FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

NATURWISSENSCHAFTLICHE
FAKULTÄT

Trends in Optimization under Uncertainty

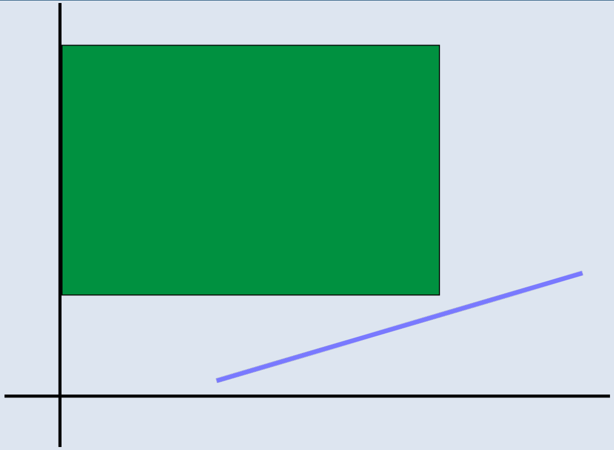
Frauke Liers
FAU Erlangen-Nürnberg

OPTIMIZATION
UNDER UNCERTAINTY
& DATA ANALYSIS

uncertain linear optimization problem

Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$

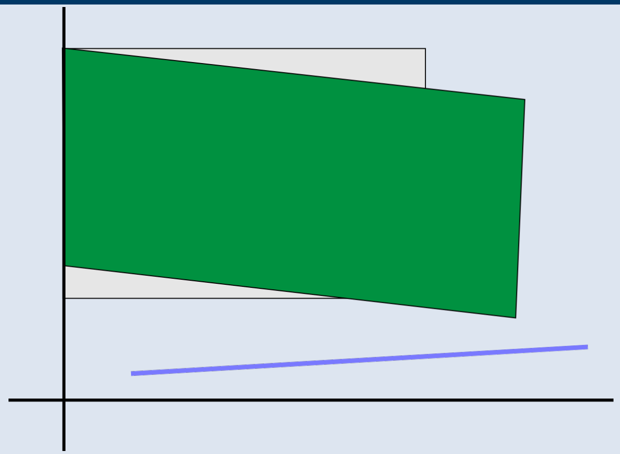
$$\left\{ \min_{x \in \mathbb{R}^n} \left\{ c^\top x \colon Ax \geq b \right\} \right\}_{(c,A,b) \in \mathcal{U}}$$



uncertain linear optimization problem

Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$

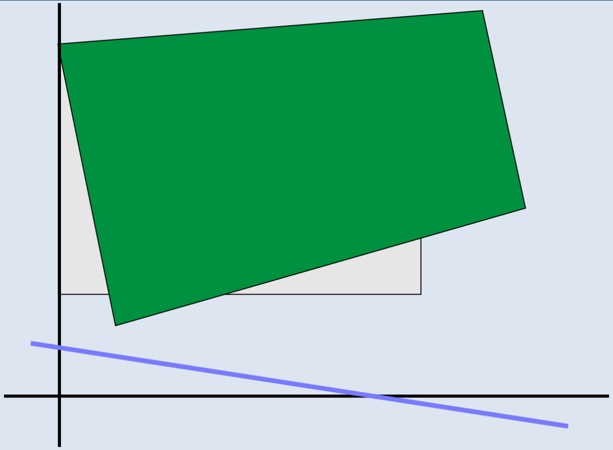
$$\left\{ \min_{x \in \mathbb{R}^n} \left\{ c^\top x \colon Ax \geq b \right\} \right\}_{(c,A,b) \in \mathcal{U}}$$



uncertain linear optimization problem

Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$

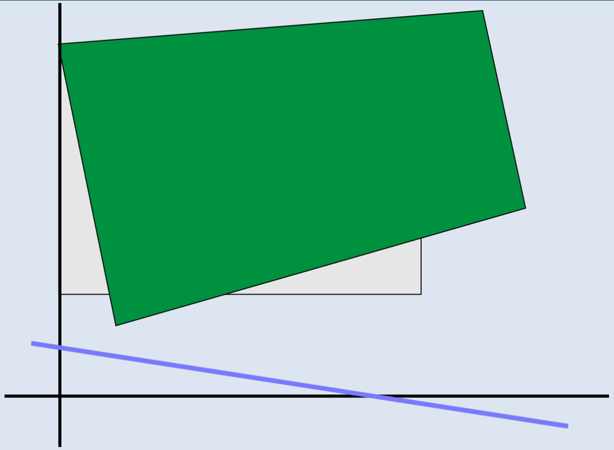
$$\left\{ \min_{x \in \mathbb{R}^n} \left\{ c^\top x \colon Ax \geq b \right\} \right\}_{(c,A,b) \in \mathcal{U}}$$



uncertain linear optimization problem

Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$

$$\left\{ \min_{x \in \mathbb{R}^n} \{ c^\top x \colon Ax \geq b \} \right\}_{(c,A,b) \in \mathcal{U}}$$



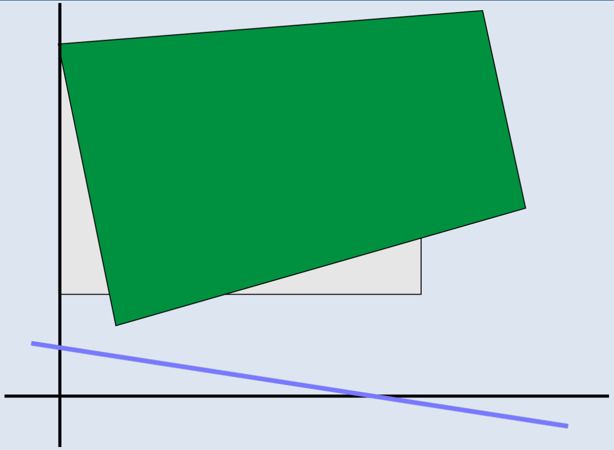
robust counterpart

$$\min_{x \in \mathbb{R}^n} \left\{ c^*(x) := \qquad \qquad \qquad \right\}.$$

uncertain linear optimization problem

Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$

$$\left\{ \min_{x\in\mathbb{R}^n} \{c^\top x \colon Ax \geq b\} \right\}_{(c,A,b)\in\mathcal{U}}$$



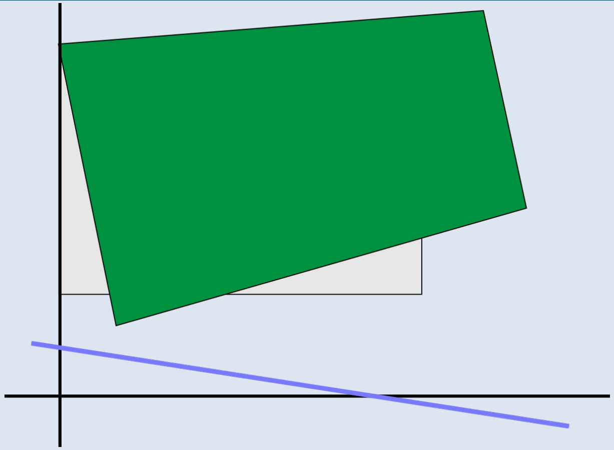
robust counterpart

$$\min_{x\in\mathbb{R}^n} \{c^*(x) := \qquad \}.$$

uncertain linear optimization problem

Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$

$$\left\{ \min_{x\in\mathbb{R}^n} \{c^\top x \colon Ax \geq b\} \right\}_{(c,A,b)\in\mathcal{U}}$$



robust counterpart

$$\min_{x\in\mathbb{R}^n} \left\{ c^*(x) := \sup_{(c,A,b)\in\mathcal{U}} \left\{ c^\top x \colon Ax \geq b \text{ for all } (c,A,b)\in\mathcal{U} \right\} \right\}.$$

Robustness: Full Protection Against Uncertainty

modelling

- input: uncertainty set U (scenarios, intervals, etc.)
- robust feasibility: solution x is feasible $\forall u \in U$ ('here-and-now')
- robust optimality: robust feasible x with best guaranteed solution value

evaluation

- robust and/or probabilistic protection? distributions known?
- mathematical tractability?
- conservatism? adjustability?

Overview

(some of the) contributions

- Soyster (1973), Kouvelis, Yu (1997), Ben-Tal, Nemirovski & co-authors Bertsimas, Sim & co-authors, den Hertog & co-authors, etc.

Overview

(some of the) contributions

- Soyster (1973), Kouvelis, Yu (1997), Ben-Tal, Nemirovski & co-authors Bertsimas, Sim & co-authors, den Hertog & co-authors, etc.
- book / survey: Ben-Tal, El Ghaoui, Nemirovski (2009), Buchheim & Kurtz (2018), Leyffer et al. (2019), Bertsimas & den Hertog (2022), etc.
- distributional robustness: surveys Rahimian & Mehrotra (2019), Lin, Fank, Gao (2022)
- robust combinatorial optimization (book Hartisch & Goerigk (2024))

Overview

(some of the) contributions

- Soyster (1973), Kouvelis, Yu (1997), Ben-Tal, Nemirovski & co-authors Bertsimas, Sim & co-authors, den Hertog & co-authors, etc.
- book / survey: Ben-Tal, El Ghaoui, Nemirovski (2009), Buchheim & Kurtz (2018), Leyffer et al. (2019), Bertsimas & den Hertog (2022), etc.
- distributional robustness: surveys Rahimian & Mehrotra (2019), Lin, Fank, Gao (2022)
- robust combinatorial optimization (book Hartisch & Goerigk (2024))

solution approaches

- reformulation to algorithmically tractable robust counterpart (duality, scenario expansion, KKT, ...)
- decomposition (cutting plane algorithms, scenario generation, ...)
- approximation (Taylor expansion, linearization, safe approximation,...)

this talk

- some introduction in reformulations
- robust (mixed-integer) nonlinear optimization via decomposition
- construction of uncertainty sets over time

Reformulation: Robust Linear Optimization

polyhedral uncertainty Ben-Tal & Nemirovski

$$(\bar{a} + Pu)^T x \leq b \qquad \forall u : Du \leq d \Leftrightarrow \tag{1}$$

$$\bar{a}^T x + \max_{\{u:Du \leq d\}} (P^T x)^T u \leq b.$$

duality trick:

$$\max_u \left\{ (P^T x)^T u \mid Du \leq d \right\} = \min_y \left\{ d^T y \mid D^T y = P^T x, y \geq 0 \right\}.$$

$$\bar{a}^T x + \min_y \left\{ d^T y \mid D^T y = P^T x, y \geq 0 \right\} \leq b.$$

If satisfied by a feasible y, then it is satisfied also for the minimum $\Rightarrow$ skip $\min \rightarrow$ x satisfies (1) iff $\exists\, y$ such that (x, y) satisfies

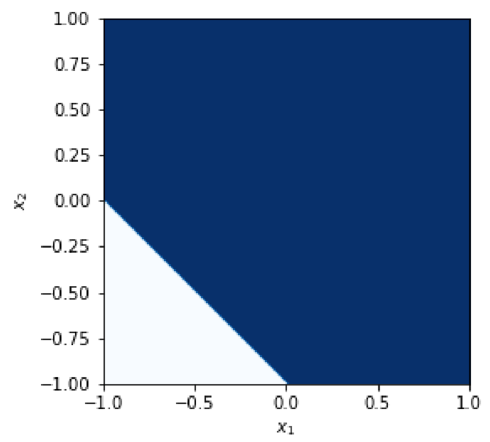
$$\bar{a}^T x + d^T y \leq b, \quad D^T y = P^T x, \; y \geq 0$$

Reformulation for Robust LPs and MIPs

conic uncertainty and integral x Ben-Tal & Nemirovski

- Assume: $\mathcal{K}$ closed, non-empty, pointed, convex cone, strong feasibility holds, i.e., $\exists \bar{u}$ with $D\bar{u} - d \in int(\mathcal{K})$. $\Rightarrow$ finite linear inequality system over dual cone $\mathcal{K}_*$
- Can be applied to *robust MIPs with tractable uncertainty sets* ✓

example ellipsoidal uncertainty set



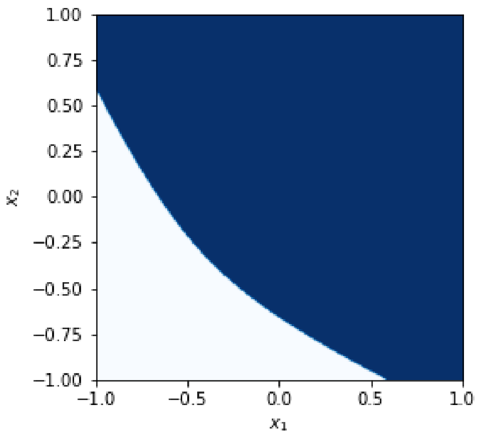
$$x_1 + x_2 \geq -1$$

Reformulation for Robust LPs and MIPs

conic uncertainty and integral x Ben-Tal & Nemirovski

- Assume: $\mathcal{K}$ closed, non-empty, pointed, convex cone, strong feasibility holds, i.e., $\exists \bar{u}$ with $D\bar{u} - d \in int(\mathcal{K})$. $\Rightarrow$ finite linear inequality system over dual cone $\mathcal{K}_*$
- Can be applied to *robust MIPs with tractable uncertainty sets* ✓

example ellipsoidal uncertainty set



$$(1+u_1)x_1 + (1+u_2)x_2 \geq -1, \; \|u\|_2 \leq \frac{1}{2}$$

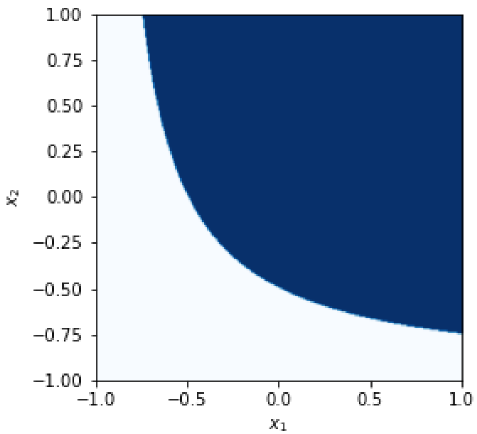
$$\Leftrightarrow x_1 + x_2 - \frac{1}{2}||x||_2 \geq -1$$

Reformulation for Robust LPs and MIPs

conic uncertainty and integral x Ben-Tal & Nemirovski

- Assume: $\mathcal{K}$ closed, non-empty, pointed, convex cone, strong feasibility holds, i.e., $\exists \bar{u}$ with $D\bar{u} - d \in int(\mathcal{K})$. $\Rightarrow$ finite linear inequality system over dual cone $\mathcal{K}_*$
- Can be applied to *robust MIPs with tractable uncertainty sets* ✓

example ellipsoidal uncertainty set



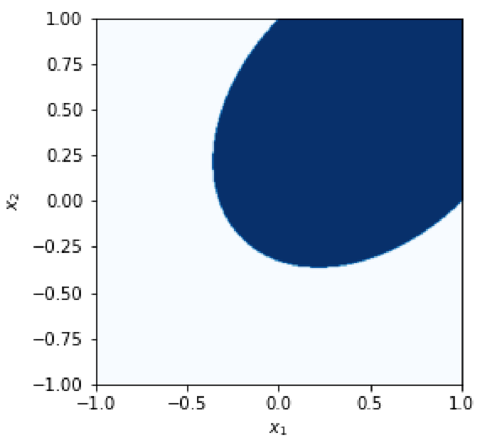
$$(1+u_1)x_1 + (1+u_2)x_2 \geq -1, \; \|u\|_2 \leq 1$$
$$\Leftrightarrow x_1 + x_2 - 1\|x\|_2 \geq -1$$

Reformulation for Robust LPs and MIPs

conic uncertainty and integral x Ben-Tal & Nemirovski

- Assume: $\mathcal{K}$ closed, non-empty, pointed, convex cone, strong feasibility holds, i.e., $\exists \bar{u}$ with $D\bar{u} - d \in int(\mathcal{K})$. $\Rightarrow$ finite linear inequality system over dual cone $\mathcal{K}_*$
- Can be applied to *robust MIPs with tractable uncertainty sets* ✓

example ellipsoidal uncertainty set



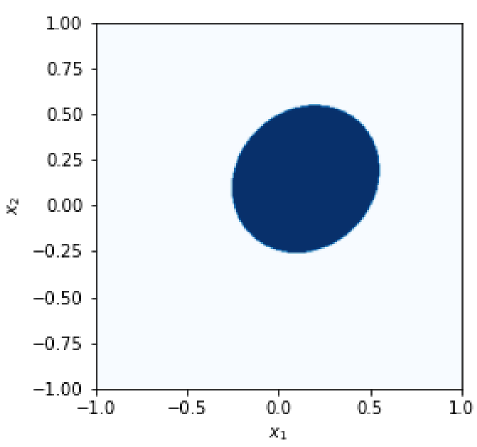
$$(1 + u_1)x_1 + (1 + u_2)x_2 \geq -1, \; ||u||_2 \leq 2$$
$$\Leftrightarrow x_1 + x_2 - 2||x||_2 \geq -1$$

Reformulation for Robust LPs and MIPs

conic uncertainty and integral x Ben-Tal & Nemirovski

- Assume: $\mathcal{K}$ closed, non-empty, pointed, convex cone, strong feasibility holds, i.e., $\exists \bar{u}$ with $D\bar{u} - d \in int(\mathcal{K})$. $\Rightarrow$ finite linear inequality system over dual cone $\mathcal{K}_*$
- Can be applied to *robust MIPs with tractable uncertainty sets* ✓

example ellipsoidal uncertainty set

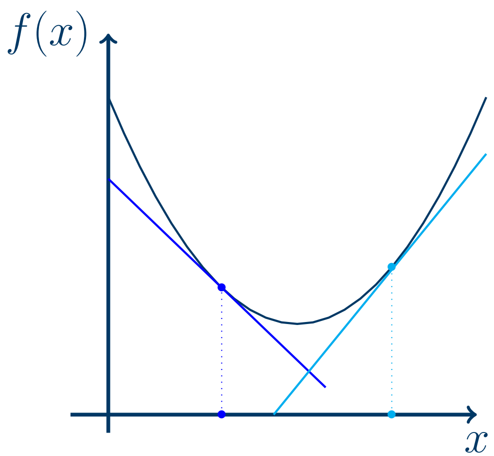


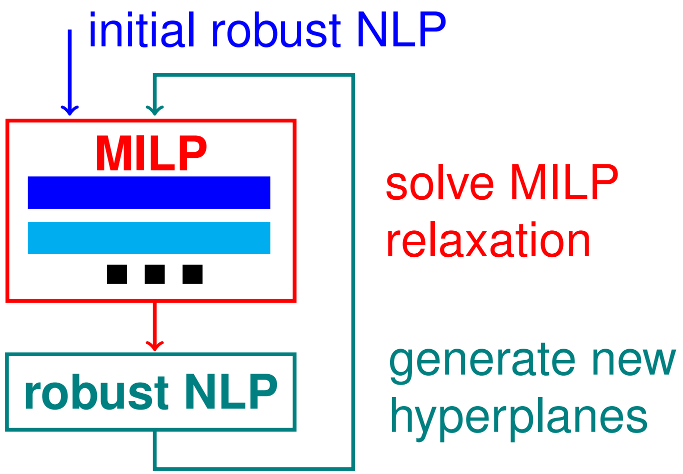
$$(1+u_1)x_1 + (1+u_2)x_2 \geq -1, \; \|u\|_2 \leq 3$$
$$\Leftrightarrow x_1 + x_2 - 3\|x\|_2 \geq -1$$

Some Trends and Research Questions

- multi-stage problems
- How to solve nonlinear robust problems with integral decisions? Leyffer, Menickelly, Munson, Vanaret, Wild (survey 2018)
- Size of uncertainty sets?

Discrete-Continuous Robust Optimization via Outer Approximation

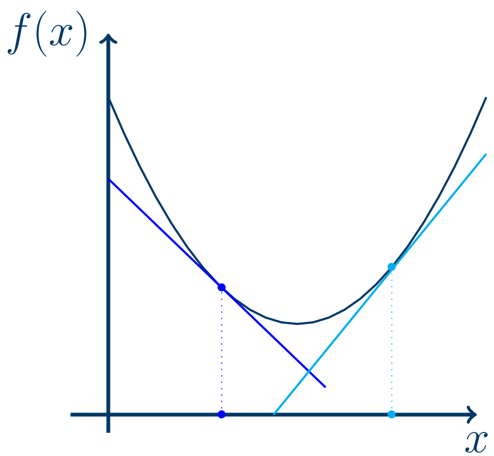


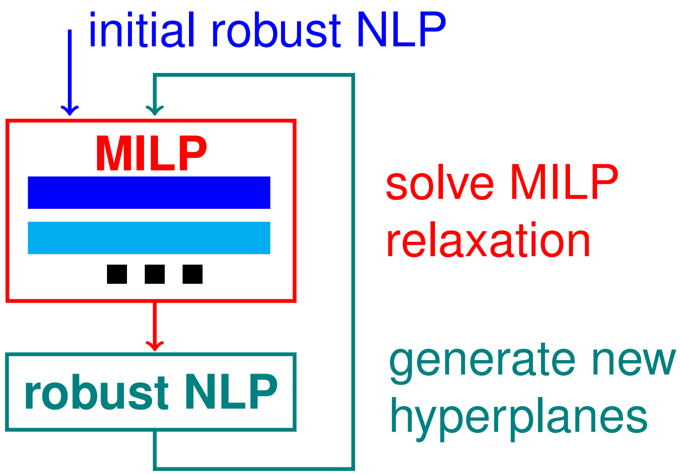


MILP $\Rightarrow$ lower bound

robust NLP $\Rightarrow$ upper bound

Discrete-Continuous Robust Optimization via Outer Approximation





MILP $\Rightarrow$ lower bound

robust NLP $\Rightarrow$ upper bound

- Solution of (nonlinear) robust subproblems?
  Kuchlbauer, L, Stingl (2022)
- Valid inequalities for master problems?
  Kuchlbauer, L, Stingl (2022)

Nonlinear Robust Optimization

formulation as minimax problem

$$\min_{x \in \mathbb{R}^n} \quad c(x)$$
$$v(x,u) \le 0 \quad \forall u \in \mathcal{U}.$$

minimax problem: $$\min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} v(x,u). \qquad \text{(RO)}$$

challenges:

- evaluation of worst case: global solution of $\max_{u \in \mathcal{U}} v(x,u)$
- nonlinear and non-convex
- few works only (Leyffer et al., 2020), no general approaches
- known reformulations need strong structural assumptions

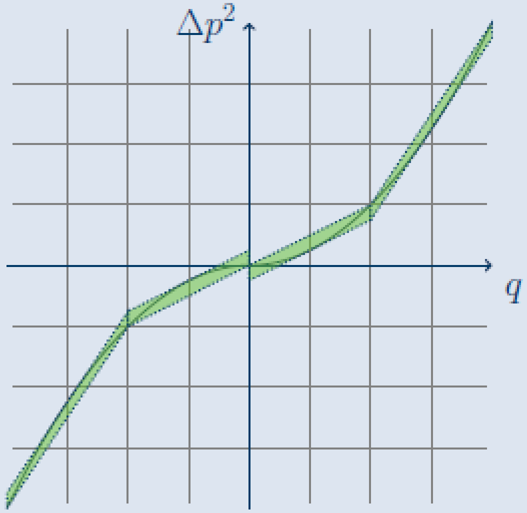
Our Solution Approach

adversarial problem

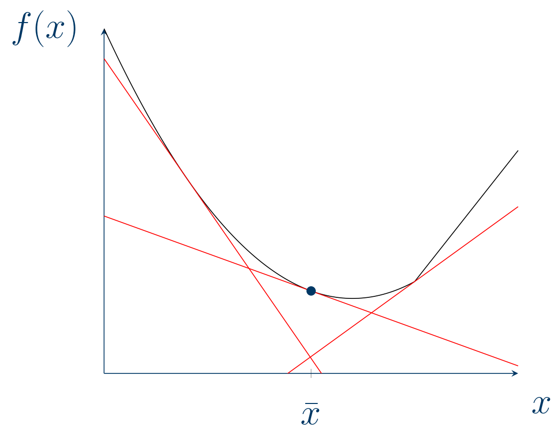
minimize the adversary's optimal value function

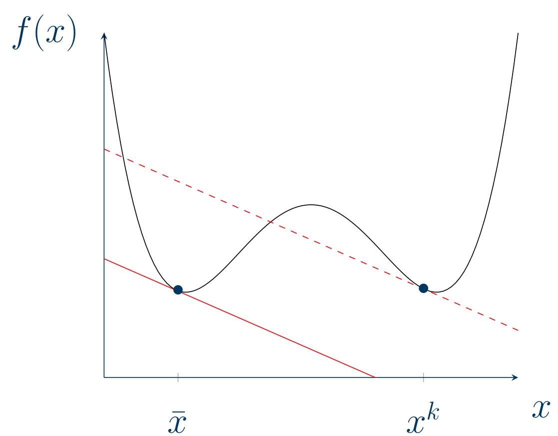
$$f(x) := \max_{u \in \mathcal{U}} v(x, u). \quad \text{(RO)}$$

- bundle method for non-smooth and non-convex function $\min_{x \in \mathbb{R}^n} f(x)$
- piecewise linear relaxation for $\max_{u \in \mathcal{U}} v(x, u)$ with guaranteed error bound Geißler et al. (2012)



Bundle Method





inner loop:

- approximate f by a piecewise linear model ϕ_k, using cutting planes
- use subgradients for nonsmooth functions
- find trial iterates around serious $\bar{x}$:

$$\min_{x^k \in \mathbb{R}^n} \phi_k(x^k) + \tau_k \|x^k - \bar{x}\|.$$

- downshift cutting planes to overcome lack of convexity

outer loop: accept trial iterate as new serious point if ϕ_k good enough

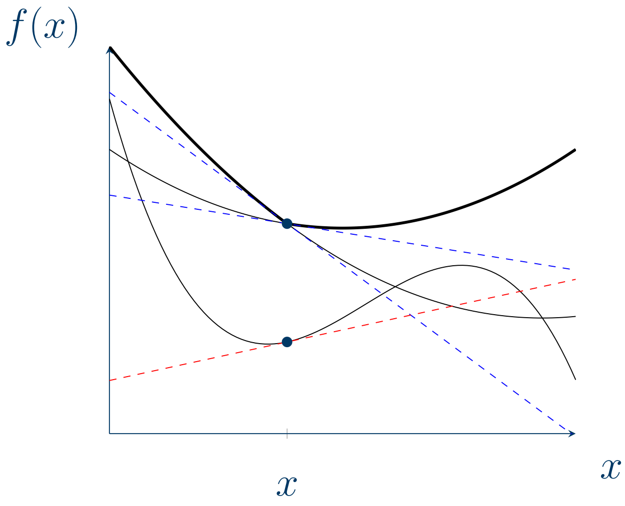
e.g., constant error in subgradient, convex: Kiwiel 2006, nonconvex: Noll 2013, Hertlein & Ulbrich (2019)

Inexactness in the Adversarial Problem

$$\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} v(x, u). \qquad \text{(RO)}$$

inexact solution u_x to the adversarial problem:

$$v(x, u_x) \geq \max_{u \in \mathcal{U}} v(x, u) - \varepsilon_x.$$



Clarke subdifferential of f at x: $\partial f(x) =$

$$\text{conv}\{\partial_x v(x, u^*) \mid u^* \in \mathcal{U},\ v(x, u^*) = \max_{u \in \mathcal{U}} v(x, u)\}.$$

$$\tilde{\partial}_a f(x) := \text{conv}\{\partial_x v(x, u) \mid u \in \mathcal{U},\ v(x, u) \geq v(x, u_x)\}.$$

approximate exact subdifferential from outside $\Rightarrow$ no constant error bound implied, no bundle concept available, need to generalize Noll!

Inexactness in the Adversarial Problem

$$\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} v(x,u). \qquad \text{(RO)}$$

inexact solution u_x to the adversarial problem:

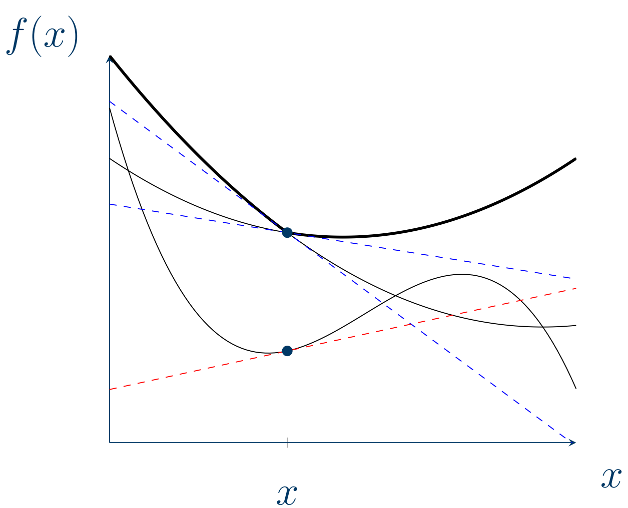
$$v(x,u_x) \geq \max_{u \in \mathcal{U}} v(x,u) - \varepsilon_x.$$

Clarke subdifferential of f at x: $\partial f(x) =$

$$\operatorname{conv}\{\partial_x v(x,u^*) \mid u^* \in \mathcal{U},\ v(x,u^*) = \max_{u \in \mathcal{U}} v(x,u)\}.$$

$$\tilde{\partial}_a f(x) := \operatorname{conv}\{\partial_x v(x,u) \mid u \in \mathcal{U},\ v(x,u) \geq v(x,u_x)\}.$$

approximate exact subdifferential from outside $\Rightarrow$ no constant error bound implied, no bundle concept available, need to generalize Noll!



fix by adaptivity!

Adaptive Approximation of Function Value

Definition

- $f : \mathbb{R}^n \to \mathbb{R}$ is approximate convex ($= LC^1$) if for every x and ε', there exists $\delta > 0$ s.t. f is ε'-convex on $B(x, \delta)$.
  (e.g., $\max_u f(x, u)$ with $f(., u) \in C^1 \forall u$)
- $f : \mathbb{R}^n \to \mathbb{R}$ is ε'-convex if for any $x, x' \in X$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda) f(x') + \varepsilon' \lambda(1 - \lambda)\|x - x'\|.$$

Lemma

Under ε'-convexity, an approximate subgradient $g_k \in \tilde{\partial}_a f(x^k)$ fulfills

$$g_k{}^T(x - x^k) \leq f(x) - f(x^k) + \varepsilon_{x^k} + \varepsilon' \|x - x^k\|.$$

$\Rightarrow$ for trial iterates x^k around a serious iterate $\bar{x}$, evaluate adversary well enough: $\varepsilon_{x^k} = \varepsilon'' \|\bar{x} - x^k\|$

$\Rightarrow$ subgradient inequality $g_k{}^T(\bar{x} - x^k) \leq f_a(\bar{x}) - f_a(x^k) + (\varepsilon' + \varepsilon'')\|\bar{x} - x^k\|$.

Convergence Result

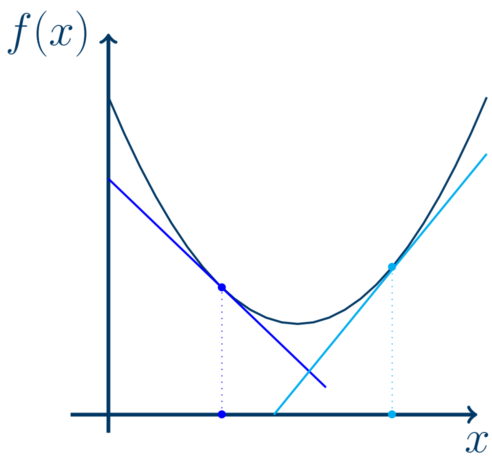
Theorem

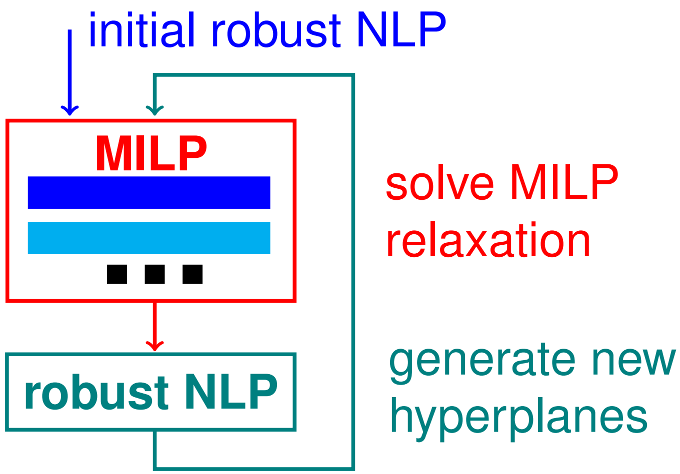
Let x_1 be s.t. $\Omega := \{x \in \mathbb{R}^n : f(x) \leq f_a(x_1)\}$ is bounded, v be approximate convex (LC^1) and $\bar{x}$ obtained by a stopping criterien or an accumulation point of serious iterates. Then, it holds that

$$0 \in \tilde{\partial}_a f(\bar{x}).$$

Optimality: $0 \in \tilde{\partial}_a f(\bar{x})$, i.e. $f(\bar{x}) \leq f(y) + \varepsilon' \|\bar{x} - y\| + \varepsilon_{\bar{x}}$, locally.

Discrete-Continuous Robust Optimization via Outer Approximation





MILP $\Rightarrow$ lower bound

robust NLP $\Rightarrow$ upper bound

- Solution of (nonlinear) robust subproblems?
  Kuchlbauer, L, Stingl (2020)
- Valid inequalities for master problems?
  Kuchlbauer, L, Stingl (2022)

Convex Mixed-Integer Robustness

$$\begin{aligned} \min_{x,y} \quad & C(x,y) \\ & G(x,y) := \max_{u \in \mathcal{U}} \sum_{i=1}^{n} V_i^+(x,y,u) \leq 0, \\ & x \in X, y \in Y \cap \mathbb{Z}^{n_y}. \end{aligned}$$

C convex in x, y, allow non-convex adversarial

bundle method $\rightarrow$ subproblem solutions and cutting planes

- generalizes OA proofs [Fletcher, Leyffer, 1994; Delfino, Oliveira, 2018; Wei et al. 2019]
- inexact cutting planes: valid, but acceptance of ε_k-feasible solutions

Theorem

The outer approximation method together with the adaptive bundle method terminates after finitely many outer approximation iterations and either detects infeasibility or outputs a solution that is ε_k-feasible and ε_{oa}-optimal.

Robust Nomination Validation in Gas Networks

Problem

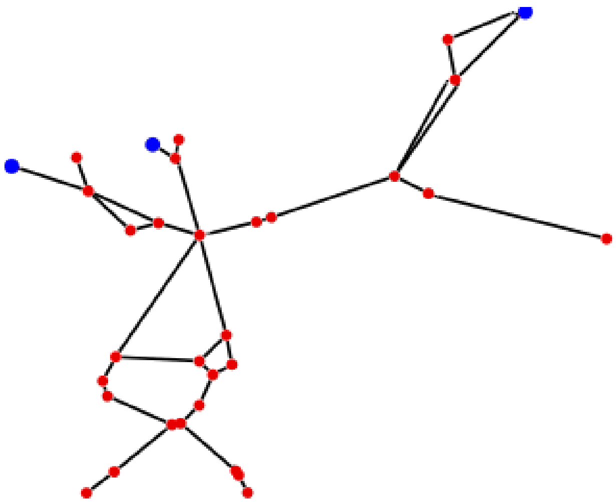
For each uncertainty, is there a configuration of the active elements leading to a feasible state?

$$
\begin{aligned}
\min_{\Delta,q,\pi} \quad & c(\Delta) \\
& Aq = d && \text{(flow conservation)} \\
& (A^T\pi)_a = -\lambda_a q_a |q_a| \quad \forall a \in \mathcal{A}_{pi} && \text{(pressure loss on pipes)} \\
& (A^T\pi)_a = \Delta_a \quad \forall a \in \mathcal{A}_{ac} && \text{(active elements)} \\
& \pi \in [\underline{\pi}, \overline{\pi}] && \text{(pressure bounds)} \\
& q \in \mathbb{R}^{|\mathcal{A}|}.
\end{aligned}
$$

uncertainties in demands d and in physical parameters λ.
in bundle: Adversary maximizes constraint violations.

Numerical Results on Realistic Instances

nodes × arcs	compr.	valves	runtime	error constr.
11×11	2	1	13	0
24×25	4	0	8	0
40×45	5	2	903	0
103×105	21	3	362	5e-5



General approach can go to realistic sizes.

Curently: Robust Chance-Constrained Optimization

Bernhard, L, Stingl

$$\min_{\Delta \in [\underline{\Delta}, \overline{\Delta}]} \quad C(\Delta)$$
$$s.t. \quad \mathbb{P}_{d \sim \zeta}(\pi_v(\Delta, d, \lambda) \in [\underline{\pi}_v, \overline{\pi}_v] \; \forall v \in V) \geq p \qquad \forall \zeta \in U.$$

- assume discrete, uncertain, probability distribution in ambiguity set U
- adaptive bundle method applicable for (approximate) robust joint CC ⇒ local solution

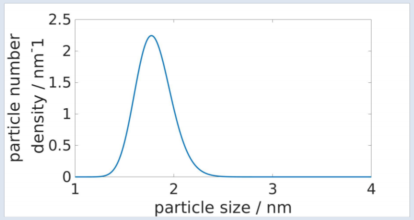
preliminary computational results: U built from confidence intervals

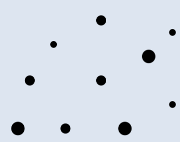
- gaslib24, gaslib40, 500 scen.: ca. 170 s, 1000 scen.: ca. 300 s
- gaslib134, 500 scen.: 172 s, 1000 scen: 400 s

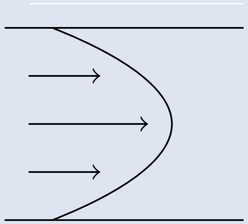
Quality Control in Nanoparticle Design

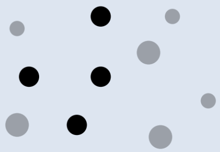
Kuchlbauer, Dienstbier, Muneer, Hedges, Stingl, L, Pflug (2024), CRC 1411

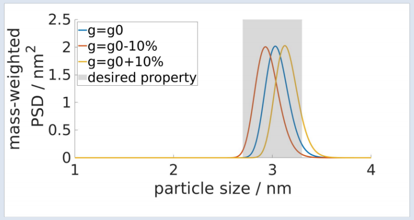
particle synthesis process is ruled by PDEs











Initial PSD — Seed — Flow tube reactor — Product Target — Final PSD

Population-Balance equation: (x particle size, t time, G_0, G_1 concentration- / size-dependent growth, q particle number density, q_0 initial particle distribution, N nucleation rate)

$$\partial_t q(t,x) + \partial_x \left(G_0(c(t)G_1(x)q(t,x)\right) = 0 \qquad \forall t > 0, x > 0,$$

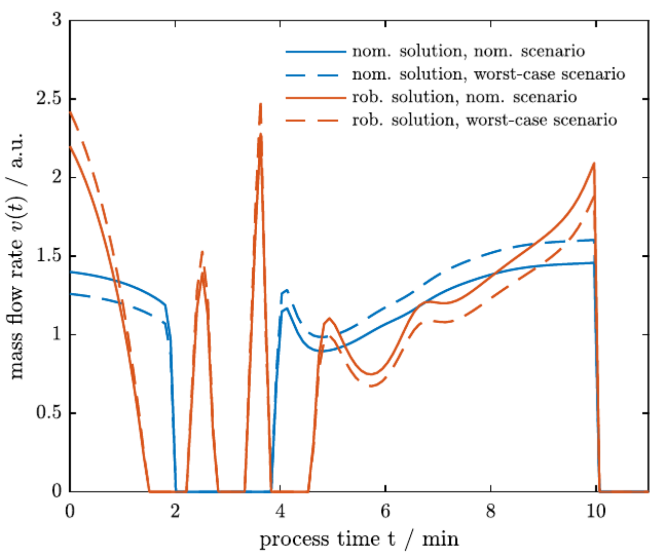
$$q(0,x) = q_0(x) = 0 \quad \forall x > 0.$$

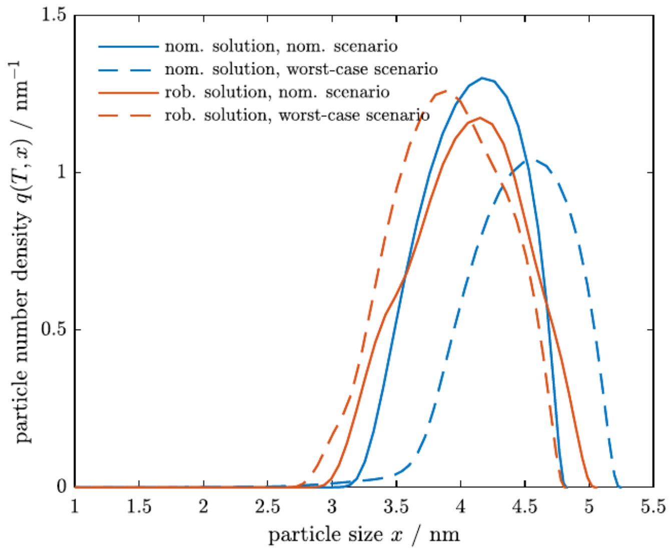
$$G_0(c(t))G_1(x_n)q(t,x_n) = N(c(t)) \qquad \forall t > 0.$$

Robust ZnO synthesis + nucleation and growth

robust protection is algorithmically tractable and not costly

- bundle method only needs a procedure that can evaluate the adversary up to a given error





⇒ bundle method can be made concrete to optimize nonlocal population balance PDEs efficiently.

this talk

- some introduction in reformulations
- robust (mixed-integer) nonlinear optimization via decomposition
- construction of uncertainty sets over time

How to learn uncertainty & protection over time?

Aigner, Bärmann, Braun, Pokutta, Schneider, Sharma, Tschuppik

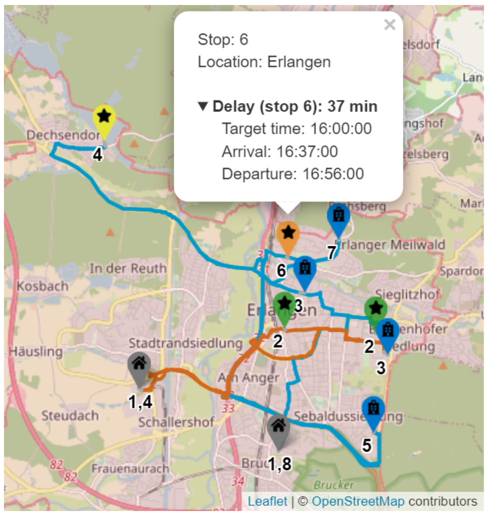
strict robust protection $\xrightarrow{\text{learn 'true' uncertainty}}$ stochastic solution

problem formulation

optimization over $\mathcal{X}$ with uncertain objective $f : \mathcal{X} \times \mathcal{S} \to \mathbb{R}$.
s discrete random vector with realizations in $\mathcal{S} = \{s^1, ..., s^{|\mathcal{S}|}\}$ and probability vector $p^* = (p_1^*, ..., p_{|\mathcal{S}|}^*)^T \in [0,1]^{|\mathcal{S}|}$.

stochastic problem $$\min_{x \in \mathcal{X}} \mathbb{E}_{s \sim p^*} [f(x, s)] := \sum_{k=1}^{|\mathcal{S}|} f(x, s^k) p_k^*$$

Motivation: Delay Minimization in Ambulances



Marker	Location	Delay [min]	
		from	to
	Pick up		15
	Pick up	16	30
	Pick up	31	60
	Pick up	61	
	Station	---	---
	Destination	---	---

Leithäuser, Büsing L, et al., BMC Medical Inform. & Decision Making (2022)

- robust logistics of (plannable) + ad-hoc transports
- vehicle routing problem with (soft + hard) time windows under uncertainty, e.g., Toth & Vigo (2014), Ibaraki et al (2005), Blauth, Agra et al (2013), Eufinger et al. (2020), ...

Distributionally robust optimization (DRO)

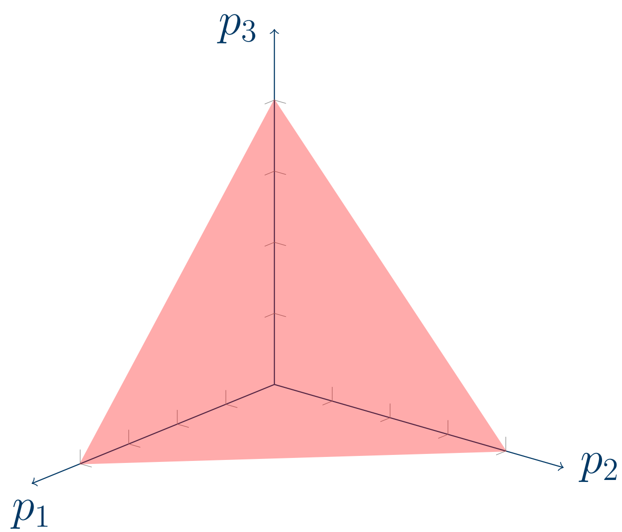
probability vector p^* for scenarios (in practice) unknown or uncertain

DRO approach

minimize worst-case expectation

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}} \mathbb{E}_{s \sim p}[f(x,s)] = \min_{x \in \mathcal{X}} \max_{s \in \mathcal{S}} f(x,s),$$

where ambiguity set $\mathcal{P}\{p \in [0,1]^{|\mathcal{S}|} \mid \sum_{k=1}^{|\mathcal{S}|} p_k = 1\}$ is the probability simplex



Data-driven distributionally robust optimization

no knowledge about p^*

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}} \mathbb{E}_{s \sim p} [f(x, s)] \quad (RO)$$

$$\mathcal{P} = \{p \in [0,1]^{|\mathcal{S}|} \mid \textstyle\sum_{k=1}^{|\mathcal{S}|} p_k = 1\}$$

full knowledge about p^*

$$\min_{x \in \mathcal{X}} \mathbb{E}_{s \sim p^*} [f(x, s)] \quad (SO)$$

$$\mathcal{P}^* = \{p^*\}$$

Data-driven distributionally robust optimization

no knowledge about p^*

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}} \mathbb{E}_{s \sim p} [f(x,s)] \quad (RO)$$

$$\mathcal{P} = \{p \in [0,1]^{|\mathcal{S}|} \mid \sum_{k=1}^{|\mathcal{S}|} p_k = 1\}$$

full knowledge about p^*

$$\min_{x \in \mathcal{X}} \mathbb{E}_{s \sim p^*} [f(x,s)] \quad (SO)$$

$$\mathcal{P}^* = \{p^*\}$$

Data-driven DRO

construct from $N > 0$ (i.i.d.) observations an ambiguity set $\mathcal{P}^N$ with

- $\mathcal{P}^* \subseteq \mathcal{P}^N \subseteq \mathcal{P}$ (often with some confidence)
- $\mathcal{P}^N \to \mathcal{P}^*$ for $N \to \infty$ (in some stochastic sense)

and solve

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}^N} \mathbb{E}_{s \sim p} [f(x,s)] \quad (DRO_N)$$

Data-driven DRO

goal: uncertainty/ambiguity sets as small as possible, but as large as necessary to guarantee solution quality

Literature

- nonlinear convex confidence regions [Bertsimas et al., 2018]
- Wasserstein balls [Esfahani and Kuhn, 2018]
- momentum based sets [Delage and Ye, 2010]
- ...

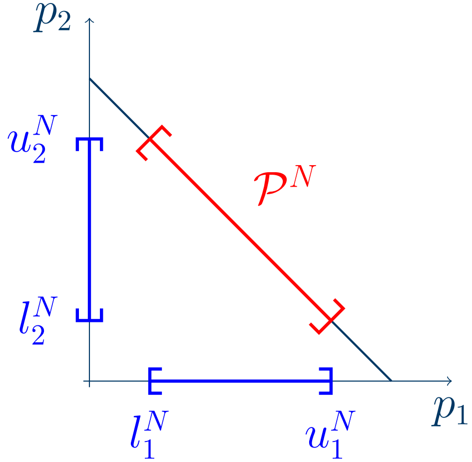
Ambiguity set with confidence intervals

$$\mathcal{P}^N = \left\{ p \in [0,1]^{|\mathcal{S}|} \mid \sum_{k=1}^{|\mathcal{S}|} p_k = 1 \,,\, l^N \leq p \leq u^N \right\}$$

construct confidence intervals s.t.

$$l_k^N \leq p_k^* \leq u_k^N \quad \forall\, k = 1, \ldots, |\mathcal{S}|,$$

with a predefined confidence probability



remark: $[l^N, u^N]$ shrinks with $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ for fixed confidence multinomial distr. Sison-Glaz ('95)

Solving DRO via reformulation

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}^N} \mathbb{E}_{s \sim p}\left[f(x,s)\right] = \min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}^N} \sum_{k=1}^{|\mathcal{S}|} f(x, s^k) p_k \quad (DRO_N)$$

reformulation

$$\begin{aligned} \min_{x,z,\alpha,\beta} \quad & z - (l^N)^\top \alpha + (u^N)^\top \beta \\ \text{s.t.} \quad & z - \alpha_k + \beta_k \geq f(x, s^k) \quad \forall k = 1, \ldots, |\mathcal{S}|, \\ & \alpha, \beta \geq 0. \\ & x \in \mathcal{X} \end{aligned}$$

Data-driven DRO over time

time-dependent ambiguity sets:

$$\mathcal{P}^t = \left\{ p \in [0,1]^{|\mathcal{S}|} \mid \sum_{k=1}^{|\mathcal{S}|} p_k = 1 \,,\, l^t \leq p \leq u^t \right\}$$

or $$\mathcal{P}^t = \left\{ p \in [0,1]^{|\mathcal{S}|} \mid \sum_{k=1}^{|\mathcal{S}|} p_k = 1 \,,\, \|p - \hat{p}^t\|_A \leq \varepsilon_t \right\}$$

Initialize $\mathcal{P}^0 = \mathcal{P}$.

DRO for every time step $t = 1, ..., T$ (rounds, e.g. days, hours, ...)

1. compute solution x^t of

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}^{t-1}} \mathbb{E}_{s \sim p} \left[f(x,s) \right], \qquad (DRO_t)$$

2. observe realization of uncertainty and update $\mathcal{P}^t$ with confidence $1 - \frac{6\delta}{\pi^2 t^2}$, where $\delta \in (0,1)$

confidence update $\sum_{t=1}^{\infty} \frac{6\delta}{\pi^2 t^2} = \delta$ implies $p^* \in \cap_{t=1}^{\infty} \mathcal{P}^t$ with a probability at least $1 - \delta$

DRO over time

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}^{t-1}} \mathbb{E}_{s \sim p} [f(x, s)] \quad (DRO_t)$$

$$\min_{x \in \mathcal{X}} \mathbb{E}_{s \sim p^*} [f(x, s)] \quad (SO)$$

Theorem

- $(DRO_t) \to (SO)$ for $t \to \infty$ (with prob. 1)
- if compact and f continuous in x, then any accumulation point of the sequence of solutions $\{x^t\}_{t=1}^{\infty}$ of (DRO_t) is almost surely an optimal solution to the problem (SO)

along the lines of proof strategy from Esfahani, Kuhn and

$$\lim_{t \to \infty} \sup_{p \in \mathcal{P}^t} \|p - p^*\| = 0 \text{ with probability 1.}$$

Data-driven DRO over time

$$\min_{x \in \mathcal{X}} \max_{p \in \mathcal{P}^{t-1}} \mathbb{E}_{s \sim p} [f(x, s)] \quad (DRO_t) \qquad\qquad \min_{x \in \mathcal{X}} \mathbb{E}_{s \sim p^*} [f(x, s)] \quad (SO)$$

Theorem

- $(DRO_t) \to (SO)$ for $t \to \infty$ (with prob. 1)
- If compact and f continuous in x, then any accumulation point of the sequence of solutions $\{x^t\}_{t=1}^{\infty}$ of (DRO_t) is almost surely an optimal solution to the problem (SO).

difficulty of solving (DRO_t) depends on f and cardinality of $\mathcal{S}$
$\hookrightarrow$ can be computationally expensive
$\hookrightarrow$ different approach: DRO via online learning

Online (distributionally) robust optimization

literature

- online first-order robust optimization[Ben-Tal et al., 2014], [Ho-Nguyen and Kilinc-Karzan, 2018]
- weak/strong learning algorithms for robust optimization [Pokutta and Xu, 2021]
- DRO Bayesian optimization [Kirschner et al., 2020]
- ...

Data-driven DRO via online learning

algorithm

1. initialize $p^0 \in \mathcal{P}^0\mathcal{P}$ and $x^0 \in$.
2. for rounds $t = 1, ..., T$:

   $\tilde{p}^t \leftarrow p^{t-1} + \eta \nabla_p \mathbb{E}_{s\sim p^t}\left[f(x^{t-1}, s)\right]$ (gradient descent step)

   $p^t \leftarrow \arg\min_{p \in \mathcal{P}^{t-1}} \frac{1}{2}\|p - p^{t-1}\|^2$ (projection step)

   $x^t \leftarrow \arg\min_{x \in \mathcal{X}} \mathbb{E}_{s\sim p^t}\left[f(x, s)\right]$ (stochastic solution)

   $\mathcal{P}^t \leftarrow$ observe data and update $\mathcal{P}^{t-1}$ with confidence $1 - \frac{6\delta}{\pi^2 t^2}$
3. Output: $(x^1, ..., x^T)$

Regret bound

dynamic regret bound

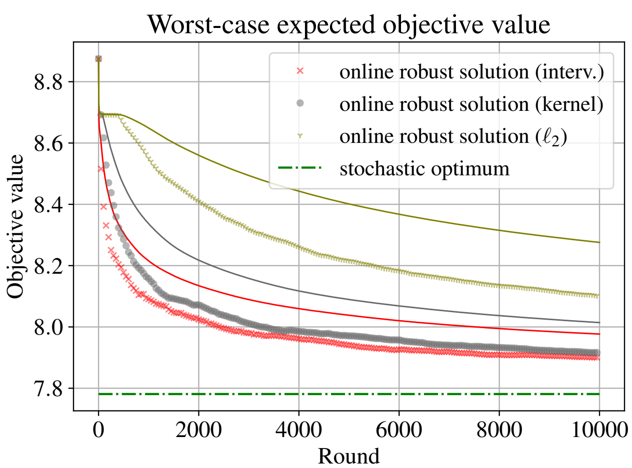
Let $f : \mathcal{X} \times \mathcal{S} \to \mathbb{R}$ be uniformly bounded, i.e., for all $(x, s) \in \mathcal{X} \times \mathcal{S}$ exists a $G > 0$, s.t. $|f(x, s)| \leq G$. Let $\eta := \sqrt{\frac{2h(T)}{G^2 T|\mathcal{S}|}}$, where $\sum_{t=1}^{T} \frac{1}{2}\|p^t - q^t\|^2 \leq h(T)$ for $p \in \mathcal{P}^{t-1}$ and $q^t \in \mathcal{P}^t$.
The output $(x^1, ..., x^T)$ of our algorithm with confidence update $\delta^t \frac{6\delta}{\pi^2 t^2}$ and $\delta \in (0, 1)$ fulfills

$$\frac{1}{T}\sum_{t=1}^{T}\left(\max_{p\in\mathcal{P}^t} \mathbb{E}_{s\sim p}\left[f(x^t, s)\right] - \min_{x\in\mathcal{X}}\max_{p\in\mathcal{P}^t} \mathbb{E}_{s\sim p}\left[f(x, s)\right]\right) \leq G\sqrt{\frac{2|\mathcal{S}|h(T)}{T}} + \frac{2G}{T},$$

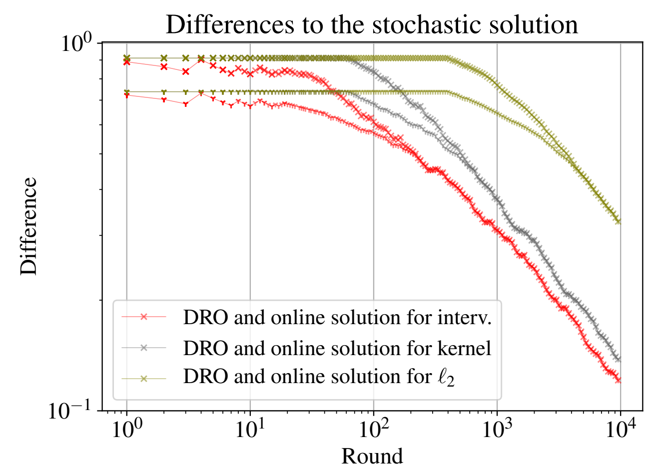
with a probability of at least $1 - \delta$.

$\hookrightarrow$ average error between worst-case performance of online solution and DRO solution decreases with a rate of $\mathcal{O}\left(\frac{\sqrt{h(T)}}{\sqrt{T}}\right)$, where $h(T) \leq \mathcal{O}(\log^2 T)$.

Numerical results



(a) evolution of objective



(b) convergence of solutions

Figure: solutions of our algorithm for instance blend2 from MIPLIB with $|\mathcal{S}| = 10$ and $T = 10000$ for different ambiguity sets

Numerical Results - Running Times

15 smallest MIPLIB, 10 smallest QPLIB instances, w. confidence intervals.

Instance	$\varnothing$**Our Method**	$\varnothing$**Reformulation**
MIP, $\lvert\mathcal{S}\rvert = 10$	52.4s	115.8s
MIP, $\lvert\mathcal{S}\rvert = 50$	57.7s	176.7s*
MIQP, $\lvert\mathcal{S}\rvert = 2$	170.2s	359.6s*

Avg time per iteration for DRO via reformulation and with our online learning algorithm. $(^*)$ At least one instance not solved within 1h.

Numerical Results - Running Times

15 smallest MIPLIB, 10 smallest QPLIB instances, w. confidence intervals.

Instance	$\varnothing$**Our Method**	$\varnothing$**Reformulation**
MIP, $\lvert\mathcal{S}\rvert = 10$	52.4s	115.8s
MIP, $\lvert\mathcal{S}\rvert = 50$	57.7s	176.7s*
MIQP, $\lvert\mathcal{S}\rvert = 2$	170.2s	359.6s*

Avg time per iteration for DRO via reformulation and with our online learning algorithm. $(^*)$ At least one instance not solved within 1h.
$\Rightarrow$ online algorithm much faster than reformulation!

Takeaways

Still many open questions in optimization under uncertainty

- nonlinear (infinite)-dimensional, integer decisions
- learn uncertainty and robust decisions efficiently
- multi-stage problems,...

Takeaways

Still many open questions in optimization under uncertainty

- nonlinear (infinite)-dimensional, integer decisions
- learn uncertainty and robust decisions efficiently
- multi-stage problems,...

Thank you for your attention.

Path length bounds

bounds $\sum_{t=1}^{T} \frac{1}{2}\|p^t - q^t\|^2 \leq h(T)$ for $p \in \mathcal{P}^{t-1}$ and $q^t \in \mathcal{P}^t$:

1. confidence Intervals:

$$h(T) = 8|\mathcal{S}| \log(\pi T)(2 + \log T)$$

2. kernel based ambiguity sets (A kernel matrix, λ smallest eigenvalue):

$$h(T) = \frac{1}{2}\left(2 + \frac{4}{\lambda}\right)^2 + \frac{32}{\lambda^2} \log \frac{\pi T}{\sqrt{6\delta}}(1 + \log T)$$

3. l_2 ambiguity sets ($A = I$):

$$h(T) = 8|\mathcal{S}| \log \frac{\pi T}{\sqrt{3\delta}}(2 + \log T)$$

$\hookrightarrow h(T) \leq \mathcal{O}(\log^2 T)$