PRAXAIR CAPD CENTER

Data-based Construction of Convex Region Surrogate (CRS) Models

Qi Zhang, Ignacio E. Grossmann

Center for Advanced Process Decision-making (CAPD),
Department of Chemical Engineering, Carnegie Mellon University

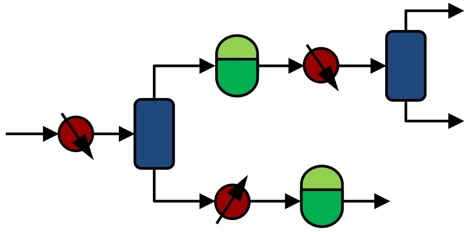
Arul Sundaramoorthy, Jose M. Pinto

Praxair Inc., Business and Supply Chain Optimization R&D

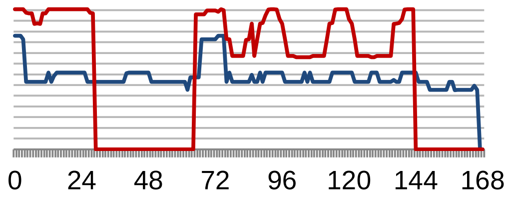
Enterprise-wide Optimization Meeting
Pittsburgh, March 2014

Integrated multiscale optimization requires **computationally efficient** and **accurate** process models.

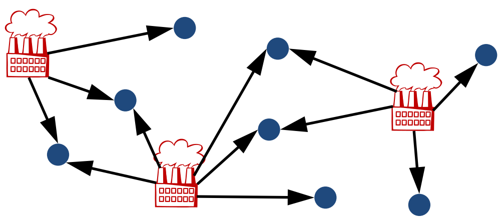
Process Synthesis



Planning and Scheduling



Supply Chain Management



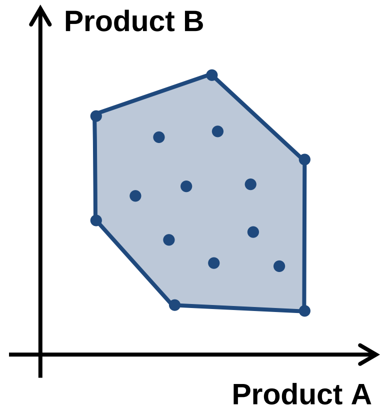
Detailed process models usually too complex to be integrated in such optimization frameworks

- **Need to construct computationally tractable but accurate surrogate models, i.e. approximate feasible region and cost correlation**
- **Require data-driven approaches suitable for the following two cases:**
  1. Existing model too difficult to be reduced but can be used to generate data
  2. No model but real process data available

Using information from given data, we approximate the process model with a union of convex regions.

Feasible region
convex

Cost function
almost linear

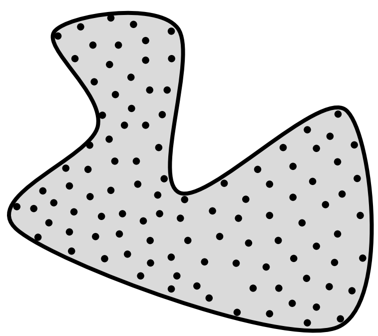


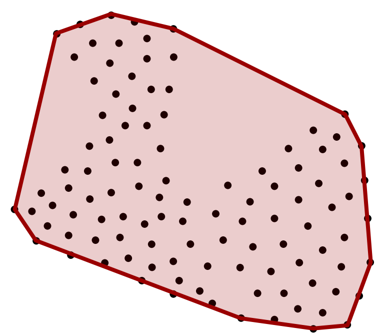
- Convex hull around all data points as feasible region[1]
- Cost correlation from linear regression
- Model remains linear and convex
- Reduce dimension by only considering relevant variables

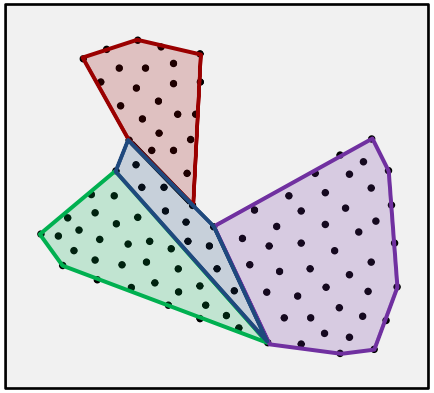
1. Karwan and Keblis (2001). *Computers and Operations Research.*

Feasible region
nonconvex

Cost function
nonlinear





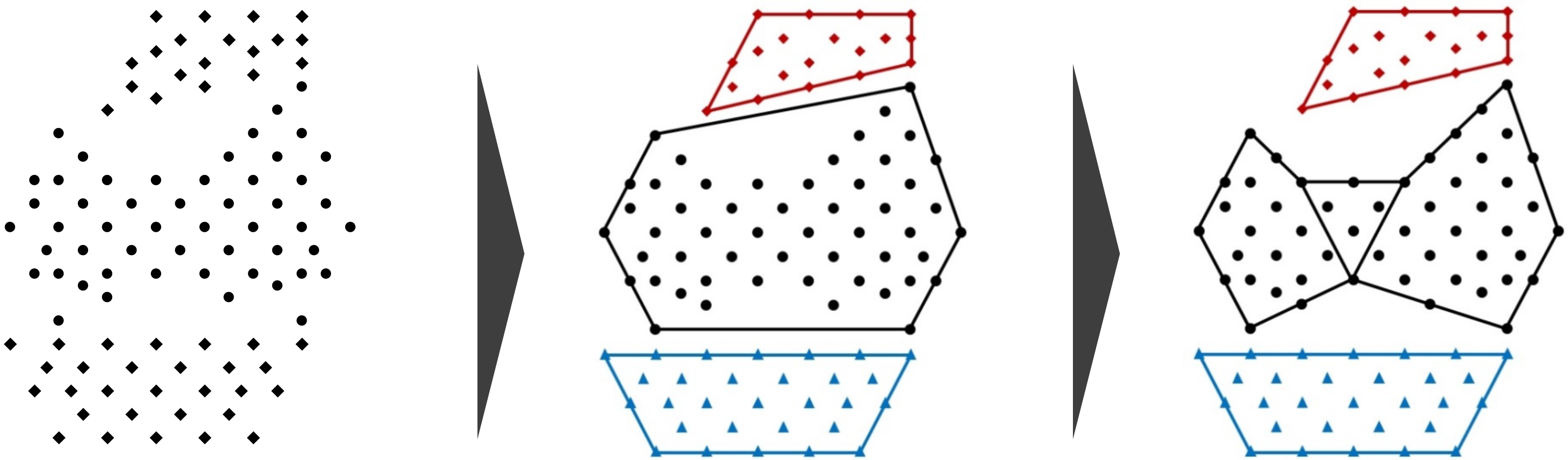


- Union of polytopes is more accurate
- Can be formulated as MILP

How do we find these convex regions?

We propose a two-phase algorithm.

- **Phase 1:** Subset assignment subject to linear parameter-cost correlation constraints
- **Phase 2:** Construction of convex regions approximating the feasible region



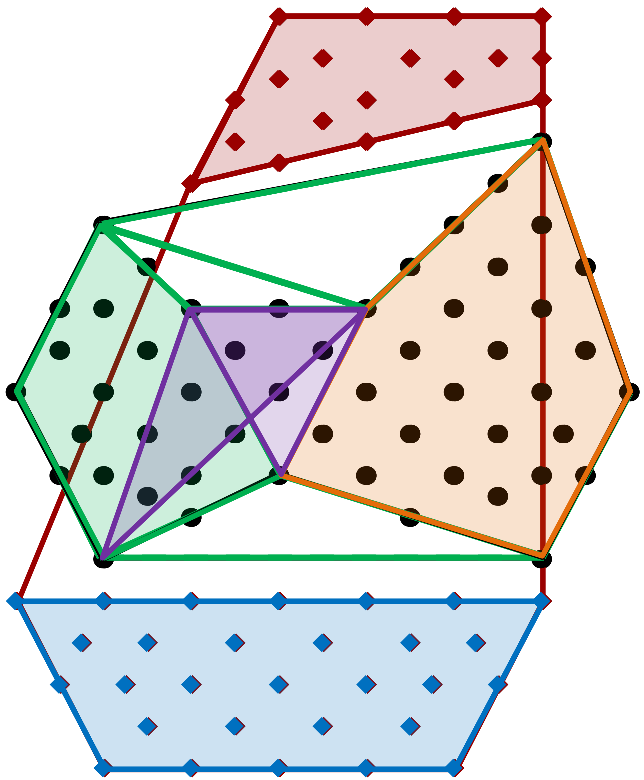
Given data points in the parameter space

At the end of Phase 1

At the end of Phase 2

Algorithm involves solving various optimization problems in an iterative framework.

Demonstration of the CRS Algorithm



Phase 1

- »» $m = 1$, infeasible → set $m = 2$, solve again
- »» feasible → construct convex hulls
- »» overlap detected → add cuts, solve again
- »» infeasible → set $m = 3$, solve again
- »» feasible → construct convex hulls
- »» no overlap → Phase 1 solution found!

Phase 2

- »» $t = 1$, examine facets, find new vertices and facets
- »» new facets created → set $t = 2$, examine facets
- »» new facets created → set $t = 3$, examine facets
- »» no new facets created → set $R = 1$, solve convex region assignment problem
- »» infeasible → set $R = 2$, solve again
- »» infeasible → set $R = 3$, solve again
- »» feasible, overlap detected → add cuts, solve again
- »» feasible, no overlap → **Solution found!**

Case Study: CRS Model of an Industrial Process.

Real process data drawn from a Praxair plant

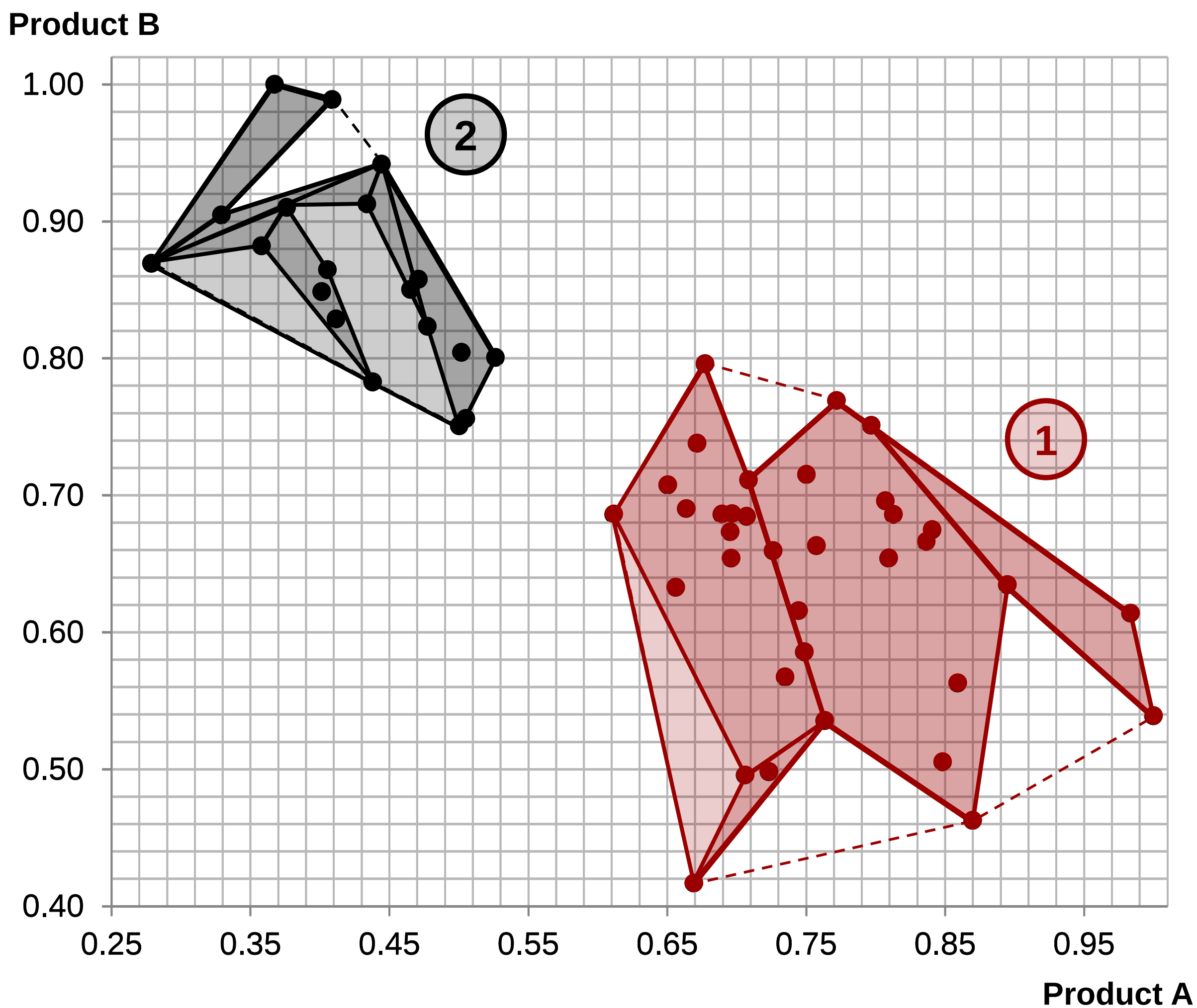
Phase 1

Coefficients for linear correlations

Set	b	c_A	c_B
1	0.900	0.062	0.000
2	0.703	0.127	0.236

Phase 2

- 51 convex regions constructed
- Specified tolerance $\epsilon = 0.06$

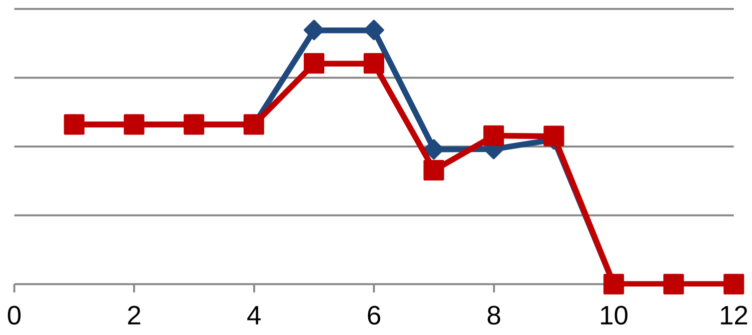


Novelty: This problem as such has not been reported before.

- **Many works on data-based modeling, but:**
  - most methods rely on nonlinear constructs to approximate nonlinearities and nonconvexities
- **Karwan and Keblis (2001).** ***Computers and Operations Research.***
  - not accurate if feasible region nonconvex and cost function nonlinear
- **Sung and Maravelias (2009).** ***AIChE Journal.***
  - not data-driven
  - makes use of the explicit model formulation

Potential Impact for Industrial Applications

- **Successfully applied to a production scheduling problem**



- With CRS model
  (CRS generated in 14 min, solved in 1.5 sec)
- With detailed nonlinear model
  (solved in 5 hr)

- **Use in more industrial applications subject to future work**
- **Need to overcome computational limitations due to**
  - higher dimensions
  - larger set of data points