NON-CONVEX ROBUST OPTIMIZATION

Boris Houska

Nonlinear Min-Max Problems

General Formulation:

$$\begin{array}{rl} \min\limits_{x} & \max\limits_{w \in W} F_0(x,w) \\ \text{subject to} & \max\limits_{w \in W} F_i(x,w) \le 0 \end{array}$$

Nonlinear Min-Max Problems

General Formulation:

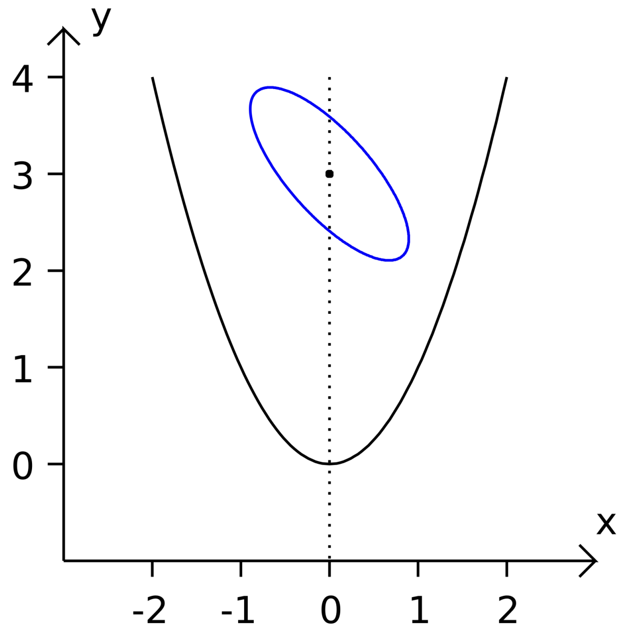
$$\begin{array}{rl} \min_{x} & \max_{w \in W} F_0(x, w) \\ \text{subject to} & \max_{w \in W} F_i(x, w) \leq 0 \end{array}$$

Aim: Solve the problem in a conservative approximation.

Overview

- **Bounding the global maximizer**
- Optimality conditions in SIP and duality
- Sequential convex bi-level programming

Robust Optimization: An Example



Robust optimization problem:

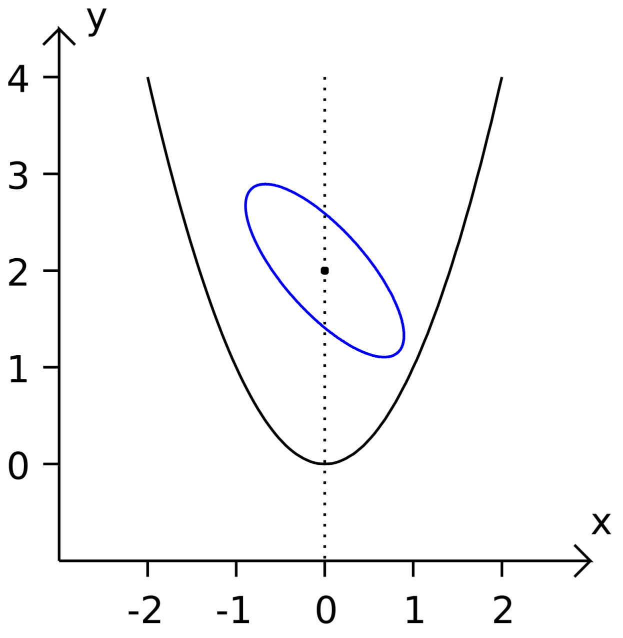
$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad (x+v)^2 - (y+w) \le 0$$

for all $(v, w) \in \mathcal{E}$.

Assumption: $\mathcal{E}$ is a given ellipsoidal uncertainty set.

Robust Optimization: An Example



Robust optimization problem:

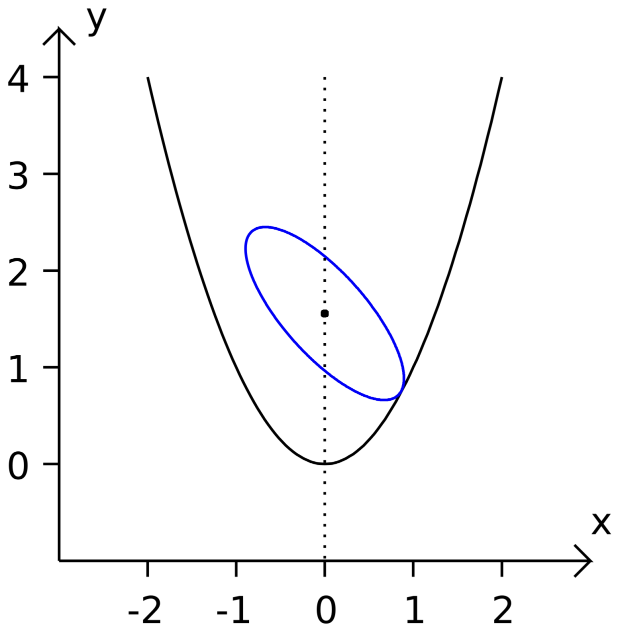
$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad (x+v)^2 - (y+w) \leq 0$$

for all $(v, w) \in \mathcal{E}$.

Assumption: $\mathcal{E}$ is a given ellipsoidal uncertainty set.

Robust Optimization: An Example



Robust optimization problem:

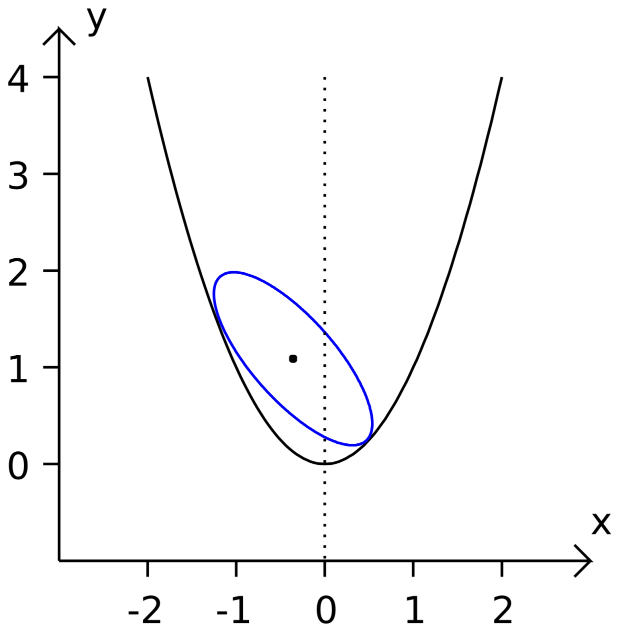
$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad (x+v)^2 - (y+w) \leq 0$$

for all $(v, w) \in \mathcal{E}$.

Assumption: $\mathcal{E}$ is a given ellipsoidal uncertainty set.

Robust Optimization: An Example



Robust optimization problem:

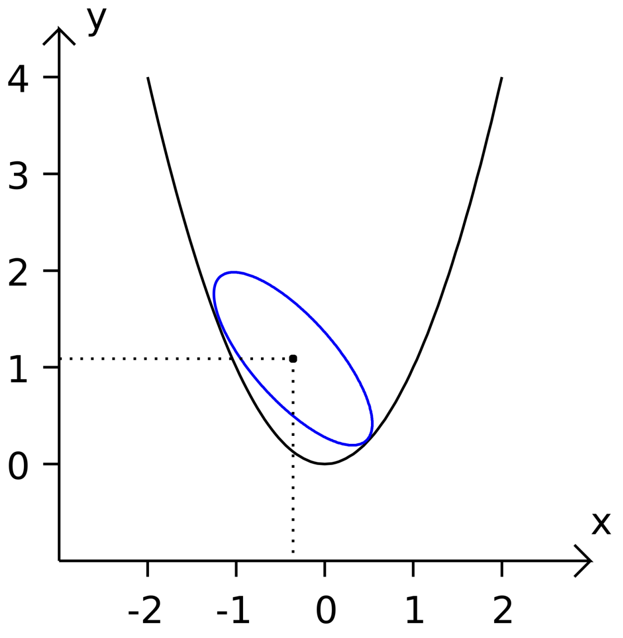
$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad (x+v)^2 - (y+w) \leq 0$$

for all $(v, w) \in \mathcal{E}$.

Assumption: $\mathcal{E}$ is a given ellipsoidal uncertainty set.

Robust Optimization: An Example



Robust optimization problem:

$$\min_{x,y} \quad y$$

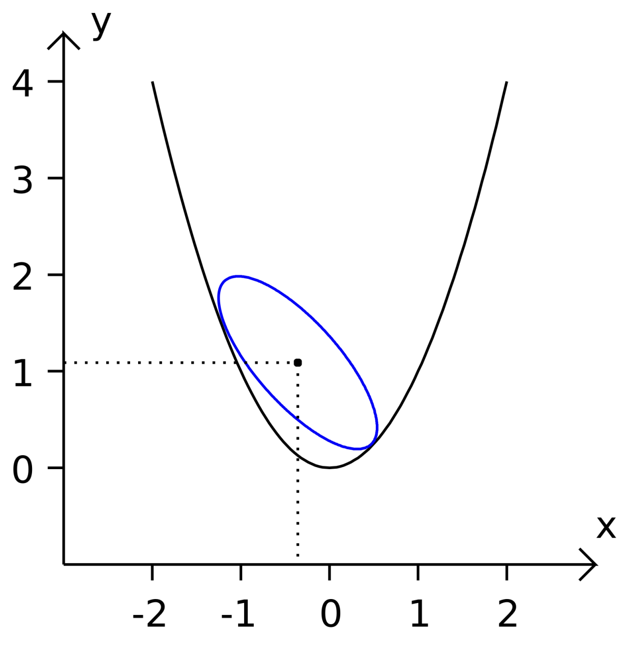
$$\text{s.t.} \quad (x+v)^2 - (y+w) \leq 0$$

for all $(v, w) \in \mathcal{E}$.

Assumption: $\mathcal{E}$ is a given ellipsoidal uncertainty set.

Question: How to find optimal solution numerically?

Robust Optimization: An Example

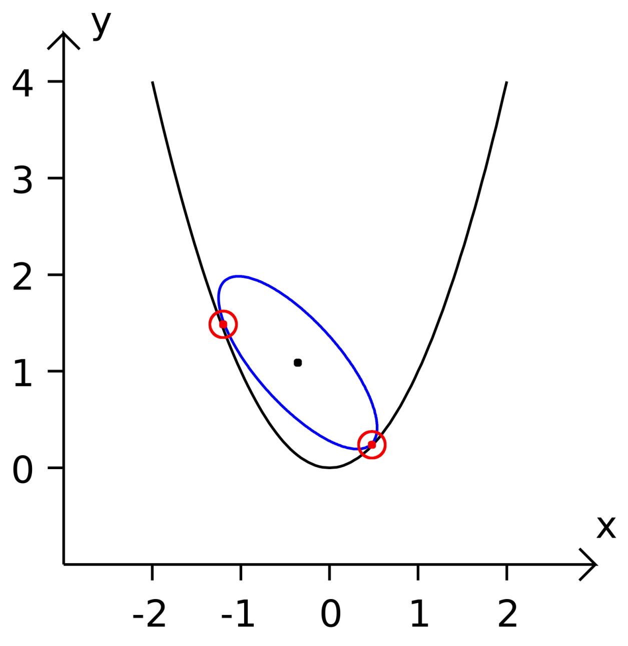


Regard as a min-max problem:

$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad \max_{(v,w)\in\mathcal{E}} (x+v)^2 - (y+w) \leq 0$$

Robust Optimization: An Example



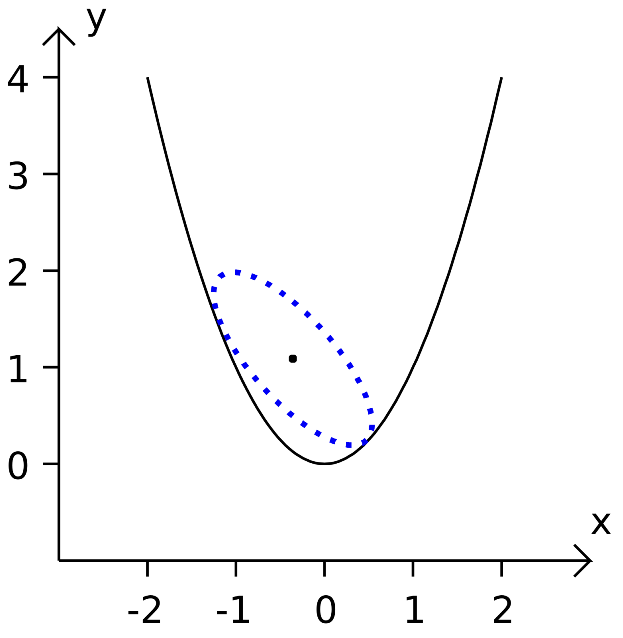
Regard as a min-max problem:

$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad \max_{(v,w)\in\mathcal{E}} (x+v)^2 - (y+w) \leq 0$$

Problem: There are two local maxima in the optimal solution.

Robust Optimization: An Example



Regard as a min-max problem:

$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad \max_{(v,w)\in\mathcal{E}} (x+v)^2 - (y+w) \leq 0$$

Problem: There are two local maxima in the optimal solution.

One Possibility: Check the inequality for all points in the ellipsoid.

Approximation Strategies: Nagy,Braatz 2004, Diehl et al. 2006

Idea:

- Choose an approximation $\widetilde{V}_i(x) \geq \max_{w \in W} F_i(x,w)$ and solve:

$$\min_x \ \widetilde{V}_0(x) \quad \text{s.t.} \quad \widetilde{V}_i(x) \leq 0 \quad \text{for all } i \in \{1, \ldots, n\}$$

Approximation Strategies: Nagy,Braatz 2004, Diehl et al. 2006

Idea:

- Choose an approximation $\widetilde{V}_i(x) \geq \max_{w \in W} F_i(x,w)$ and solve:

$$\min_x \; \widetilde{V}_0(x) \quad \text{s.t.} \quad \widetilde{V}_i(x) \;\leq\; 0 \quad \text{for all } i \in \{1,\ldots,n\}$$

Assumption: We have a function $\overline{\lambda}_i : \mathbb{R}^{n_x} \to \mathbb{R}^+$ with

$$\forall w \in \mathcal{W}: \quad \lambda_{\max}\left(\frac{\partial^2}{\partial w^2} F_i(x,w) \right) \;\leq\; 2\,\overline{\lambda}_i(x)\;.$$

Approximation Strategies: Nagy,Braatz 2004, Diehl et al. 2006

Idea:

- Choose an approximation $\widetilde{V}_i(x) \geq \max_{w \in W} F_i(x,w)$ and solve:

$$\min_x \widetilde{V}_0(x) \quad \text{s.t.} \quad \widetilde{V}_i(x) \leq 0 \quad \text{for all } i \in \{1,\ldots,n\}$$

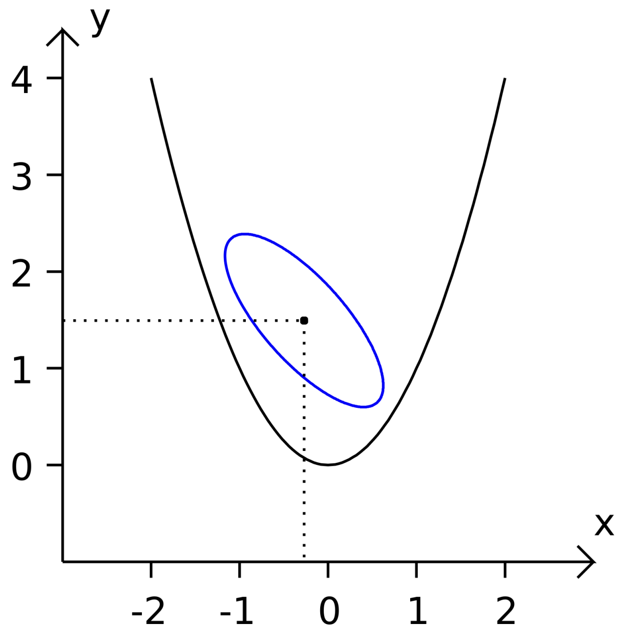
Assumption: We have a function $\overline{\lambda}_i : \mathbb{R}^{n_x} \to \mathbb{R}^+$ with

$$\forall w \in \mathcal{W} : \quad \lambda_{\max}\left(\frac{\partial^2}{\partial w^2} F_i(x,w)\right) \leq 2\,\overline{\lambda}_i(x) \, .$$

Example: $W := \left\{ w \mid w^T w \leq 1 \right\}$. Linear approximation:

$$\widetilde{V}_i(x) := L_i(x) := F_i(x,0) + \left\| \frac{\partial F(x,0)}{\partial w} \right\|_2 + \overline{\lambda}_i(x) \, .$$

Example: The "Best" Conservative Linear Approximation



Consider once more the problem:

$$\min_{x,y} \quad y$$

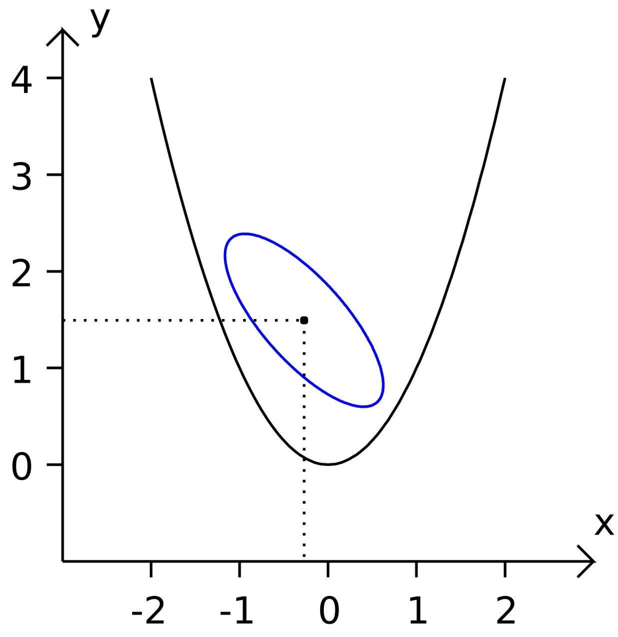
$$\text{s.t.} \quad \max_{(v,w)\in\mathcal{E}} (x+v)^2 - (y+w) \leq 0$$

Notation: $L =$ Chol. factor of Σ^{-1}.

The "best" conservative linear approximation solves the CP:

$$\min_{x,y} \quad y \quad \text{s.t.} \quad x^2 - y + \left\| \begin{pmatrix} 2L_{1,1}x - L_{1,2} \\ -L_{2,2} \end{pmatrix} \right\|_2 + L_{1,1}^2 \;\leq\; 0\,.$$

Example: The "Best" Conservative Linear Approximation



Consider once more the problem:

$$\min_{x,y} \quad y$$

$$\text{s.t.} \quad \max_{(v,w)\in\mathcal{E}} (x+v)^2 - (y+w) \leq 0$$

Notation: $L =$ Chol. factor of Σ^{-1}.

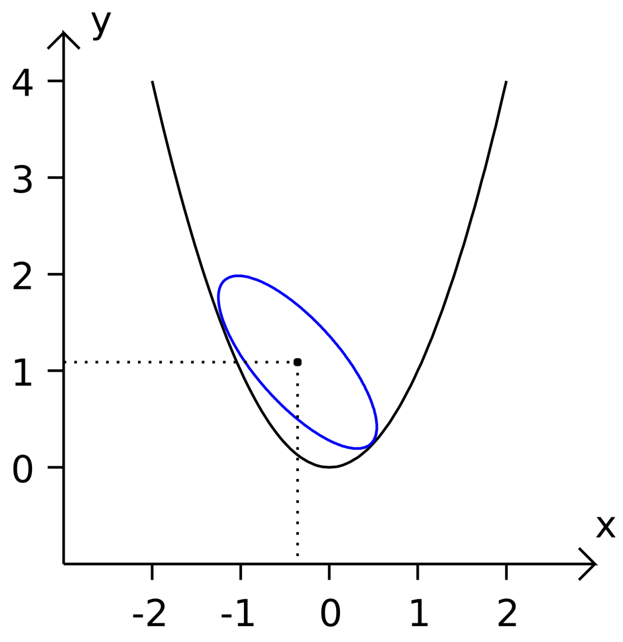
We find a conservative solution by the linear approximation:

$$\left(\widetilde{x}^*, \widetilde{y}^*\right) \approx (-0.27\,,\, 1.49) \neq (-0.35\,,\, 1.08)\ .$$

But we have lost 38% of optimality.

Can we do better?

Find the solution $(x^*, y^*) \; = \; (\,-0.35...\,,\, 1.08...\,)$ by convex optimization:



Define:

$$Q := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad q(x) := \begin{pmatrix} 2x \\ -1 \end{pmatrix}$$

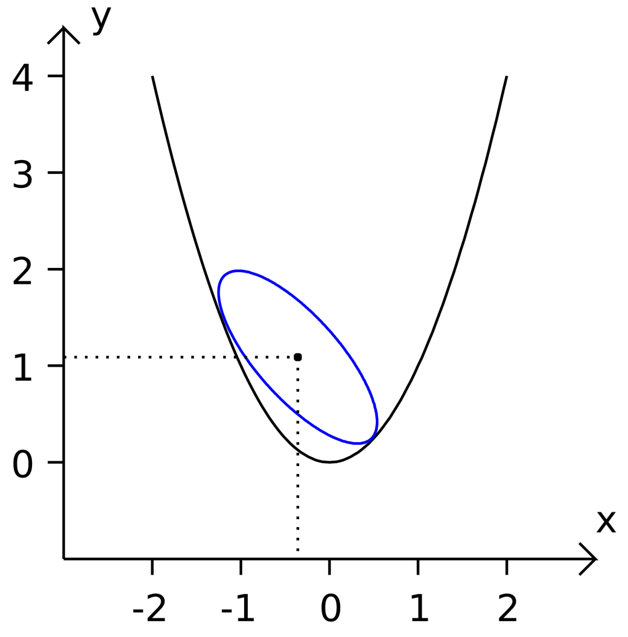
and

$$\Sigma := \begin{pmatrix} 0.8 & -0.6 \\ -0.6 & 0.8 \end{pmatrix}^{-1}$$

$$\min_{x,y,\lambda > 0.8} \quad y \quad \text{s.t.} \quad x^2 - y + \frac{1}{4} q(x)^T \left(\lambda\Sigma - Q\right)^{-1} q(x) + \lambda \;\leq\; 0\,.$$

Can we do better?

Find the solution $(x^*, y^*) = (-0.35\ldots, 1.08\ldots)$ by convex optimization:



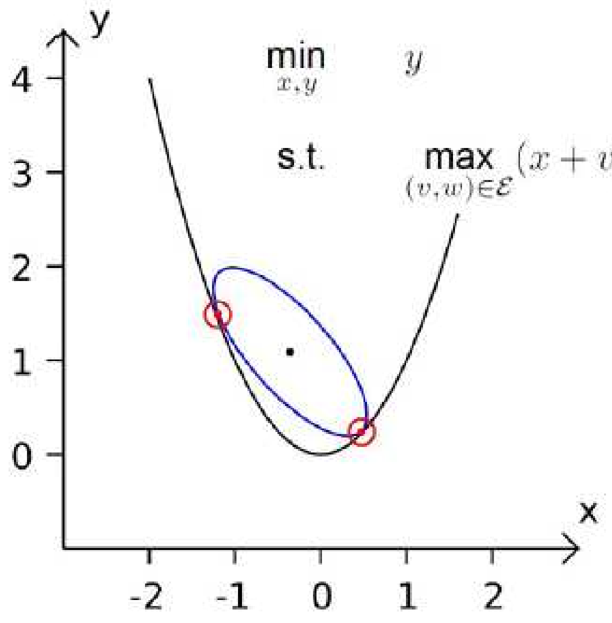
Alternative formulation as LMI:

$$\min_{x,y,\lambda} \quad y$$
$$\text{s.t.} \quad \begin{pmatrix} y-\lambda & q(x)^T & x \\ q(x) & \lambda\Sigma - Q & 0 \\ x & 0 & 1 \end{pmatrix} \succeq 0$$

$$\min_{x,y,\lambda>0.8} \quad y \quad \text{s.t.} \quad x^2 - y + \frac{1}{4} q(x)^T \left(\lambda\Sigma - Q\right)^{-1} q(x) + \lambda \le 0\,.$$

Overview: Bounding the Global Maximizer

Example:



$$\min_{x,y} \quad y$$
$$\text{s.t.} \quad \max_{(v,w)\in\mathcal{E}} (x+v)^2 - (y+w) \leq 0$$

Methods:

- Branch & Bound
- Linear Approximation
- Linear Matrix Inequalities
- Lagrangian Duality

[1] **Bhattacharjee**,et al. Global solution of semi-infinite programs. 2005.
[2] **Stein** et al. The adaptive convexification algorithm for semi-infinite programming [..]. 2012.
[3] **Nagy** and **Braatz**. Distributional uncertainty analysis using [..] polynomial chaos expansions. 2007.
[4] **Diehl** et al. An approximation technique for robust nonlinear optimization. 2006.
[5] **Lasserre**. Moments, Positve Polynomials and Their Applications. Imperial College Press, 2009.
[6] **Ben-Tal.** Robust Optimization. Princeton University Press, 2009.
[7] **Polak** et al. On the use of augmented Lagrangians [..] semiinfinite min-max problems. 2005.

Dual Lagrangian Overestimation Techniques

Question:

- Can we do better than linear approximation?

Dual Lagrangian Overestimation Techniques

Question:

- Can we do better than linear approximation?

Main Idea:

- Define the dual Lagrangian over-estimator as

$$\forall x \in \mathbb{R}^{n_x} : \quad M_i(x) := \min_{\lambda_i \geq \overline{\lambda}_i(x)} d_i(x, \lambda_i) \; \geq \; \min_{\lambda \geq 0} d_i(x, \lambda_i)$$

where d_i is defined to be the following dual Lagrangian function:

$$d_i(x, \lambda_i) \quad := \quad \max_{w \in \mathcal{W}} F_i(x, w) - \lambda_i w^T w + \lambda_i \; .$$

Dual Lagrangian Overestimation Techniques

Question:

- Can we do better than linear approximation?

Main Idea:

- Define the dual Lagrangian over-estimator as

$$\forall x \in \mathbb{R}^{n_x} : \quad M_i(x) := \min_{\lambda_i \geq \overline{\lambda}_i(x)} d_i(x, \lambda_i) \geq \min_{\lambda \geq 0} d_i(x, \lambda_i)$$

where d_i is defined to be the following dual Lagrangian function:

$$d_i(x, \lambda_i) \quad := \quad \max_{w \in \mathcal{W}} F_i(x, w) - \lambda_i w^T w + \lambda_i \; .$$

Proposition:

- The function $d_i(x, \cdot)$ is concave for $\lambda_i \geq \overline{\lambda}_i(x)$.

Dual Lagrangian Overestimation Techniques

Theorem [Houska,Diehl 2010]

- The functions M_i are conservative upper bounds on the robust counterpart functions V_i and always less conservative that the best linear approximation:

$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) \ \leq \ M_i(x) \ \leq \ L_i(x) \ .$$

Dual Lagrangian Overestimation Techniques

Theorem [Houska,Diehl 2010]

- The functions M_i are conservative upper bounds on the robust counterpart functions V_i and always less conservative that the best linear approximation:

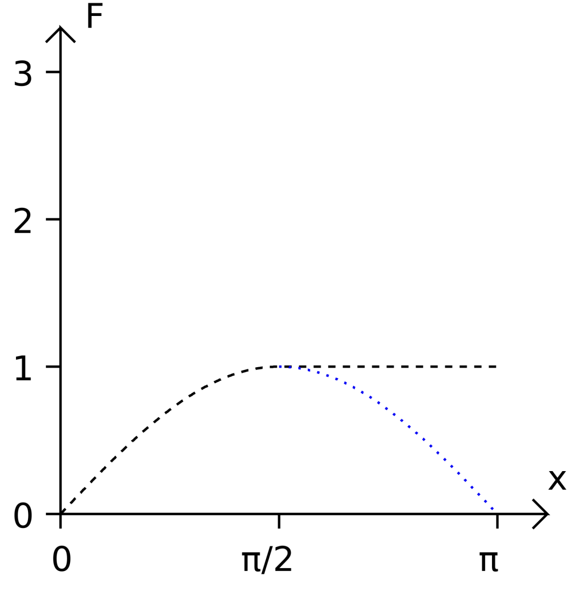
$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) \ \leq \ M_i(x) \ \leq \ L_i(x) \ .$$

- If F_i is a quadratic form in w than the approximation is exact, i.e., we have

$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) \ = \ M_i(x) \ .$$

What about the Non-Quadratic Case?

Example: Consider the function $F(x,w) \; := \; \sin(\, x\, w\,)$



A Hessian upper bound is given by

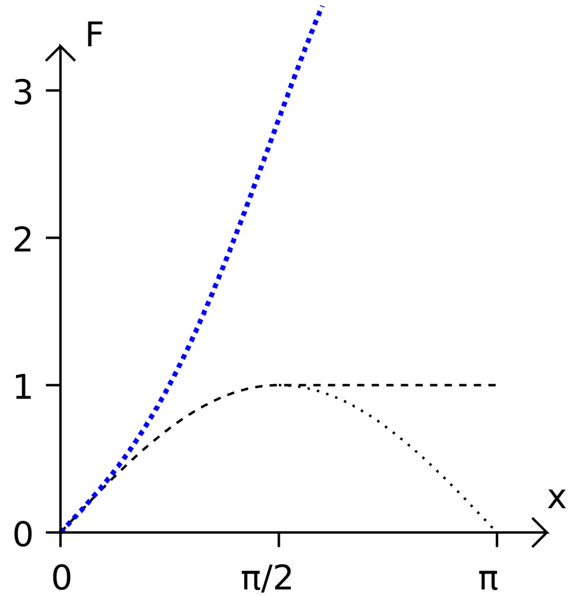
$$\overline{\lambda}(x) \;:=\; \begin{cases} \frac{x^2}{2}\sin(\,|x|\,) & \text{if } |x| \le \frac{\pi}{2} \\ \frac{x^2}{2} & \text{otherwise.} \end{cases}$$

The exact robust counterpart function is given by

$$V(x) \;=\; \begin{cases} \sin(\,|x|\,) & \text{if } |x| \le \frac{\pi}{2} \\ 1 & \text{otherwise.} \end{cases}$$

What about the Non-Quadratic Case?

Example: Consider the function $F(x,w) \; := \; \sin(\, x\, w\,)$



A Hessian upper bound is given by

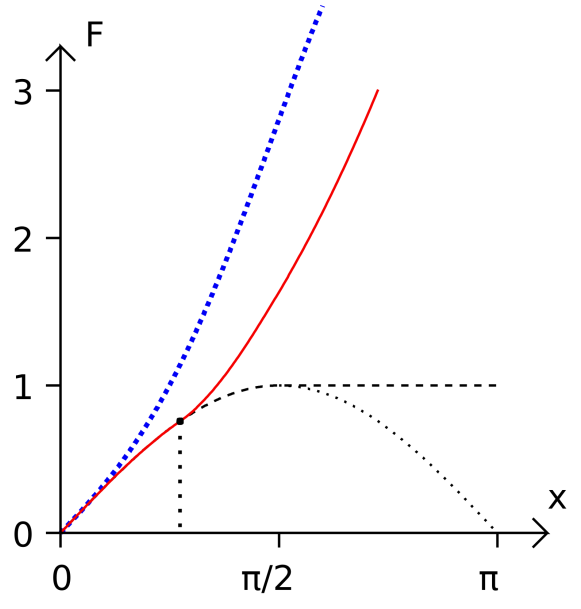
$$\overline{\lambda}(x) \; := \; \begin{cases} \frac{x^2}{2} \sin(\,|x|\,) & \text{if } |x| \leq \frac{\pi}{2} \\ \frac{x^2}{2} & \text{otherwise.} \end{cases}$$

The best linear approximation is given by

$$L(x) \; = \; \begin{cases} |x| + \frac{x^2}{2} \sin(\,|x|\,) & \text{if } |x| \leq \frac{\pi}{2} \\ |x| + \frac{x^2}{2} & \text{otherwise.} \end{cases}$$

What about the Non-Quadratic Case?

Example: Consider the function $F(x,w) \; := \; \sin(\, x\, w\,)$



A Hessian upper bound is given by

$$\overline{\lambda}(x) \; := \; \begin{cases} \frac{x^2}{2} \sin(\,|x|\,) & \text{if } |x| \leq \frac{\pi}{2} \\ \frac{x^2}{2} & \text{otherwise.} \end{cases}$$

The dual Lagrangian overestimate M satisfies

$$V(x) \; \leq \; M(x) \; \leq \; L(x) \; .$$

Moreover, we have $V(x) \; = \; M(x)$ for $|x| < 0.86$.

Bounds on the Duality Gap

Theorem [Houska 2011]

- With $W := \{w \mid w^T w \leq \gamma^2\}$ and

$$\overline{\gamma}_i(x) \; := \; \min_{w \in W} \; \left\| \left(\frac{\partial^2 F_i(x,w)}{\partial w^2} - 2\overline{\lambda}_i(x) I \right)^{-1} \frac{\partial F_i(x,0)}{\partial w} \right\|_2 \; .$$

The approximation function M_i always satisfies

$$|\, M_i(x) - V_i(x) \,| \; \leq \; \overline{\lambda}_i(x) \max \left\{ \, 0, \, \gamma^2 - \overline{\gamma}_i(x)^2 \, \right\} \; .$$

Overview

- Bounding the global maximizer
- **Optimality conditions in SIP and duality**
- Sequential convex bi-level programming

Cooperation Partners

Joint work with M. Diehl, O. Stein, and P. Steuermann:



- A lifting method for generalized semi-infinite programs based on lower level Wolfe duality

PDF of the Paper available at:

- Optimization Online (December, 2011)
- Computational Optimization and Applications, 2012.

Problem Formulation

Generalized Semi-Infinite Programming:

$$\min_{x \in M} f(x) \quad \text{with} \quad M = \{x \in X |\ g(x,y) \le 0 \text{ for all } y \in Y(x)\}$$

and

$$Y(x) = \{y \in \mathbb{R}^m |\ v_\ell(x,y) \le 0,\ 1 \le \ell \le s\}.$$

Problem Formulation

Generalized Semi-Infinite Programming:

$$\min_{x \in M} f(x) \quad \text{with} \quad M = \{x \in X \mid g(x,y) \leq 0 \text{ for all } y \in Y(x)\}$$

and

$$Y(x) = \{y \in \mathbb{R}^m \mid v_\ell(x,y) \leq 0,\ 1 \leq \ell \leq s\}.$$

Assumptions:

- $X \subseteq \mathbb{R}^n$ is a closed set.
- The functions f, g, and v are twice continuously differentiable.
- The functions $-g(x,\cdot)$, $v_\ell(x,\cdot)$, $1 \leq \ell \leq s$ are convex on $Y(x)$.

Equivalent Min-Max Formulation

Min-Max Formulation:

$$\min_{x} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X \\ \max\limits_{y,\, v(x,y) \le 0} g(x,y) \ \le \ 0 \ . \end{cases}$$

Equivalent Min-Max Formulation

Min-Max Formulation:

$$\min_x \; f(x) \quad \text{s.t.} \quad \begin{cases} x \in X \\ \max\limits_{y, v(x,y) \leq 0} g(x,y) \; \leq \; 0 \, . \end{cases}$$

Standard definitions for lower level problem:

- LICQ satisfied iff $D_y v^{\mathrm{act}}(\bar{x}, \bar{y})$ has full rank.
- SCC satisfied iff: $v(\bar{x}, \bar{y}) - \bar{\gamma} \; < \; 0$.
- SOSC satisfied iff: $p^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) p < 0$ for all $p \in T \setminus \{0\}$.
- Tangential Cone:

$$T := \{ p \mid \forall i \in \mathbb{A}_\mathrm{s} : \; D_y v_i(\bar{x}, \bar{y}) p = 0 \; \wedge \; \forall i \in \mathbb{A}_\mathrm{w} : \; D_y v_i(\bar{x}, \bar{y}) p \leq 0 \} \; .$$

Overview

Aim of this section:

- Discuss three ways to solve GSIPs:

WAY 1: MPCC

Idea:

- Write lower level optimality conditions into constraints:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X,\ g(x,y) \le 0, \\ \nabla_y \mathcal{L}(x,y,\gamma) = 0, \\ 0 \le\ -v(x,y)\ \perp\ \gamma \ge 0 \end{cases}$$

with $\mathcal{L}(x,y,\gamma)\ :=\ g(x,y) - \gamma^\intercal v(x,y)$.

WAY 1: MPCC

Idea:

- Write lower level optimality conditions into constraints:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X,\ g(x,y) \leq 0, \\ \nabla_y \mathcal{L}(x,y,\gamma) = 0, \\ 0 \leq\ -v(x,y)\ \perp\ \gamma \geq 0 \end{cases}$$

with $\mathcal{L}(x,y,\gamma)\ :=\ g(x,y) - \gamma^\intercal v(x,y)$.

Pros/Cons:

+ Equivalent to original generalized semi-infinite problem, if the lower level maximizer is a KKT point.
- MPCCs are computationally demanding.

WAY 2: Dual Reformulation

Idea:

- Assume that strong duality holds.
- Introduce dual variables γ and use Wolfe duality:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad x \in X,\ \mathcal{L}(x,y,\gamma) \leq 0,\ \nabla_y \mathcal{L}(x,y,\gamma) = 0,\ \gamma \geq 0\ .$$

WAY 2: Dual Reformulation

Idea:

- Assume that strong duality holds.
- Introduce dual variables γ and use Wolfe duality:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad x \in X,\ \mathcal{L}(x,y,\gamma) \leq 0,\ \nabla_y \mathcal{L}(x,y,\gamma) = 0,\ \gamma \geq 0\ .$$

Pros/Cons:

\+ We end up with a “well-posed” standard optimization problem.

\- The dual reformulation is not necessarily equivalent to the original problem if $Y(x) \neq \mathbb{R}^m$.

WAY 2: Dual Reformulation

Idea:

- Assume that strong duality holds.
- Introduce dual variables γ and use Wolfe duality:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad x \in X,\ \mathcal{L}(x,y,\gamma) \leq 0,\ \nabla_y \mathcal{L}(x,y,\gamma) = 0,\ \gamma \geq 0\ .$$

Counter example:

- Choose $g(x,y) = -x - y^3 - 3y^2 + 16y - 12$, $v_1(y) = -y$, and $v_2(y) = y - 1$.

WAY 3: Including the Index Constraint

Idea:

- Start with dual reformulation but include the index set constraint:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X, \\ g(x,y) - \gamma^{\mathsf{T}} v(x,y) \;=\; \mathcal{L}(x,y,\gamma) \le 0, \\ \nabla_y \mathcal{L}(x,y,\gamma) = 0, \\ v(x,y) \le 0, \;\; \gamma \ge 0. \end{cases}$$

- We call this problem **LWPC**.

WAY 3: Including the Index Constraint

Idea:

- Start with dual reformulation but include the index set constraint:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X, \\ g(x,y) - \gamma^{\mathsf{T}} v(x,y) \; = \; \mathcal{L}(x,y,\gamma) \leq 0, \\ \nabla_y \mathcal{L}(x,y,\gamma) = 0, \\ {\color{red} v(x,y) \leq 0}, \;\; \gamma \geq 0. \end{cases}$$

- We call this problem **LWPC**.

Pros/Cons:

\+ Formulation is equivalent to the original GSIP.

\- We still assume the existence of KKT points...

WAY 3: Including the Index Constraint

Idea:

- Start with dual reformulation but include the index set constraint:

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X, \\ g(x,y) - \gamma^\intercal v(x,y) \; = \; \mathcal{L}(x,y,\gamma) \leq 0, \\ \nabla_y \mathcal{L}(x,y,\gamma) = 0, \\ \color{red}{v(x,y) \leq 0}, \;\; \gamma \geq 0. \end{cases}$$

- We call this problem **LWPC**.

Question:

- Does the seemingly redundant index set constraint lead to a violation of constraint qualifications?

Constraint Qualifications (Part I)

Theorem We can prove that LICQ for LWPC if

1. The matrix $D_y v^{\mathrm{act}}(\bar{x}, \bar{y})$ has full rank (lower level LICQ).
2. The condition $v(\bar{x}, \bar{y}) - \bar{\gamma} < 0$ holds (lower level strict complementarity).
3. The inequality $p^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) p < 0$ holds for all $p \in T \setminus \{0\}$ (lower level second order condition).
4. $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $D_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ do not hold simultaneously (upper level LICQ).

Constraint Qualifications (Part I)

Theorem We can prove that LICQ for LWPC if

1. The matrix $D_y v^{\text{act}}(\bar{x}, \bar{y})$ has full rank (lower level LICQ).
2. The condition $v(\bar{x}, \bar{y}) - \bar{\gamma} < 0$ holds (lower level strict complementarity).
3. The inequality $p^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) p < 0$ holds for all $p \in T \setminus \{0\}$ (lower level second order condition).
4. $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $D_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ do not hold simultaneously (upper level LICQ).

Question:

- Can we still prove the existence of KKT points if we do not have lower level strict complementarity?

Example

Consider the convex min-max problem (solution: $(\bar{x}, \bar{y}, \bar{\gamma})^\intercal = 0$):

$$\min_{x} \ -x \qquad \text{s.t.} \qquad \max_{y \geq 0} \ x - y^2 \ \leq \ 0 \ .$$

Example

Consider the convex min-max problem (solution: $(\bar{x}, \bar{y}, \bar{\gamma})^\intercal = 0$):

$$\min_{x} \; -x \qquad \text{s.t.} \qquad \max_{y \geq 0} \; x - y^2 \leq 0 \, .$$

$$\text{LWPC}: \qquad \min_{x,y,\gamma} \; -x \quad \text{s.t.} \quad \begin{cases} 0 \geq x - y^2 + \gamma y \\ 0 = -2y + \gamma \\ 0 \geq -y \\ 0 \geq -\gamma \end{cases}$$

Example

$$\text{LWPC}: \qquad \min_{x,y,\gamma} \; -x \quad \text{s.t.} \quad \begin{cases} 0 & \geq \quad x - y^2 + \gamma y \\ 0 & = \quad -2y + \gamma \\ 0 & \geq \quad -y \\ 0 & \geq \quad -\gamma \end{cases}$$

Jacobian of the active constraints does not have full rank:

$$J \quad = \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} .$$

...but we have $J\xi \; = \; (-1,-1,-1)^\intercal \; < \; 0$ for $\xi \; := \; (-1,1,1)$.

Constraint Qualifications (Part II)

Definition:

$$T := \{p \in \mathbb{R}^m \mid \forall i \in \mathbb{A}_{\mathrm{s}} : \; D_y v_i(\bar{x}, \bar{y}) p = 0 \; \wedge \; \forall i \in \mathbb{A}_{\mathrm{w}} : \; D_y v_i(\bar{x}, \bar{y}) p \leq 0\}$$

Constraint Qualifications (Part II)

Definition:

$$T := \{p \in \mathbb{R}^m \mid \forall i \in \mathbb{A}_{\mathrm{s}} : \ D_y v_i(\bar{x}, \bar{y})p = 0 \ \wedge \ \forall i \in \mathbb{A}_{\mathrm{w}} : \ D_y v_i(\bar{x}, \bar{y})p \leq 0\}$$

Theorem: MFCQ for LWPC holds at $(\bar{x}, \bar{y})$, if

1. There exist a vector $\xi_1 \in \mathbb{R}^{n_x}$ with $D_y v^{\mathrm{act}}(\bar{x}, \bar{y})\xi_1 \ < \ 0$.
2. The inequality $p^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})p < 0$ holds for all $p \in T \setminus \{0\}$.
3. $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ do not hold simultaneously.

Constraint Qualifications (Part II)

Definition:

$$T := \{p \in \mathbb{R}^m \mid \forall i \in \mathbb{A}_\mathrm{s} : \ D_y v_i(\bar{x}, \bar{y})p = 0 \ \wedge \ \forall i \in \mathbb{A}_\mathrm{w} : \ D_y v_i(\bar{x}, \bar{y})p \leq 0\}$$

Theorem: MFCQ for LWPC holds at $(\bar{x}, \bar{y})$, if

1. There exist a vector $\xi_1 \in \mathbb{R}^{n_x}$ with $D_y v^{\mathrm{act}}(\bar{x}, \bar{y})\xi_1 \ < \ 0$.
2. The inequality $p^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})p < 0$ holds for all $p \in T \setminus \{0\}$.
3. $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ do not hold simultaneously.

No strict complementarity assumption needed!

Numerical Examples

Example: Design Centering Problem

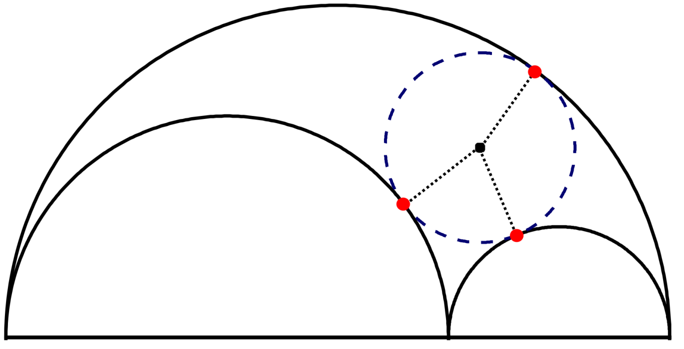
$$
\begin{aligned}
\min_{x \in R^3} \quad & -x_3 \\
\text{s.t.} \quad & \left\{
\begin{array}{lll}
0 & \geq & 1 - (x_1 + y_1)^2 - (x_2 + y_2)^2 \qquad\qquad \forall y \in Y(x) \\
0 & \geq & \frac{1}{4} - \left(x_1 - \frac{3}{2} + y_1\right)^2 - (x_2 + y_2)^2 \quad \forall y \in Y(x) \\
0 & \geq & \left(x_1 - \frac{1}{2} + y_1\right)^2 - (x_2 + y_2)^2 - \frac{9}{4} \quad \forall y \in Y(x) \\
0 & \geq & -x_2 \; ,
\end{array}
\right.
\end{aligned}
$$

where $Y(x)$ depends explicitly on the third component of x:

$$
Y(x) := \left\{ \; y \in \mathbb{R}^2 \;\; \middle| \;\; y^T y \; \leq \; x_3 \; \right\} \; .
$$

Numerical Examples

- Solution found with LWPC formulation + SQP solver:



- Starting at the “poor” guess

$$x^0 := (1,\,1,\,0.1)^T \;, \quad \gamma^0 := (1,\,1,\,1)^T \;,$$
$$y_1^0 := (-0.1,\,-0.1)^T \;, \quad y_2^0 := (0.1,\,-0.1)^T \;, \quad y_3^0 := (0.1,\,0.1)^T \;,$$

Full-step BFGS-SQP needs 56 iterations (KKT-tolerance: 10^{-9}).

Overview

- Bounding the global maximizer
- Optimality conditions in SIP and duality
- **Sequential convex bi-level programming**

Sequential Convex Bilevel Programming

Now: $$\min_{x \in \mathbb{R}^{n_x}} \quad \max_{w_0 \in \mathcal{B}} \quad H_0(x, w_0)$$

$$\text{subject to} \quad \max_{w_i \in \mathcal{B}} \quad H_i(x, w_i) \leq 0$$

$$\mathcal{B} := \left\{ w \in \mathbb{R}^{n_w} \mid B(x, w) \leq 0 \right\}$$

Assumptions: The functions -H and B are:

a) convex in w and
b) twice continuously differentiable in x and w.

Min-Max versus MPCC

$$\min_{x \in \mathbb{R}^{n_x}} \quad \max_{w_0 \in \mathcal{B}} \quad H_0(x, w_0)$$
$$\text{subject to} \quad \max_{w_i \in \mathcal{B}} \quad H_i(x, w_i) \leq 0$$

KKT

Conditions

$$\mathcal{B} = \left\{ w \in \mathbb{R}^{n_w} \mid B(w) \leq 0 \right\}$$

$$L_j(x, w, \lambda) = H_j(x, w) - \lambda^T B(w)$$

$$\begin{aligned} \underset{x, w, \lambda}{\text{minimize}} \quad & H_0(x, w_0) \\ \text{subject to} \quad & 0 \geq H_i(x, w_i) \\ & 0 = \nabla_w L_j(x, w_j, \lambda_j) \\ & 0 \geq B(w_j) \\ & 0 \leq \lambda_j \\ & 0 = \sum_{k=0}^{n} \lambda_k^T B_k(w_k) \end{aligned}$$

Say: H convex in x …

(Maximum over convex functions is convex)

Non-Convex ?

Optimality Conditions

Surprising (but well-known):

$$0 = \frac{\partial}{\partial x} K(x^*, \chi^*, w^*, \lambda^*) \qquad 0 = \frac{\partial}{\partial w} L_j(x^*, w_j^*, \lambda_j^*)$$

$$0 \geq L_i(x^*, w_i^*, \lambda_i^*) \qquad 0 \geq B(w_j^*)$$

$$0 \geq \chi_i^* \qquad 0 \leq \lambda_j^*$$

$$0 = \sum_{k=1}^{n} \chi_k^* L_k(x^*, w_k^*, \lambda_k^*) \qquad 0 = \sum_{k=0}^{n} \lambda_k^{*T} B(w_k^*)$$

$$K(x, \chi, w^*, \lambda^*) := L_0(x, w_0^*, \lambda_0^*) - \sum_{k=1}^{n} \chi_k L_k(x, w_k^*, \lambda_k^*)$$

Only first order derivatives needed!

(Remark: ELICQ = MPCC-LICQ and EMFCQ = MPCC-MPCQ.)

[1] **Jongen** et al.. Generalized semi-infinite optimization: a first order optimality condition [..], 1998.
[2] **Hettich** and **Kortanek**. Semi infinite programming: theory, methods, and application. 1993.

Sequential Convex Bi-Level Programming

1) Initialize x and w and evaluate derivatives of H and B.

Repeat:

2) Solve the **convex** min-max QCQP

$$\min_{\Delta x} \max_{\Delta w_0 \in \mathcal{B}_0^{\text{lin}}} \left\{ H^0 + L_x^0 \Delta x + \left(\frac{\Delta w_0^T L_{ww}^0}{2} + \Delta x^T L_{xw}^0 + H_w^0 \right) \Delta w_0 + \frac{\Delta x^T K_{xx} \Delta x}{2} \right\}$$

$$\text{s.t.} \max_{\Delta w_i \in \mathcal{B}_i^{\text{lin}}} \left\{ H^i + L_x^i \Delta x + \left(\frac{\Delta w_i^T L_{ww}^i}{2} + \Delta x^T L_{xw}^i + H_w^i \right) \Delta w_i \right\} \leq 0$$

with $\mathcal{B}_j^{\text{lin}} := \left\{ \Delta w_j \mid B_w^i \Delta w_i + B^i \leq 0 \right\}$

3) Line-Search & Update:

$$z^+ = z + \alpha \, \Delta z := (x + \alpha \Delta x, \chi + \alpha \Delta \chi, w + \alpha \Delta w, \lambda + \alpha \, \Delta \lambda)$$

4) Evaluate derivatives of H and B at the new point z^+ .

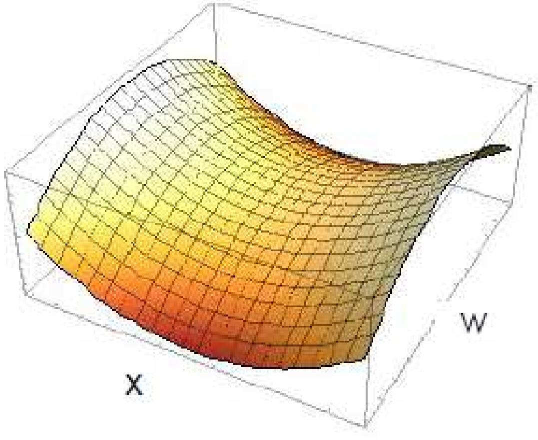
How to construct a merit function?

Quadratically convergent if (Dennis-Moré condition)

$$\left\| K_{xx} - L^0_{xx} + \sum_{k=1}^{n} \chi_k L^k_{xx} \right\| \leq \mathbf{O}\left(\|\Delta z\|\right)$$

Quadratic convergence with only second derivatives, like for standard NLP, but for min-max optimization

Globalization Problem:



We go simultaneously

“downhill“ (min) and “uphill“ (max).

How to construct a merit function?

Globalization Strategy

Notation:

$$\Phi_U(x, w, \lambda) := L_0(x, w_0, \lambda_0) + \sum_{k=1}^{n} \hat{\chi}_k \, \pi_k(L_k(x, w_k, \lambda_k)) \qquad \pi_k(s) = \max\{0, s_k\}$$

$$\Phi_L^j(x, w_j, \lambda_j) := \left| \frac{\partial L_j(x, w_j, \lambda_j)}{\partial w} \right| \hat{\rho}_j + \hat{\lambda}_j^T \pi(B(x, w_j))$$

$$\Phi(x, w, \lambda) := \Phi_U(x, w, \lambda) + \Phi_L^0(x, w_0, \lambda_0) + \sum_{k=1}^{n} \hat{\chi}_k \Phi_L^k(x, w_k, \lambda_k)$$

Theorem

If $\hat{\chi} > \left|\chi^{\dagger}\right|, \; \hat{\rho}_k > \frac{3}{2}|\rho_k|, \; \hat{\lambda}_j > 0$, then the min-max QCQP

a) finds a strict decent direction of the merit function

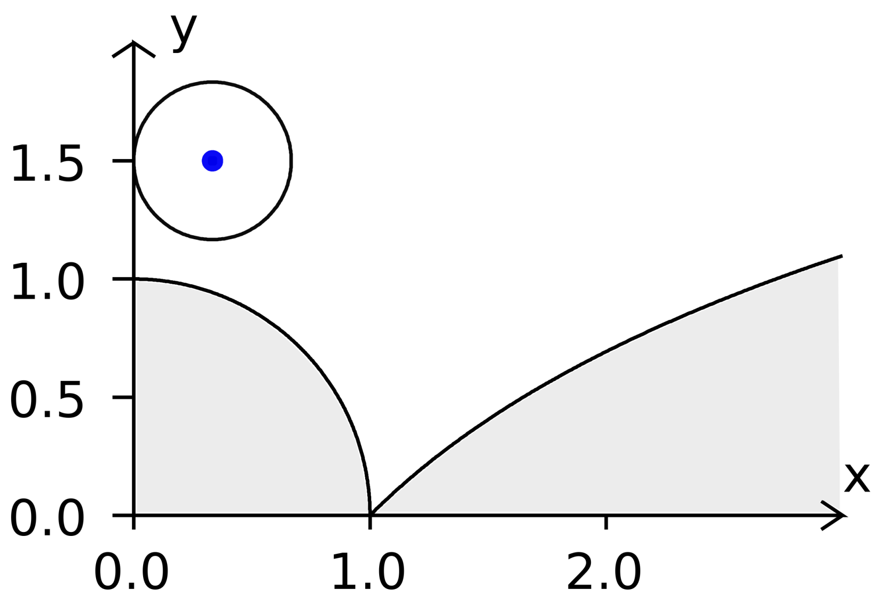
or b) is infeasible

or c) we are at a stationary point of the min-max problem.

Can be used as basis for any globlization routine, e.g., Linesearch.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

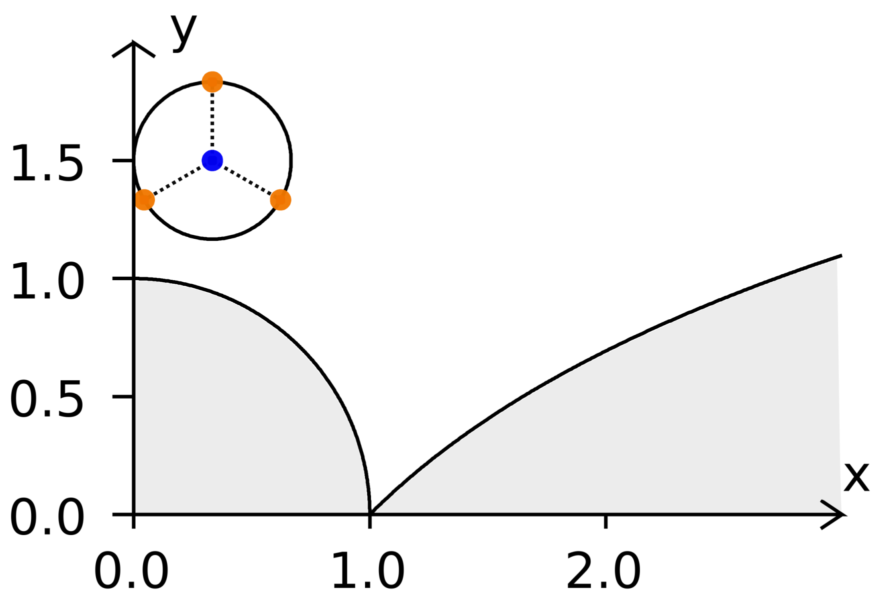
$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{cases}$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

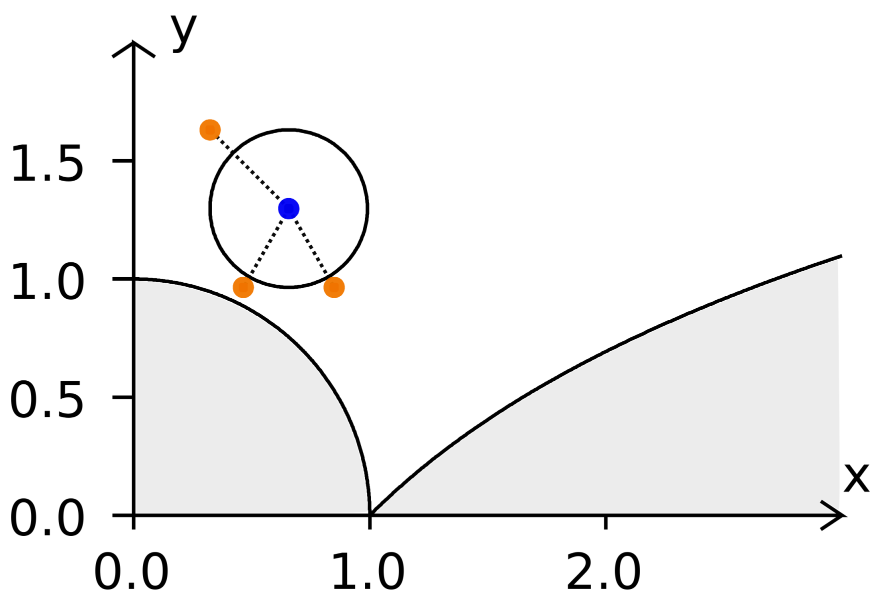
$$\left\{ \begin{aligned} 0 &\geq -x + w \\ 0 &\geq 1 - (x+w)^2 - (y+v)^2 \\ 0 &\geq \log(x+w) - (y+v) \end{aligned} \right.$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

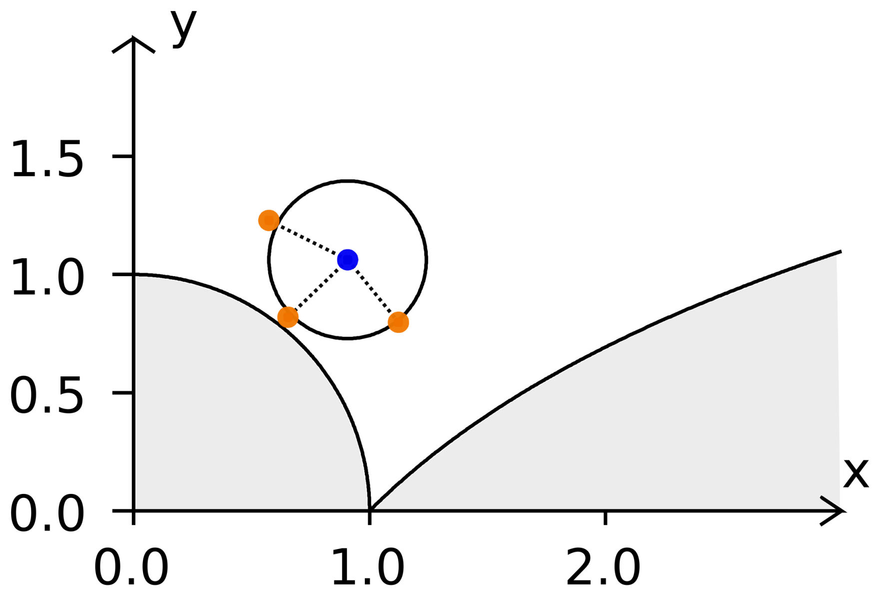
$$\left\{ \begin{array}{l} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{array} \right.$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

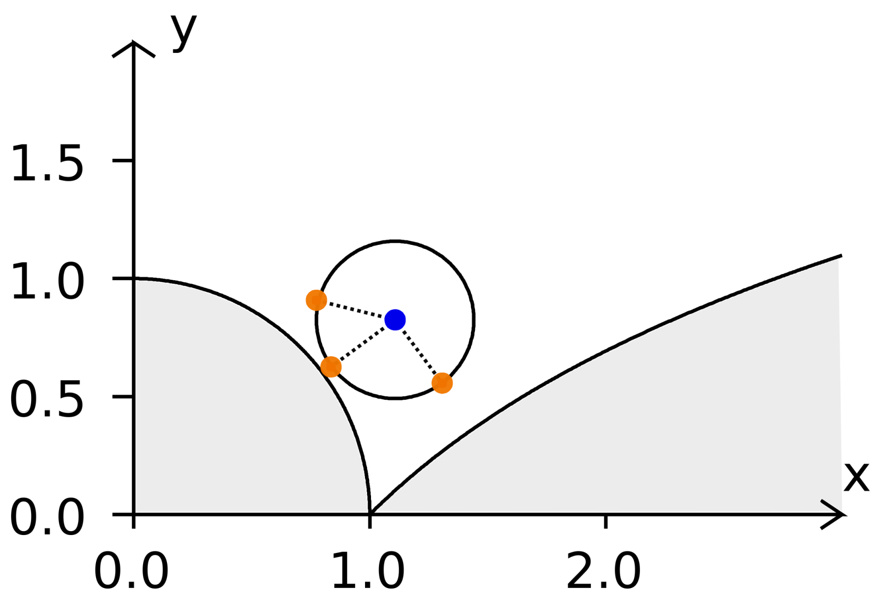
$$\left\{ \begin{array}{l} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{array} \right.$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

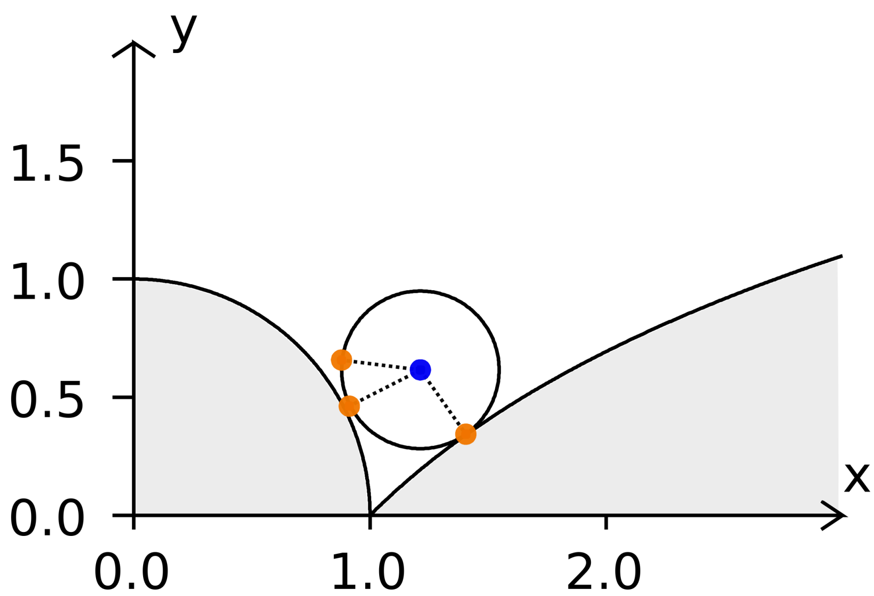
$$\left\{ \begin{array}{l} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{array} \right.$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

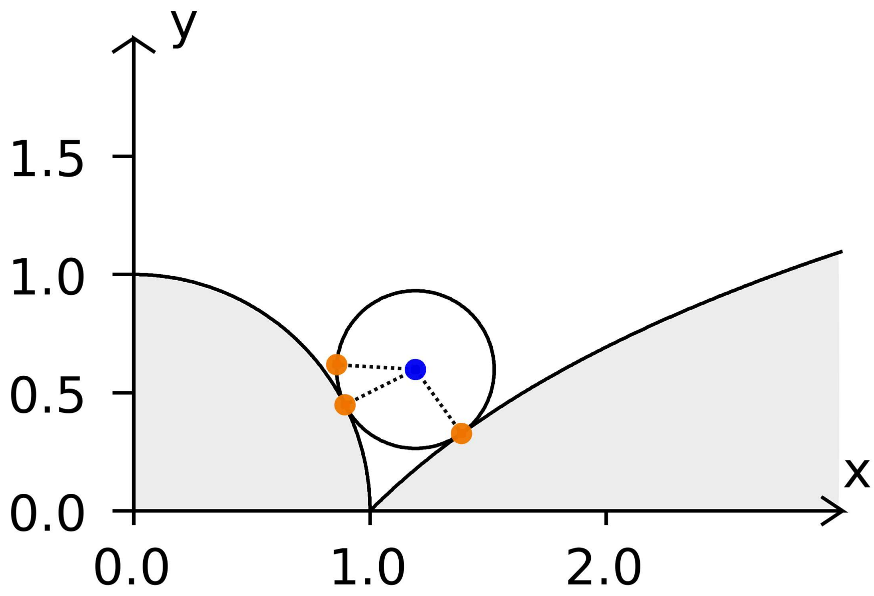
$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{cases}$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

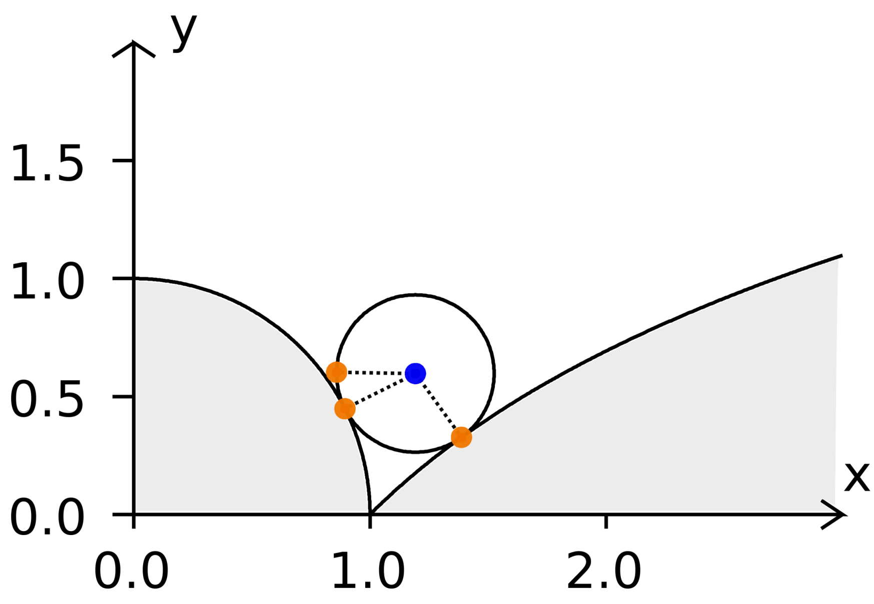
$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{cases}$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

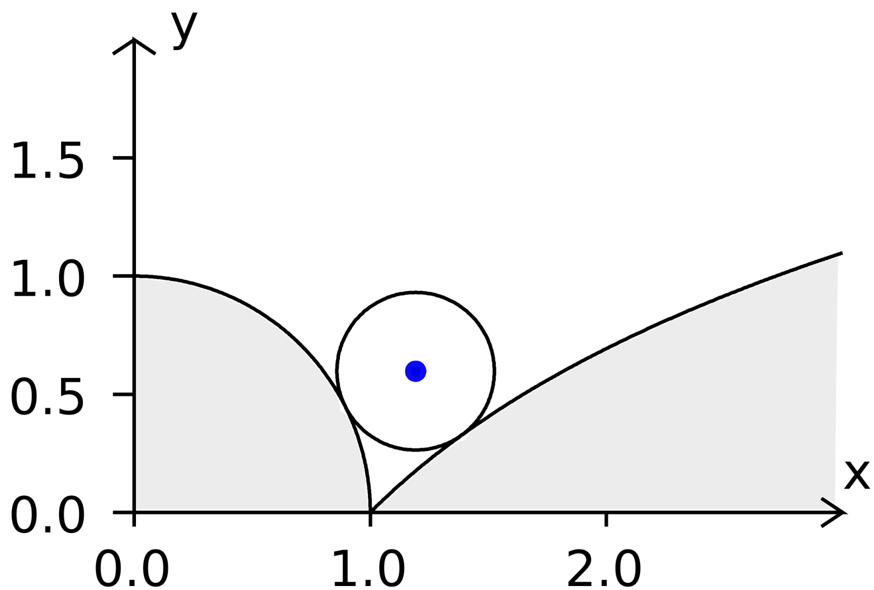
$$\begin{cases} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{cases}$$

Uncertainty set:

- A ball with radius $r = \frac{1}{3}$, i.e. $B(v,w) := v^2 + w^2 - r^2 \leq 0$.

Sequential Convex Bilevel Prog.: An Example

Robust Optimization with a SCBP Method:



$$\min_{x,y} \quad \left(x - \tfrac{1}{2}\right)^2 + y^2$$

subject to

$$\left\{ \begin{array}{l} 0 \geq -x + w \\ 0 \geq 1 - (x+w)^2 - (y+v)^2 \\ 0 \geq \log(x+w) - (y+v) \end{array} \right.$$

Iteration	1	2	3	4	5	6	7	8
$-\log_{10}$(KKT-TOL)	0.3	0.5	0.7	1.0	1.5	3.4	7.0	12.1

Summary

- Linear Approximation Techniques
- Lagrange Dual Relaxations
- Optimality Conditions in SIP
- Min-Max versus MPCC formulations
- Sequential Convex Bi-Level Programming