Data analysis and visualisation for process design

Eric S Fraga

Centre for Computational Science
Centre for Process Systems Engineering
Department of Chemical Engineering
University College London

IChemE CAPE Subject Group meeting
Outline

1. Process design & optimisation

2. Case study
   - Analysis
   - Optimisation

3. Conclusions
Automated process design

Problem characteristics

- Complex non-linear, non-convex, discontinuous & noisy models.
- Small, possibly non-convex, feasible regions.
- Ill- or un-defined objective functions and constraint equations outside feasible regions.

Solution requirements

- Need robust optimisation methods.
- User must be able to understand the results.
- User must have confidence in the results.
Multi-objective optimisation

- Many problems are inherently multi-criteria.
- Combining these criteria into a single objective function is not always satisfactory.
- Direct solution of multi-criteria problems requires the generation of a *Pareto* front.

⇒ Requires global optimisation which is non-trivial!
Outline

1. Process design & optimisation

2. Case study
   - Analysis
   - Optimisation

3. Conclusions
Problem definition

- Wish to separate components in a stream to achieve very high purity Benzene for recycle back to main process which produces Chlorobenzene.
- Feed to purification section is:

<table>
<thead>
<tr>
<th>Component</th>
<th>Flow ( \text{kmol s}^{-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Benzene ( C_6H_6 )</td>
<td>0.97</td>
</tr>
<tr>
<td>2. Chlorobenzene ( C_6H_5Cl )</td>
<td>0.01</td>
</tr>
<tr>
<td>3. Di-Chlorobenzene ( p-C_6H_4Cl_2 )</td>
<td>0.01</td>
</tr>
<tr>
<td>4. Tri-Chlorobenzene ( C_6H_3Cl_3 )</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Pressure 1 atm

Temperature 313 K
Process flowsheet

The process structure, obtained by the Jacaranda automated design procedure, consists of three distillation units:

The post-design optimisation problem has 9 design variables (3×unit), two non-convex & discontinuous objective functions and is tightly constrained.
The optimisation model

\[
\min_x \{ f_1(x), f_2(x) \}
\]
\[
g(x) \leq 0
\]

where \( x \in \mathbb{R}^9 \) specify the reflux ratio, the recovery and the operating pressure respectively for each unit, \( f_1 \) and \( f_2 \) are the capital and operating costs of the process and \( g(x) \) the constraints:

\[
g \supset \begin{cases} 
0 \text{ or } 1 \\
\max_{d=1,2} \left\{ 0.98 - \frac{F_{d,top,benzene}(x)}{F_{d,top}(x)} \right\} \\
\max_{d=1,2} \left\{ F_{d,top,chlorobenzene}(x) - 0.005 \right\} \\
\max_{d=1,2,3} \left\{ T_{d,reboiler,dew}(x) - 503.5, \\
251.33 - T_{d,condenser,bubble}(x) \right\} 
\end{cases}
\]

Indicator of validity
Benzene purity
Chlorobenzene loss
Utility constraints

(\( d \) is the distillation unit index.)
Outline

1. Process design & optimisation

2. Case study
   - Analysis
   - Optimisation

3. Conclusions
Problem characteristics

Model validity

- $g_1(x)$ is an indicator function so the constraint $g_1(x) \leq 0$ cannot be handled directly by conventional mathematical programming techniques.
- Size of $A_0 = \{x : 0 \leq x_i \leq 1, g_1(x) \leq 0\}$ determined using Monte Carlo methods: $\text{vol}(A_0) = 0.873$.
- Therefore, model is valid almost everywhere.

Feasible region

- With $10^6$ trial points, only 6 were feasible!
- We surmise that the feasible region forms a thin layer around a surface in the hypercube defined by bounds on the variables.
Data analysis & visualisation

- 300 initial random points fed to optimizer minimising penalty function

\[ P(x) = \sum_{i=1}^{4} g_i^2(x) \]

- Present visualisation of the 9-dimensional domain using multi-dimensional distance preserving mapping.

- Interested in the boundaries of the domain so present the 512 vertices of the hypercube.
Data analysis & visualisation

- 300 initial random points fed to optimizer minimising penalty function

\[ P(x) = \sum_{i=1}^{4} g_i^2(x) \]

- Present visualisation of the 9-dimensional domain using multi-dimensional distance preserving mapping.

- Interested in the boundaries of the domain so present the 512 vertices of the hypercube.

- Vertices closest to feasible points highlighted and labelled.
Data analysis & visualisation

- 300 initial random points fed to optimizer minimising penalty function
  \[ P(x) = \sum_{i=1}^{4} g_i^2(x) \]

- Present visualisation of the 9-dimensional domain using multi-dimensional distance preserving mapping.

- Interested in the boundaries of the domain so present the 512 vertices of the hypercube.

- Vertices closest to feasible points highlighted and labelled.

- Include feasible points and re-orient display.
Reduced dimension search space

- Vertices of hypercube closest to feasible region tell us that feasible solutions should have (normalised) values close to

\[ x_3 = x_6 = x_9 = 0 \quad \text{(pressures)} \]
\[ x_5 = x_8 = 1 \quad \text{(recoveries)} \]

- Therefore, process optimisation could concentrate on \( x_1 \) (recovery), \( x_2, x_4 \) and \( x_7 \) (reflux ratios).
- However, we should not limit ourselves to 4d space alone.
- Consider further analysis.
Principle component analysis indicates that variables $x_3$ and $x_8$ have a non-negligible contribution.

As we have already identified $x_1$, $x_2$, $x_4$ and $x_7$ as key variables,

... consider, therefore, solving the 6d problem instead of the 4d one.
Outline

1. Process design & optimisation

2. Case study
   - Analysis
   - Optimisation

3. Conclusions
Generating the Pareto set

- Due to discontinuities and noise in objective functions, we compared several direct search methods: Hooke & Jeeves best overall.
- Bi-criterial problem solved using

\[ f(x, \lambda) = \lambda f_1(x) + (1 - \lambda) f_2(x) \]

for \( 0 \leq \lambda \leq 1 \), starting with \( \lambda = 0.5 \) and working outwards.
- Validate by starting from \( \lambda = 0 \) and \( \lambda = 1 \) working inwards.
Iterative optimisation procedure

- Start with 4-dimensional problem and generate an initial Pareto set.
Iterative optimisation procedure

- Start with 4-dimensional problem and generate an initial Pareto set.
- Solve the 6d problem using the 4d set as initial guesses.

![Graph showing operating cost vs capital cost](image)
Iterative optimisation procedure

- Start with 4-dimensional problem and generate an initial Pareto set.
- Solve the 6d problem using the 4d set as initial guesses.
- Solve the 9d problem using the 6d set as initial guesses.
Iterative optimisation procedure

- Start with 4-dimensional problem and generate an initial Pareto set.
- Solve the $6d$ problem using the $4d$ set as initial guesses.
- Solve the $9d$ problem using the $6d$ set as initial guesses.
- Solve $9d$ problem using both $6d$ & $9d$ sets.
Iterative optimisation procedure

- Start with 4-dimensional problem and generate an initial Pareto set.
- Solve the 6d problem using the 4d set as initial guesses.
- Solve the 9d problem using the 6d set as initial guesses.
- Solve 9d problem using both 6d & 9d sets.
Iterative optimisation procedure

- Start with 4-dimensional problem and generate an initial Pareto set.
- Solve the 6d problem using the 4d set as initial guesses.
- Solve the 9d problem using the 6d set as initial guesses.
- Solve 9d problem using both 6d & 9d sets.
Visualising the Pareto set

- Visualisation can also help understand the results.
- Multi-dimensional view shows the position of Pareto set solutions with respect to the hypercube vertices.
- We can see solutions approaching two different vertices of the hypercube.
- These solutions vary over the full range of reflux ratios.
Outline

1. Process design & optimisation

2. Case study
   - Analysis
   - Optimisation

3. Conclusions
Summary

- Optimisation problems in design are typically complex.
- Standard optimisation techniques often not suitable alone.
- Visualisation and data analysis techniques may be key to gaining insight into the problems and their solutions.
- Furthermore, these techniques may provide the basis for new targeted optimisation methods.
Acknowledgements

The following have contributed to the work presented in this seminar:

Ms Ausra Mackutė, Lithuania
Professor Antanas Žilinskas, Lithuania

and financial support from the EPSRC and the British Council is gratefully acknowledged.

Reference

Antanas Žilinskas, Eric S. Fraga, and Ausra Mackutė.
Data analysis and visualisation for robust multi-criteria process optimisation.