Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Parametric optimization and uncertainty

Optimization Under Uncertainty

- Consider uncertain variables/parameters $\boldsymbol{y}$, e.g., :
  - $\boldsymbol{y}$ is a model parameter that can at best be estimated,
  - $\boldsymbol{y}$ results from a stochastic process, which is not yet resolved.

- $\boldsymbol{y}$ can affect objective function and constraints

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{0} \end{aligned} \qquad \text{(PAR)}$$

  - $\boldsymbol{y}$ is not an optimization variable
  - if $\boldsymbol{y}$ is fixed then (PAR) is normal NLP in $\boldsymbol{x}$

- Alternative notions to handle uncertainty:
  - Parametric Optimization
  - Stochastic Programming
  - Robust Optimization

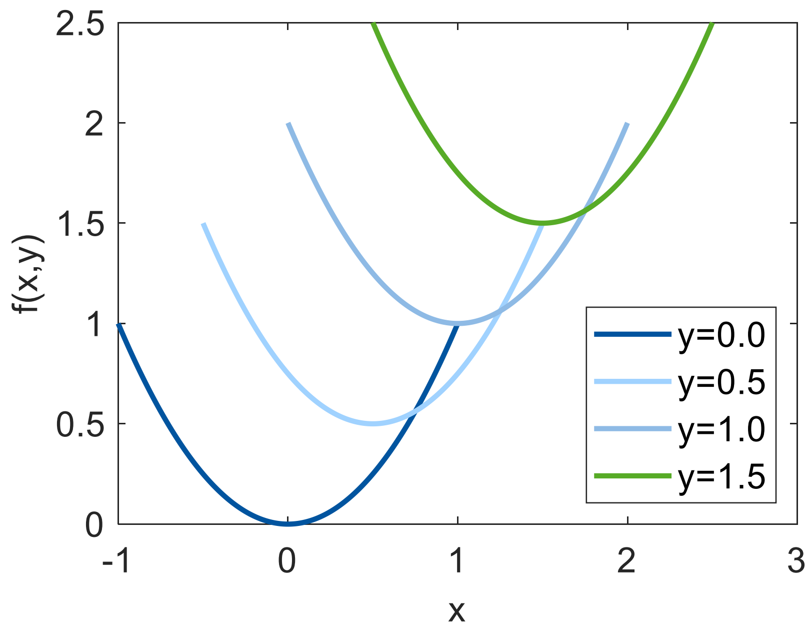
Parametric Optimization

- Parametric optimization is finding solution as function of parameter

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y}) \le \boldsymbol{0} \end{aligned} \qquad \text{(PAR)}$$

- The optimal solution point and optimal objective value depend on parameters: $\boldsymbol{x}^*(\boldsymbol{y})$ and $f^*(\boldsymbol{y}) = f(\boldsymbol{x}^*(\boldsymbol{y}), \boldsymbol{y})$

- Example:
  - $\min_x \; (x-y)^2 + y$
  - $x^*(y) = y$
  - $f^*(y) = f(x^*(y), y) = (y-y)^2 + y = y$



Parametric Optimization and Uncertainty

$$\begin{aligned} f^*(\boldsymbol{y}) = \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{0} \end{aligned} \qquad \text{(PAR)}$$

- Suppose that $\boldsymbol{y}$ is uncertain parameter

- Parametric optimization is only useful under the following conditions:
  - The uncertainty is realized before the decision variables must be fixed
  - Once the uncertainty is realized, it is not practical to solve (PAR) for fixed $\boldsymbol{y}$

  These conditions hold for example in online control and resource allocation.

- Solution approaches typically identify parameter regions for which the solution stays qualitatively same.
  - So called critical region (CR), e.g., no active set change

Example: Parametric Quadratic Optimization

Consider convex quadratic parametric optimization problem

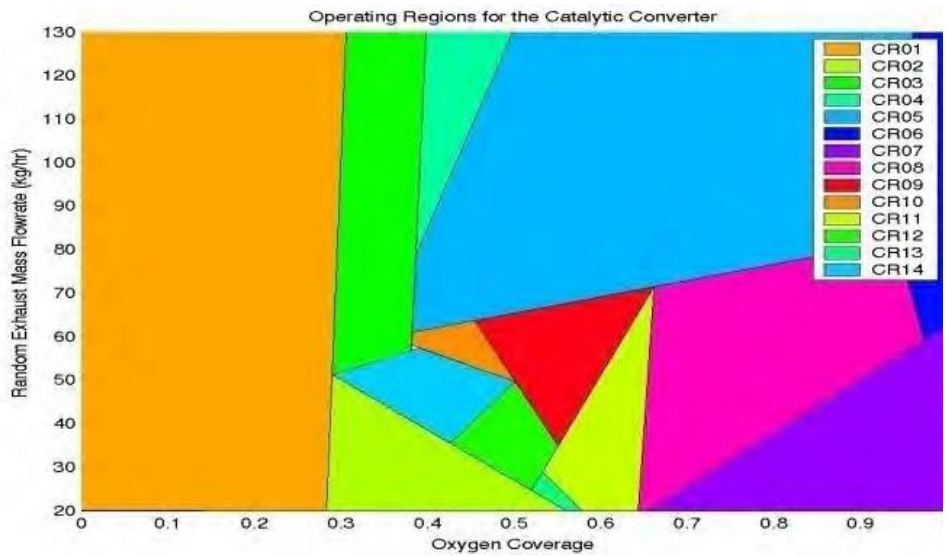
$$f^*(\boldsymbol{y}) = \min_{\boldsymbol{x} \in X} \frac{1}{2} \boldsymbol{x}^T \boldsymbol{G} \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{x} \quad \text{with} \quad \boldsymbol{y} \in Y$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{F}\boldsymbol{y}$$

- Solve for $\mathbf{y} = \bar{\boldsymbol{y}}$
- Recall that 1st order KKT are necessary and sufficient
  - Under mild assumptions
- Recall sensitivity analysis via KKT conditions
  - Yields $\boldsymbol{x}^*(\cdot)$ as an affine function of $\boldsymbol{y}$.
  - $\boldsymbol{x}^*(\cdot)$ seizes to be valid once one of the inequalities in the KKT conditions is violated.

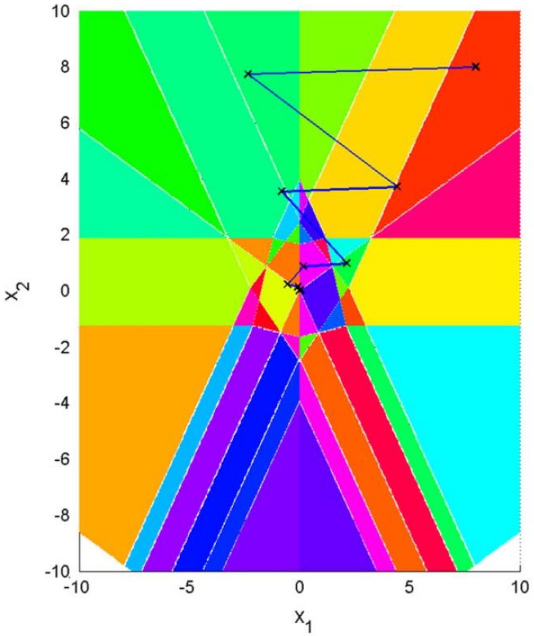
Example: Parametric Quadratic Optimization – Algorithm

1. Iterate $k = 1, 2, \ldots$
2. Solve (QPAR) with $\boldsymbol{y} \in Y$ as free variable
3. Fix $\bar{\boldsymbol{y}}$ to the solution and solve (QPAR) for $\boldsymbol{y} = \bar{\boldsymbol{y}}$
4. Derive parametric solution for current critical region $CR^{(k)}$
5. Derive constraints to define $CR^{(k)}$
6. If Y is covered with critical regions, STOP
7. Go to step 1 with $CR^{(k)}$ removed from the feasible set

$$
\begin{aligned}
f^*(\boldsymbol{y}) = \min_{\boldsymbol{x} \in X} \quad & \frac{1}{2}\boldsymbol{x}^T \boldsymbol{G}\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{F}\boldsymbol{y}
\end{aligned}
\qquad \text{(QPAR)}
$$



Picture from: Pistikopoulos, OSE 2012



Picture from: Oberdieck, Pistikopoulos, Automatica 2015

Check Yourself

- What is parametric optimization and how does it relate to uncertainty?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Introduction to optimization under uncertainty

Aachener Verfahrenstechnik | RWTH AACHEN UNIVERSITY

Optimization Under Uncertainty

- Consider uncertain variables/parameters $\boldsymbol{y}$, e.g., :
  - $\boldsymbol{y}$ is a model parameter that can at best be estimated,
  - $\boldsymbol{y}$ results from a stochastic process, which is not yet resolved.

- $\boldsymbol{y}$ can affect objective function and constraints

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y}) \leq \mathbf{0} \end{aligned}$$

  - $\boldsymbol{y}$ is not an optimization variable
  - if $\boldsymbol{y}$ is fixed then we have normal NLP in $\boldsymbol{x}$

- Alternative notions to handle uncertainty:
  - Parametric Optimization
  - Stochastic Programming
  - Robust Optimization

Alternatives to Parametric Optimization

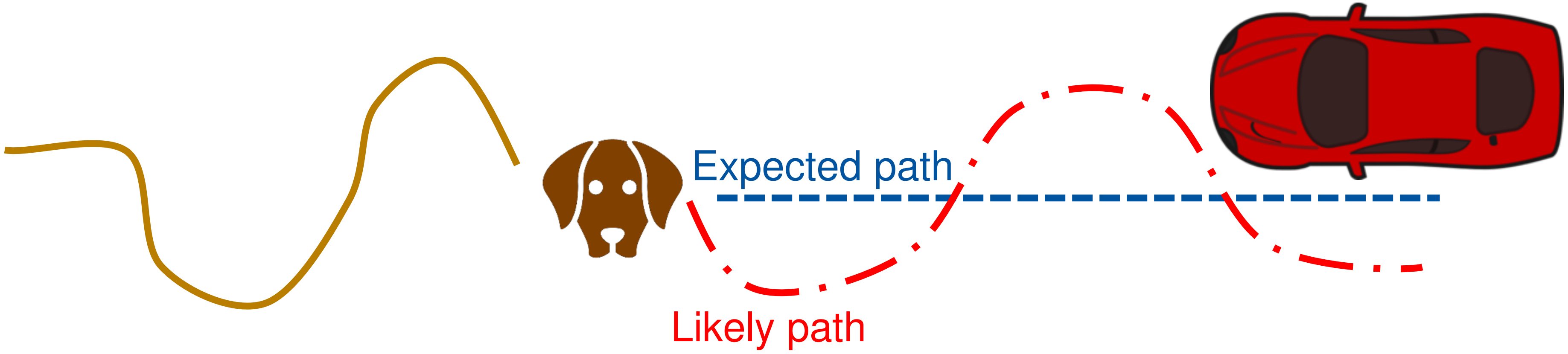
Different notions to account for uncertainty:

- Stochastic approaches consider probability measures over the set of possible uncertainty realizations
- Robust approaches consider worst-case uncertainty realization

	Stochastic	Robust
Assumption	Known probability for $\boldsymbol{y}$	$\mathbf{y}$ bounded, $\mathbf{y} \in Y$
Objective	e.g., $\min\limits_{\mathbf{x}} \mathbb{E}_Y(f(\mathbf{x}, \mathbf{y}))$	$\min\limits_{\mathbf{x}} \max\limits_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$
	Optimality in probabilistic sense	Optimality for the worst case
Constraint	e.g., $\mathbb{P}_Y(\mathbf{c}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}) \geq \alpha,\ \alpha \in (0,1)$	$\mathbf{c}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \forall \mathbf{y} \in Y$
	Chance for feasibility	Guaranteed feasibility

- Compared to neglecting uncertainties, both approaches result in more reliable and conservative solutions

Example: Optimization under Uncertainty (1)



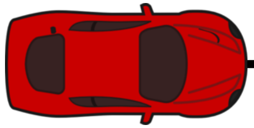
Inspired by Sam L. Savage, Stanford University

Nominal optimization:
- Consider expected path, drive straight. Likely result: 

Stochastic approach:
- Consider likely paths, drive to the side and slow down. Likely result: 

Robust approach:
- Consider all possible paths, stop the car. Guaranteed result: 

Example: Optimization under Uncertainty (2)



50% no traffic, 45min

40% moderate traffic, 60min

10% much traffic, 120min

Minimize time at airport (T@A), missed flight ⇒ 3h wait

- Nominal optimization:
  - Leave 58.5 min before flight: avg. T@A: 90min, 50% missed flights
- Stochastic approach:
  - Leave 60 min before flight: avg. T@A: 19.5min, 10% missed flights
- Robust approach:
  - Leave 120 min before flight: avg. T@A: 61.5min, 0% missed flights

Check Yourself

- What problems can arise if uncertainty is neglected in the solution of optimization problems?
- Name the basic notions of how uncertainty can be addressed. How do the approaches relate to each other?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Two-stage stochastic optimization

Aachener Verfahrenstechnik

RWTH AACHEN UNIVERSITY

Two-Stage Stochastic Programming

Consider the following problem description:

- $\boldsymbol{y} \in Y$ is an uncertain parameter with some probability distribution.
- $\boldsymbol{x} \in X$ is a decision variable that must be fixed before the uncertainty is realized with constraints $\boldsymbol{c}^{\mathrm{U}}(\boldsymbol{x}) \leq \mathbf{0}$.
- $\boldsymbol{z} \in Z$ is a decision variable that can be fixed after the uncertainty is realized with constraints $\boldsymbol{c}^{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leq \mathbf{0}$.
- An objective function $f : X \times Y \times Z \to \mathbb{R}$ is to be minimized.
- f is separable such that $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = f^{\mathrm{U}}(\boldsymbol{x}) + f^{\mathrm{L}}(\boldsymbol{y}, \boldsymbol{z})$.

- The corresponding two-stage stochastic program can be written as

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f^{\mathrm{U}}(\boldsymbol{x}) + \mathbb{E}_Y(F(\boldsymbol{x}, \boldsymbol{y})) \\ \text{s.t.} \quad & \boldsymbol{c}^{\mathrm{U}}(\boldsymbol{x}) \leq \mathbf{0} \end{aligned} \qquad \text{(ST1)}$$

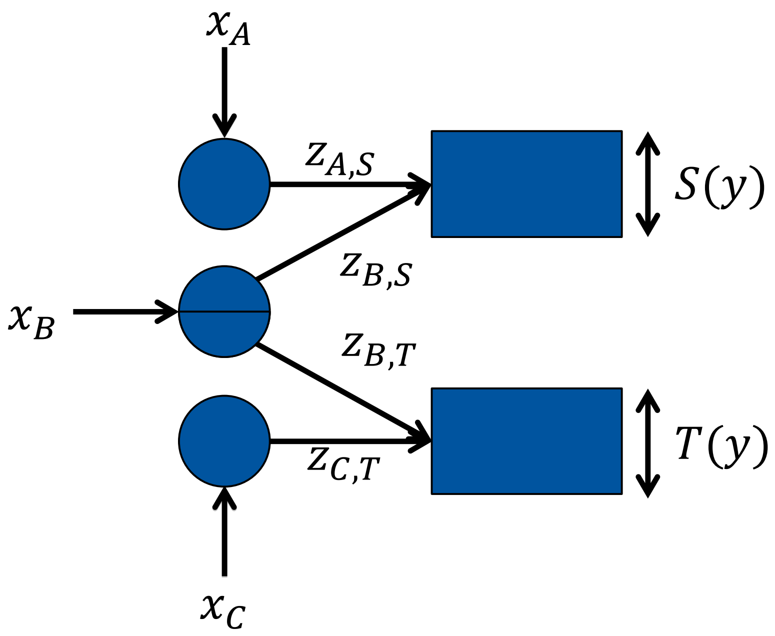
$$\begin{aligned} F(\boldsymbol{x}, \boldsymbol{y}) = \min_{\boldsymbol{z}} \quad & f^{\mathrm{L}}(\boldsymbol{y}, \boldsymbol{z}) \\ \text{s.t.} \quad & \boldsymbol{c}^{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leq \mathbf{0} \end{aligned} \qquad \text{(ST2)}$$

- (ST1) determines the fist-stage decision process before the uncertainty is realized.
- (ST2) describes the second-stage decision process for a specific uncertainty scenario.

Two-Stage Stochastic Programming Example: Plant design

Determine sizing (x_A, x_B, x_C) of different production tracks to maximize profit under uncertain demand $(S(y), T(y))$.

1. Decide on investment into tracks A, B, C: (x_A, x_B, x_C).
2. Realize uncertainty y and obtain demand: $S(y)$ and $T(y)$.
3. Decide on production of products S and T from tracks A, B, C: $(z_{A,S}, z_{B,S}, z_{B,T}, z_{C,T})$.



$$\min_{\boldsymbol{x}} \; d_A x_A + d_B x_B + d_C x_C + \mathbb{E}_Y(F(x,y))$$
$$\text{s.t.} \; x_A + x_B + x_C \le x_{max}$$
$$x_A, x_B, x_C \ge 0$$

$$F(x,y) = \min_{z} -d_S\left(z_{A,S} + z_{B,S}\right) - d_T\left(z_{B,T} + z_{C,T}\right)$$
$$\text{s.t.} \; z_{A,S} \le x_A \qquad z_{A,S} + z_{B,S} \le S(y)$$
$$z_{B,S} + z_{B,T} \le x_B \qquad z_{B,T} + z_{C,T} \le T(y)$$
$$z_{C,T} \le x_C$$

A Single-Stage Reformulation

$$\begin{aligned} &\min_{\boldsymbol{x}} f^{\mathrm{U}}(\boldsymbol{x}) + \mathbb{E}_Y(F(\boldsymbol{x},\boldsymbol{y})) \\ &\text{s.t.}\ \ \boldsymbol{c}^{\mathrm{U}}(\boldsymbol{x}) \le \boldsymbol{0} \end{aligned} \qquad \text{(ST1)}$$

$$\begin{aligned} F(\boldsymbol{x},\boldsymbol{y}) = &\min_{\boldsymbol{z}}\ f^{\mathrm{L}}(\boldsymbol{y},\boldsymbol{z}) \\ &\text{s.t.}\ \boldsymbol{c}^{\mathrm{L}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \le \boldsymbol{0} \end{aligned} \qquad \text{(ST2)}$$

- If $\boldsymbol{y}$ has a continuous or large discrete distribution, $\mathbb{E}_Y(F(\boldsymbol{x},\boldsymbol{y}))$ is difficult if not impossible to evaluate exactly.
- The distribution can be approximated by finitely many scenarios $s \in \mathcal{S}$ with $\mathbf{y} = \boldsymbol{y}_s > \mathbf{0}$ and probability of occurrence P_s such that $\sum_{s\in\mathcal{S}} P_s = 1$
- Then, the two-stage stochastic program can be approximated by the single-stage program

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{z}_s} f^{\mathrm{U}}(\boldsymbol{x}) + \sum_{s\in\mathcal{S}} P_s \cdot f^{\mathrm{L}}(\boldsymbol{y}_s,\boldsymbol{z}_s) \\ &\text{s.t.}\ \boldsymbol{c}^{\mathrm{U}}(\boldsymbol{x}) \le \boldsymbol{0} \\ &\qquad \boldsymbol{c}^{\mathrm{L}}(\boldsymbol{x},\boldsymbol{y}_s,\boldsymbol{z}_s) \le \boldsymbol{0}, \forall s \in \mathcal{S} \end{aligned}$$

- The resulting problem is potentially very large and challenging to solve.
- More efficient approaches have been proposed in literature.

Check Yourself

- What is the essential feature of a two-stage stochastic program?
- When can a two-stage stochastic program be reformulated exactly into a single-stage program?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Introduction to Semi Infinite Programs

Semi-Infinite Optimization Problems

What does the following mean?

$$\min_{\boldsymbol{x} \in X} \; f(\boldsymbol{x}) \qquad \text{(SIP)}$$
$$\text{s.t. } \; c(\boldsymbol{x}, y) \leq 0, \forall y \in Y$$

- A point $\boldsymbol{x} \in X$ is feasible if and only if $c(\boldsymbol{x}, y) \leq 0$ holds for all possible values of $y \in Y$.
- If $|Y| = \infty$, we speak of a semi-infinite optimization problem (SIP)
- For simplicity of notation we take a single SIP constraint, and a single uncertain variable $y \in Y \subset R$
- SIP has finitely many variables and infinitely many constraints
- SIPs date back to at least 1960s
- SIPs are useful for worst-case optimization="robust optimization"
- Many generalizations exist, including to $Y(\boldsymbol{x})$, existence constraints
- It is important special case of hierarchical problems (bilevel, trilevel)

Semi-Infinite Optimization Problems: A Useful Reformulation

$$\min_{\boldsymbol{x}\in X} \; f(\boldsymbol{x}) \qquad \text{(SIP)}$$
$$\text{s.t.} \;\; c(\boldsymbol{x}, y) \leq 0, \forall y \in Y$$

- If $\max_{y\in Y} c(\boldsymbol{x}, y)$ exists, we can rewrite (SIP) as

$$\min_{\boldsymbol{x}\in X} \; f(\boldsymbol{x})$$
$$\text{s.t.} \;\; 0 \geq \max_{y\in Y} c(\boldsymbol{x}, y)$$

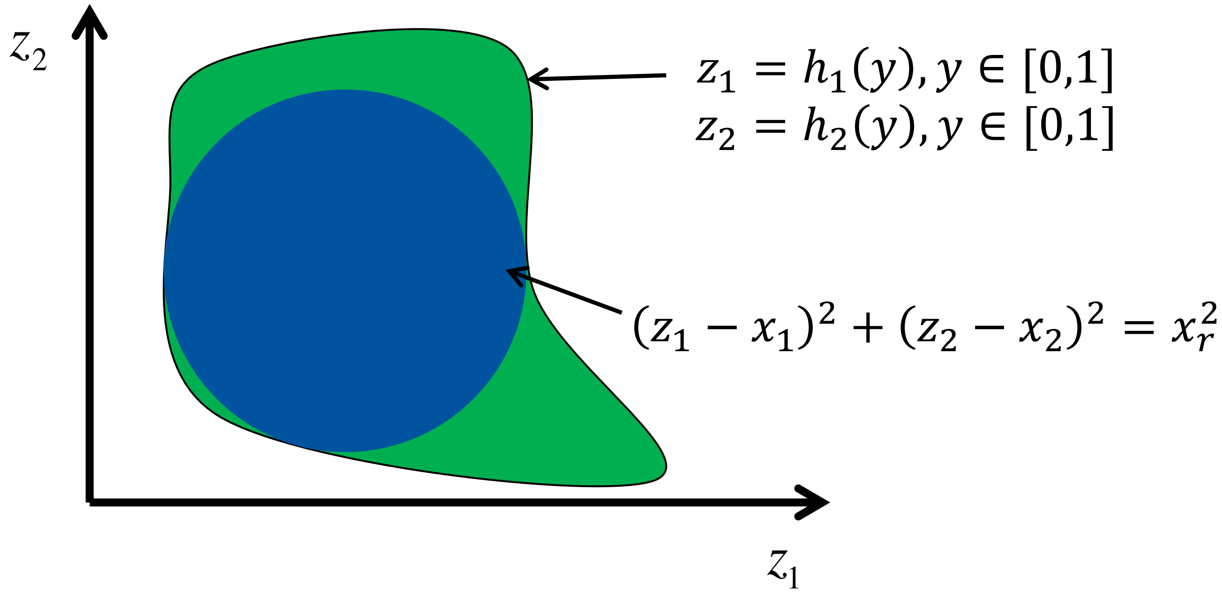
- We call $\max_{y\in Y} c(\boldsymbol{x}, y)$ the lower-level problem (LLP)
- A point $\boldsymbol{x} \in X$ is feasible if and only if the global solution to LLP is ≤ 0

Design Centering as Semi-Infinite Program

- Fit the largest Shape A that fits into Shape B

- Example: circle
  - Variables: center coordinates & radius $\boldsymbol{x} = (x_1, x_2, x_r)$
  - Objective: maximal radius $\Rightarrow \max_{\boldsymbol{x}} x_r$
  - Constraint: Shape A shall fit into Shape B

$$\Rightarrow (h_1(y) - x_1)^2 + (h_2(y) - x_2)^2 \geq x_r^2,\ \forall y \in [0,1]$$



- Real-life applications:
  - Diamond cutting: Stein, Optimization with Multivalued Mappings: Theory, Applications and Algorithms, 2006
  - Model reduction: Oluwole et. al, Combustion and Flame, 2006

Example: Path Constraints in Dynamic Optimization

- Dynamic optimization: infinite # variables
  - Handled by control-vector parametrization
- States uniquely determined as function of time by model
- Path constraints hold for $\forall\, t \rightarrow$ semi-infinite problem
- Standard algorithms & solvers enforce the path constraints for a finite # times:
  - $[0, t_f]$ is discretized to $\{t_1, t_2, \dots, t_N\} \subset [0, t_f]$
  - Relaxation that may result in violations
- Motivates the use of SIP techniques[1,2]

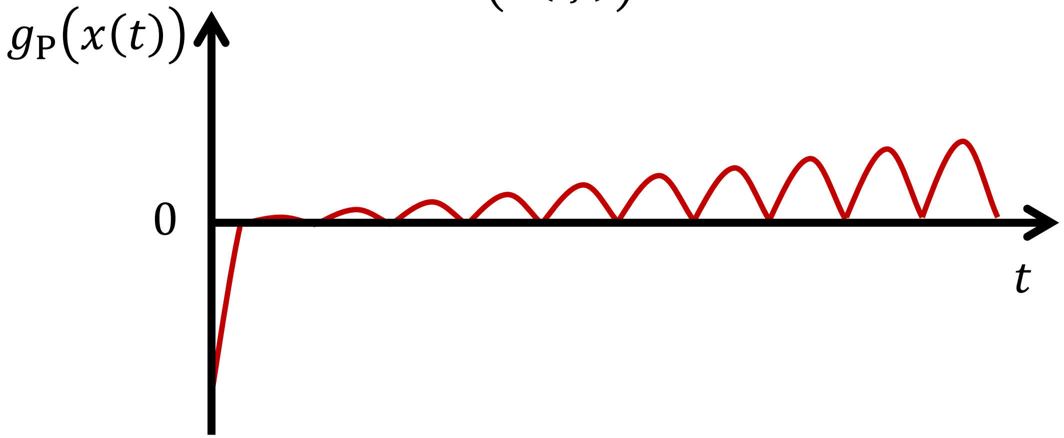
$$\min_{\boldsymbol{x}(\cdot),\boldsymbol{u}(\cdot)} \Phi\left(\boldsymbol{x}(t_f)\right)$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}}(t) = \boldsymbol{f}\left(\boldsymbol{x}(t), \boldsymbol{u}(t)\right), t \in [t_0, t_f]$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0$$

$$\boldsymbol{g}_{\mathrm{P}}\left(\boldsymbol{x}(t), \boldsymbol{u}(t)\right) \leq \boldsymbol{0}\ \forall t \in [t_0, t_f]$$

$$\boldsymbol{g}_{\mathrm{T}}\left(\boldsymbol{x}(t_f)\right) \leq \boldsymbol{0}$$



[1] Polak, SIAM Review, 1987
[2] Fu et. int Mitsos, Automatica, 2015

Example: Optimal Operation Under Parametric Uncertainty

$$\min_{\boldsymbol{x}(\cdot),\boldsymbol{u}(\cdot)} \Phi\left(\boldsymbol{x}(t_f)\right)$$

s.t. $\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{y}), t \in [t_0, t_f]$

$\boldsymbol{x}(t_0) = \boldsymbol{x}_0(\boldsymbol{y})$

$\boldsymbol{g}_{\mathrm{P}}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{y}) \leq \mathbf{0}\ \forall t \in [t_0, t_f]$

$$\min_{\boldsymbol{u}(\cdot)} \Phi\left(\boldsymbol{x}(t_f)\right)$$

s.t. $\boldsymbol{g}_{\mathrm{P}}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{y}) \leq \mathbf{0}\ \forall t \in [t_0, t_f]$

Where $\boldsymbol{x}(\cdot)$ is the solution of

$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{y}), t \in [t_0, t_f]$

$\boldsymbol{x}(t_0) = \boldsymbol{x}_0(\boldsymbol{y})$

- Recall sequential interpretation
- Worst-case formulation by introducing parametric solution of states

$$\min_{\boldsymbol{u}(\cdot)} \Phi\left(\boldsymbol{x}(t_f)\right)$$

$\boldsymbol{g}_{\mathrm{P}}\left(\boldsymbol{x}_{\boldsymbol{y}}(t), \boldsymbol{u}(t), \boldsymbol{y}\right) \leq \mathbf{0}\ \forall \boldsymbol{y}, \forall t \in [t_0, t_f]$

Where $\boldsymbol{x}(\cdot)$ is the solution of nominal $\boldsymbol{y} = \bar{\boldsymbol{y}}$

$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \bar{\boldsymbol{y}}), t \in [t_0, t_f]$

$\boldsymbol{x}(t_0) = \boldsymbol{x}_0(\bar{\boldsymbol{y}})$

and $\boldsymbol{x}_{\boldsymbol{y}}(\cdot)$ is the solution for given $\boldsymbol{y}$

$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{y}), t \in [t_0, t_f]$

$\boldsymbol{x}(t_0) = \boldsymbol{x}_0(\boldsymbol{y})$

[1] Puschke et. int Mitsos, Comp&Chem Eng, 2018

Check Yourself

- For which kind of constraint functions is an SIP particularly difficult to solve? Why is this the case?
- What are applications of SIP?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Basic solution methods for Semi Infinite Programs

Aachener Verfahrenstechnik
RWTH AACHEN UNIVERSITY

Semi-Infinite Optimization Problems: A Useful Reformulation

$$\begin{aligned} \min_{\boldsymbol{x}\in X}\ & f(\boldsymbol{x}) \\ \text{s.t.}\ & c(\boldsymbol{x},y) \le 0, \forall y \in Y \end{aligned} \qquad \text{(SIP)}$$

- If $\max_{y\in Y} c(\boldsymbol{x},y)$ exists, we can rewrite (SIP) as

$$\begin{aligned} \min_{\boldsymbol{x}\in X}\ & f(\boldsymbol{x}) \\ \text{s.t.}\ & 0 \ge \max_{y\in Y} c(\boldsymbol{x},y) \end{aligned}$$

- We call $\max_{y\in Y} c(\boldsymbol{x},y)$ the lower-level problem (LLP)
- A point $\boldsymbol{x} \in X$ is feasible if and only if the global solution to LLP is ≤ 0

Intuitive Solution Approach as Nested Problem?

- Treat semi-infinite constraint as black box function

$$c^*(\boldsymbol{x}) = \max_{y \in Y} c(\boldsymbol{x}, y)$$

- Use local solver to solve

$$\min_{\boldsymbol{x} \in X} \; f(\boldsymbol{x})$$
$$\text{s.t. } c^*(\boldsymbol{x}) \leq 0$$

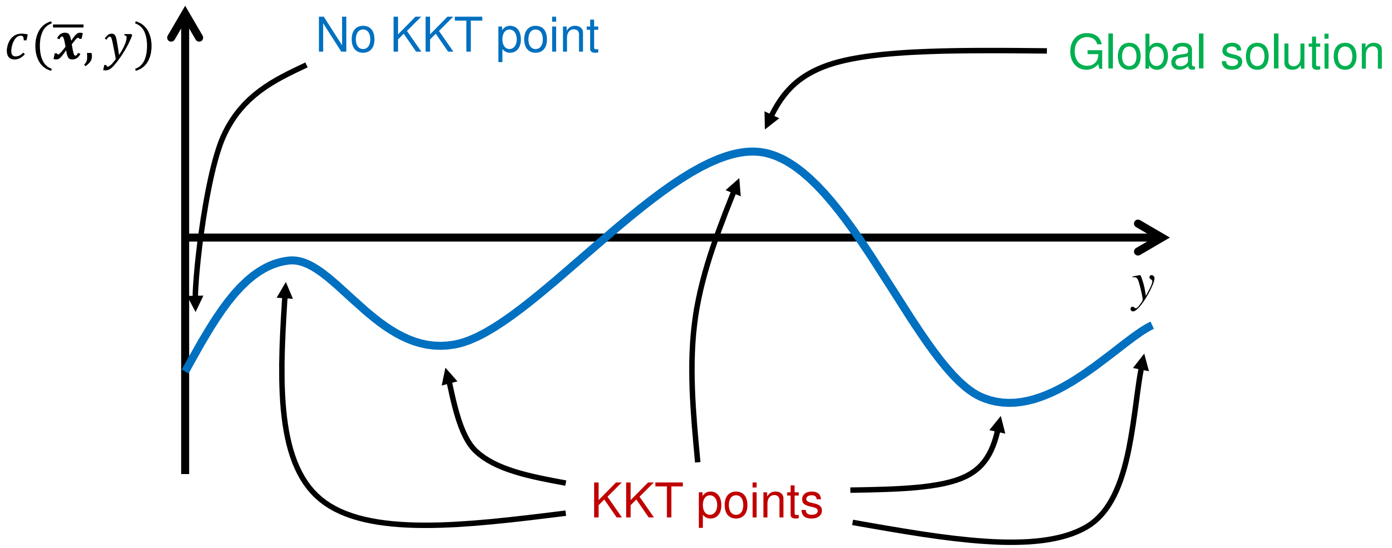
- For any particular $\boldsymbol{x} \in X$ obtain $c^*(\boldsymbol{x})$ by solving (LLP) globally

- Problems
  - Gradient-based methods are not applicable: $c^*(\cdot)$ is not differentiable
  - Gradient-free local solvers evaluate many points $\boldsymbol{x} \in X$ requiring many global solutions of (LLP)
  - Parametric solution of $c^*(\cdot)$ is extremely expensive

- Nonsmooth local solvers of interest

Intuitive Solution Approach using Optimality Conditions of LLP?

- Replace the lower-level problem with its KKT conditions
- The resulting problem is finite and can be solved by NLP methods
  - Recall that the KKT conditions have nonsmoothness
- In general wrong since KKT conditions are **not sufficient for the global** maximization of LLP



- Relaxation of the SIP: resulting point is not necessarily feasible in the SIP

Solution Methods of SIP: Local Reduction and Discretization

1. Local reduction

- For a point $\bar{\boldsymbol{x}} \in X$, find all KKT points of the LLP
- Track all KKT points
- The resulting problem is finite and can be solved by NLP methods

2. Discretization

- Replace Y with a finite discretization $Y^D \subset Y$
- The resulting problem is a finite approximation of the SIP and can be solved by NLP methods
- The resulting problem is a relaxation
- The discretization Y^D is populated to better approximate the SIP
- The relaxation can be tightened by using also KKT conditions of LLP

Check Yourself

- For which kind of constraint functions is an SIP particularly difficult to solve? Why is this the case?
- Why not solve the nested problem directly?
- Name the basic approaches to the solution of SIPs.
- How does the discretized version of an SIP relate to the original problem?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Deterministic global solution of Semi Infinite Programs

Aachener Verfahrenstechnik
RWTH AACHEN UNIVERSITY

Semi-Infinite Optimization Problems: A Useful Reformulation

$$\min_{\boldsymbol{x}\in X} f(\boldsymbol{x}) \quad \text{s.t.} \quad c(\boldsymbol{x},y) \leq 0, \forall y \in Y \qquad \text{(SIP)}$$

- If $\max_{y\in Y} c(\boldsymbol{x},y)$ exists, we can rewrite (SIP) as

$$\min_{\boldsymbol{x}\in X} f(\boldsymbol{x}) \quad \text{s.t.} \quad 0 \geq \max_{y\in Y} c(\boldsymbol{x},y)$$

- We call $\max_{y\in Y} c(\boldsymbol{x},y)$ the lower-level problem (LLP)

- Goal: solve the SIP globally using discretization-based lower & upper bounding procedures
  - up to optimality gap $f^{UBD} - f^{LBD} \leq \varepsilon^f$
  - Assuming an SIP Slater point, continuous functions and compact X, Y

Mitsos, Optimization, 2011

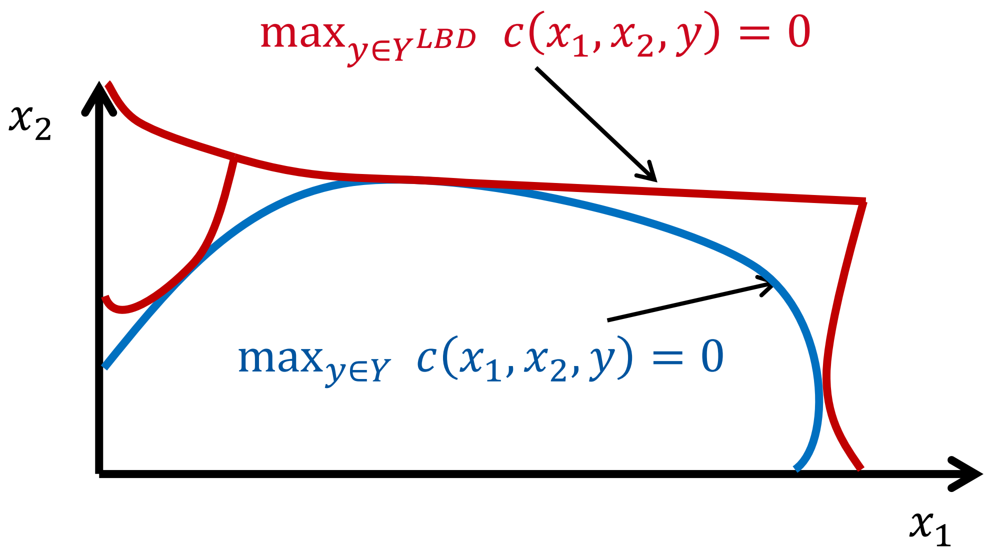
Discretization Approach from Blankenship & Falk

- Approximate Y with $Y^{LBD} \subset Y$, $|Y^{LBD}| < \infty$ to lower bounding problem (LBD)
  - The LLP is restricted and thus the SIP relaxed
  - Outer approximation

$$\begin{aligned} \min_{\boldsymbol{x} \in X} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c(\boldsymbol{x}, y) \leq 0, \ \forall y \in Y^{LBD} \end{aligned} \qquad \text{(LBD)}$$

- Solve (LBD) and obtain $\bar{\boldsymbol{x}}$

- Solve (LLP) **globally** for $\boldsymbol{x} = \bar{\boldsymbol{x}}$ to alternatively obtain
  - $c^*(\bar{\boldsymbol{x}}) \leq 0$, $\bar{\boldsymbol{x}}$ is SIP feasible
  - $\bar{y}$: $c(\bar{\boldsymbol{x}}, \bar{y}) > 0$



Blankenship and Falk, J. Optim. Theory Appl.,1976

Lower Bounding Procedure

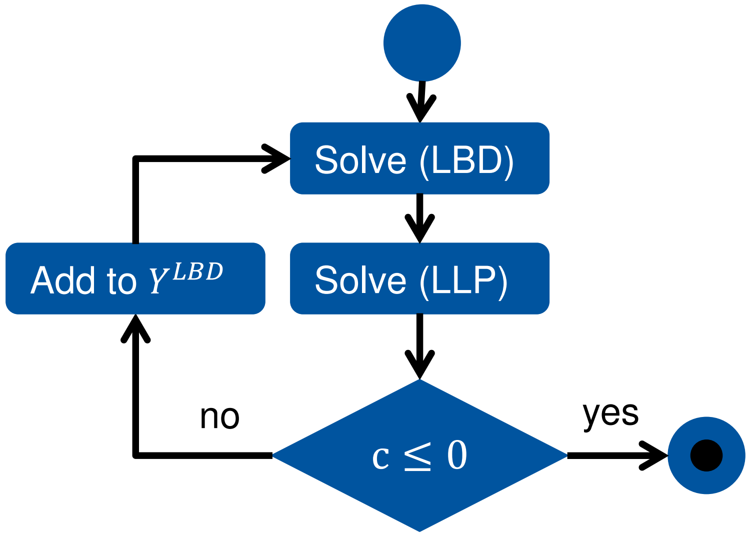
- (LBD) is a relaxation of (SIP) for any $Y^{LBD} \subset Y$
- The global solution of (LBD) gives a lower bound on the globally optimal objective value of (SIP)

- Algorithm:

1. Initialize $Y^{LBD} \subset Y, f^{LBD} \leftarrow -\infty$

2. Solve (LBD) globally to obtain $\bar{x}$
   - Set $f^{LBD} \leftarrow f(\bar{\boldsymbol{x}})$

3. Solve (LLP) globally for $\bar{\boldsymbol{x}}$ to obtain $\bar{y}$
   - If $c(\bar{\boldsymbol{x}}, \bar{y}) \leq 0$ then set $\boldsymbol{x}^* \leftarrow \bar{\boldsymbol{x}}$ and terminate ($\boldsymbol{x}^*$ is feasible)
   - Else set $Y^{LBD} \leftarrow \{\bar{y}\} \cup Y^{LBD}$ and go to step 2



- The algorithm converges to the global solution to (SIP) in the limit
- Finite generation of a feasible point is not guaranteed

Discretization with Guaranteed Feasibility

- Recall (LBD)

$$\begin{aligned} \min_{\boldsymbol{x}\in X} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c(\boldsymbol{x},y) \leq 0, \forall y \in Y^{LBD} \end{aligned} \qquad \text{(LBD)}$$

- Restrict ($\varepsilon > 0$) the right-hand-side of the constraint to obtain the upper bounding problem (UBD)[1]

$$\begin{aligned} \min_{\boldsymbol{x}\in X} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c(\boldsymbol{x},y) \leq -\varepsilon, \forall y \in Y^{UBD} \end{aligned} \qquad \text{(UBD)}$$

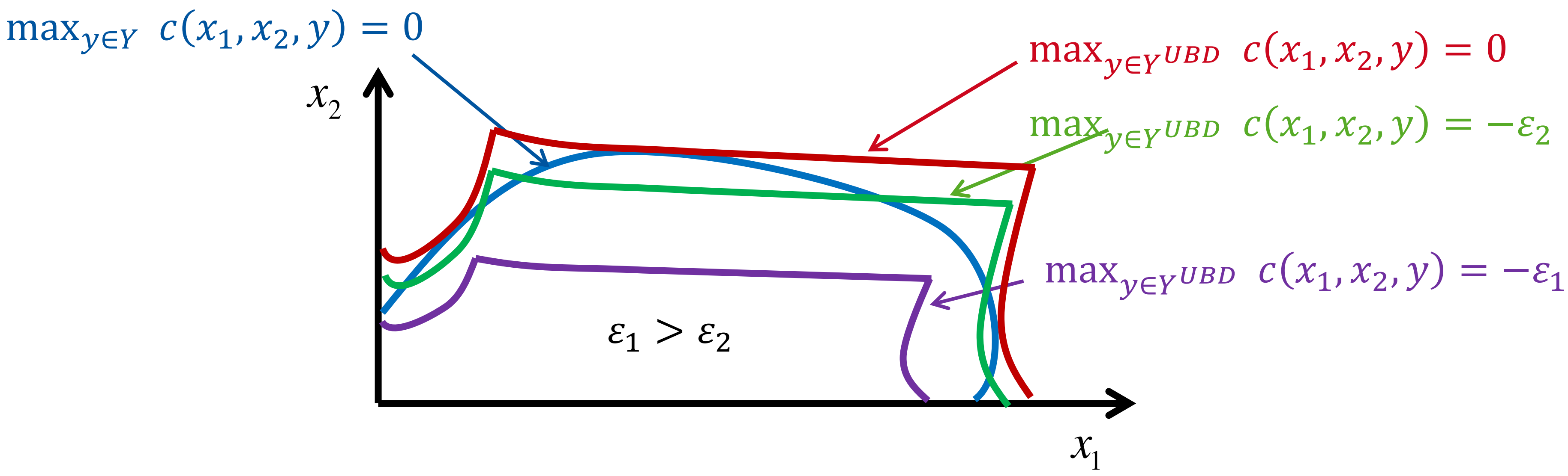
- (UBD) is neither a relaxation nor a restriction of (SIP)

[1] Mitsos, Optimization, 2011

Upper Bounding Problem

- (UBD) is neither a relaxation nor a restriction of (SIP)

$$\min_{\boldsymbol{x}\in X} \quad f(\boldsymbol{x}) \qquad \text{(UBD)}$$
$$\text{s.t.} \quad c(\boldsymbol{x}, y) \le -\varepsilon, \forall y \in Y^{UBD}$$



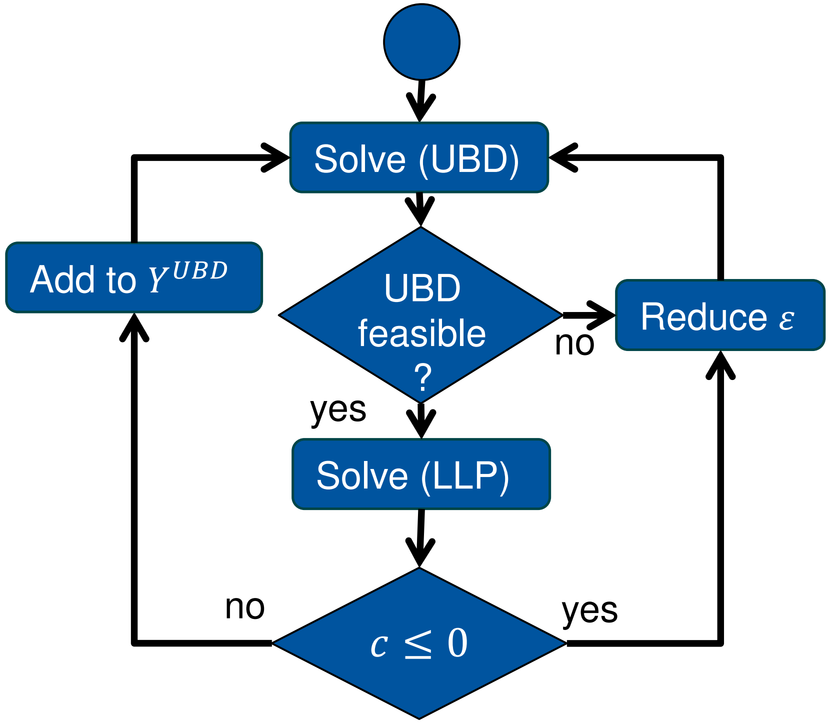
- By populating Y^{UBD}, the relaxation is made tighter → better approximate (SIP)
- By reducing $\varepsilon > 0$, the restriction is less severe → better approximate (SIP)

[1] Mitsos, Optimization, 2011

Simplified Upper Bounding Procedure

- Algorithm:

1. Initialize $Y^{UBD} \subset Y, f^{UBD} \leftarrow \infty, \varepsilon > 0, r > 1$
2. Solve (UBD) globally to obtain $\bar{\boldsymbol{x}}$
3. Solve (LLP) globally for $\bar{\boldsymbol{x}}$ to obtain $\bar{y}$
   - If $c(\bar{\boldsymbol{x}}, \bar{y}) \leq \mathbf{0}$ then set $\boldsymbol{x}^* \leftarrow \bar{\boldsymbol{x}}$ and $\varepsilon \leftarrow \varepsilon / r$
   - Else set $Y^D \leftarrow \{\bar{y}\} \cup Y^{UBD}$
   - Go to step 2



- Finite convergence to global optimum if SIP-Slater point exists
  – If for $\varepsilon^f > 0$, $\exists \boldsymbol{x}^S \in X$: $c(\boldsymbol{x}^S, y) < 0$, $\forall y \in Y$, $f(\boldsymbol{x}^S) \leq f^* + \varepsilon^f$ the procedure finitely produces an ε^f-optimal, feasible point

Check Yourself

- How does the discretized version of an SIP relate to the original problem?
- Describe a global solution method for SIP

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Parameter estimation in thermodynamics as bilevel/SIP

Aachener Verfahrenstechnik

RWTH AACHEN UNIVERSITY

Parameter Estimation for Thermodynamic Property Models

- Accurate prediction of phase separation behavior is needed, e.g., in separation design.
- Excess-Gibbs free energy models are used for liquid phases $G(T, P, \boldsymbol{x}, \boldsymbol{q})$
  - NRTL, UNIQUAC, Wilson, … :
- EOS models are used for vapor phases
  - No adjustment of parameters for binary interaction
- Binary interaction parameters $\boldsymbol{q}$ estimated via equilibrium experiments
- Parameters must be fit for model equations in equilibrium state.
  - Necessary equilibrium criterion: Isopotential (equality of chemical potential between phases)
  - Necessary and sufficient equilibrium criterion: $\min_{\boldsymbol{x}} G(T, P, \boldsymbol{x}, \boldsymbol{q})$
    - Specialized criteria exist: Gibbs tangent plane (Michelsen 1980) supporting hyperplane (Mitsos & Barton AIChEJ 2007)

Parameter Estimation: Classical approach

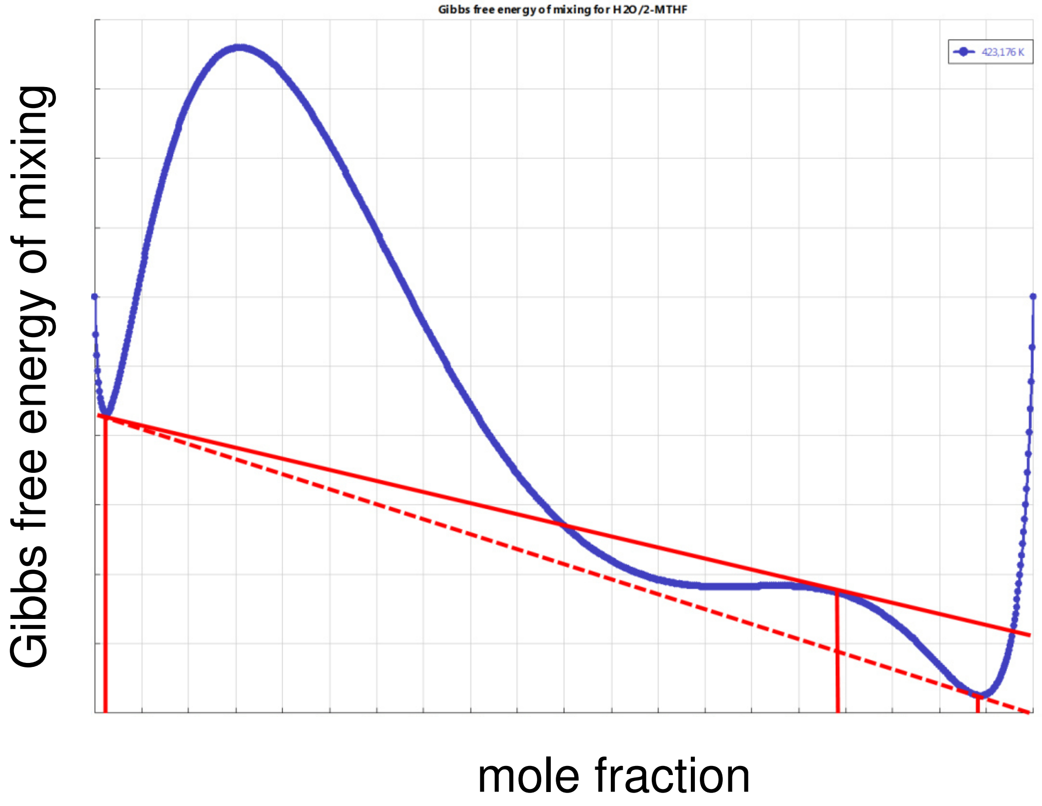
- Measure phase compositions $x^{m,i,k}$ at different temperatures T^i and constant pressure P.

- Minimize prediction-measurement discrepancy

- Enforce isopotential for thermodynamic equilibrium

- $$\min_{\boldsymbol{q}} LS(x^{p,i,k} - x^{m,i,k})$$
  $$\mu_1^1(T^i, P, x^{p,i,1}, \boldsymbol{q}) = \mu_1^k(T^i, P, x^{p,i,k}, \boldsymbol{q}) \ \forall\, i, k$$
  $$\mu_2^1(T^i, P, x^{p,i,1}, \boldsymbol{q}) = \mu_2^k(T^i, P, x^{p,i,k}, \boldsymbol{q}) \ \forall\, i, k$$

- Problem: isopotential is only **necessary**, not sufficient.



Glass, et int., Mitsos, *Fluid Phase Equilibria* 433 (2017): 212-225.

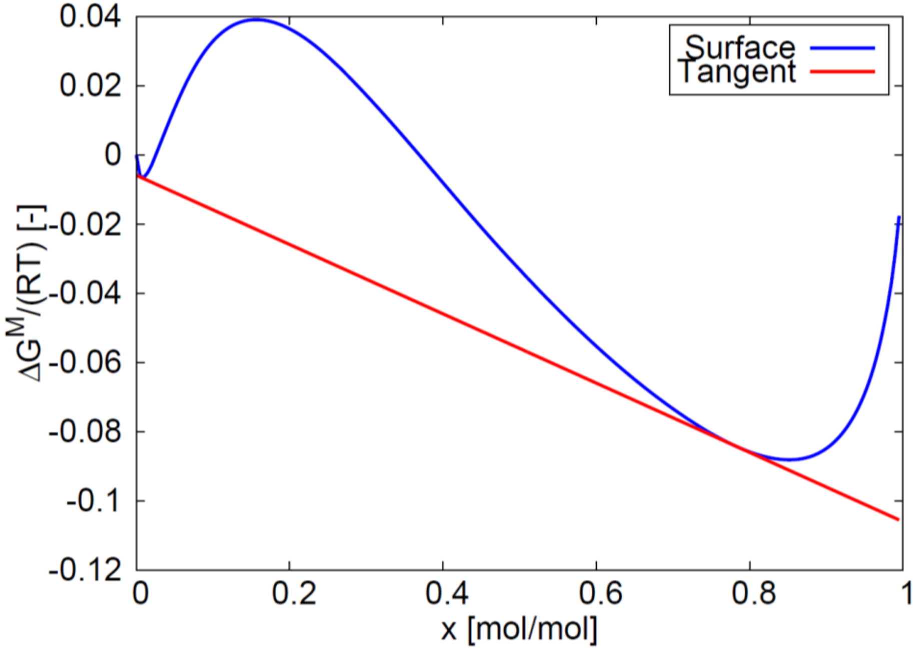
Parameter Estimation: Bilevel Formulation

- Developed by Mitsos, Bollas & Barton CES 2009
  - Extended by Glass & Mitsos
  - Implemented in BOARPET

- Introduce tangent plane criterion

$$\Delta g\left(T^i, P^i, x^{p,i,1}, \boldsymbol{q}\right) + \left.\frac{\delta \Delta g}{\delta x}\right|_{T^i, P^i, x^{p,i,1}, \boldsymbol{q}} \left(x - x^{p,i,1}\right) \leq \Delta g\left(T^i, P^i, x, \boldsymbol{q}\right) \quad \forall x \in [0,1]$$

- Additional constraints:
  - correct number of phase splits
  - correct number of phases for each split



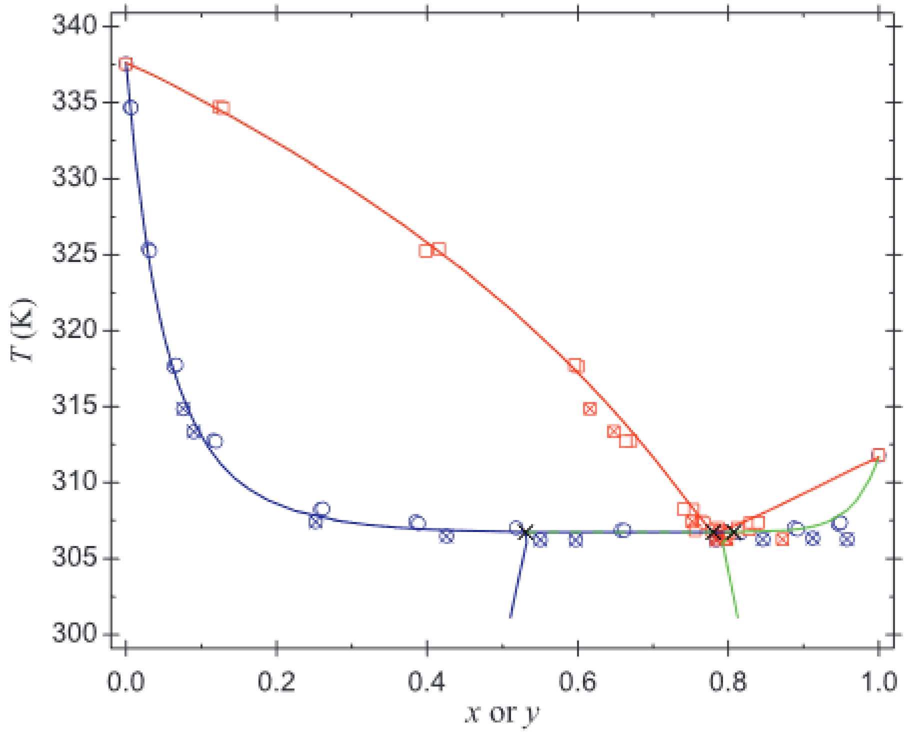
Glass, et int., Mitsos, *Fluid Phase Equilibria* 433 (2017): 212-225.

https://www.avt.rwth-aachen.de/cms/AVT/Forschung/Software/~kvkz/BOARPET/

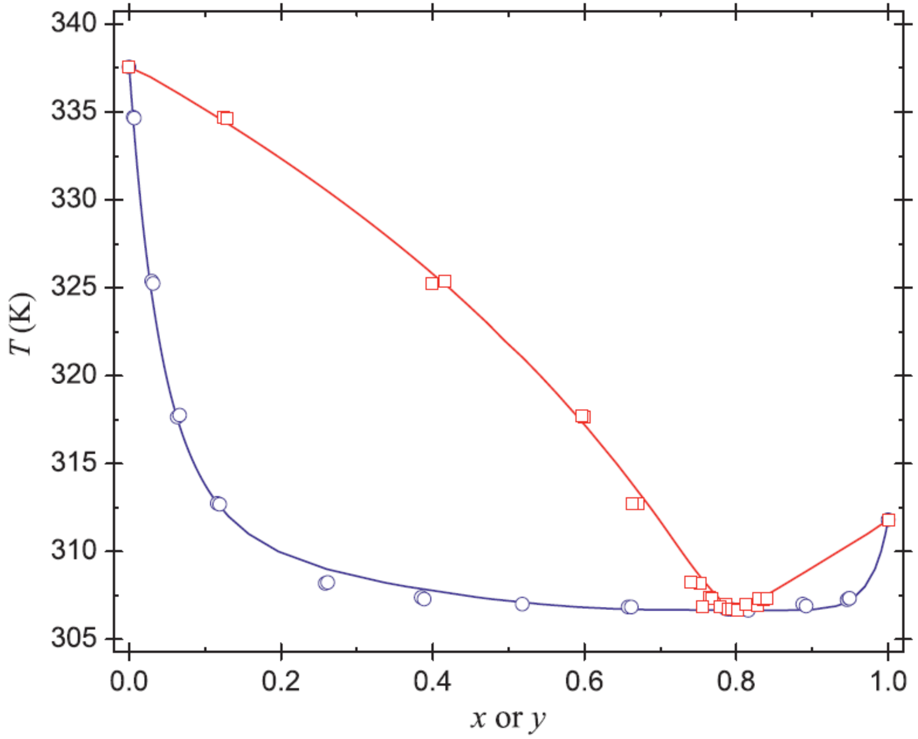
Parameter Estimation: VLE of 2-methyl-2-butene – methanol

- Experimental data shows homogeneous azeotrope

$\gamma - \phi$ method predicts heterogeneous azeotrope (spurious LLE split).



Bilevel formulation gives qualitatively and quantitatively correct fit.

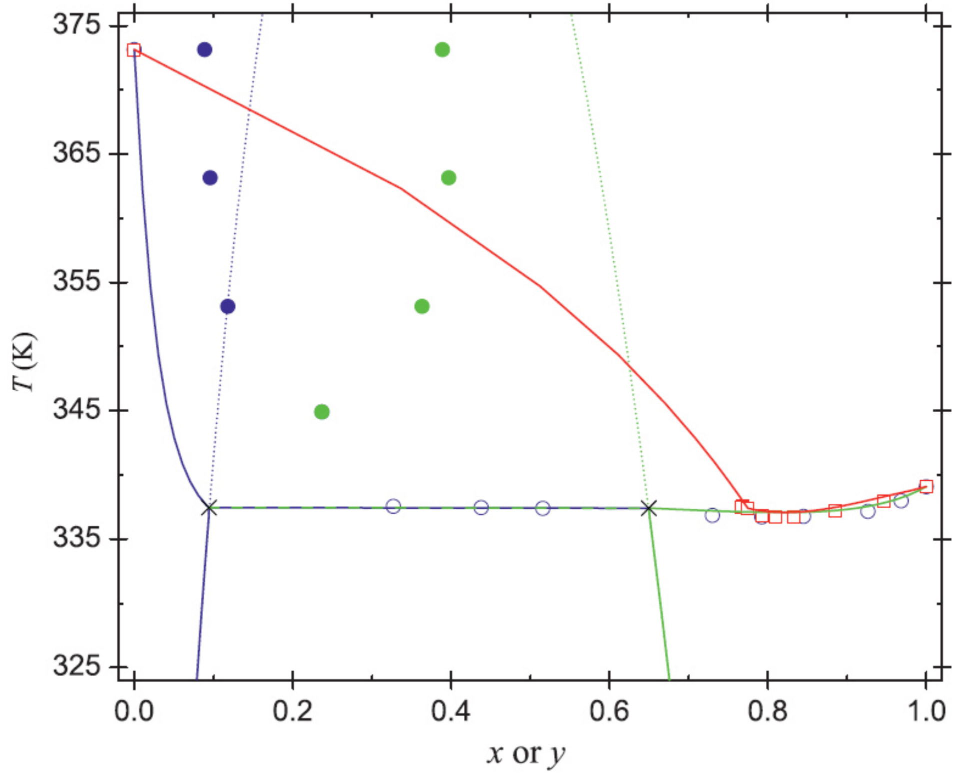


Bollas, Barton & Mitsos, *Chemical Engineering Science 2009*

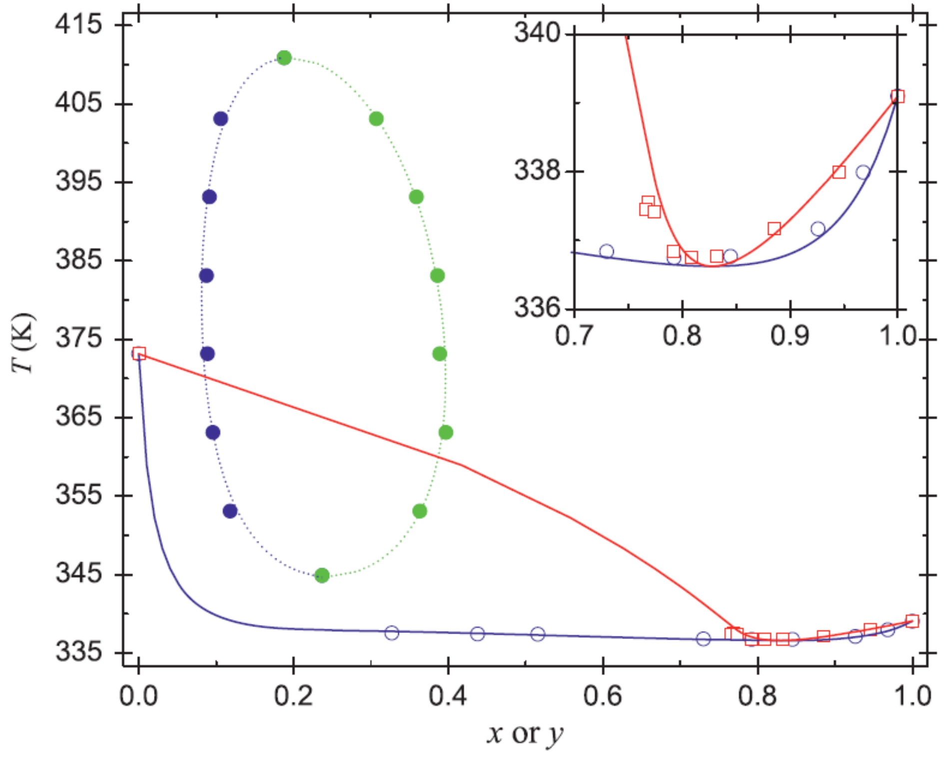
Parameter Estimation: VLE of tetrahydrofuran-water

- Experimental data shows homogeneous azeotrope, LCST, and UCST

$\gamma - \phi$ method predicts spurious liquid phase split. LCST is missed



Proposed formulation gives qualitatively and quantitatively correct fit.



Bollas, Barton & Mitsos, *Chemical Engineering Science 2009*

Check Yourself

- How does parameter estimation in thermodynamics work?