- Inexact Jacobians from TM (Walther, B., 2014)
- Derivative Free Models from TM (Eason, B., 2015, 2018)

Trust Region Filter with Reduced Models

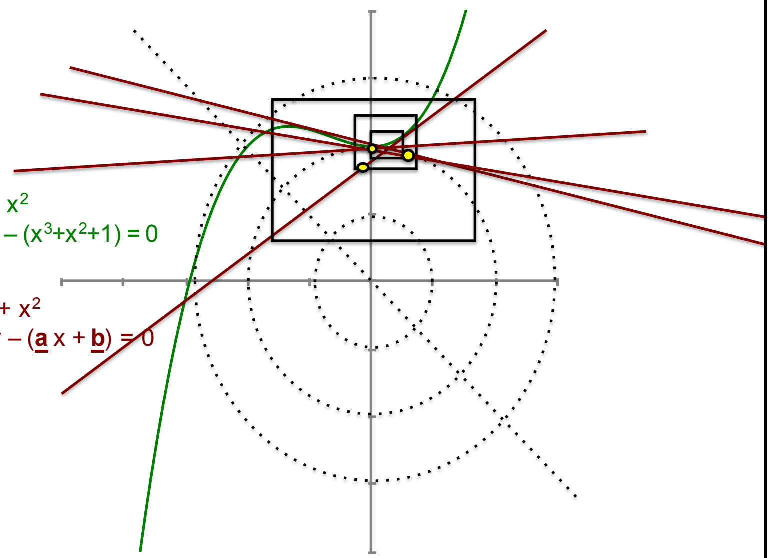
1. Initialize constants, starting guesses, trust region and filter.
2. Generate RM that is κ-fully linear on trust region.
3. Check for Compatibility. If infeasible, add (f_k, θ_k) to filter. Go to Step 9.
4. Check for convergence of TRSP. If satisfied, reduce Δ. Go to Step 2.
5. Solve TRSP to yield $x_k + s_k$. If infeasible, add (f_k, θ_k) to filter. Go to Step 9.
6. If $x_k + s_k$ is not acceptable to filter, reduce Δ. Go to step 5.
7. If s_k is f-type step, improve trust region Δ. Else, add (f_k, θ_k) to filter, update Δ.
8. Set k = k + 1, so to step 2.
9. Restoration: compute some x_{k+1} acceptable to filter (heuristics). Go to step 7.

- Algorithm is globally convergent to KKT point of TM.
- Modified proof from Fletcher et al. (2002) for nonlinear TRSP
- Steps 2-4 (criticality phase from CSV (2009))

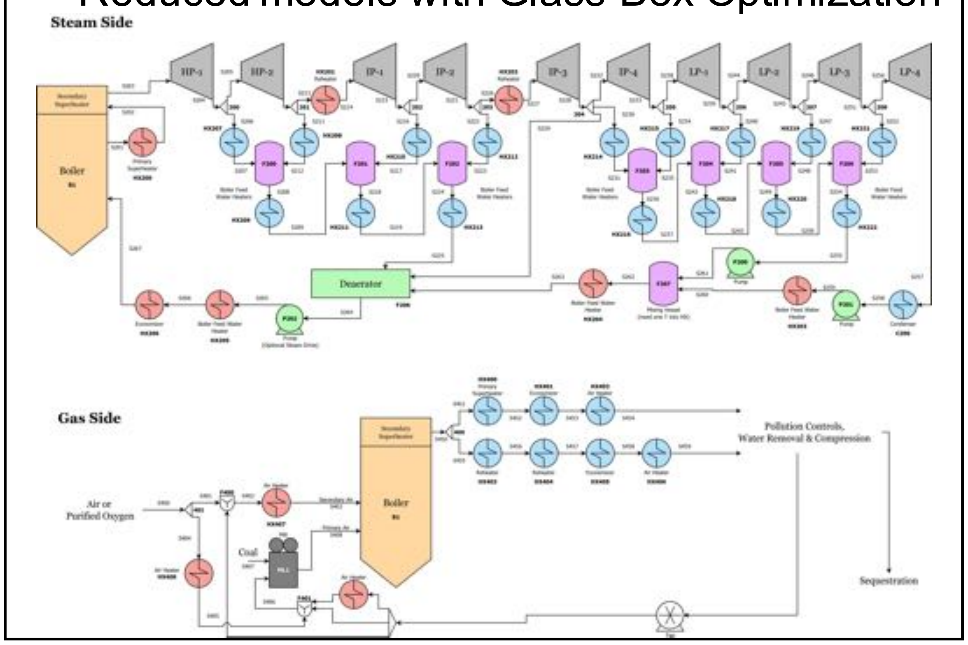
Toy Problem with Trust Region Strategy

(TM) Min $y^2 + x^2$
s.t. $y - (x^3 + x^2 + 1) = 0$

(RM) Min $y^2 + x^2$
s.t. $y - (\underline{\mathbf{a}}\, x + \underline{\mathbf{b}}) = 0$



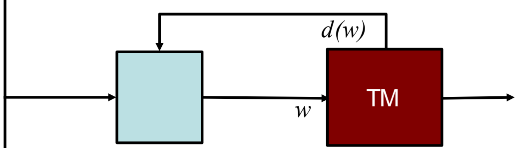
Reduced models with Glass Box Optimization



Lifting to Separate Glass and Black Box (TM) Models (Eason, B., 2016)

$$\begin{aligned} \min \quad & f(z, w, d(w)) \\ \text{s.t.} \quad & h(z, w, d(w)) = 0 \\ & g(z, w, d(w)) \le 0 \end{aligned}$$

$$x = \begin{bmatrix} w \\ y \\ z \end{bmatrix}$$

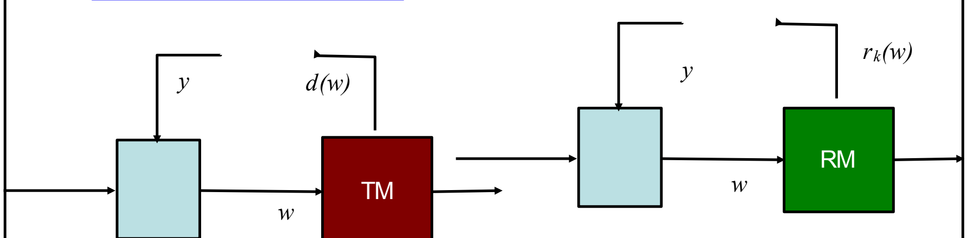


Lifting to Separate Glass and Black Box (TM) Models (Eason, B., 2016)

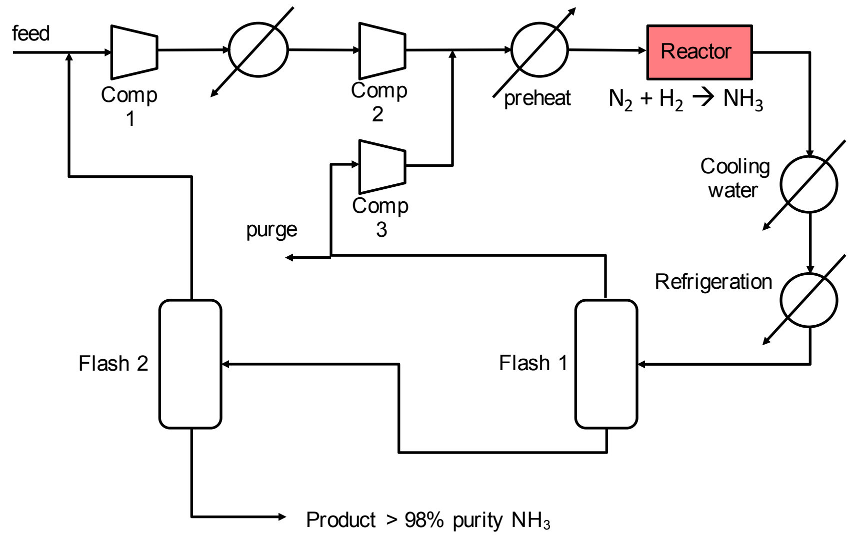
$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \le 0 \\ & y = d(w) \end{aligned}$$

$$x = \begin{bmatrix} w \\ y \\ z \end{bmatrix}$$

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \le 0 \\ & y = r_k(w) \end{aligned}$$



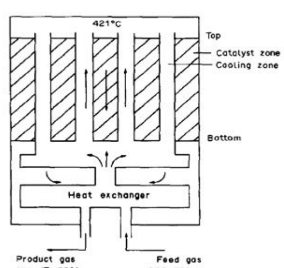
Ammonia (Haber-Bosch) Process Optimization



Ammonia Reactor Model (TM)

Kinetics	Component Balance

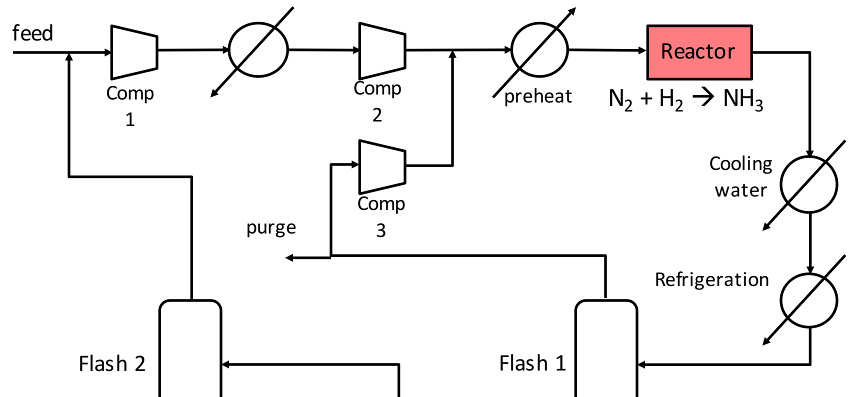
$$rate^1_{t,j} = exp(k^1_{t,j})$$
$$rate^2_{t,j} = exp(k^2_{t,j})$$
$$k^1_{t,j} = 9.7923 - 20800/(RT^g_{t,j})$$
$$k^2_{t,j} = 37.7858 - 47400/(RT^g_{t,j})$$
$$\mathcal{R}_{t,j} = rate^1_{t,j}\frac{P_{N2,t,j}(P_{H2,t,j})^{1.5}}{P_{NH3,t,j}} - rate^2_{t,j}\frac{P_{NH3,t,j}}{(P_{H2,t,j})^{1.5}}$$



$$\dot{T}^f_{t,j} = \frac{-S_1}{WC_{pf}}U(T^g_{t,j} - T^f_{t,j})$$
$$\dot{T}^g_{t,j} = \frac{-S_1}{WC_{pg}}U(T^g_{t,j} - T^f_{t,j}) + \frac{\Delta HS_2}{WC_{pg}}\mathcal{R}_{t,j}$$
$$\dot{N}_{t,j} = -\mathcal{R}_{t,j}$$

$$N_{H_2,t,j} = N^0_{H2} - 3(N^0_{N2} - N_{N_2,t,j})$$
$$N_{NH3,t,j} = N^0_{NH3} + 2(N^0_{N2} - N_{N_2,t,j})$$
$$N_{Ar,t,j} = N^0_{Ar}$$
$$N_{CH4,t,j} = N^0_{CH4}$$

Ammonia (Haber-Bosch) Process Optimization

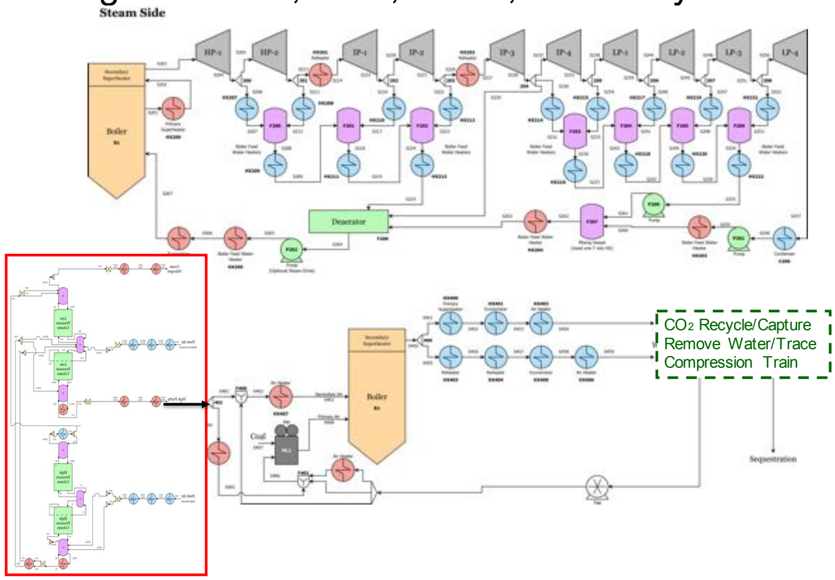


	NLP	Filter	Filter
Reduced Model	TM	Lin. Simplex Model	Lin. Exact gradient (FOC)
Objective	$8.9x10⁶	$8.9x10⁶	$8.9x10⁶
Iterations (NLPs)	1*	36	15
Reactor TM evals	*****	288	120

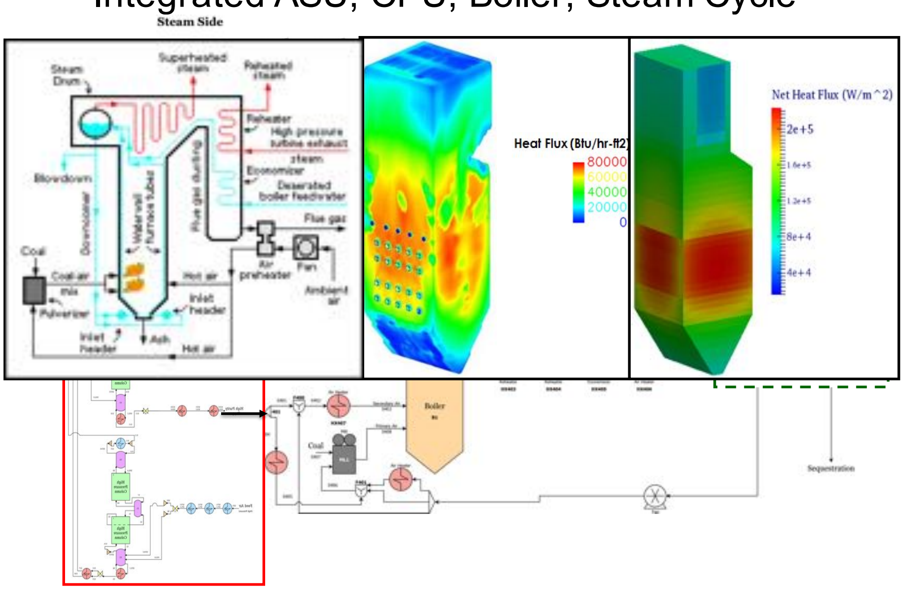
RM-based Optimization Case Studies

- Optimization of IGCC Processes with Reduced Order CFD Models,(Lang, Zitney, B., 2011)
- Fuel Cell Optimization Using Reduced Order Methods, (Jain, Jhon, B., 2011)
- Trust Region Optimization of Pressure Swing Adsorption with POD Models (Agarwal, B., 2013)
- Model-Based Control for Optimal Transitions in Polyethylene Solution Polymerization, (J. Shi, I. Hamdan, B., 2013)
- Process Optimization for Oxidative Coupling of Methane (Bremer, Esche, Wozny et al., 2014)
- Multi-scale Optimization for Process Systems Engineering, (B., Lin, Lang, 2014)
- Oxycombustion with CFD Models (Dowling, Eason, B., 2015)
- **Integrated Oxycombustion with CFD Models (Zhu, Eason, B., 2016)**

Oxycombustion Optimization:
Integrated ASU, CPU, Boiler, Steam Cycle



Oxycombustion Optimization:
Integrated ASU, CPU, Boiler, Steam Cycle



Oxycombustion Optimization
(Dowling et al., 2015)

Max **Thermal Efficiency**
s.t. Steam cycle connectivity
Heat exchanger model
Pump model
Fixed isentropic efficiency turbine model
Correlation models for ASU and CPU
Hybrid boiler model with fixed fuel rate
Steam thermodynamics

Standard supercritical steam cycle, double reheat

Use reduced models with trust region method
➔ rigorous optimum

Solved in GAMS 24.2.1 with CONOPT 3
Trust region algorithm in MATLAB R2013a

31

•Oxycombustion Compared to Air-Fired Power Plants

	Air-fired	Oxy-fired	Oxy w/IRRCs
Flue gas temperature (K)	1600	1600	1600
Steam exit temperature (K)	835	835	835
Steam exit pressure (bar)	223	223	223
Fuel rate, HHV (MW)	1325.5	1325.5	1325.5
ASU + CPU Power (MW)	N/A	**114.3**	**96.9**
Net Power (MWe)	**515.5**	**440.0**	**457.4**
Efficiency (HHV)	**38.9%**	**33.2%**	**34.5%**

5.7% penalty for oxy-fired configuration
4.4% penalty for oxy-fired with Recuperative Rankine Cycles

Summary and Conclusions

Equation-Oriented Optimization

- Fast NLP tools, parallelized, with sensitivity
- Powerful modeling tools, esp. for structured large-scale models
- Experience-based modeling and initialization

Validate the TM Optimum

- RM requirements: κ-fully linear property
- Trust region filter methods with criticality
- Global Convergence Proofs to TM Optimum

Multi-scale Optimization

- Heterogeneous truth models (CFD, DFT, MD…)
- Wealth of EO-based RMs (data-driven, short-cut, POD…)

Huge Potential for Process Optimization Applications

Many Thanks to:

Colleagues

Dr. Anshul Agarwal
Prof. Alex Dowling
Dr. Ravi Kamath
Prof. Carl Laird
Yi-dong Lang
Dr. Weijie Lin
Dr. David Miller
Prof. Andreas Wächter
Dr. Wei Wan
Prof. Victor Zavala
Dehao Zhu

Research Support

Center for Advanced Process Decision-Making (CAPD@CMU)

Dow Corporation University Partnership Initiative

US Department of Energy
- National Energy Technology Laboratory
- Sandia National Laboratories