Multipolar Robust Optimization

Walid Ben-Ameur (Telecom SudParis, Institut Polytechnique de Paris)
Adam Ouorou (Orange Labs), Guanglei Wang (Alibaba, China)
Mateusz Zotkiewicz (Warsaw University of Technology, Poland)

Agenda

1. Robust Optimization
2. Adjustable Robust Optimization
3. Multipolar Robust Optimization

Uncertainty

- Data can be random

- Data can be difficult to measure/observe
    - Heisenberg's uncertainty principle

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

Stochastic approaches

- Requires a good knowledge of the probability distribution of data

- Generally leads to intractable problems

- Chance-constrained programming variant:

  - Impose that some constraints are satisfied with some probability

  - Few cases are tractables: (e,g.)

$$a \sim \mathcal{N}(\bar{a}, \Sigma);\, \Sigma = L^t L$$

$a^t x \leq b$ holds with probability at least 1-ε

$$\overline{a}^t x + \phi^{-1}(1 - \varepsilon)\|Lx\|_2 \leq b$$

Robust optimization 1.

- Data u belong to an uncertainty set $\mathcal{H}$ (convex compact).

Robust counterpart

$$\min_{x} \max_{u \in \mathcal{H}} f(x,u)$$

$$g_i(x,u) \le 0, \forall u \in \mathcal{H}, \forall i \in I$$

  - Make the worst-case as good as possible
  - Several robustness measures:

$$f^*(u) = \min_{x} f(x,u) \text{ subject to } g_i(x,u) \le 0, \forall i \in I$$

Absolute Deviation

$$\min_{x} \max_{u \in \mathcal{H}} f(x,u) - f^*(u)$$

Relative deviation

$$\min_{x} \max_{u \in \mathcal{H}} \frac{f(x,u) - f^*(u)}{|f^*(u)|}$$

$$g_i(x,u) \le 0, \forall u \in \mathcal{H}, \forall i \in I$$

Robust optimization 2.

- We can assume that the objective function is deterministic (by adding constraints of type $f(x, u) \leq t$
- We can assume that uncertainty is constraint-wise (Ben-Tal, El-Ghaoui, Nemirovski, 2009)
  - Example: Robust Linear Programming

$$\min_{Ax \leq b, \forall A \in \mathcal{H}} c^t x$$

$$\Leftrightarrow$$

$$\min_{a_i{}^t x \leq b_i, \forall a_i \in \mathcal{H}_i \equiv Proj_i(\mathcal{H})} c^t x$$

Robust optimization 3.

History:

- Soyster, A.L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. Oper. Res. 21, 1154–1157.
- Mulvey, J.M., Vanderbei, R.J., Zenios, S.A., 1995. Robust optimization of large-scale systems. Oper. Res. Lett. 43 (2), 264–281.
- Kouvelis, P., Yu, G., 1997. Robust Discrete Optimization and its Applications. Kluwer Academic Publishers.
- El-Ghaoui, L. and Lebret, H. (1997). Robust solutions to least-square problems to uncertain data matrices. *SIAM Journal on Matrix Analysis and Applications*, 18:1035–1064.
- Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research*, 23(4):769–805.
- Fingerhut, J. A., Suri, S., and Turner, J. S. (1997). Designing least-cost nonblocking broadband networks. *Journal of Algorithms*, 24(2):287–309.

Robust optimization 4.

- How to solve ?:
  - By convex reformulation:

$$\begin{aligned}&Min\ c^t x\\ &{a_i}^t x \le b_i, \forall a_i \in \mathcal{H}_i, \forall i \in I\end{aligned}$$

$$\mathcal{H}_i : \{a_i, D_i\ a_i \le d_i\}$$

$$\begin{aligned}&Max\ {a_i}^t x\\ &D_i\ a_i \le d_i\end{aligned} \quad = \quad \begin{aligned}&Min\ \ {d_i}^t y_i\\ &{D_i}^t y_i = x, y_i \ge 0\end{aligned}$$

$${a_i}^t x \le b_i, \forall a_i \in \mathcal{H}_i \iff \begin{aligned}&Max\ {a_i}^t x \ \ \le b_i\\ &D_i\ a_i \le d_i\end{aligned} \iff \exists y_i \ge 0, {D_i}^t y_i = x, {d_i}^t y_i \le b_i$$

$$\begin{aligned}&Min\ c^t x\\ &{a_i}^t x \le b_i, \forall a_i \in \mathcal{H}_i\end{aligned} \iff \boxed{\begin{aligned}&Min\ c^t x\\ &{D_i}^t y_i = x, \forall i \in I\\ &{d_i}^t y_i \le b_i, \forall i \in I\\ &y_i \ge 0\ , \forall i \in I\end{aligned}}$$

Ben-Tal, A. and Nemirovski, A. (1998)

El-Ghaoui, L. and Lebret, H. (1997)

Robust optimization 5.

- How to solve ?:
  - By convex reformulation:

$$Min\ c^t x$$
$$a_i{}^t x \leq b_i, \forall a_i \in \mathcal{H}_i, \forall i \in I \qquad \mathcal{H}_i : \{a_i, \|a_i - \overline{a}_i\|_2 \leq \rho_i\}$$

$$\begin{matrix} Max\ a_i{}^t x \\ \|a_i - \overline{a}_i\|_2 \leq \rho_i \end{matrix} = \max_{\|u_i\| \leq 1} \overline{a}_i{}^t\ x + \rho_i u_i{}^t x = \overline{a}_i{}^t x + \rho_i \|x\|_2$$

$$\begin{matrix} Min\ c^t x \\ a_i{}^t x \leq b_i, \forall a_i \in \mathcal{H}_i \end{matrix} \Leftrightarrow \begin{matrix} Min\ c^t x \\ \overline{a}_i{}^t x + \rho_i \|x\|_2 \leq b_i, \forall i \in I \end{matrix}$$

Robust optimization 6.

- How to solve ?:
  - By convex reformulation:

$$\min_{x} \max_{u \in \mathcal{H}} f(x, u)$$
$$g_i(x, u) \leq 0, \forall u \in \mathcal{H}, \forall i \in I$$

- Roughly speaking, convex reformulation can be found mainly when functions f and g_i are convex in x and concave in u

- Exceptions: if g_i is quadratic (not necessarily concave) and $\mathcal{H}_i$ is an ellipsoid (a convex set defined by one convex quadratic constraint), then a convex reformulation exists (strong duality still holds here) (Ben-Tal, El-Ghaoui, Nemirovski, 2009)

Robust optimization 7.

- How to solve ?:
  - By constraint generation:

$$Min\ c^t x$$
$$a_i{}^t x \leq b_i, \forall a_i \in \mathcal{H}_i, \forall i \in I$$

- Let $\mathcal{H}'_i$ be a finite subset of $\mathcal{H}_i$
- Repeat
  - Solve $x' = argmin\ c^t x$
    $$a_i{}^t x \leq b_i, \forall a_i \in \mathcal{H}'_i, \forall i \in I$$
  - For each $i \in I$, $a_i{}' = argmax\ a_i{}^t x$
    $$\forall a_i \in \mathcal{H}_i$$
    - If $a'_i{}^t x > b_i$, add $a_i{}'$ to $\mathcal{H}'_i$,

Ben-Ameur and Kerivin, (2001,2003,2005)

Robust optimization 8.

- How to solve ?:

$$Min\ c^t x$$
$$a_i{}^t x \leq b_i, \forall a_i \in \mathcal{H}_i, \forall i \in I$$

  - By constraint generation:

    - More efficient than convexification for some uncertainty sets (for example a polytope having an exponential number of facets, while being well described by a separation oracle).
    - Polynomial-time if we use the ellipsoid method (equivalence of separation and optimization)

Robust optimization 9.

- How to solve ?:
  - By duality approach of Beck, Ben-Tal (2009):
    - Under the constraint-wise uncertainty condition, Primal worst equals dual Best !!

$$
\begin{array}{c} Max\ c^t x \\ {a_i}^t x \le b_i, \forall a_i \in \mathcal{H}_i, \forall i \in I \\ x \ge 0 \end{array}
\quad = \quad
\begin{array}{c} Min\ b^t y \\ \exists a_i \in \mathcal{H}_i : \sum_i y_i a_i \ge c \\ \mathrm{y} \ge 0 \end{array}
$$

Pessimistic primal

Optimistic dual

- Result holds for the general problem under convex-concave conditions and Slater's qualification

$$
\min_x \max_{u_0 \in \mathcal{H}_0} f\,(x, u_0)
$$

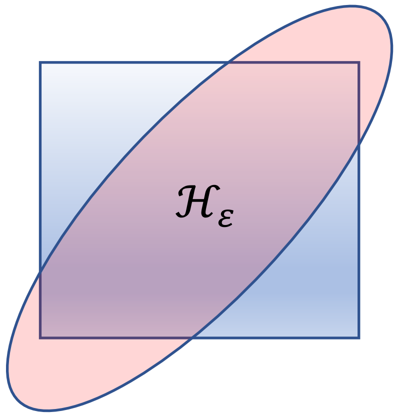
$$
g_i(x, u_i) \le 0, \forall u_i \in \mathcal{H}_i, \forall i \in I
$$

Robust optimization 10.

- How to choose the uncertainty set ?:

  - Using data and statistical tests….
    - Choose $\mathcal{H}_\varepsilon$ such that $a^t x \leq b, \forall a \in \mathcal{H}_\varepsilon$, implies that Pr($a^t x \leq b$)$\geq 1 - \varepsilon$
      - Example: assume that $a = \bar{a} + Pu, E(u) = 0, \|u\|_\infty \leq 1, u_1, u_2,$ … independent
        - $\mathcal{H}_\varepsilon = \left\{ \bar{a} + Pu, \|u\|_\infty \leq 1, \|u\|_2 \leq \sqrt{2 * ln\left(\frac{1}{\varepsilon}\right)} \right\}$ gives the wanted probability

          Ben-Tal, El-Ghaoui, Nemirovski, 2009
        - Etc.



  - Sampling:
    - Knowing the distribution of data, by choosing a largely enough number of samples, we can guarantee that the probability to have of a violation is less than some constant (see the precise statement in Calafiore and Campi, 2005; Campi and Gratti, 2008)

Robust optimization 11.

- Robust discrete optimization:

  - Generally difficult
    - Example : given two edge weight scenarios, compute an s-t path minimizing the maximum of the two total weights...

  - Some easy special cases

    Bertsimas&Sim, 2003

    - $S \subseteq \{0,1\}^n$, $\mathcal{H}_\Gamma = \{u: \sum_i u_i \leq \Gamma, 0 \leq u_i \leq 1\}$

$$\min_{x \in S} \max_{u \in \mathcal{H}_\Gamma} \sum_i (c_i + u_i d_i) x_i$$

can be solved by solving n+1 nominal problems with modified costs and can be approximated with the same ratio

    - Some generalizations (Knapsack constraints: Minoux 2009; Poss 2018)...

Adjustable robust optimization 1.

- Multistage optimization: here and now variables x and wait and see variables y (recourse variables) that can be adjusted once the scenario is revealed

$$A = A(u) \equiv P_0 + \sum_i P_i u_i\,, \qquad R \text{ recourse matrix}$$

$$\min c^t x$$
$$A(u)x + Ry(u) \leq b, \forall u \in \mathcal{H}$$

- Fully adjustable robust optimization is coNP-hard in general

  Gupta, Kleinberg, Kumar, Rastogi, Yuenber, 2001

  Ben-Tal, Goryashko, Guslitzer, Nemirovski, 2004

- Fully adjustable is easy to compute if the set of extreme points of $\mathcal{H}$ is polynomially bounded (each point of $\mathcal{H}$ is a convex combination of extreme points)

Adjustable robust optimization 2.

- Under constraint-wise uncertainty, static is as good as fully adjustable
  (Ben-Tal, Goryashko, Guslitzer, Nemirovski, 2004)

$$A(u) \equiv P_0 + \sum_i P_i u_i$$

- Fully adjustable robust optimization is more efficient than static robust optimization

Fully adjustable robust counterpart		Static robust counterpart
$\min c^t x$ $A(u)x + Ry(u) \leq b, \forall u \in \mathcal{H}$	$\leq$	$\min c^t x$ $A(u)x + Ry \leq b, \forall u \in \mathcal{H}$

Adjustable robust optimization 3.

- Affine adjustability: Assume that the wait and see variables affinely depend on uncertainty (Ben-Tal, Goryashko, Guslitzer, Nemirovski, 2004) (related to first-order decisions in multi-stage stochastic optimization, Garstka and Wets, 1974)

$$y(u) = Mu + l$$

$$\min_{x,M,l} c^t x$$

$$A(u)x + R(Mu + l) \leq b, \forall u \in \mathcal{H}$$

Affinely adjustable robust counterpart

$$A(u) \equiv P_0 + \sum_i P_i u_i$$

- Similar to a static robust problem where x, M, l are here and now variables (solved as before)

- If R is not fixed, for example, R is affine in u, the problem is generally intractable (exception: if $\mathcal{H}$ is an ellipsoid) (Ben-Tal, Goryashko, Guslitzer, Nemirovski, 2004)

Adjustable robust optimization 4.

- Polynomial decision rules: (Bertsimas, D., Iancu, D. A., and Parrilo, P. A., 2011)

$$y(u) = P(u), P \text{ is a polynomial}$$

  - The complexity of the robust counterpart problem (find the best monomials coefficients) is related to testing the positivity of a polynomial.

    - A positive polynomial can be expressed as a sum of squares (not a priori degree-bounded)…

      - The robust counterpart is approximated by sums of squares of degree no larger than a fixed constant (can be represented by a semidefinite programming, Lasserre, 2001)

Adjustable robust optimization 5.

- Fourier-Motzkin elimination: (Zhen, Hertog, Sim, 2018)
  - Eliminate some adjustable variables (but increase the number of constraints)…

$$\min c^t x$$
$$A(u)x + Ry(u) \leq b, \forall u \in \mathcal{H}$$

$$\sum_j A_{ij}(u)x_j + \sum_k R_{ik} y_k(u) \leq b_i$$

If $R_{ik_0} > 0, y_{k_0} \leq \frac{1}{R_{ik_0}}\left[b_i - \sum_j A_{ij}(u)x_j - \sum_{k\neq k_0} R_{ik} y_k(u)\right]$

If $R_{i'k_0} < 0, y_{k_0} \geq \frac{1}{R_{i'k_0}}\left[b_{i'} - \sum_j A_{i'j}(u)x_j - \sum_{k\neq k_0} R_{i'k} y_k(u)\right]$

$$\frac{1}{R_{i'k_0}}\left[b_{i'} - \sum_j A_{i'j}(u)x_j - \sum_{k\neq k_0} R_{i'k} y_k(u)\right] \leq \frac{1}{R_{ik_0}}\left[b_i - \sum_j A_{ij}(u)x_j - \sum_{k\neq k_0} R_{ik} y_k(u)\right]$$

Adjustable robust optimization 6.

- Duality approach of Bertsimas and De Ruiter 2016:
  - Transform an adjustable robust primal problem to an adjustable robust dual problem (with another uncertainty set)

$$A(u) \equiv P_0 + \sum_i P_i u_i$$

$$\mathcal{H}_{primal} = \{a: Da \leq d\}$$

$$\min c^t x$$
$$A(u)x + Ry(u) \leq b, \forall u \in \mathcal{H}_{primal}$$

$$\mathcal{H}_{dual} = \{\omega: R^t\omega \leq 0, 1^t\omega = 1, \omega \geq 0\}$$

$$\min c^t x$$
$$\omega^t P_i x + {D^t}_i \lambda(\omega) \geq 0, \forall i, \forall \omega \in \mathcal{H}_{dual}$$
$$d^t \lambda(\omega) + b^t \omega - \omega^t P_0 x \leq b, \forall \omega \in \mathcal{H}_{dual}$$

- Some similarities with a paper by Kuhn, Wieseman and Georghioi, 2011

Adjustable robust optimization 7.

- Sampling for multi-stage robust optimization: Vayanos, Kuhn and Rusten, 2012
  - Randomly select a finite subset of scenarios $\mathcal{H}' \subseteq \mathcal{H}$ with a sufficiently large cardinality (based on Calafiore and Campi, 2005; Campi and Gratti, 2008)
  - Choose a set of basis functions (a set of functions whose linear hull is dense in the set of continuous functions)
  - Solve the adjustable robust problem considering $\mathcal{H}'$ and optimizing the linear coefficients

$$\min c^t x$$
$$A(u)x + Ry(u) \leq b, \forall u \in \mathcal{H}$$

$$y_i(u) \approx \sum_k \beta_k{}^i \varphi_k(u)$$

$$\min_{x,\beta_k} c^t x$$
$$A(u^j)x + \sum_k \varphi_k(u^j) R\beta_k \leq b, \forall u^j \in \mathcal{H}'$$

Adjustable robust optimization 8.

- Multi-static/Finite adjustability: Partition the uncertainty set into several subsets and consider a fixed recourse y for each subset…

Ben-Ameur 2007; Ben-Ameur&Zotkiewicz 2009,2011

Bertsimas and Caramanis, 2010; Bertsimas and Dunning, 2016

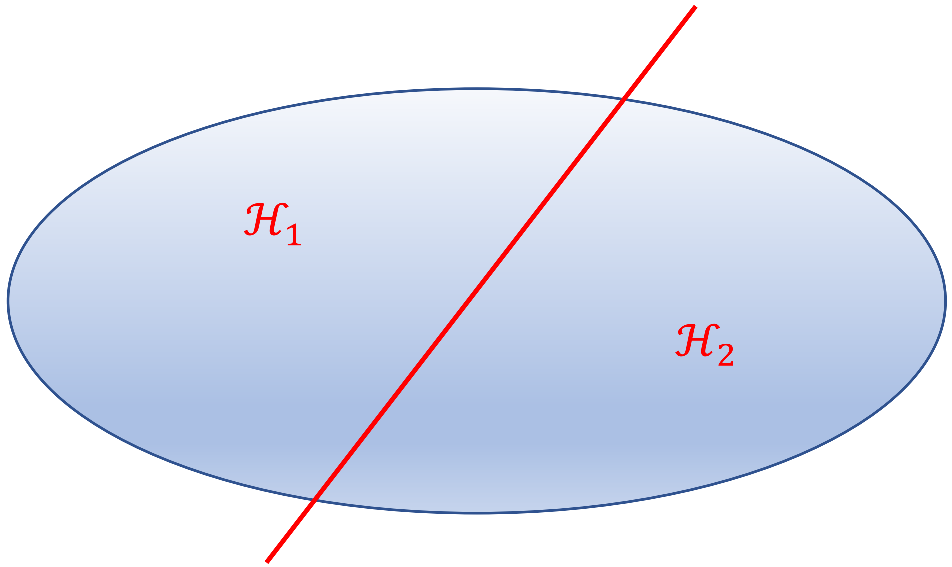
Postek and den Hertog, 2016

$$\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$$

$$\min c^t x$$

$$A(u)x + Ry_1 \leq b, \forall u \in \mathcal{H}_1$$

$$A(u)x + Ry_2 \leq b, \forall u \in \mathcal{H}_2$$

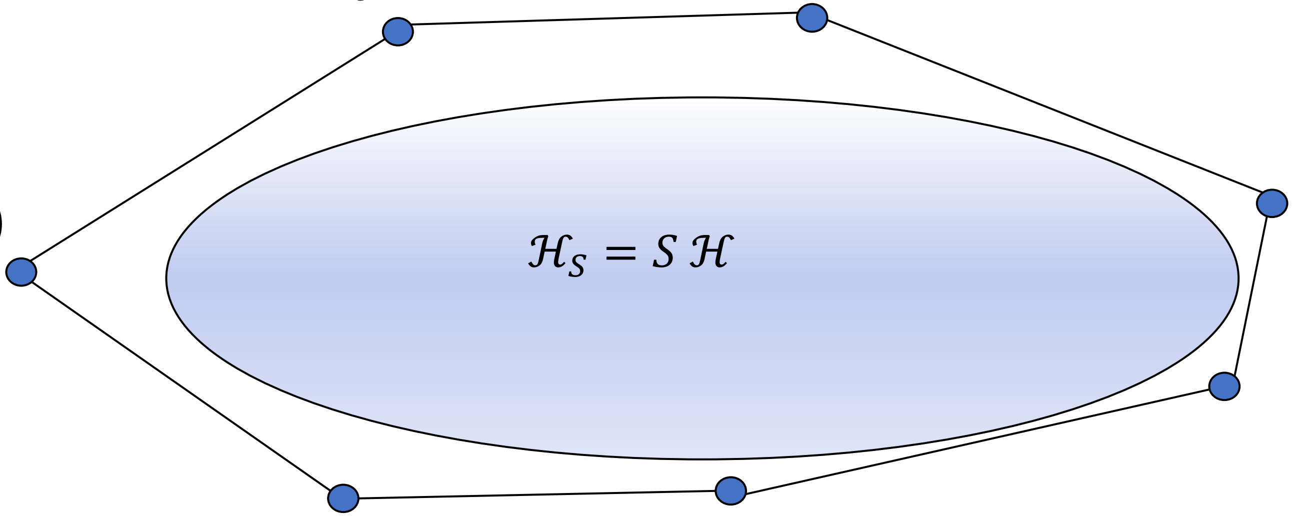


- We can also consider a more general decision rules for each subset (affine for example)

Multipolar robust optimization 1.

- Ingredients:
  - Shadow matrix S: either related to observations or just used to reduce complexity $\mathcal{H}_S = S\,\mathcal{H}$
  - Pole-set : Ω such that $\mathcal{H}_S \subseteq conv(\Omega)$

$$\min c^t x$$
$$A(u)x + Ry(u) \leq b, \forall u \in \mathcal{H}$$



Multipolar Robust Counterpart

$$\min c^t x$$
$$A(u)x + R \sum_{\omega \in \Omega} \lambda_\omega(u) y_\omega \leq b,$$
$$\forall u \in \mathcal{H}, \forall\, \lambda_\omega(u) \geq 0, \sum_{\omega \in \Omega} \lambda_\omega(u) = 1, Su = \sum_{\omega \in \Omega} \lambda_\omega(u)\,\omega$$

Multipolar robust optimization 2.

- Static is a special case of multipolar: S null matrix and Ω contains only the null vector as a pole

- Fully-adjustable is a special case of multipolar: take $S = I$ and Ω be the set of extreme points of $\mathcal{H}$

Theorem: Affine is a special case of multipolar: $S = I$ and conv(Ω) is any simplex containing $\mathcal{H}$

Corollary: If the uncertainty set is a simplex, then the affine adjustable approach is equivalent to the fully adjustable robust approach

(generalizing a result of Bertsimas&Goyal, 2012 where only right-hand-side uncertainty is considered)

Multipolar robust optimization 3.

- **Tractability:**

$$\min c^t x$$

$$A(u)x + R\sum_{\omega\in\Omega}\lambda_\omega(u)y_\omega \le b, \forall u \in \mathcal{H}, \forall\, \lambda_\omega(u) \ge 0, \sum_{\omega\in\Omega}\lambda_\omega(u) = 1, Su = \sum_{\omega\in\Omega}\lambda_\omega(u)\,\omega$$

Constraint generation

Convex reformulation using strong duality

$$\max A_i(u)x + R_i \sum_{\omega\in\Omega}\lambda_\omega(u)y_\omega - b_i$$

$$u \in \mathcal{H}$$

$$\lambda_\omega(u) \ge 0,$$

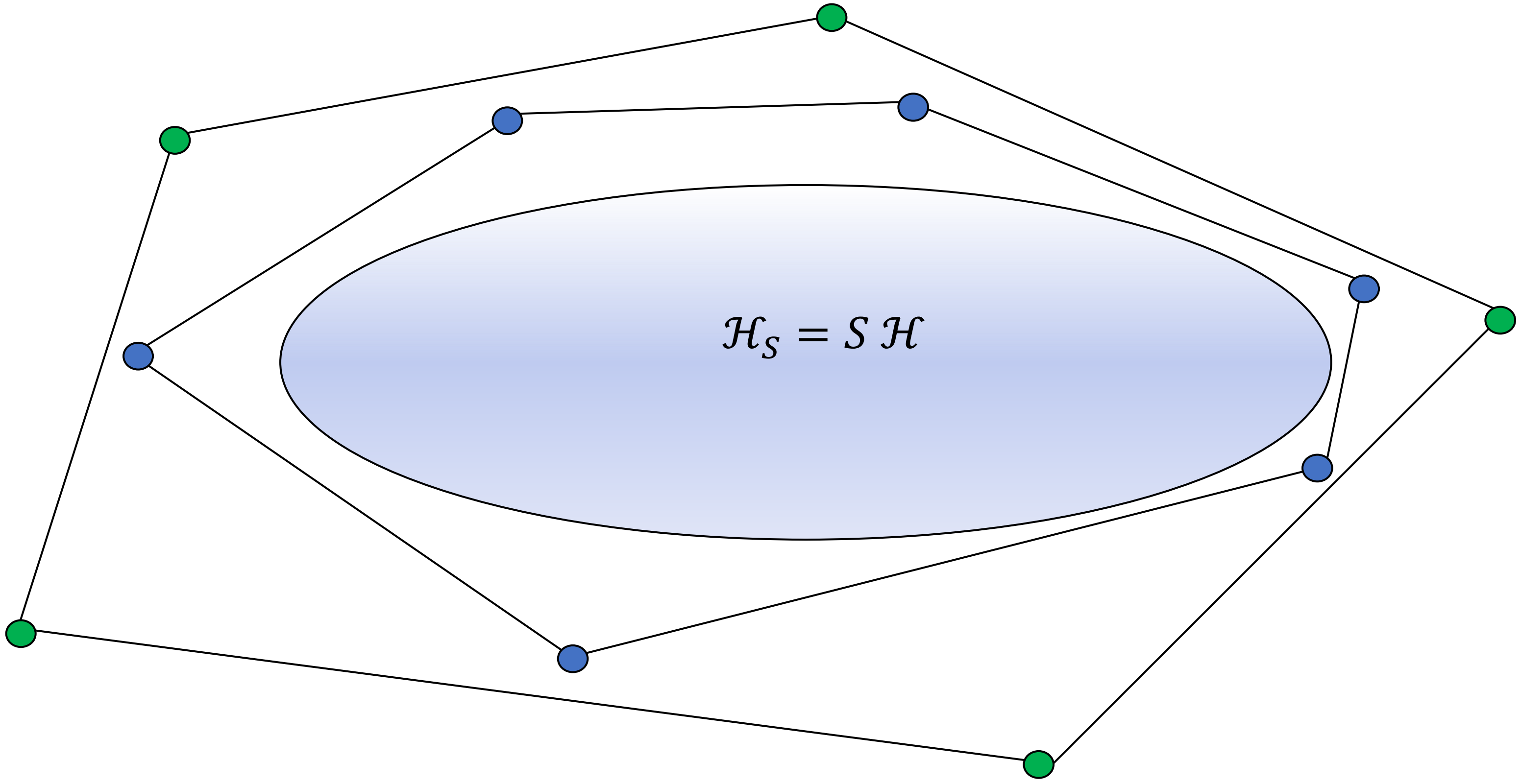
$$\sum_{\omega\in\Omega}\lambda_\omega(u) = 1,$$

$$Su = \sum_{\omega\in\Omega}\lambda_\omega(u)\,\omega$$

Multipolar robust optimization 4.

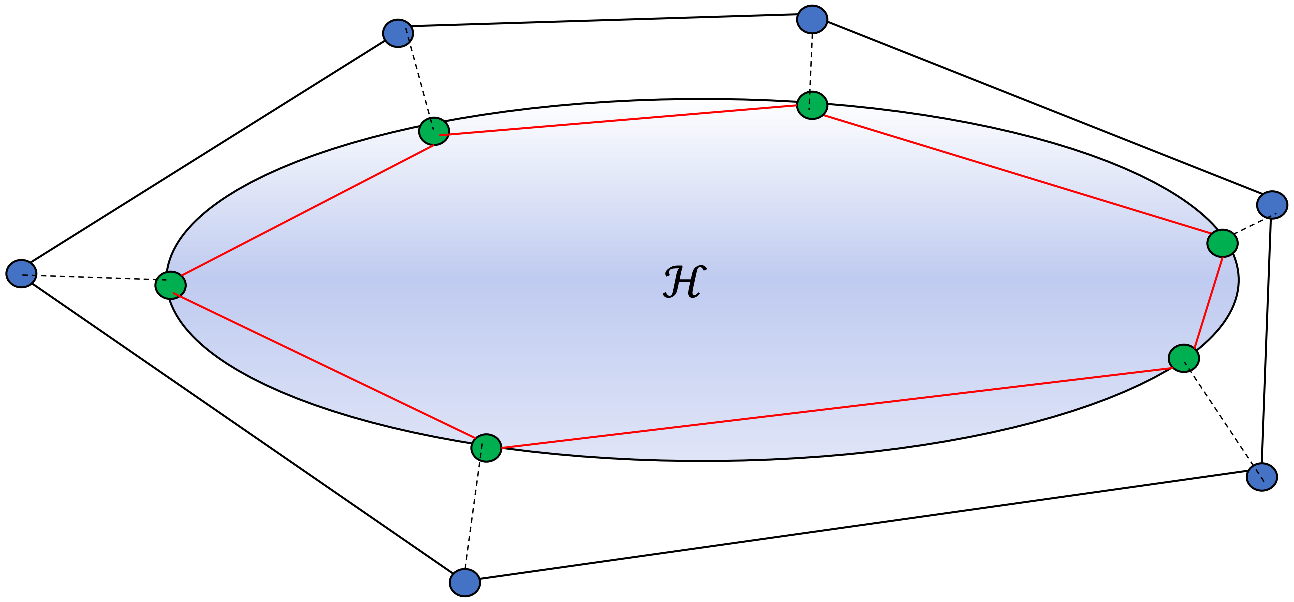
- Monotonicity: If $\mathcal{H}_S \subseteq conv(\Omega) \subseteq conv(\Omega')$ then

$$MRC(\Omega) \leq MRC(\Omega')$$



Multipolar robust optimization 5.

- Convergence: if $S = I$, under mild assumptions, the solution converges to an optimal fully-adjustable solution



Multipolar robust optimization 6.

- Pole-set construction:

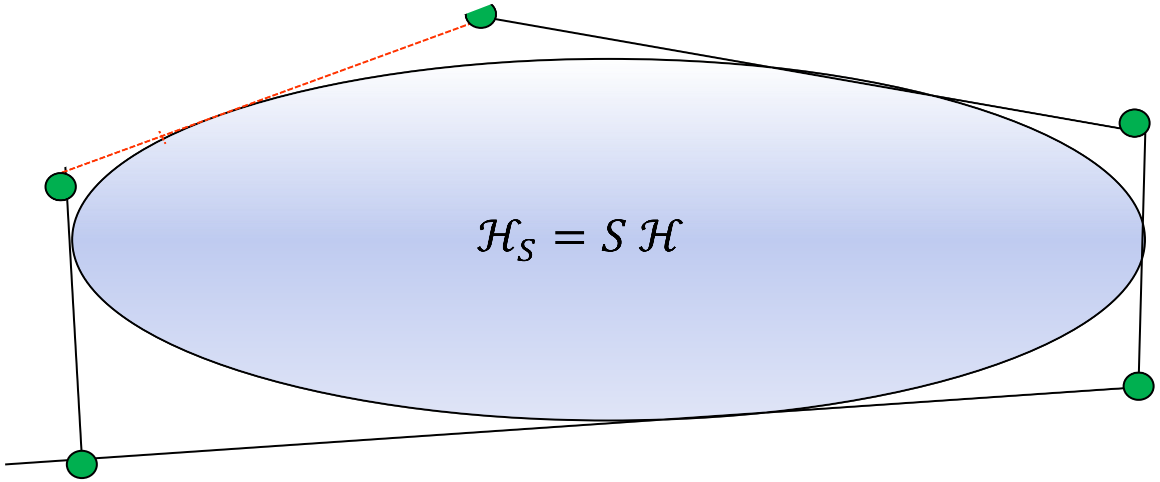
  - Generate a set Ω' of $1 + \dim(\mathcal{H}_S)$ affinely independent points. Compute σ and h such that $\sigma \times conv(\Omega') + h \supset \mathcal{H}_S$

Theorem: the smallest value of σ and the corresponding translate vector h can be computed efficiently for any convex set $\mathcal{H}_S$

(Generalizes a result of Nevskii 2011 established for hypercubes)

Multipolar robust optimization 7.

- Pole-set construction
  - The pole-set is updated by a tightening procedure



Multipolar robust optimization 8.

- Numerical example: lobbying problem

$$\min u$$

$$\sum_{i=1}^{m} r_i y_i(t) \leq u, \forall t \in D$$

$$\sum_{j=1}^{n} Q_{ij} t_j + y_i(t) \geq 0, \forall i = 1,\ldots,m$$

$$y_i(t) \geq 0, \forall i = 1,\ldots,m$$

$$Q_{ij} \in [-1,1]$$

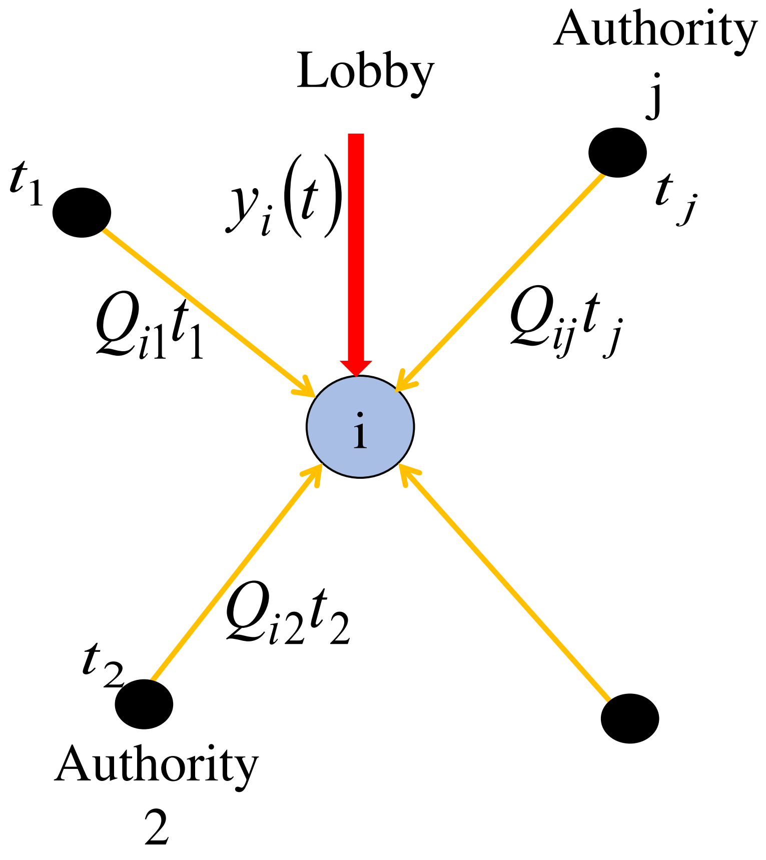


Table 1 The pole-sets of hypercubes

Hypercube	$\|\Omega_0\|$	$\|\Omega_1\|$	$\|\Omega_2\|$	$\|\Omega_3\|$	#ext.
H_9	10	32	162	387	512
H_{10}	11	36	112	322	1,024
H_{12}	13	44	144	449	4,046

Table 3 The multipolar robust values with different pole-sets (hypercube uncertainty sets)

(m,n)	static	affine/$\Pi_{H_n}(\Omega_0)$	$\Pi_{H_n}(\Omega_1)$	$\Pi_{H_n}(\Omega_2)$	$\Pi_{H_n}(\Omega_3)$	$\Pi^*_{H_n}$
(10,9)	24.84	12.42	12.18(9.45)	10.55(73.62)	10.16(88.98)	9.88
(10,10)	25.50	12.75	11.53(58.65)	10.96(86.06)	10.70(98.56)	10.67
(10,12)	30.66	15.33	14.63(35.18)	13.71(81.41)	13.43(95.48)	13.34
(20,9)	50.75	25.37	23.82(46.69)	22.08(99.10)	22.06(99.70)	22.05
(20,10)	50.88	25.44	23.56(16.95)	20.74(42.38)	18.58(61.86)	14.35
(20,12)	59.79	29.89	27.54(23.64)	25.40(45.17)	23.58(63.48)	19.95

Table 5 Lower bounds related to hypercubes

(m,n)	$\Pi_{H_n}(\Gamma_0)$	$\Pi_{H_n}(\Gamma_1)$	$\Pi_{H_n}(\Gamma_2)$	$\Pi_{H_n}(\Gamma_3)$	$\Pi^*_{H_n}$
(10,9)	6.88	8.18	9.52	**9.65**	9.88
(10,10)	7.11	8.12	**9.62**	8.34	10.67
(10,12)	**10.50**	8.88	9.48	9.57	13.34
(20,9)	20.08	16.05	18.70	**21.98**	22.05
(20,10)	11.88	11.88	12.44	**12.92**	14.35
(20,12)	14.73	16.65	16.96	**19.07**	19.95

Conclusion

Multipolar robust optimization

- Encompasses previous main approaches (Static, Affine, Fully adjustable)
- Makes the link between discrete approaches (through pole recourses) and continuous approaches (through convex combinations)
- Allows some control of complexity
- Works even with partial information (through shadow matrix)