

## Practical Robust Optimization - an introduction -

Dick den Hertog

Tilburg University, Tilburg, The Netherlands

NGB/LNMB Seminar
"Back to school, learn about the latest developments in Operations Research"

January 17, 2013



- Motivation for Robust Optimization (RO)
  - Flaw of using nominal values
  - Why RO can make a difference
- Robust Optimization Methodology
- Deriving tractable Robust Counterparts
- Practical issues
- Practical example
- Adjustable Robust Optimization
- Concluding remarks



## CentER Google results ...

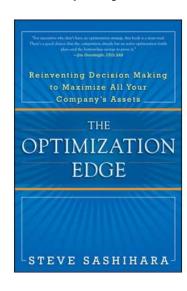
- "Robust Optimization" 186,000 hits
- "... and logistics" 30,000 hits
- " ... and supply chain" − 34,000 hits
- "... and production" 67,000 hits
- " ... and energy" 74,000 hits
- " ... and finance" 35,000 hits
- "... and engineering" 116,000 hits





### What sets great optimization apart?

"The Optimization Edge" (2011) by Steve Shashihara



One of the eight differentiators for "great optimization" mentioned:

Good optimization gives you the best choice based on the data. Great optimization is "robust" and resilient in the face of change and data errors.





### Uncertainty in optimization problems

Optimization problems often contain uncertain parameters, due to

- measurement/rounding errorse.g. temperature, current inventory
- estimation errorse.g. demand, cost or prices
- implementation errorse.g. length, depth, width, voltage

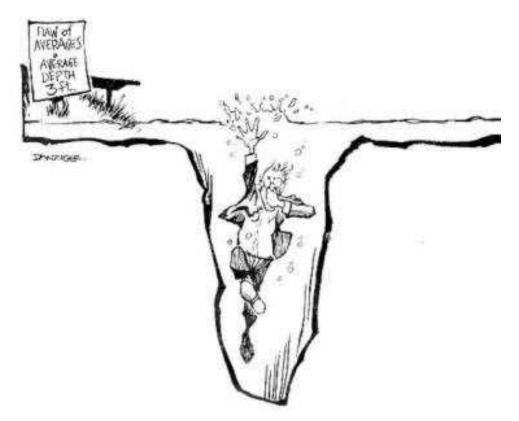
Flaw of using nominal values in optimization problems.

RO: find a solution that is robust against this uncertainty in the parameters.





Optimization based on nominal values often lead to (severe) infeasibilities.



Taken from "Flaw of averages" (2009,2012), Sam Savage.





Optimization based on nominal values often lead to (severe) infeasibilities.

Consider the constraint:  $a^T x \leq b$ , where

- lacksquare a is uncertain,  $\bar{a}$  is nominal value.
- $\blacksquare$   $a = \bar{a} + \rho \zeta$ , and  $-1 \le \zeta_i \le 1$ .
- lacksquare  $\zeta$  is uniformly distributed.

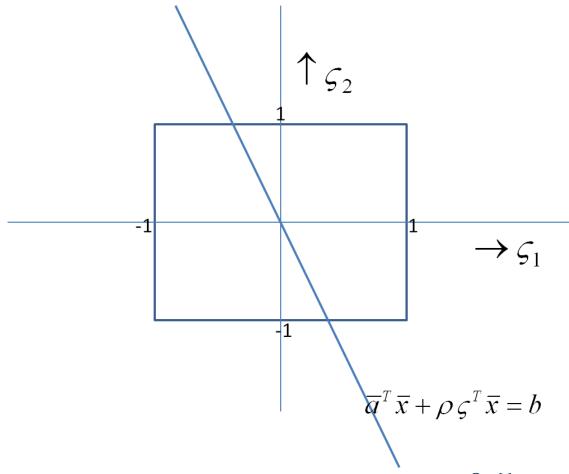
#### Assume:

this constraint is binding in the optimal nominal solution  $\bar{x}$  (for  $a=\bar{a}$ ).





Constraint: 
$$a^Tx = (\bar{a} + \rho\zeta)^Tx = \bar{a}^Tx + \rho\zeta^Tx \leq b$$
.



Probability of infeasibility = 0.5!





Optimization based on nominal values often lead to (severe) infeasibilities.

Consider the binding constraints:  $(a^k)^T x \leq b_k$ , k = 1, ..., N.

- $\blacksquare$   $a^k$  is uncertain,  $\bar{a}^k$  is nominal value.
- $a^k = \bar{a}^k + \rho^k \zeta^k$ , and  $-1 \le \zeta_i^k \le 1$ .
- lacksquare  $\zeta^k$  are independent and uniformly distributed.

Assume: these N constraints are binding in the optimal solution (for  $a=\bar{a}$ ).

Probability of infeasibility  $=1-\left(\frac{1}{2}\right)^N$ .



Optimization based on nominal values often lead to (severe) infeasibilities.

Consider the constraints:  $3.000x_1 - 2.999x_2 + ... \le 1$ .

#### Suppose:

- lacksquare  $x_1 = x_2 = 1000$  is optimal
- the number 3.000 is uncertain
- $\blacksquare$  maximal deviation is 1%.

Then LHS of constraint can be

$$3.030 * 1,000 - 2.999 * 1,000 + \dots = 31 > 1$$





Consider PILOT4 from NETLIB (1,000 variables, 410 constraints).

#### Constraint # 372:

```
\bar{a}^T x = -15.79081 x_{826} - 8.598819 x_{827} - 1.88789 x_{828} - 1.362417 x_{829} - 1.526049 x_{830} -0.031883 x_{849} - 28.725555 x_{850} - 10.792065 x_{851} - 0.19004 x_{852} - 2.757176 x_{853} -12.290832 x_{854} + 717.562256 x_{855} - 0.057865 x_{856} - 3.785417 x_{857} - 78.30661 x_{858} -122.163055 x_{859} - 6.46609 x_{860} - 0.48371 x_{861} - 0.615264 x_{862} - 1.353783 x_{863} -84.644257 x_{864} - 122.459045 x_{865} - 43.15593 x_{866} - 1.712592 x_{870} - 0.401597 x_{871} +x_{880} - 0.96049 x_{898} - 0.946049 x_{916} > b = 23.387405
```

Most coefficients are "ugly reals" (like - 15.79081).

Highly unlikely that coefficients are known to this accuracy.

Only exception: the coefficient 1 at  $x_{880}$  - it perhaps reflects the structure of the problem and might be exact.





Suppose: accuracy is 0.1%:

$$|a_i^{true} - \bar{a}_i| \le 0.001 |\bar{a}_i|.$$

Worst case: the constraint can be violated by as much as 450%:

$$\min_{a^{true} \text{ satisfies f(x)}} (a^{true})^T \bar{x} - b = -128.8 \approx -4.5b.$$





#### Assume "random uncertainty":

$$a_i^{true} = \bar{a}_i + \epsilon_i |\bar{a}_i|, \qquad \epsilon_i \sim \mathsf{Uniform}[-0.001, 0.001]$$

#### Based on 1,000 simulations:

$$V = \max \left[ \frac{b - (a^{true})^T \bar{x}}{|b|}, 0 \right].$$

> The nominal solution is highly "unreliable"





#### Among 90 NETLIB LP problems:

- In 19 problems
  - 0.01% perturbation  $\longrightarrow$  more than 5%-violations of (some of) the constraints
- In 13 of these 19 problems
  0.01%-perturbation → more than 50% violations of the constraints.
- In 6 of these 13 problems

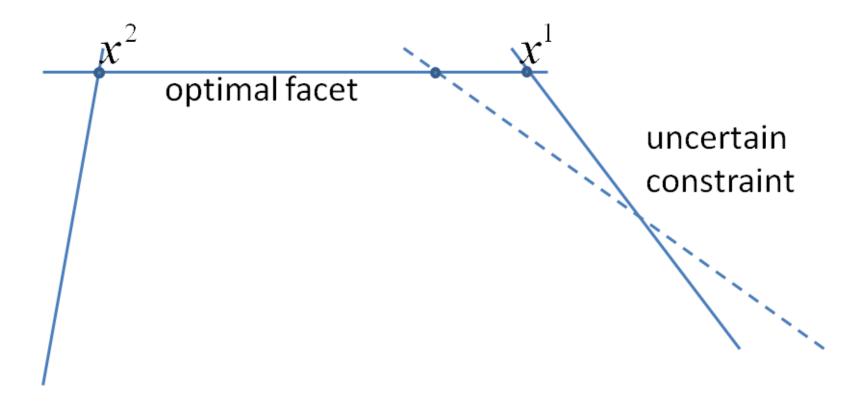
  constraint violaton was over 100%; in one problem even 210,000%.





### Why RO can make a difference ...

#### Extreme situation:



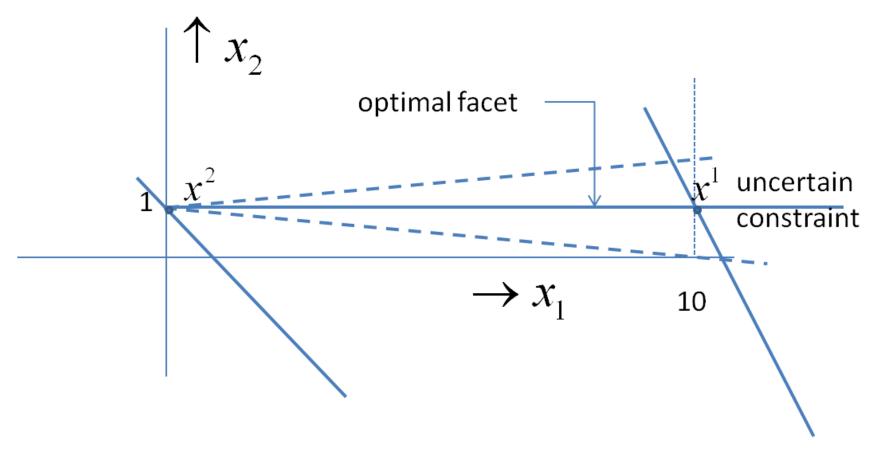
Take solution  $x^2$  on the optimal facet, which is much more robust!





### Why RO can make a difference ...

Extreme situation:  $x_2 - \zeta x_1 = 1$ , where  $-0.1 \le \zeta \le 0.1$ .



Take  $x^2$  on optimal facet, which is (much) more robust!

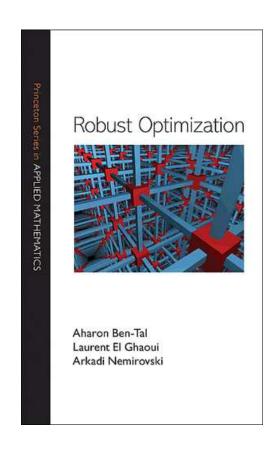


### History of Robust Optimization

- Early work by Soyster (1973) and Kouvelis (1997).
- Started with papers in 1997-1998: Ben-Tal, El Ghaoui, Nemirovski.
- Since 2000 many papers on RO.
- Budget uncertainty set (Bertsimas and Sim, 2004).
- Adjustable Robust Optimization (Ben-Tal et al. 2004).
- Book: Robust Optimization (2009) by Ben-Tal, El Ghaoui, and Nemirovski.
- Practical relevance: many applications!







Important chapters: Preface, Chapter 1 and Chapter 14.





### Relation with sensitivity analysis (SA)

- SA is post analysis ("post mortem").
- SA only analyzes infinitesimal changes (shadowprices, reduced costs).
- SA only analyzes changes in one or a few parameters.

#### **Robust Optimization:**

- takes data uncertainty into account already at the modelling stage;
- to "immunize" solutions against uncertainty.





### Relation with Stochastic Programming

"Large" data uncertainty is modelled in a stochastic fashion and then processed via Stochastic Programming techniques.

#### Disadvantages:

- Often difficult to specify reliably the distribution of uncertain data.
- Expectations/probabilities often difficult to calculate.
- Feasible region often nonconvex.

#### **Robust Optimization:**

- does not assume stochastic nature of the uncertain data;
- remains computationally tractable.





### Basic assumptions in RO

- The optimization variables represent "here and now" decisions.
- 2. The decision maker is fully responsible for decisions made (only) when the actual data is within the uncertainty set.
- 3. The constraints of the uncertain problem in question are "hard".

#### Later on:

- [1] will be alleviated by "adjustable robust optimization";
- [3] will be alleviated by "globalized robust optimization".





### Introduction to robust linear optimization

#### Robust counterpart:

$$(RC) \qquad \max\{c^T x : Ax \le b, \, \forall A \in U\},\$$

where  $x \in \mathbb{R}^n$ , and U is a given uncertainty region.

We may assume w.l.o.g.:

- objective is certain
- RHS is certain
- lacksquare U is closed and convex
- constraintwise uncertainty.





Hence, we focus on

$$(a+B\zeta)^T x \le \beta, \quad \forall \zeta \in Z,$$

#### where

- $oldsymbol{\zeta} \in \mathbb{R}^L$  is the primitive uncertain vector
- $\blacksquare$   $B \in \mathbb{R}^{n \times L}$
- lacksquare Z is the uncertainty region; often  $|L| \ll n$ .

Example: factor model in portfolio problems.

This is a semi-infinite optimization problem.

Goal: obtain computationally tractable reformulation!





### Tractable robust counterparts

$$(a+B\zeta)^T x \le \beta, \quad \forall \zeta \in Z.$$

Uncertainty region		Robust Counterpart	Tractability
Box	$     \zeta  _{\infty} \le \rho$	$  a^T x + \rho    B^T x   _1 \le \beta$	LP
Ball/ellipsoidal	$     \zeta  _2 \le \rho$	$  a^T x + \rho   B^T x  _2 \le \beta$	CQP
Polyhedral	$D\zeta + d \ge 0$	$\begin{cases} a^T x + d^T y \le \beta \\ D^T y = -B^T x \\ y \ge 0 \end{cases}$	LP
Cone (closed, convex, pointed)	$D\zeta + d \in K$	$\begin{cases} a^T x + d^T y \le \beta \\ D^T y = -B^T x \\ y \in K^* \end{cases}$	Conic Opt.





### RC for box uncertainty

$$(a + B\zeta)^T x \le \beta \qquad \forall \zeta : \|\zeta\|_{\infty} \le 1. \tag{1}$$

This is equivalent to:

$$\max_{\zeta:\|\zeta\|_{\infty} \le 1} (a + B\zeta)^T x = a^T x + \max_{\zeta:\|\zeta\|_{\infty} \le 1} (B^T x)^T \zeta \le \beta.$$

We have:

$$\max_{\zeta:\|\zeta\|_{\infty} \le 1} (B^T x)^T \zeta = \max_{\zeta:|\zeta_i| \le 1} \sum_i (B^T x)_i \zeta_i = \sum_i |(B^T x)_i| = \|B^T x\|_1.$$

Hence, (1) is equivalent to:

$$a^T x + \|B^T x\|_1 \le \beta.$$





### RC for ball uncertainty

$$(a+B\zeta)^T x \le \beta \qquad \forall \zeta : \|\zeta\|_2 \le 1. \tag{2}$$

This is equivalent to:

$$a^T x + \max_{\zeta: \|\zeta\|_2 \le 1} (B^T x)^T \zeta \le \beta,$$

Intermezzo:  $\max_{\zeta} \{d^{T}\zeta : \|\zeta\|_{2} \leq 1\}$  or  $\max \{d^{T}\zeta : \zeta^{T}\zeta \leq 1\}$ 

Lagrange:  $d = \lambda \zeta$ , and  $\lambda = ||d||_2$ .

Optimal objective value:  $\frac{d^Td}{\lambda} = \|d\|_2$ .

Hence (2) is equivalent to:

$$a^T x + \|B^T x\|_2 \le \beta.$$





### RC for polyhedral uncertainty

$$(a+B\zeta)^T x \le \beta \qquad \forall \zeta : D\zeta + d \ge 0.$$

This is equivalent to:

$$a^T x + \max_{\zeta : D\zeta + d \ge 0} (B^T x)^T \zeta \le \beta. \tag{3}$$

Note that by duality

$$\max\{(B^T x)^T \zeta : D\zeta + d \ge 0\} = \min\{d^T y : D^T y = -B^T x, \ y \ge 0\}.$$

Hence (3) is equivalent to

$$a^T x + \min_{y} \{ d^T y : D^T y = -B^T x, \ y \ge 0 \} \le \beta,$$

or

$$a^T x + d^T y \le \beta$$
,  $D^T y = -B^T x$ ,  $y \ge 0$ .



### Budget uncertainty region

Often used uncertainty region (Bertsimas and Sim, 2004):

$$\{\zeta : \|\zeta\|_{\infty} \le 1, \text{ number of elements } \ne 0 \text{ at most } \Gamma\}$$

which is equivalent to (only for linear uncertain constraints):

$$\{\zeta : \|\zeta\|_{\infty} \le 1, \|\zeta\|_{1} \le \Gamma\}.$$

This is a polyhedral uncertainty set.

Hence, (RC) can be reformulated as LP.





Accuracy in uncertain data is 0.1%, in the worst case.

#### **Robust Counterpart:**

```
-15.79081x_{826} - 8.598819x_{827} - 1.88789x_{828} - 1.362417x_{829} - 1.526049x_{830} \\ -0.031883x_{849} - 28.725555x_{850} - 10.792065x_{851} - 0.19004x_{852} - 2.757176x_{853} \\ -12.290832x_{854} + 717.562256x_{855} - 0.057865x_{856} - 3.785417x_{857} - 78.30661x_{858} \\ -122.163055x_{859} - 6.46609x_{860} - 0.48371x_{861} - 0.615264x_{862} - 1.353783x_{863} \\ -84.644257x_{864} - 122.459045x_{865} - 43.15593x_{866} - 1.712592x_{870} - 0.401597x_{871} \\ +x_{880} - 0.96049x_{898} - 0.946049x_{916} \\ -0.001\sum_{i}|\bar{a}_{i}x_{i}| \ge b = 23.387405
```

#### Robust solution:

- does not violate the constraints
- increase in objective value is less than 1%!

Remember: nominal solution yields a 450% violation!





### Planning of Air Traffic Controllers

#### Master thesis by Dori van Hulst (Quintiq)



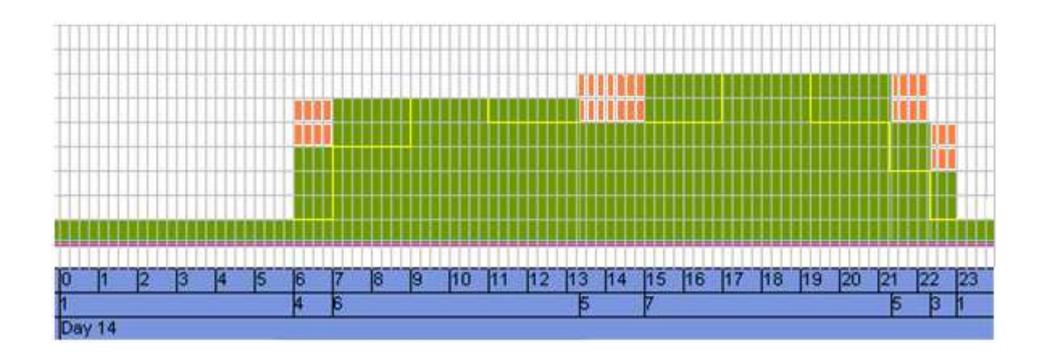
- Planning is done 2-3 months ahead.
- Different "shift types" can be used.
- Under capacity is dangerous, hence costly.
- Nominal prolbem is a MIP model.
- Uncertainty in demand = # planes in air corridors.





### Planning of Air Traffic Controllers

#### Uncertainty in demand per day:



This is polyhedral uncertainty!





### Planning of Air Traffic Controllers

#### Results:

	Costs Under =4.2		Costs Under =1.9	
	Nominal	Robust	Nominal	Robust
Labour costs nominal curve	433	762	426	675
Labour costs worst case	2054	882	1179	795
Average labour costs	1293	759	830	694
Average std per day of av. costs	13.2	2.4	6.2	2.2
Average under-utilization (hours)	67	6	67	14
Percentage of times better	14%	86%	34%	66%





### Why ellipsoidal uncertainty?

- Resulting Robust Counterpart is tractable (namely CQP).
- 2. Much less conservative than box uncertainty.
- Arises naturally from statistical considerations:
   Confidence sets in regression
   Confidence set for covariance

. . .

4. Is a safe approximation for chance constraint.

1, 2 and 4 also hold for budget uncertainty!

Tilburg • University



### Safe approximation of chance constraint

Suppose:  $\zeta_i$  are stochastic and independent, with known support, say [-1,1], and zero mean.

Consider:  $U_{\Omega} = \{ \zeta : ||\zeta||_2 \leq \Omega \}$  (ball)

$$(RC) \quad (a+B\zeta)^T x \le \beta, \ \forall \zeta \in U_{\Omega}. \tag{4}$$

If x satisfies (4) then:

Prob 
$$((a + B\zeta)^T x \le \beta) \ge 1 - \exp(-\Omega^2/2) =: 1 - \epsilon.$$

Example:  $\Omega = 7.44 \Longrightarrow \epsilon = 10^{-12}$ .

Same result for  $U_{\Omega}=\{\zeta: \|\zeta\|_2\leq \Omega, \|\zeta\|_{\infty}\leq 1\}$  (ball-box).





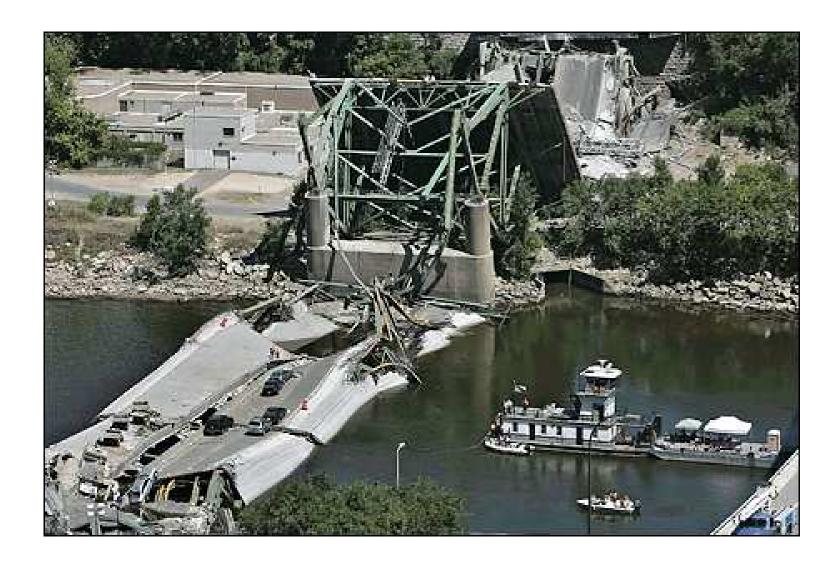
### Is Robust Optimization too pessimistic?

- In many cases the constraint is strict, e.g. safety restrictions.
- User can adapt level of protection.
- Overly pessimistic solutions may be caused by modeling errors.
- Use Globalized Robust Counterparts.
- Robust solutions often perform well on the average.
- Chance constraints can be approximated.





# CentER Bridge that was not robust ...





# CentER Recipe for RO

- Step 0.
  - Check whether nominal solution is robust.
  - Check whether SP can solve your problem.
- Step 1.
  - Determine uncertain parameters.
  - Determine uncertainty region.
- Step 2.
  - Derive tractable robust counterpart (exact or approxim.).
- Step 3.
  - Solve tractable robust counterpart.
- Step 4.
  - Check robustness of robust solution.





#### Some of the decisions $x_i$ :

- should be made when actual data becomes partially known
- can depend on the corresponding portions of the data.

#### **Examples:**

- Inventory problem with uncertain demand. e.g. replenishment orders  $x_t$  of day t usually can depend on actual demands at days 1, ..., t-1.
- Brachytherapy with inaccuracy in positioning needles. e.g. position of needle k can depend on actual position of needles 1, ..., k-1.





## Multistage problems

Consider the following constraint of a multi-stage problem:

$$(ARC)$$
  $(a+B\zeta)^T x + d^T y \le \beta, \quad \forall \zeta \in Z,$ 

#### where

- x is non-adjustable
- y is adjustable (fixed recourse).

ARC stands for Adjustable Robust Counterpart.

Hence:  $y = y(\zeta)$ .

This leads to an NP-hard problem!





### Linear Decision Rule

$$(ARC) \quad (a+B\zeta)^T x + d^T y \le \beta, \quad \forall \zeta \in Z,$$

Linear decision rule  $y = u + V\zeta$  is often very effective:

$$(AARC) \quad (a+B\zeta)^T x + d^T (u+V\zeta) \le \beta, \quad \forall \zeta \in Z,$$

or equivalently

$$(AARC) \quad a^T x + d^T u + (B^T x + V^T d)^T \zeta \le \beta, \quad \forall \zeta \in Z.$$

This problem is linear in x and  $\zeta$ ; hence previous results apply! AARC stands for Affinely Adjustable Robust Counterpart.





## Example: inventory problems

$$\begin{cases} \min \sum_{t=1}^{T} c_t x_t + \sum_{t=1}^{T} h_t I_t \\ I_t = I_{t-1} + x_t - d_t, \quad \forall t \\ I_t \ge 0, \quad \forall t \\ 0 \le x_t \le U_t, \quad \forall t \\ \dots \end{cases}$$

Demand  $d_t$  is uncertain: we assume interval uncertainty.  $x_t$  and  $I_t$  are adjustable: linear decision rules

$$x_t = u_t + \sum_{i=1}^{t-1} v_{it} d_i, \quad I_t = w_t + \sum_{i=1}^{t-1} z_{it} d_i$$

One can vary the "information base"!



#### AARC shows often very good behaviour (Ben-Tal et al., 2004):

Unc.ty (%)	Opt(ARC)	Opt(AARC)	Opt(RC)
10	13531.8	13531.8 (+0.0%)	15033.4 (+11.1%)
20	15063.5	15063.5 (+0.0%)	18066.7 (+19.9%)
30	16595.3	16595.3 (+0.0%)	21100.0 (+27.1%)
40	18127.0	18127.0 (+0.0%)	24300.0 (+34.1%)
50	19658.7	19658.7 (+0.0%)	27500.0 (+39.9%)
60	21190.5	21190.5 (+0.0%)	30700.0 (+44.9%)
70	22722.2	22722.2 (+0.0%)	33960.0 (+49.5%)





## Globalized robust counterpart

$$(a+B\zeta)^T x \leq \beta + \alpha \operatorname{dist}(\zeta, Z_1) \quad \forall \zeta \in Z_2,$$

where  $Z_1 \subset Z_2$ .

#### Example:

- $Z_1 = \{\zeta : \|\zeta\|_2 \le \rho_1\}, \ Z_2 = \{\zeta : \|\zeta\|_\infty \le \rho_2\}, \rho_1 \le \rho_2.$
- $extbf{dist}(\zeta, Z_1) = \min_{\zeta' \in Z_1} \|\zeta \zeta'\|_2.$

#### Globalized Robust Counterpart (GRC):

$$\begin{cases} a^T x + \rho_2 ||v||_1 + \rho_1 ||B^T x - v||_2 \le \beta \\ ||B^T x - v||_2 \le \alpha \end{cases}$$

This is a system of conic quadratic constraints.





#### Advantage of Globalized Robust Counterpart:

- Relaxes assumption [3]: "Constraints of uncertain problem are hard".
- Smooth behaviour.
- Often better 'average behaviour'.
- GRC is still tractable for many choices of  $Z_1$ ,  $Z_2$ , and distance function.





RCs of two equivalent problems may not be equivalent.

#### **Example 1:**

$$(RP)$$
  $(2+\zeta)x_1 \le 1 \quad \forall \zeta : |\zeta| \le 1$ 

$$(\overline{RP}) \quad \begin{cases} (2+\zeta)x_1 + s = 1 & \forall \zeta : |\zeta| \le 1 \\ s \ge 0 \end{cases}$$

Nominal problems are equivalent, but:

Feasible set for (RP):  $x_1 \le 1/3$ Feasible set for (RP):  $x_1 = 0$ .

Try to avoid equality constraints (e.g. using elimination)!





#### **Example 2:**

$$(RP) |x_1 - \zeta| + |x_2 - \zeta| \le 2 \quad \forall \zeta : |\zeta| \le 1$$

or

$$\begin{cases}
y_1 + y_2 \le 2 \\
y_1 \ge x_1 - \zeta & \forall \zeta : |\zeta| \le 1 \\
y_1 \ge \zeta - x_1 & \forall \zeta : |\zeta| \le 1 \\
y_2 \ge x_2 - \zeta & \forall \zeta : |\zeta| \le 1 \\
y_2 \ge \zeta - x_2 & \forall \zeta : |\zeta| \le 1
\end{cases}$$

x=(1,-1) is feasible for (RP) but not for  $(\overline{RP})!$ 

Be careful with "splitting up" constraints!





#### **Example 2:**

$$(RP) |x_1 - \zeta| + |x_2 - \zeta| \le 2 \quad \forall \zeta : |\zeta| \le 1$$

#### Correct reformulation:

$$(\overline{RP}) \begin{cases} x_1 - \zeta + x_2 - \zeta \le 2 & \forall \zeta : |\zeta| \le 1 \\ x_1 - \zeta - (x_2 - \zeta) \le 2 & \forall \zeta : |\zeta| \le 1 \\ -(x_1 - \zeta) + x_2 - \zeta \le 2 & \forall \zeta : |\zeta| \le 1 \\ -(x_1 - \zeta) - (x_2 - \zeta) \le 2 & \forall \zeta : |\zeta| \le 1. \end{cases}$$

See Gorissen et al. (2012).



- RC reformulations also valid for MIP problems.
- Structure (e.g. network, total unimodularity) often destroyed by RC.
- Analyze robustness of nominal and robust optimal solution.
- There are no methods known that can deal with integer adjustable variables.
- RO methodology implemented in modelling packages as Yalmip, ROME, AIMMS.

## CentER Extensions

- Tractable robust counterpart for other uncertainty regions,
   e.g. defined by separable convex functions.
- Tractable robust counterpart for problems with certain types of non-affine uncertain parameters.

Example: 
$$(1 + \zeta_1^2)x_1 + \zeta_1\zeta_2x_2 \le 5, \ \forall \zeta : \|\zeta\|_2 \le 1.$$

- Tractable (approximations of) robust counterpart for (conic) quadratic, SDP and other nonlinear problems.
- Tractable robust counterpart for certain classes of nonlinear decision rules.
- Globalized robust counterpart for nonlinear optimization problems.





#### Hundreds of papers on applications in the following categories:

- Circuit design
- Signal processing / signal estimation
- Communication
- Control
- Structural optimization / material design / shape design
- Mechanical Engineering
- Optics
- Applied Physics
- Chemical engineering
- Medical applications
- Computer Science





- Machine learning
- Agriculture
- Water resources / water management / hydrology
- Energy / Environment
- Vehicle routing
- Inventory / Supply chain management
- Facility location
- Transportation / civil engineering
- Revenue Management
- Maintenance
- Finance
- Design of Experiments
- Statistics





### Application: Cancer treatment



- Treatment plan optimization:
  - tumor should get prescribed dose
  - healthy organs should be spared.
- Formulated as a large LP probem.
- Uncertainties in location of tumor / organs.
   e.g. because of breathing
- Robust solution much better than nominal solution.
- See e.g.: Olafsson and Wright (2006).



# CentER Conclusions

- Practical optimization problems often contain uncertain parameters.
- RO is a tractable way to obtain robust solutions: Provides a natural way for modelling uncertainty. Robust counterpart is often tractable.
- The basic paradigm of RO is "worst-case", but also often improves average behaviour.
- Adjustable RO is an efficient way to solve multistage problems.
- RO implemented in modelling software packages.

