Robust Optimization: The Need, The Challenge, The Achievements

May 26, 27, 2018

Dept. of Statistical Inference and Mathematics,

The Institute of Statistical Mathematics, Tokyo

Aharon Ben-Tal

Technion—Israel Institute of Technology

DATA UNCERTAINTY IN OPTIMIZATION

♣ **Consider a generic optimization problem of the form**

$$\min_x \{f(x;\zeta) : F(x;\zeta) \in \mathbf{K}\}$$

• $x \in \mathbf{R}^n$: **decision vector** • $\zeta \in \mathbf{R}^M$: **data** • $\mathbf{K} \subset \mathbf{R}^m$: **closed convex set**

♠ **More often than not the data** ζ **is** *uncertain* – **not known exactly when problem is solved.**

Sources of data uncertainty:

• **part of the data is measured/estimated** $\Rightarrow$ *estimation errors*

• **part of the data (e.g., future demands/prices) does not exist when problem is solved** $\Rightarrow$ *prediction errors*

• **some components of a solution cannot be implemented exactly as computed** $\Rightarrow$ *implementation errors* **which in many models can be mimicked by appropriate data uncertainty**

Example

Effect of implementation errors

Antenna array (Ben-Tal and Nemirovski (2002))

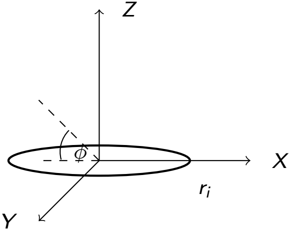
1. We consider an optimization problem with 40 circular antennas.
2. Each antenna has its diagram $D_i(\phi)$ - a plot of intensity of signal sent to different directions.
3. The diagram of the set of 40 antennas is the sum of their diagrams .

$$D(\phi) = \sum_{i=1}^{n} x_i D_i(\phi)$$

4. To the i-th antenna we can send a different amount of power x_i.
5. **Objective:** Set the x_i's in such a way that the diagram has the desired shape.

Application - antenna array optimization

Consider a circular antenna:



Energy sent in angle ϕ is characterized by *diagram*

Diagram of a single antenna:

$$D_i(\phi) = \frac{1}{2} \int_0^{2\pi} \cos\left(\frac{2\pi i}{40} \cos(\phi) \cos(\theta) \right) d\theta$$

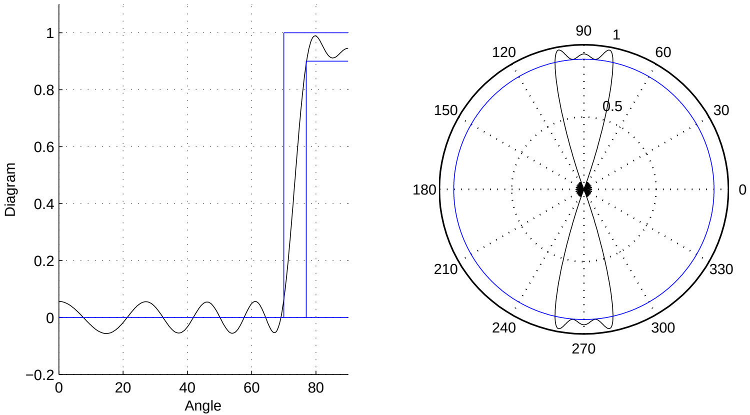
Diagram of n antennas

$$D(\phi) = \sum_{i=1}^{n} x_i D_i(\phi)$$

x_i - power assigned to antenna i

Objective: construct $D(\phi)$ as close as possible to the desired $D^*(\phi)$ using the antennas available.

Desired diagram graphically



Antenna array (Ben-Tal and Nemirovski (2002))

Problem conditions:

- for $77^\circ < \phi \leq 90^\circ$ the diagram is nearly uniform:

$$0.9 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 77^\circ < \phi \leq 90^\circ$$

- for $70^\circ < \phi \leq 77^\circ$ the diagram is bounded:

$$-1 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 70^\circ < \phi \leq 77^\circ$$

- we minimize the maximum absolute diagram value over $0^\circ < \phi \leq 70^\circ$:

$$\min \max_{0^\circ < \phi \leq 70^\circ} \left| \sum_{i=1}^{n} x_i D_i(\phi) \right|$$

Optimization problem to be solved

$$
\begin{aligned}
\min \quad & \tau \\
\text{s.t.} \quad & -\tau \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq \tau, \quad 0 \leq \phi \leq 70^\circ \\
& -1 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 70^\circ \leq \phi \leq 77^\circ \\
& 0.9 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 77^\circ \leq \phi \leq 90^\circ
\end{aligned}
$$

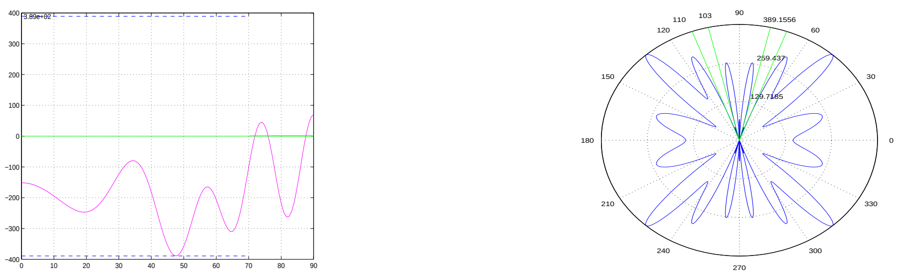
Typically, decisions x_i suffer from implementation error z_i:

$$x_i \mapsto \widetilde{x}_i = (1 + z_i) x_i$$

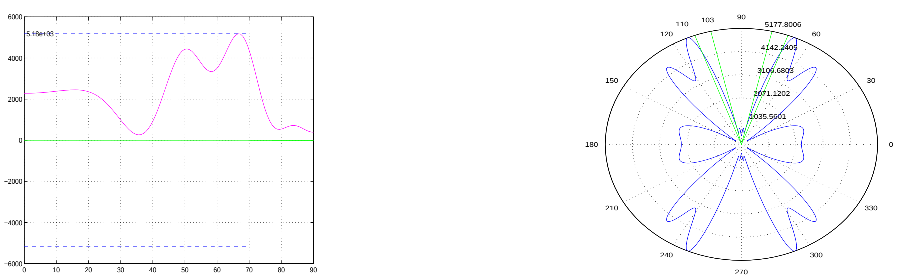
We want each constraint to hold with probability at least $1 - \epsilon$!

Antenna Design (continued)

Dream: no implementation errors
Sidelobe level 0.018



Reality: $x_j \mapsto (1+\epsilon_j)x_j$
$[\epsilon_j \sim$ **Uniform**$[-0.001, 0.001]]$
100-diagram sample: Sidelobe level $\in [97, 536]$



Reality: $x_j \mapsto (1+\epsilon_j)x_j$
$[\epsilon_j \sim$ **Uniform**$[-0.02, 0.02]]$
100-diagram sample: Sidelobe level $\in [2469, 11552]$

Example

Effect of data inaccuracy

Data Uncertainty in Optimization

♣ **Consider a real-world LP program `PILOT4` from the `NETLIB` library (1,000 variables, 410 constraints). The constraint # 372 is:**

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

♠ **Most of the coefficients are "ugly reals" (like -15.79081 or -84.644257). It is highly unlikely that the corresponding real-life parameters are known to high accuracy, so that the ugly coefficients can be thought of as uncertain – not known exactly.**

The only exception is the coefficient 1 at x_{880} – it perhaps reflects the structure of the problem and might be exact.

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

(?) What happens with the constraint, evaluated at the nominal solution x^{n} as reported by CPLEX, when the accuracy in the uncertain data is 0.1%:

$$
|a_i^{\mathrm{true}} - a_i^{\mathrm{n}}| \leq 0.001 |a_i^{\mathrm{n}}| \qquad (*)
$$

- **In the worst case, the constraint can be violated by as much as 450%:**

$$
\min_{a^{\mathrm{true}}} \left\{ [a^{\mathrm{true}}]^T x^{\mathrm{n}} \,|\, a^{\mathrm{true}} \text{ satisfies } (*) \right\} - b < -128.2 \approx 4.5|b|.
$$

- **Assuming "random uncertainty":**

$$a_i^{\mathrm{true}} = a_i^{\mathrm{n}} + \epsilon_i |a_i^{\mathrm{n}}|, \quad \epsilon_i \sim \textbf{Uniform}[-0.001, 0.001]$$

and running 1,000 simulations, we come to the results as follows:

Prob$\{V > 0\}$	**Prob**$\{V > 150\%\}$	**Mean**(V)
0.50	**0.18**	**125%**

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

$\Rightarrow$ **The nominal solution is highly "unreliable" – small perturbations of (clearly uncertain!) data entries can make the solution heavily infeasible...**

♣ **Among 90 `NETLIB` LP problems,**

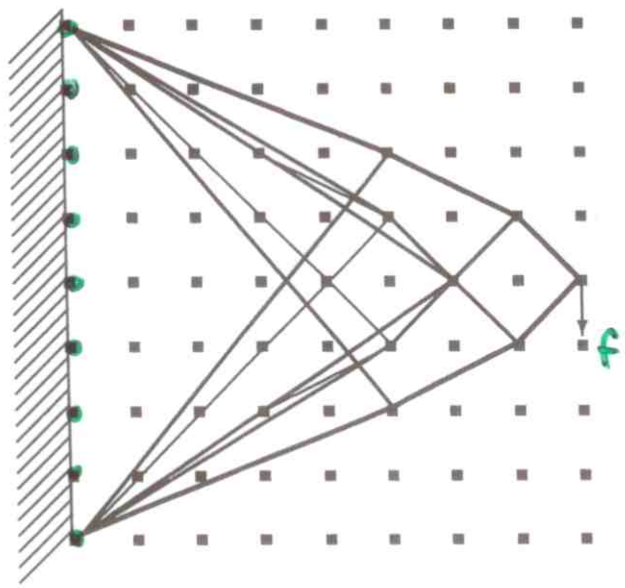
- **In 19 problems 0.01%-perturbations of "clearly uncertain" data result in more than 5%-violations of (some of) the constraints as evaluated at the nominal optimal solution;**
- **In 13 of these 19 problems, 0.01%-perturbations of "clearly uncertain" data result in more than 50%-violations of the constraints!**

Example

Effect of uncertain predictions

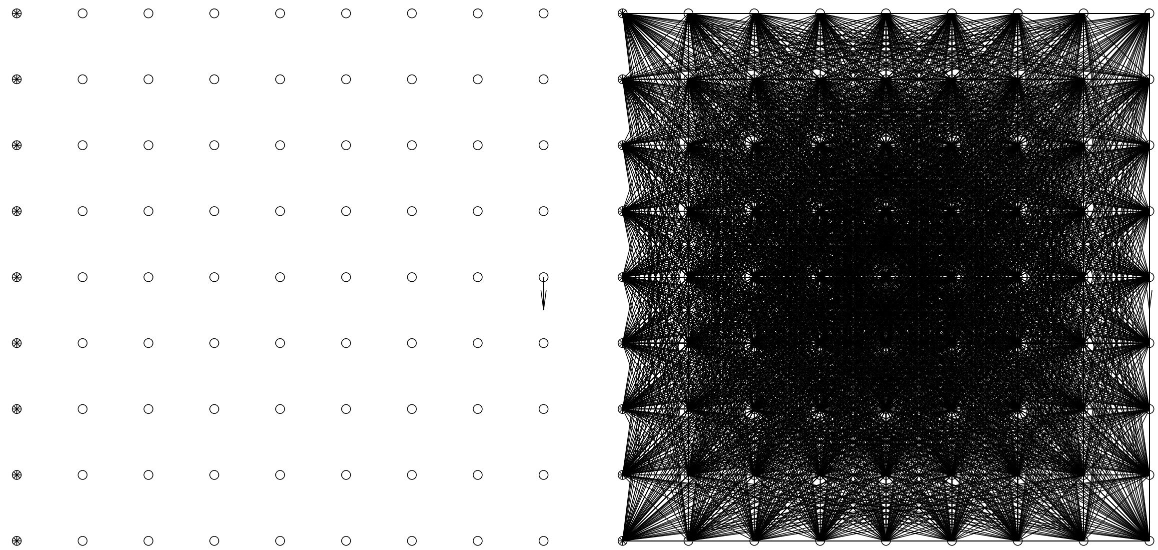
TRUSS TOPOLOGY DESIGN

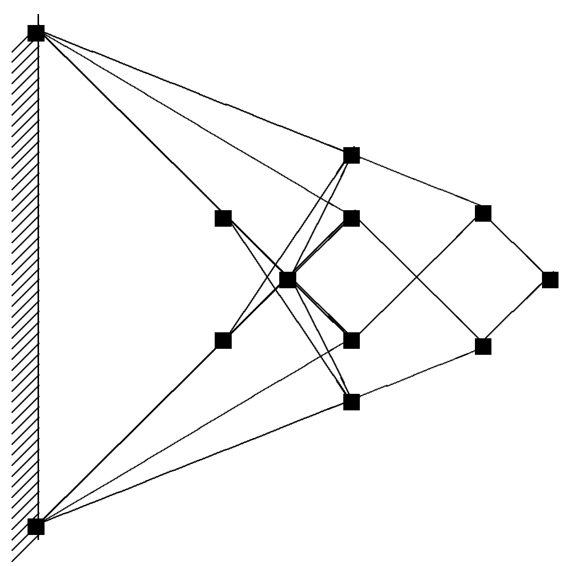
Engineering Formulation: Given total volume of a truss (construction comprised of thin elastic bars linked to each other, like electricity mast or Eifel Tower), find the truss which is most rigid with respect to a given set of external loads.



A Truss

Truss Topology Design





The simplest TTD problem is

$$\text{Compliance} = \frac{1}{2} f^T x \quad \to \quad \min$$

s.t.

$$\begin{array}{rcl}
\underbrace{\left[\sum_{i=1}^{m} t_i b_i b_i^T\right]}_{A(t)\succeq 0} x & = & f \\
\sum_{i=1}^{m} t_i & \leq & w \\
t & \geq & 0
\end{array}$$

- **Data:**

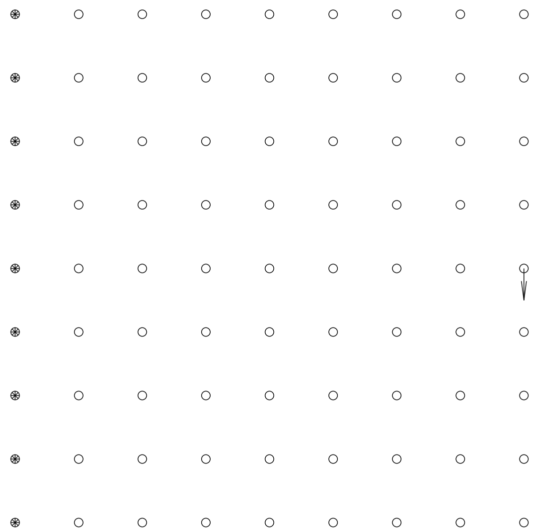
— $b_i \in \mathbb{R}^n$, n – **# of nodal degrees of freedom (for a** $10 \times 10 \times 10$ **ground structure,** $n \approx 3,000$**)**

— m – **# of tentative bars (for** $10 \times 10 \times 10$ **ground structure,** $m \approx 500,000$**)**

- **Design variables:** $t \in \mathbb{R}^m$, $x \in \mathbb{R}^n$

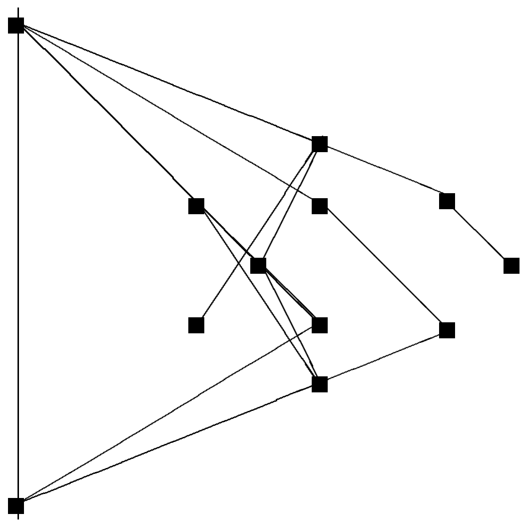
Can we trust the truss?

<u>Example:</u> **Assume we are designing a planar truss – a cantilever; the 9×9 nodal structure and the only load of interest f^* are as shown on the picture:**



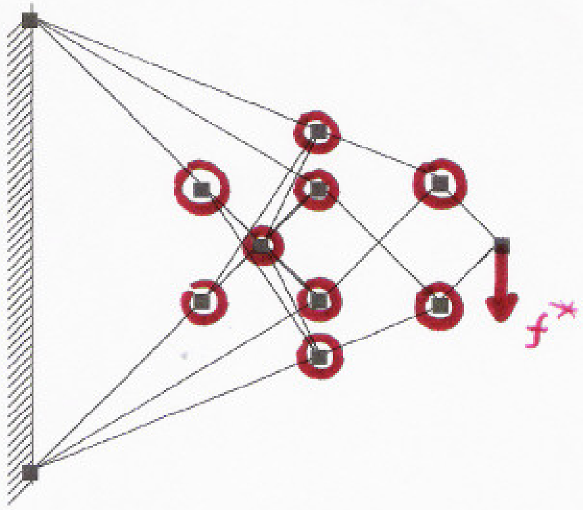
9×9 ground structure and the load of interest

The optimal single-load design yields a nice truss as follows:



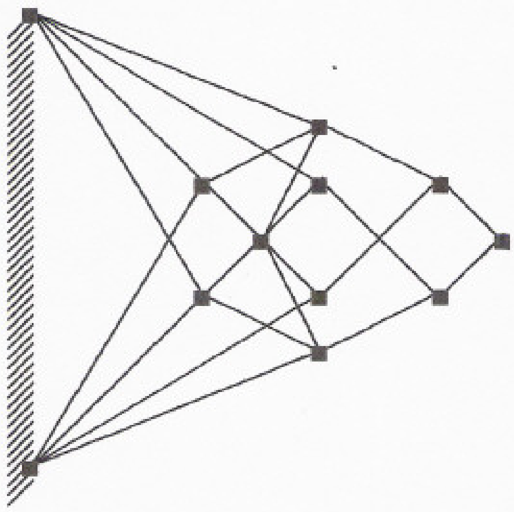
Optimal cantilever (single-load design)
the compliance is 1.000

• **Passing from the single-load to the robust design, we modify the result as follows:**



Optimal cantilever (single-load design)

$\Downarrow$



"Robust" cantilever

Compliances	**Design**	
	Single-load	**Robust**
Compliance w.r.t. f^*	**1.000**	**1.0024**
max **compliance w.r.t. loads** f**:** $\| f \| \leq 0.1 \| f^* \|$	**32000**	**1.03**

"NON-ADJUSTABLE" ROBUST OPTIMIZATION:
Robust Counterpart of Uncertain Problem

$$\boxed{\min_x \{f(x,\zeta) : F(x,\zeta) \in \mathbf{K}\}} \qquad (\mathbf{U})$$

♣ **The initial ("Non-Adjustable") Robust Optimization paradigm (Soyster '73, B-T&N '97–, El Ghaoui et al. '97–, Bertsimas&Sim '03–,...) is based on the following tacitly accepted assumptions:**

A.1. All decision variables in (U) represent "here and now" decisions which should get specific numerical values as a result of solving the problem and *before* the actual data "reveals itself".

A.2. The uncertain data are "unknown but bounded": one can specify an appropriate (typically, bounded) *uncertainty set* $\mathcal{U} \subset \mathbf{R}^M$ of possible values of the data. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within this set.

A.3. The constraints in (U) are "hard" – we cannot tolerate violations of constraints, even small ones, when the data is in $\mathcal{U}$.

$$\boxed{\begin{array}{c}\min_x \{f(x,\zeta) : F(x,\zeta) \in \mathbf{K}\} \\ \zeta \in \mathcal{U}\end{array}} \qquad (\mathbf{U})$$

♠ **Conclusions:**

• **The only meaningful candidate solutions are the** *robust* **ones – those which remain feasible whatever be a realization of the data from the uncertainty set:**

$$x \text{ robust feasible } \Leftrightarrow F(x,\zeta) \in \mathbf{K} \ \forall \zeta \in \mathcal{U}$$

• **"Robust optimal" solution to be used is a robust solution with the smallest possible** *guaranteed* **value of the objective, that is, the optimal solution of the optimization problem**

$$\min_{x,t} \{t : f(x,\zeta) \leq t,\ F(x,\zeta) \in \mathbf{K} \ \forall \zeta \in \mathcal{U}\} \qquad \mathbf{(RC)}$$

called the *Robust Counterpart* **of** $(\mathbf{U})$**.**

♠ **With traditional modelling methodology,**

• **"large" data uncertainty is modelled in a stochastic fashion and then processed via Stochastic Programming techniques**

> **<u>Fact:</u>** *In many cases, it is difficult to specify reliably the distribution of uncertain data and/or to process the resulting Stochastic Programming program.*

♠ **The ultimate goal of** *Robust Optimization* **is to take into account data uncertainty already at the modelling stage in order to "immunize" solutions against uncertainty.**

• **In contrast to Stochastic Programming, Robust Optimization does not assume stochastic nature of the uncertain data (although can utilize, to some extent, this nature, if any).**

Optimization Problems with Uncertain Data

♣ **A generic optimization problem is of the form**

$$\min_x \{f(x;\zeta) \mid F(x;\zeta) \le 0\} \qquad (\mathbf{P}_\zeta)$$

- x **is the design vector**
- f, F **are specified by the description of the problem**
- ζ **is a finite-dimensional vector specifying the data.**

Example 1. Linear Programming:

$$\min_x \{c^T x \mid Ax \le b\} \qquad [\zeta = (c, A, b)]$$

Example 2. Convex Quadratic Programming:

$$\min_x \{c^T x \mid x_i^T A_i^T A_i x_i - 2b_i^T x_i + c_i \le 0,\ i = 1, ..., m\} \qquad [\zeta = (c, \{A_i, b_i, c_i\}_{i=1}^m)]$$

Example 3. Conic Quadratic Programming:

$$\min_x \{c^T x \mid \| A_i x - b_i \|_2 \le c_i^T x - d_i,\ i = 1, ..., m\} \qquad [\zeta = (c, \{A_i, b_i, c_i, d_i\}_{i=1}^m)]$$

Example 4. Semidefinite Programming:

$$\min_x \left\{c^T x \mid A_0 + \sum_{j=1}^{\dim x} x_i A_i \succeq 0\right\} \qquad [\zeta = (c, A_0, ..., A_{\dim x})]$$

Semi-Infinite Conic Programs

♣ Conic Program:

$$\min_x \{c^T x : Ax - b \in \mathbf{K}\} \qquad \textbf{(C)}$$

• (c, A, b) – **problem's data**

• **closed pointed convex cone $\mathbf{K}$, $\mathrm{int}\,\mathbf{K} \neq \emptyset$, in a Euclidean space – problem's structure**

Examples:

- **Linear Programming: $\mathbf{K} = \mathbf{R}^n_+$**
- **Conic Quadratic Programming: $\mathbf{K}$ is a direct product of Lorentz cones**
$$\mathbf{L}^k = \{y \in \mathbf{R}^k : y_k \geq \sqrt{y_1^2 + ... + y_{k-1}^2}\}$$
- **Semidefinite Programming: $\mathbf{K} = \mathbf{S}^n_+$ is the cone of positive semidefinite matrices in the space $\mathbf{S}^n$ of $n \times n$ symmetric matrices**

♣ Semi-Infinite Conic Program:

$$\min_x \{c^T x : Ax - b \in \mathbf{K} \; \forall [A, b] \in \mathcal{U}\}$$

where $\mathcal{U}$ is a given "uncertainty set" (assumed to be convex and compact).

$$\min_x \left\{ c^T x : Ax - b \in \mathbf{K}\ \forall [A,b] \in \mathcal{U} \right\} \qquad \textbf{(S)}$$

♣ **The main mathematical question associated with semi-infinite problem (S) is:**

> **(?) When and how (S) can be reformulated as a "computationally tractable" optimization problem?**

The answer to (?) clearly depends on the interplay between the geometries of the cone K and of the uncertainty set $\mathcal{U}$.

Intractability

Consider a (nearly linear) constraint:

$$\|Px - p\|_1 \le 1\,, \quad \forall\, p \in \mathcal{U} \tag{1}$$

where

$$\mathcal{U} = \{p = B\zeta : \|\zeta\|_\infty \le 1\}$$

(a polyhedral set). $\qquad B \succeq 0$ given matrix

Check whether $x = 0$ is robust feasible, i.e. the validity of the inequality

$$\|B\zeta\|_1 \le 1 \quad \forall\, \zeta : \|\zeta\|_\infty \le 1\,. \tag{2}$$

Since $\|u\|_1 = \max\{y^T u \mid \|y\|_\infty \le 1\}$, (2) is equivalent to

$$\max_{y,\zeta} \left\{ y^T B\zeta \mid \|y\|_\infty \le 1,\ \|\zeta\|_\infty \le 1 \right\} \le 1\,. \tag{3}$$

Maximum is achieved at $y = \zeta^*$ so (3) is equivalent to

$$\max\{y^T By \mid \|y\|_\infty \le 1\} \le 1\,.$$

The problem on the lhs (maximizing a nonnegative quadratic form over the unit box) is known to be NP-hard. In fact, it is NP-hard to compute this maximum with an accuracy better than 4%.

---

*Suppose $(\bar{y}, \bar{\zeta})$, $\bar{y} \ne \bar{\zeta}$ is optimal. Then $\left(\frac{\bar{y}+\bar{\zeta}}{2}\,,\ \frac{\bar{y}+\bar{\zeta}}{2}\right)$ is feasible and $\left(\frac{\bar{y}+\bar{\zeta}}{2}\right)^T B \left(\frac{\bar{y}+\bar{\zeta}}{2}\right) > \bar{y}^T B\bar{\zeta}$ which is a contradiction to the optimality of $\bar{y}, \bar{\zeta}$.

A Short Introduction to Conic Optimization

Conic optimization program

• **Let $\mathbf{K} \subset \mathbf{R}^m$ be a cone defining a good vector inequality $\geq_{\mathbf{K}}$ (i.e., $\mathbf{K}$ is a closed pointed cone with a nonempty interior).**

A generic conic problem associated with $\mathbf{K}$ is an optimization program of the form

$$\min_x \left\{ c^T x : Ax - b \geq_{\mathbf{K}} 0 \right\}. \qquad \textbf{(CP)}$$

Conic Duality Theorem

• **A conic problem**

$$\min_x \left\{ e^T s : s \in [\mathcal{L} + f] \cap \mathbf{K} \right\}$$

is called strictly feasible, if its feasible plane intersects the interior of the cone K.

$$\begin{array}{rl} \textbf{(P):} & \min_x \left\{ c^T x : Ax - b \geq_{\mathbf{K}} 0 \right\} \\ \textbf{(D):} & \max_y \left\{ b^T y : A^T y = c, y \geq_{\mathbf{K}_*} 0 \right\} \end{array}$$

• **Conic Duality Theorem. Consider a conic problem (P) along with its dual (D).**

1. Symmetry: The duality is symmetric: the problem dual to dual is (equivalent to) the primal;

2. Weak duality: The value of the dual objective at any dual feasible solution is $\leq$ the value of the primal objective at any primal feasible solution;

3. Strong duality in strictly feasible case: If one of the problems (P), (D) is strictly feasible and bounded, then the other problem is solvable, and the optimal values in (P) and (D) are equal to each other.

If both (P), (D) are strictly feasible, then both problems are solvable with equal optimal values.

$$\min_x \left\{ c^T x : Ax - b \geq_{\mathbf{K}} 0 \right\}. \qquad \textbf{(CP)}$$

Examples:

- **Linear Programming**

$$\min_x \left\{ c^T x : Ax - b \geq 0 \right\} \qquad (LP)$$

(K is a nonnegative orthant)

- **Conic Quadratic Programming:**

$$\min_x \left\{ c^T x : \|D_\ell x + d_\ell\|_2 \leq e_\ell^T x + f_\ell,\ \ell = 1, ..., k \right\}$$
$$\Updownarrow$$
$$\min_x \left\{ c^T x : Ax - b \equiv \begin{bmatrix} \begin{bmatrix} D_1 x + d_1 \\ e_1^T x + f_1 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} D_k x + d_k \\ e_k^T x + f_k \end{bmatrix} \end{bmatrix} \geq_{\mathbf{K}} 0 \right\}, \qquad (CQP)$$

$$\mathbf{K} = \mathbf{L}^{m_1} \times ... \times \mathbf{L}^{m_k}$$

is a direct product of Lorentz cones

- **Semidefinite Programming:**

$$\min_x \left\{ c^T x : \mathcal{A}x - B \equiv x_1 A_1 + ... + x_n A_n - B \succeq 0 \right\} \qquad (SDP)$$
$$\left[P \succeq Q \Leftrightarrow P \geq_{\mathbf{S}^m_+} Q \right]$$

Conic Quadratic Problem

Primal

$$\min_x c^T x$$

$$\|D_i x - d_i\|_2 \leq p_i^T x - q_i \qquad i = 1, \ldots, k$$

$$Rx = r$$

Dual

$$\max_{v,y;u} r^T v + \sum_{i=1}^{K} (d_i y_i + q_i u_i)$$

$$R^T v + \sum (D_i^T y_i + p_i u_i) = c$$

$$\|y_i\| \leq u_i \qquad i = 1, \ldots, k$$

Program dual to an SDP program

$$\min_x \left\{ c^T x \mid \mathcal{A}x - B \equiv \sum_{j=1}^n x_j A_j - B \succeq 0 \right\} \qquad \textbf{(SDPr)}$$

According to our general scheme, the problem dual to (SDPr) is

$$\max_Y \{\langle B, Y\rangle \mid \mathcal{A}^* Y = c, \ Y \succeq 0\} \qquad \textbf{(SDDl)}$$

(recall that $\mathbf{S}^m_+$ is self-dual!).

It is easily seen that the operator $\mathcal{A}^*$ conjugate to $\mathcal{A}$ is given by

$$\mathcal{A}^* Y = (\mathrm{Tr}(YA_1), ..., \mathrm{Tr}(YA_n))^T : \mathbf{S}^m \to \mathbf{R}^n.$$

Consequently, the dual problem is

$$\max_Y \{\mathrm{Tr}(BY) \mid \mathrm{Tr}(YA_i) = c_i, \ i = 1, ..., n, \ Y \succeq 0\} \qquad \textbf{(SDDl)}$$

• Example: **The problem with nonlinear objective and constraints**

	minimize $\sum_{\ell=1}^{n} x_\ell^2$
(a)	$x \geq 0;$
(b)	$a_\ell^T x \leq b_\ell,\ \ell = 1, ..., n;$
(c)	$\\|Px - p\\|_2 \leq c^T x + d;$
(d)	$x_\ell^{\frac{\ell+1}{\ell}} \leq e_\ell^T x + f_\ell,\ \ell = 1, ..., n;$
(e)	$x_\ell^{\frac{l}{l+3}} x_{l+1}^{\frac{1}{l+3}} \geq g_\ell^T x + h_\ell,\ \ell = 1, ..., n-1;$
(f)	$\mathrm{Det} \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_1 & x_2 & \cdots & x_{n-1} \\ x_3 & x_2 & x_1 & \cdots & x_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n-1} & x_{n-2} & \cdots & x_1 \end{pmatrix} \geq 1;$
(g)	$1 \leq \sum_{\ell=1}^{n} x_\ell \cos(\ell\omega) \leq 1 + \sin^2(5\omega)\ \forall \omega \in \left[-\frac{\pi}{7}, 1.3\right]$

can be converted, in a systematic way, into an equivalent problem

$$\min_x \left\{ c^T x : Ax - b \succeq 0 \right\},$$

" $\succeq$ " **being one of our 3 standard vector inequalities, so that seemingly highly diverse constraints of the original problem allow for unified treatment.**

- **<u>Lemma on Schur Complement.</u> A symmetric block matrix**

$$A = \begin{pmatrix} P & Q^T \\ Q & R \end{pmatrix}$$

with positive definite R is positive (semi)definite if and only if the matrix

$$P - Q^T R^{-1} Q$$

is positive (semi)definite.

<u>Proof.</u> A **is** $\succeq 0$ **if and only if**

$$\inf_v \begin{pmatrix} u \\ v \end{pmatrix}^T \begin{pmatrix} P & Q^T \\ Q & R \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \geq 0 \quad \forall u. \qquad (*)$$

When $R \succ 0$**, the left hand side** $\inf$ **can be easily computed and turns to be**

$$u^T (P - Q^T R^{-1} Q) u.$$

Thus, $(*)$ **is valid if and only if**

$$u^T (P - Q^T R^{-1} Q) u \geq 0 \quad \forall u,$$

i.e., if and only if

$$P - Q^T R^{-1} Q \succeq 0.$$

Lecture 2

Robust Solutions of Uncertain Linear Optimization Problems

When treating an uncertain linear inequality $a^T x \le b$ we'll use

$$\mathcal{U} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a^0 \\ b^0 \end{pmatrix} + \sum \zeta_\ell \begin{pmatrix} a^\ell \\ b^\ell \end{pmatrix} \mid \zeta \in Z \right\}$$

where Z is represented by linear conic inequalities:

$$Z = \{\zeta \mid \exists\, u : P\zeta + Qu + p \in K\}$$

where K is closed convex pointed cone.

Efficient Representation of Sets

$\hat{X} \subset \mathbb{R}^n \times \mathbb{R}^k$ *represents* $X \subset \mathbb{R}^n$, if

$$X = \{x \in \mathbb{R}^n \mid \exists\, u \in \mathbb{R}^k : (x, u) \in \hat{X}\}$$

(the projection of $\hat{X}$ onto the space of the x-variables is exactly X)

<u>Example</u>

$$X = \left\{ x \in \mathbb{R}^n : \sum_{j=1}^{n} |x_j| \le 1 \right\} .$$

Straightforward representation of X requires the 2^n linear inequalities

$$\pm x_1 \pm x_2 \pm \cdots \pm x_n \le 1 \, .$$

Alternatively, X can be represented by

$$\hat{X} = \Big\{ (x, u) \in \mathbb{R}^n \times \mathbb{R}^n : \sum u_i \le 1,\; -u_j \le x_i \le u_j, \,\forall\, j \Big\}$$

requiring only $2n + 2$ linear inequalities.

The Constraint-wise Nature of the RC

$$\left.\begin{array}{l} x_1 \geq \zeta_1 \\ x_2 \geq \zeta_2 \end{array}\right\} \ \forall \begin{pmatrix}\zeta_1 \\ \zeta_2\end{pmatrix} \in \mathcal{U} = \left\{ \begin{pmatrix}\zeta_1 \\ \zeta_2\end{pmatrix} \middle| \begin{array}{l} \zeta_1 \geq 0 \\ \zeta_1 + \zeta_2 \leq 1 \end{array} \right\}$$

$$\Leftrightarrow \qquad \left.\begin{array}{l} x_1 \geq \max_{\zeta_1 \in \mathcal{U}} \zeta_1 = 1 \\ x_2 \geq \max_{\zeta_2 \in u} \zeta_2 = 1 \end{array}\right\}$$

Same RC under uncertainty

$$\hat{\mathcal{U}} = \mathcal{U}_1 \times \mathcal{U}_2 \qquad \mathcal{U}_i = \{\zeta_i \mid 0 \leq \zeta_i \leq 1\}$$

$\mathcal{U}_i$ = proj. of $\mathcal{U}$ on ζ_i-space.

Conclusion The RC of uncertain linear inequalities under uncertainty set u is the same when u is extended to the direct product

$$\hat{\mathcal{U}} = \mathcal{U}_1 \times \mathcal{U}_2 \cdots \times \mathcal{U}_m$$

of its projections onto the spaces of the data of respective constraints.

Moreover, each $\mathcal{U}_i$ can be replaced by its closed convex hull.

Focus on a *single* uncertainty-affected linear inequality—a family

$$\left\{a^T x \le b\right\}_{[a;b]\in\mathcal{U}}$$

of linear inequalities with the data varying in the uncertainty set

$$\mathcal{U} = \left\{[a;b] = \left[a^0; b^0\right] + \sum_{\ell=1}^{L} \zeta_\ell \left[a^\ell; b^\ell\right];\ \zeta \in \mathcal{Z}\right\}$$

and on "tractable representation" of the RC

$$a^T x \le b \,\forall \left([a;b] = \left[a^0; b^0\right] + \sum_{\ell=1}^{L} \zeta_\ell \left[a^\ell; b^\ell\right];\ \zeta \in \mathcal{Z}\right) \qquad (2)$$

of this uncertain inequality

Tractable Representation of (2): Simple Cases

Example

$$\mathcal{Z} = \mathrm{Box}_1 \equiv \left\{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1\right\}.$$

In this case, (2) reads

$$\begin{array}{rl} & [a^0]^T x + \sum\limits_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x \le b^0 + \sum\limits_{\ell=1}^{L} \zeta_\ell b^\ell \qquad \forall\,(\zeta : \|\zeta\|_\infty \le 1) \\ \Leftrightarrow & \sum\limits_{\ell=1}^{L} \zeta_\ell [[a^\ell]^T x - b^\ell] \le b^0 - [a^0]^T x \qquad \forall\,(\zeta : |\zeta_\ell| \le 1,\ \ell = 1, \dots, L) \\ \Leftrightarrow & \max\limits_{-1 \le \zeta_\ell \le 1} \left[\sum\limits_{\ell=1}^{L} \zeta_\ell [[a^\ell]^T x - b^\ell]\right] \le b^0 - [a^0]^T x \end{array}$$

The concluding maximum in the chain is clearly $\sum_{\ell=1}^{L} |[a^\ell]^T x - b^\ell|$, so (2) becomes

$$[a^0]^T x + \sum_{\ell=1}^{L} |[a^\ell]^T x - b^\ell| \le b^0,$$

which in turn admits a representation by a system of linear inequalities:

$$\begin{cases} -u_\ell \leq [a^\ell]^T x - b^\ell \leq u_\ell, \quad \ell = 1, \ldots, L, \\ [a^0]^T x + \sum_{\ell=1}^{L} u_\ell \leq b^0 \,. \end{cases}$$

Example

$$\mathcal{Z} = \mathrm{Ball}_\Omega = \left\{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \leq \Omega\right\} \,.$$

In this case, (2) reads

$$\begin{array}{rl} & [a^0]^T x + \sum\limits_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x \leq b^0 + \sum\limits_{\ell=1}^{L} \zeta_\ell b^\ell \;\forall\, (\zeta : \|\zeta\|_2 \leq \Omega) \\ \Leftrightarrow & \max\limits_{\|\zeta\|_2 \leq \Omega} \left[\sum\limits_{\ell=1}^{L} \zeta_\ell [[a^\ell]^T x - b^\ell]\right] \leq b^0 - [a^0]^T x \\ \Leftrightarrow & \Omega \sqrt{\sum\limits_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2} \leq b^0 - [a^0]^T x, \end{array}$$

a conic quadratic constraint.

The RC of a Linear Inequality-General Case

$$Z = \{\zeta \in \mathbb{R}^L : \exists\, u \in \mathbb{R}^K : P\zeta + Qu + p \in \underbrace{K}_{\text{closed convex pointed cone in } \mathbb{R}^N}\}$$

$$\mathcal{U} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a^o \\ b^o \end{pmatrix} + \sum \zeta_\ell \begin{pmatrix} a^\ell \\ b^\ell \end{pmatrix} \,\middle|\, \zeta \in Z \right\}$$

Uncertain linear constraint

$$a^T x \leq b \;\; \forall \begin{pmatrix} a \\ b \end{pmatrix} \in \mathcal{U}$$

becomes
$$\underbrace{(a^o)^T x - b^o}_{d[x]} + \sum_{\ell=1}^{L} \zeta_\ell \underbrace{[(a^\ell)^T x - b^\ell]}_{c_\ell[x]} \leq 0 \quad \forall\, \zeta \in Z$$

$\Leftrightarrow$
$$\sup_{\zeta \in Z} c^T[x]\zeta + d[x] \leq 0$$

$\Leftrightarrow$
$$\sup_{\zeta, u} \{c^T[x]\zeta \mid P\zeta + Qu + p \in K\} \leq -d[x]\,. \qquad (*)$$

Theorem *Let the perturbation set* Z *be given by*

$$Z = \{\zeta \in \mathbb{R}^L \mid \exists\, u \in \mathbb{R}^K : P\zeta + Qu + p \in K\}$$

where K *is a closed convex pointed cone in* $\mathbb{R}^N$ *which is either polyhedral, or is such that*

$$\exists\, \bar{\zeta}, \bar{u} : P\bar{\zeta} + Q\bar{u} + p \in int\ K\,.$$

Consider the robust counterpart of a linear inequality:

$$a^T x \le b \quad \forall \begin{pmatrix} a \\ b \end{pmatrix} \in \mathcal{U}$$

where

$$\mathcal{U} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a^o \\ b^o \end{pmatrix} + \sum_{\ell=1}^{L} \zeta_\ell \begin{pmatrix} a^\ell \\ b^\ell \end{pmatrix} \,\middle|\, \zeta \in Z \right\}.$$

Then a vector $x \in \mathbb{R}^n$ is robust feasible *if and only if* $\exists\, y \in \mathbb{R}^L$, which together with x satisfies the following linear/conic inequalities:

$$
\begin{aligned}
& p^T y + (a^o)^T x \le b_0 \\
& Q^T y = 0 \\
& (P^T y)_\ell + (a^\ell)^T x = b^\ell, \quad \ell = 1, 2, \ldots, L \\
& y \in K_*
\end{aligned}
$$

where $K_* = \{y \mid y^T v \ge 0, \forall\, v \in K\}$ is the dual cone of K.

Illustration: Single-Period Portfolio Selection

There are 200 assets. Asset #200 ("money in the bank") has yearly return $r_{200} = 1.05$ and zero variability. The yearly returns r_ℓ, $\ell = 1, \dots, 199$ of the remaining assets are independent random variables taking values in the segments $[\mu_\ell - \sigma_\ell, \mu_\ell + \sigma_\ell]$ with expected values μ_ℓ; here

$$\mu_\ell = 1.05 + 0.3\frac{(200-\ell)}{199},\ \sigma_\ell = 0.05 + 0.6\frac{(200-\ell)}{199},\ \ell = 1, \dots, 199\,.$$

The goal is to distribute \$1 between the assets in order to maximize the return of the resulting portfolio, the required risk level being $\varepsilon = 0.5\%$.

We want to solve the uncertain LO problem

$$\max_{y,t}\left\{t : \sum_{t=1}^{199} r_\ell y_\ell + r_{200} y_{200} - t \geq 0, \sum_{\ell=0}^{200} y_\ell = 1,\ y_\ell \geq 0\, \forall\, \ell\right\},$$

where y_ℓ is the capital to be invested into asset $\#\ell$.

The uncertain data are the returns r_ℓ, $\ell = 1,\dots,199$; their natural parameterization is

$$r_\ell = \mu_\ell + \sigma_\ell \zeta_\ell \,,$$

where ζ_ℓ, $\ell = 1,\dots,199$, are independent random perturbations with zero mean varying in the segments $[-1,1]$. Setting $x = [y; -t] \in \mathbb{R}^{201}$, the problem becomes

$$\left\{\begin{array}{ll} \text{minimize} & x_{201} \\ \text{subject to} & \\ (a) & \left[a^0 + \sum_{\ell=1}^{199} \zeta_\ell a^\ell\right]^T x - \left[b^0 + \sum_{\ell=1}^{199} \zeta_\ell b^\ell\right] \le 0 \\ (b) & \sum_{j=1}^{200} x_\ell = 1 \\ (c) & x_\ell \ge 0,\ \ell = 1,\dots,200 \end{array}\right. \tag{4}$$

where

$$\begin{array}{rcl} a^0 & = & [-\mu_1; -\mu_2; \dots; -\mu_{199}; -r_{200}; -1]; a^\ell = \sigma_\ell \cdot [0_{\ell-1,1}; 1; 0_{201-\ell,1}],\ \ell = 1,\dots,199; \\ b^\ell & = & 0,\ \ell = 0,1,\dots,199\,. \end{array}$$

The only uncertain constraint in the problem is the linear inequality (a). We consider 3 perturbation sets along with the associated robust counterparts of problem (4).

1. *Box RC* which ignores the information on the stochastic nature of the perturbations affecting the uncertain inequality and uses the only fact that these perturbations vary in $[-1, 1]$. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\}$;

2. *Ball-Box* with the safety parameter $\Omega = \sqrt{2\ln(1/\varepsilon)} = 3.255$, which ensures that the optimal solution of the associated RC (a CQ prob.) satisfies (a) with probability at least $1 - \varepsilon = 0.995$. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\}, \ \|\zeta\|_2 \leq 3.255\}$;

3. *Budgeted uncertainties* with the uncertainty budget $\gamma = \sqrt{2\ln(1/\varepsilon)}\sqrt{199} = 45.921$, which results in the same probabilistic guarantees as for the Ball-Box RC. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\}, \ \|\zeta\|_1 \leq 45.921\}$;

Results

Box RC. The associated RC is the LP

$$\max_{y,t}\left\{ t : \begin{array}{l} \sum_{\ell=1}^{199}(\mu_\ell - \sigma_\ell)y_\ell + 1.05y_{200} \geq t \\ \sum_{\ell=1}^{200} y_\ell = 1,\ y \geq 0 \end{array} \right\};$$

as it should be expected, this is nothing but the instance of our uncertain problem corresponding to the worst possible values $r_\ell = \mu_\ell - \sigma_\ell,\ \ \ell = 1, \ldots, 199$, of the uncertain returns. Since these values are less than the guaranteed return for money, the robust optimal solution prescribes to keep our initial capital in the bank with guaranteed yearly return 1.05.

Ball-Box RC. The associated RC is the conic quadratic problem

$$\max_{y,z,w,t} \left\{ t : \begin{array}{l} \sum_{\ell=1}^{199} (\mu_\ell y_\ell + 1.05 y_{200} - \sum_{\ell=1}^{199} |z_\ell| - 3.255 \sqrt{\sum_{\ell=1}^{199} w_\ell^2} \geq t \\ z_\ell + w_\ell = y_\ell, \quad \ell = 1, \ldots, 199, \ \sum_{\ell=1}^{200} y_\ell = 1, \ y \geq 0 \end{array} \right\}.$$

The robust optimal value is 1.1200, meaning 12.0% profit with risk as low as $\varepsilon = 0.5\%$.

Example 1: Synthesis of Antennae array

♣ **The diagram of an antenna. Consider a (monochromatic) antenna placed at the origin. The electric field generated by the antenna at a remote point $r\delta$ (δ is a unit direction) is**

$$E = a(\delta) r^{-1} \cos\left(\phi(\delta) + t\omega - 2\pi r/\lambda\right) + o(r^{-1})$$

• t: time • ω: frequency • λ: wavelength

• It is convenient to aggregate $a(\delta)$ and $\phi(\delta)$ into a single complex-valued function – the *diagram* of the antenna

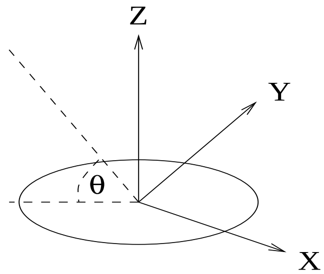
$$D(\delta) = a(\delta)(\cos(\phi(\delta)) + i \sin(\phi(\delta))).$$

- **The directional density of the energy sent by the antenna is proportional to $|D(\cdot)|^2$**
- **The diagram $D(\cdot)$ of a complex antenna comprised of several antenna elements is the sum of the diagrams $D_i(\cdot)$ of the elements:**

$$D(\delta) = D_1(\delta) + ... + D_N(\delta)$$

Robust Antenna Design via Robust LP

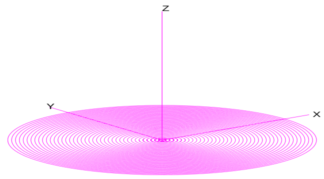
- **A ring-type element:**

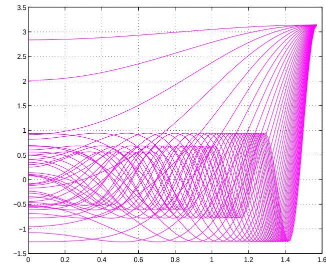


$$D_r(\theta) = \frac{1}{2}\int_0^{2\pi} \cos\left(\frac{2\pi r}{\lambda}\cos(\theta)\cos(\phi)\right) d\phi$$

[diagram depends on the altitude angle θ only]

• **Example:**





40 concentric rings **Diagrams of the rings**

• **<u>Data:</u> 40 concentric rings in the XY-plane with the radii** $r_j = \frac{j}{40}$, $j = 1, ..., 40$, **wavelength** $\lambda = 0.25$

• **<u>Target:</u> To minimize the maximum of the diagram modulus in the "angle of no interest"** $0 \leq \theta \leq 70^{\text{o}}$ – **the "side-lobe level"**

$$\max_{0 \leq \theta \leq 70^{\text{o}}} \Big| \underbrace{\sum_{j=1}^{40} x_j D_{r_j}(\theta)}_{D(\theta)} \Big|$$

under the restrictions that

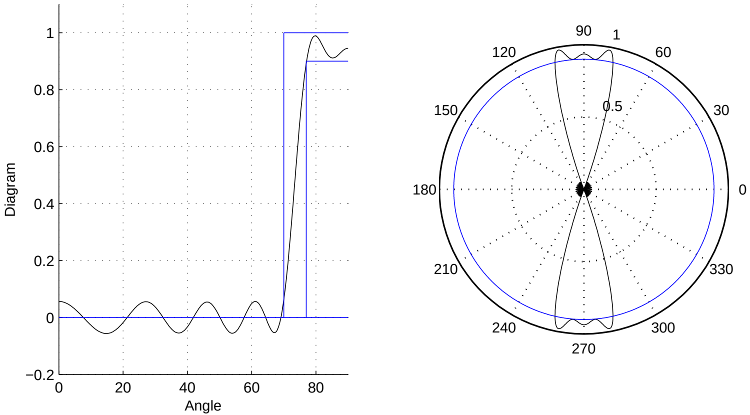
(1) The diagram in the "angle of interest" $77^{\text{o}} \leq \theta \leq 90^{\text{o}}$ **is nearly uniform:**

$$77^{\text{o}} \leq \theta \leq 90^{\text{o}} \Rightarrow 0.9 \leq D(\theta) \leq 1$$

(2) The diagram outside the angle of interest is not too large:

$$|D(\theta)| \leq 1 \quad \forall \theta.$$

Desired diagram graphically



Antenna Design (continued)

• *Nominal Design* **is given by the optimal solution to the LP program**

$$\min_{x,\tau}\left\{\tau : \begin{array}{rcl} -\tau \le \sum_{j=1}^{40} D_{r_j}(\theta_i)x_j & \le & \tau \qquad\quad 0 \le \theta_i < 70^{\rm o} \\ -1 \le \sum_{j=1}^{40} D_{r_j}(\theta_i)x_j & \le & 1, \qquad 70^{\rm o} \le \theta_i < 77^{\rm o} \\ 0.9 \le \sum_{j=1}^{40} D_{r_j}(\theta_i)x_j & \le & 1, \qquad 77^{\rm o} \le \theta_i \le 90^{\rm o} \end{array}\right\} \quad \textbf{(Nom)}$$

where $\{\theta_i\}_{i=1}^N$ **is a "fine grid of angles".**

• **In reality, the computed weights** x_j **are characteristics of physical devices and are therefore affected by implementation errors.**

The nominal design can be highly sensitive to the implementation errors!

How it works? – Antenna Example

$$\min_{x,\tau}\left\{\tau : -\tau \le D_*(\theta_\ell) - \sum_{j=1}^{10} x_j D_j(\theta_\ell) \le \tau,\ \ell = 1,...,L\right\}$$

$$\Updownarrow$$

$$\boxed{\min_{x,\tau}\{\tau : Ax + \tau a + b \ge 0\} \qquad \textbf{(LP)}}$$

• **The influence of "implementation errors"**

$$x_j \mapsto (1+\epsilon_j)x_j$$

is <u>**as if**</u> **there were no implementation errors, but the part** A **of the constraint matrix was uncertain and known "up to multiplication by a diagonal matrix with diagonal entries from** $[0.999, 1.001]$**":**

$$\mathcal{U}_{\rm ini} = \{A = A^{\rm nom}{\rm Diag}(1+\epsilon_1, ..., 1+\epsilon_{10}) : |\epsilon_j| \le 0.001\} \qquad \textbf{(U)}$$

Note that

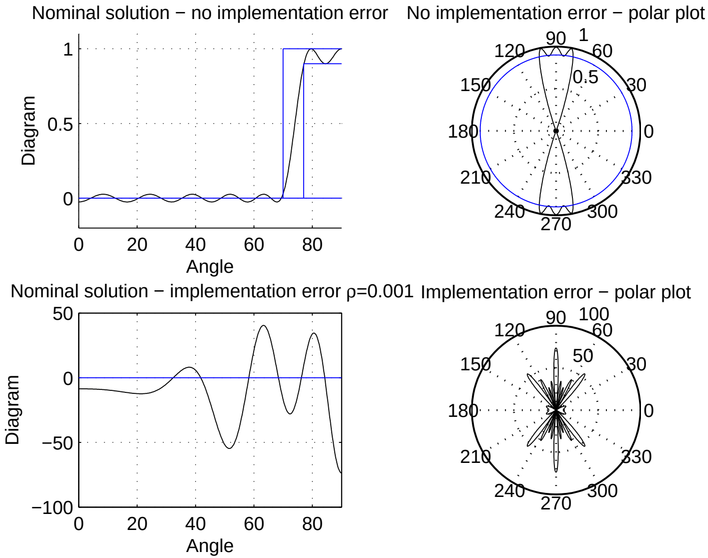
> *As far as a particular constraint is concerned, the uncertainty is an interval one with* $\delta A_{ij} = 0.001|A_{ij}|$. *The remaining coefficients (and the objective) are certain.*

♣ **To improve reliability of our design, we could replace the uncertain LP program (LP), (U) with its robust counterpart, which is nothing but an explicit LP program.**

However, to work directly with $\mathcal{U}_{\rm ini}$ **would be "too conservative" – we would ignore the fact that the implementation errors are random and independent, so that the probability for all of them to take simultaneously the "most unfavourable" values is negligibly small.**

Let us try to define the uncertainty set in a smarter way.

Nominal solution - dream and reality



♣ **Applying the outlined methodology to our Antenna example:**

$$\min_{x,\tau}\left\{\tau : -\tau \le D_*(\theta_\ell) - \sum_{j=1}^{10} x_j D_j(\theta_\ell) \le \tau,\ 1 \le \ell \le 120\right\} \qquad \textbf{(LP)}$$

$$\Downarrow$$

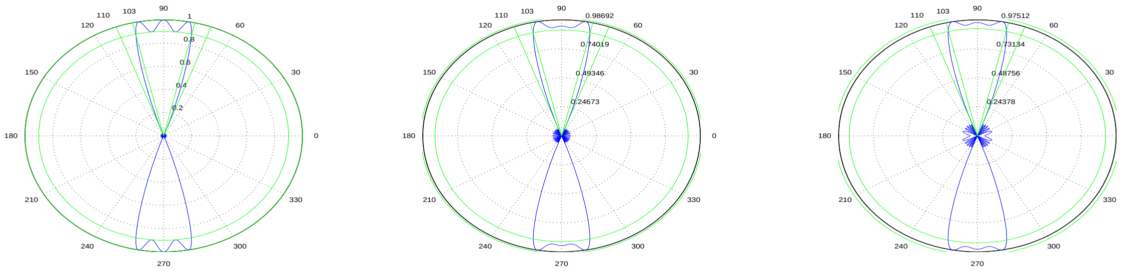
$$\boxed{\begin{array}{rcl} & \min\limits_{x,\tau} \tau & \\ D_*(\theta_\ell) - \sum\limits_{j=1}^{10} x_j D_j(\theta_\ell) + \kappa\sigma\sqrt{\sum\limits_{j=1}^{10} x_j^2 D_j^2(\theta_\ell)} & \le & \tau \\ D_*(\theta_\ell) - \sum\limits_{j=1}^{10} x_j D_j(\theta_\ell) - \kappa\sigma\sqrt{\sum\limits_{j=1}^{10} x_j^2 D_j^2(\theta_\ell)} & \ge & -\tau \\ 1 \le \ell \le 120 & & \end{array}} \qquad \textbf{(RC)}$$

$$[\sigma = 0.001]$$

we get a *robust design*.

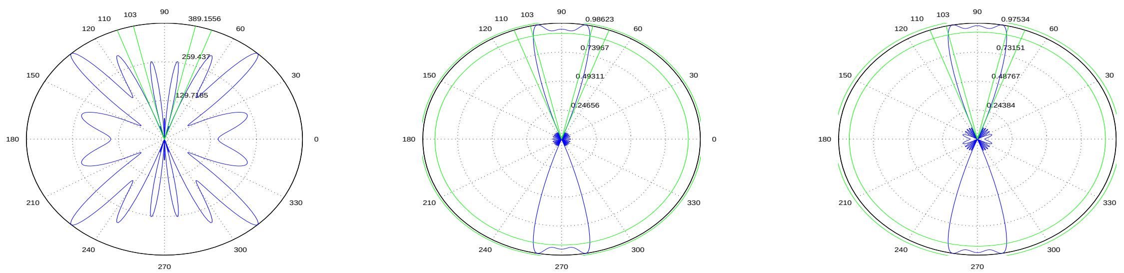
Summary on Antenna Design

No implementation errors



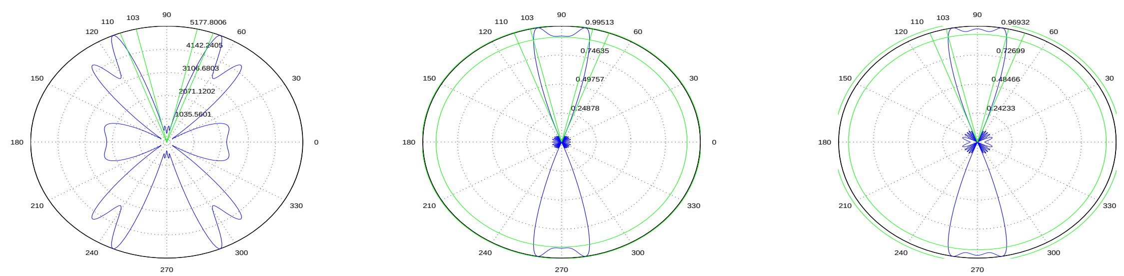
Nominal | Ellipsoidal Robust | Interval Robust

Implementation errors 0.1%



Nominal | Ellipsoidal Robust | Interval Robust

Implementation errors 2%



Nominal | Ellipsoidal Robust | Interval Robust

Example 2: NETLIB Case Study: Diagnosis

♣ `NETLIB` is a collection of about 100 not very large LPs, mostly of real-world origin. To motivate the methodology of our "case study", here is constraint # 372 of the `NETLIB` problem `PILOT4`:

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

The related *nonzero* **coordinates in the optimal solution x^* of the problem, as reported by `CPLEX`, are:**

$$
\begin{array}{lll}
x^*_{826} = 255.6112787181108 & x^*_{827} = 6240.488912232100 & x^*_{828} = 3624.613324098961 \\
x^*_{829} = 18.20205065283259 & x^*_{849} = 174397.0389573037 & x^*_{870} = 14250.00176680900 \\
x^*_{871} = 25910.00731692178 & x^*_{880} = 104958.3199274139 &
\end{array}
$$

This solution makes (C) an equality within machine precision.

♣ Most of the coefficients in (C) are "ugly reals" like -15.79081 or -84.644257. We definitely may believe that these coefficients characterize technological devices/processes, and as such *hardly are known to high accuracy*. **Thus, "ugly coefficients" may be assumed to be uncertain and to coincide with the "true" data within accuracy of 3-4 digits. The only exception is the coefficient `1` of $\mathtt{x}_{880}$, which perhaps reflects the structure of the problem and is exact.**

$$
\begin{array}{rcl}
a^T x & \equiv & -15.79081 x_{826} - 8.598819 x_{827} - 1.88789 x_{828} - 1.362417 x_{829} - 1.526049 x_{830} \\
& & -0.031883 x_{849} - 28.725555 x_{850} - 10.792065 x_{851} - 0.19004 x_{852} - 2.757176 x_{853} \\
& & -12.290832 x_{854} + 717.562256 x_{855} - 0.057865 x_{856} - 3.785417 x_{857} - 78.30661 x_{858} \\
& & -122.163055 x_{859} - 6.46609 x_{860} - 0.48371 x_{861} - 0.615264 x_{862} - 1.353783 x_{863} \\
& & -84.644257 x_{864} - 122.459045 x_{865} - 43.15593 x_{866} - 1.712592 x_{870} - 0.401597 x_{871} \\
& & + x_{880} - 0.946049 x_{898} - 0.946049 x_{916} \\
& \geq & b \equiv 23.387405
\end{array}
\qquad (C)
$$

$$
\begin{array}{lll}
x^*_{826} = 255.6112787181108 & x^*_{827} = 6240.488912232100 & x^*_{828} = 3624.613324098961 \\
x^*_{829} = 18.20205065283259 & x^*_{849} = 174397.0389573037 & x^*_{870} = 14250.00176680900 \\
x^*_{871} = 25910.00731692178 & x^*_{880} = 104958.3199274139 &
\end{array}
$$

♣ **Assume that the uncertain entries of a are 0.1%-accurate approximations of unknown entries in the "true" data $\tilde{a}$, how would this uncertainty affect the validity of the constraint** *evaluated at the nominal solution x^*?*

• **The worst case, over all 0.1%-perturbations of uncertain data, violation of the constraint is as large as 450% of the right hand side!**

• **In the case of** *random* **and** *independent* 0.1% **perturbations of the uncertain coefficients, the statistics of the "relative constraint violation"**

$$
V = \frac{\max[b - \tilde{a}^T x^*, 0]}{b} \times 100\%
$$

also is disastrous:

Prob$\{V > 0\}$	**Prob**$\{V > 150\%\}$	**Mean**(V)
0.50	**0.18**	**125%**

Relative violation of constraint # 372 in `PILOT4`
(1,000-element sample of 0.1% perturbations of the uncertain data)

♣ **We see that** *quite small (just 0.1%) perturbations of "obviously uncertain" data coefficients can make the "nominal" optimal solution x^* heavily infeasible and thus – practically meaningless.*

♣ **In our Case Study, we choose a "perturbation level" ϵ (taking values 1%, 0.1%, 0.01%), and, for every one of the** `NETLIB` **problems, measure the "reliability index" of the nominal solution at this perturbation level, specifically, as follows.**

1. **We compute the optimal solution x^* of the program by** `CPLEX`**.**
2. **For every one of the <u>*in*</u>equality constraints**

$$a^T x \leq b \tag{$*$}$$

   - **we split the right hand side coefficients a_j into "certain" (rational fractions p/q with $|q| \leq 100$) and "uncertain" (all the rest). Let J be the set of all uncertain coefficients of $(*)$.**
   - **we define the** *reliability index* **of $(*)$ as**

$$\frac{a^T x^* + \epsilon \sqrt{\sum\limits_{j \in J} a_j^2 (x_j^*)^2} - b}{\max[1, |b|]} \times 100\% \qquad \textbf{(I)}$$

   Note that *the reliability index is of order of typical violation* **(measured in percents of the right hand side)** *of the constraint, as evaluated at x^*, under independent random perturbations, of relative magnitude ϵ, of the uncertain coefficients*.

3. **We treat the nominal solution as** *unreliable*, **and the problem - as** *bad*, **the level of perturbations being ϵ, if the worst, over the inequality constraints, reliability index is worse than** 5%.

♣ <u>The results</u> of the Diagnosis phase of our Case Study are as follows.

From the total of 90 NETLIB problems we have processed,

• in 27 problems the nominal solution turned out to be unreliable at the largest ($\epsilon = 1\%$) level of uncertainty;

• 19 of these 27 problems are already bad at the 0.01%-level of uncertainty, and in 13 of these 19 problems, 0.01% perturbations of the uncertain data can make the nominal solution more than 50%-infeasible for some of the constraints.

Problem	**Size**[a]	$\epsilon = 0.01\%$		$\epsilon = 0.1\%$		$\epsilon = 1\%$	
		#bad[b]	**Index**[c]	**#bad**	**Index**	**#bad**	**Index**
80BAU3B	2263×9799	**37**	**84**	**177**	**842**	**364**	**8,420**
25FV47	822×1571	**14**	**16**	**28**	**162**	**35**	**1,620**
ADLITTLE	57×97			**2**	**6**	**7**	**58**
AFIRO	28×32			**1**	**5**	**2**	**50**
BNL2	2325×3489					**24**	**34**
BRANDY	221×249					**1**	**5**
CAPRI	272×353			**10**	**39**	**14**	**390**
CYCLE	1904×2857	**2**	**110**	**5**	**1,100**	**6**	**11,000**
D2Q06C	2172×5167	**107**	**1,150**	**134**	**11,500**	**168**	**115,000**
E226	224×282					**2**	**15**
FFFFF800	525×854					**6**	**8**
FINNIS	498×614	**12**	**10**	**63**	**104**	**97**	**1,040**
GREENBEA	2393×5405	**13**	**116**	**30**	**1,160**	**37**	**11,600**
KB2	44×41	**5**	**27**	**6**	**268**	**10**	**2,680**
MAROS	847×1443	**3**	**6**	**38**	**57**	**73**	**566**
NESM	751×2923					**37**	**20**
PEROLD	626×1376	**6**	**34**	**26**	**339**	**58**	**3,390**
PILOT	1442×3652	**16**	**50**	**185**	**498**	**379**	**4,980**
PILOT4	411×1000	**42**	**210,000**	**63**	**2,100,000**	**75**	**21,000,000**
PILOT87	2031×4883	**86**	**130**	**433**	**1,300**	**990**	**13,000**
PILOTJA	941×1988	**4**	**46**	**20**	**463**	**59**	**4,630**
PILOTNOV	976×2172	**4**	**69**	**13**	**694**	**47**	**6,940**
PILOTWE	723×2789	**61**	**12,200**	**69**	**122,000**	**69**	**1,220,000**
SCFXM1	331×457	**1**	**95**	**3**	**946**	**11**	**9,460**
SCFXM2	661×914	**2**	**95**	**6**	**946**	**21**	**9,460**
SCFXM3	991×1371	**3**	**95**	**9**	**946**	**32**	**9,460**
SHARE1B	118×225	**1**	**257**	**1**	**2,570**	**1**	**25,700**

[a] **# of linear constraints (excluding the box ones) plus 1 and # of variables**

[b] **# of constraints with index $> 5\%$**

[c] **The worst, over the constraints, reliability index, in %**

How it works? `NETLIB` Case Study

♣ **We solved the Robust Counterparts of the bad `NETLIB` problems, assuming interval uncertainty in "ugly coefficients" of** *in***equality constraints and** *no uncertainty in equations*. **It turns out that**

- **Reliable solutions do exist, except for 4 cases corresponding to the highest ($\epsilon = 1\%$) perturbation level.**
- **The "price of immunization" in terms of the objective value is surprisingly low: when $\epsilon \leq 0.1\%$, it never exceeds 1% and it is less than 0.1% in 13 of 23 cases. Thus,** *passing to the robust solutions, we gain a lot in the ability of the solution to withstand data uncertainty, while losing nearly nothing in optimality.*

Lecture 3

Robust Conic Quadratic Problems

Theorem [S-Lemma] *Let A, B be symmetric $n \times n$ matrices, and assume that the quadratic inequality*

$$x^T A x \geq 0 \tag{1}$$

is strictly feasible: there exists $\bar{x}^T A \bar{x} > 0$. then the quadratic inequality

$$x^T B x \geq 0 \tag{2}$$

is a consequence of (1) if and only if it is a linear consequence of (1), i.e., if and only if there exists a nonnegative λ such that

$$B \succeq \lambda A \, .$$

The RC of a Quadratically Constrained Problem

$$\min_{x \in \mathbb{R}^n} \left\{ \gamma^T x : x^T A^T A x \leq 2 b^T x + c, \quad \forall (A, b, c) \in \mathcal{U} \right\}, \tag{UQ}$$

with single ellipsoid uncertainty, namely

$$\mathcal{U} = \left\{ (A, b, c) = (A^0, b^0, c^0) + \sum_{\ell=1}^{L} u_\ell (A^\ell, b^\ell, c^\ell) : \|u\|_2 \leq 1 \right\}. \tag{1}$$

Theorem 1 *A robust counterpart of (UQ) with the uncertainty set $\mathcal{U}$ given by (1) is equivalent to the SDP problem*

$$\min_{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}} \gamma^T x$$

$$s.t. \quad \left(\begin{array}{c|ccc|c} c^0 + 2x^T b^0 - \lambda & \frac{1}{2} c^1 + x^T b^1 & \cdots & \frac{1}{2} c^L + x^T b^L & (A^0 x)^T \\ \hline \frac{1}{2} c^1 + x^T b^1 & \lambda & & & (A^1 x)^T \\ \vdots & & \ddots & & \vdots \\ \frac{1}{2} c^L + x^T b^L & & & \lambda & (A^L x)^T \\ \hline A^0 x & A^1 x & \cdots & A^L x & I_m \end{array} \right) \succeq 0, \quad \textit{(RUQ)}$$

Approximate S-Lemma

- Let $Q_1, \dots, Q_L$ be positive semidefinite matrices with positive definite sum; let A be a symmetric matrix, and let a be a vector. Let

$$\mathrm{Opt}(\rho) = \max_x \left\{ x^T A x + 2a^T x : x^T Q_\ell x \le \rho^2,\ \ell = 1, \dots, L \right\}$$

- In general, computing $\mathrm{Opt}(\rho)$ is NP-hard. However, we can use Semidefinite Relaxation scheme to bound $\mathrm{Opt}(\rho)$ from above:

$$\begin{aligned}
\mathrm{Opt}(\rho) &= \max_x \{x^T A x + 2a^T x : x^T Q_\ell x \le \rho^2, \\
&\qquad\qquad \ell = 1, \dots, L\} \\
&= \max_{x,t} \{x^T A x + 2ta^T x : x^T Q_\ell x \le \rho^2, \\
&\qquad\qquad \ell = 1, \dots, L,\ t^2 \le 1\}
\end{aligned}$$

$$y = (x, t),\ Y = yy^T = \begin{pmatrix} x \\ t \end{pmatrix} (x, t) = \begin{bmatrix} xx^T & tx \\ tx & t^2 \end{bmatrix}$$

$$\underset{\underset{\mathrm{tr}\, A x x^T}{\shortparallel}}{x^T A x} + 2ta^T x = \mathrm{tr} \begin{bmatrix} A & a^T \\ a & 0 \end{bmatrix} \begin{bmatrix} xx^T & tx \\ tx & t^2 \end{bmatrix}$$

$$Y = \begin{bmatrix} xx^T & tx \\ tx & t^2 \end{bmatrix}$$

$$\leq \max_Y \left\{ Tr\left(\underbrace{\left[\begin{array}{c|c} A & a^T \\ \hline a & \end{array}\right]}_{R} Y \right) : \begin{array}{l} Tr\left(\overbrace{\left[\begin{array}{c|c} Q_\ell & \\ \hline & \end{array}\right]}^{R_\ell} Y \right) \leq \rho^2, \\ \qquad \ell = 1, \dots, L, \\ Tr\left(\underbrace{\left[\begin{array}{c|c} & \\ \hline & 1 \end{array}\right]}_{R_0 = dd^T} Y \right) \leq 1, \\ Y \succeq 0 \end{array} \right\}$$

$$\equiv SDP(\rho) \qquad (1)$$

Theorem 4 *The set* R_ρ *of* (x,λ) *satisfying* $\lambda \geq 0$ *and*

$$SDP(\rho)\ \begin{bmatrix} c[x]-\sum_{k=1}^K \lambda_k & (-b_\rho[x]-d_\rho)^T & a[x]^T \\ -b_\rho[x]-d_\rho & \sum_{k=1}^K \lambda_k Q_k & -A_\rho[x]^T \\ a[x] & -A_\rho[x] & I_M \end{bmatrix} \succeq 0$$

is an approximate robust counterpart of the set $\mathcal{X}_\rho$ *of robust feasible solutions of* (UQC)*, i.e. (12) holds:*

$$u^T Q_k u \leq 1\,,\ k=1,\ldots,K\ \Rightarrow \tag{12}$$

$$u^T A_\rho[x]^T A_\rho[x]u + 2u^T(A_\rho[x]^T a[x] - b_\rho[x] - d_\rho) \leq c[x] - a[x]^T a[x].$$

The level of conservativeness of $\mathcal{X}_\rho$ is

$$\Omega \leq 9.19\sqrt{\log K} \qquad (K \text{ is the number of ellipsoids})$$

i.e.

$$\mathrm{opt}(\rho) \leq SDP(\rho) \leq \mathrm{opt}(\Omega\rho)\,.$$

For box uncertainty set $\|u\|_\infty| \leq\ |$

$$\Omega \leq \frac{\pi}{2}\,.$$

Robust Semidefinite Optimization

Uncertain Semidefinite Programming

- For an Uncertain Semidefinite problem

  (USD):

$$\left\{ \min_x \left\{ c^T x : \mathcal{A}[x] \equiv A_0 + \sum_{j=1}^{n} x_j A_j \succeq 0 \right\} : [A_0, \ldots, A_n] \in \mathcal{U} \right\}$$

  the RC can be NP-hard already in the simplest cases when $\mathcal{U}$ is a box or an ellipsoid.

- The strongest generic result on tight computationally tractable approximations of Uncertain Semidefinite constraints deals with the case of structured norm-bounded perturbations

$$\mathcal{U} = \left\{ \begin{array}{l} \mathcal{A}[x] = \mathcal{A}^n[x] + \sum_{\ell=1}^{L} [L_\ell^T \Delta_\ell R_\ell[x] + R_\ell^T[x] \Delta_\ell^T L_\ell] : \\ \Delta_\ell \in \mathbb{R}^{d_\ell \times d_\ell}, \ \|\Delta_\ell\| \le \rho\,, \ \Delta_\ell = \delta_\ell I_{d_\ell}, \ \ell \in I^S \end{array} \right\}$$

  - $A^n[x]$: symmetric $m \times m$ matrix affinely depending on x
  - $R_\ell[x] : d_\ell \times m$ matrix affinely depending on x

(RC):

$$\left\{ \begin{array}{l} \mathcal{A}^n[x] + \sum\limits_{\ell=1}^{L} [L_\ell^T \Delta_\ell R_\ell[x] + R_\ell^T[x] \Delta_\ell^T L_\ell] \succeq 0 \\ \forall \{\Delta_\ell\}_{\ell=1}^L : \Delta_\ell \in \mathbb{R}^{d_\ell \times d_\ell},\ \|\Delta_\ell\| \le \rho,\ \Delta_\ell = \delta_\ell I_{d_\ell},\ \ell \in I^S \end{array} \right\}$$

Theorem 1 (Ben-Tal, Nemirovski, Roos '02) The Robust Counterpart (RC) of an Uncertain LMI with structured norm-bounded perturbations admits a ϑ-tight approximation which is an explicit semidefinite program. The tightness factor ϑ depends solely on the maximum of sizes d_ℓ of scalar perturbation blocks

$$\mu = \max\{d_\ell : \ell \in I^S\}$$

specifically,

$$\vartheta \le \frac{\pi\sqrt{\mu}}{2} .$$

If there are no scalar perturbation blocks ($I^S = \phi$), or all scalar perturbation blocks are of size 1 ($d_\ell = 1$), then

$$\vartheta = \frac{\pi}{2} .$$

In the case of a single perturbation block ($L = 1$), $\vartheta = 1$, i.e. (RC) is equivalent to an explicit single LMI.

Unstructured Norm-Bounded Perturbations

Definition 1 *We say that uncertain LMI*

$$A_\zeta(y) \equiv A^n(y) + \sum_{\ell=1}^{L} \zeta_\ell A_\ell(y) \succeq 0\,, \tag{1}$$

is with unstructured norm-bounded perturbations, if

1. *The perturbation set $\mathcal{Z}$ is the set of all $p \times q$ matrices ζ with the usual matrix norma $\|\cdot\|_{2,2}$ not exceeding 1;*

2. *$\mathcal{A}_\zeta(y)$ can be represented as*

$$\mathcal{A}_\zeta(y) \equiv \mathcal{A}^n(y) + \left[L^T(y)\zeta R(y) + R^T(y)\zeta^T L(y)\right]\,, \tag{2}$$

where both $L(\cdot)$, $R(\cdot)$ are affine and at least one of these matrix-valued functions is independent of y.

We have proved the following statement:

Theorem: 1 *The RC*

$$\mathcal{A}^n(y)+L^T(y)\zeta R+R^T\zeta^T L(y) \succeq 0,\ \forall\,(\zeta \in \mathbb{R}^{p\times q} : \|\zeta\|_{2,2} \leq \rho) \tag{3}$$

of uncertain LMI (1) with unstructured norm-bounded uncertainty (2) (where, w.l.o.g., we assume that $R \neq 0$ *and* R *is independent of* y*) and uncertainty level* ρ *can be represented equivalently by the LMI*

$$\left[\begin{array}{c|c} \lambda I_p & \rho L(y) \\ \hline \rho L^T(y) & \mathcal{A}^n(y) - \lambda R^T R \end{array}\right] \succeq 0 \tag{4}$$

in variables y, λ*.*

Example: Truss Topology Design

A *truss* is a mechanical construction, like railroad bridge, electric mast or Eiffel Tower, comprised of thin elastic *bars* linked with each other at *nodes*. Some of the nodes are partially or completely fixed. An *external load* is a collection of external forces acting at the nodes; under such a load, the nodes slightly move, thus causing elongations and compressions in bars, until the construction achieves an equilibrium, where the tensions caused in the bars as a result of their deformations compensate the external forces. The *compliance* is the potential energy capacitated in the truss at the equilibrium as a result of deformations of the bars.

A mathematical model of the outlined situation is as follows.

$$TTD \quad \left[\begin{array}{c} \min\limits_{x \in R^M,\, t \in R^N} f^T x \\ A(t)x = f \\ t \in T \end{array} \right]$$

where

$$A(t) = \sum_{i=1}^{N} t_i b_i b_i^T$$

is the ***stiffness matrix*** of the truss;
b_i is a vector given in terms of the material property of bar i (Young modulus) and the "nominal" (i.e. in the unloaded truss) position of the nodes.

$$T = \left\{ t \in \mathbb{R}^N : \quad t \geq 0, \quad \sum t_j \leq w \right\}$$

t_i represent *volume* of bar i.

$f^T x$ is the *compliance* — the potential energy capacitated in the truss at the equilibrium due to *external force* $f \in \mathbb{R}^M$ acting on the nodes.

An equivalent formulation of a TTD problem is as follows:

$$\begin{aligned}\mathrm{Compl}_f(A) &= \min_x \left\{ \frac{1}{2} f^T x \mid A(t)x = f \right\} \\ &= \max_x \left\{ f^T x - \frac{1}{2} x^T A(t) x \right\}\end{aligned}$$

(TTD) is equivalent to

$$\min_{t\in T} \mathrm{Compl}_f(A) = \min_{t\in T} \max_x \left\{ f^T x - \frac{1}{2} x^T A(t) x \right\}$$

$$\Leftrightarrow \quad \min_{\substack{t\in T\\ \tau\in R}} \tau$$

$$f^T x - \frac{1}{2} x^T A(t) x \preceq \tau \quad \forall x$$

$$\Leftrightarrow \quad \min_{t,\tau} \{ \tau \mid 2\tau s^2 - 2s\, f^T (sx) + \frac{1}{2} (xs)^T A(t)(sx) \succeq 0 \quad \forall x, s \}$$

$$\Leftrightarrow \quad \min_{t,\tau} \{ \tau \mid 2\tau s^2 - 2s\, f^T v + \frac{1}{2} v^T A(t) v \succeq 0, \quad \forall v, s \}$$

$$\Leftrightarrow \quad \min_{t,\tau} \left\{ \tau \mid \begin{pmatrix} 2\tau & -f^T \\ -f & A \end{pmatrix} \succeq 0 \right\}$$

$$\Leftrightarrow \quad \boxed{\min_{t\in T} \left\{ \tau \mid \begin{pmatrix} 2\tau & f \\ f & A \end{pmatrix} \succeq 0 \right\}} \quad \text{(SDP)}$$

$$\min_{t,T} \left\{ \tau \,\middle|\, \begin{pmatrix} 2\tau & f^T \\ f & A(t) \end{pmatrix} \succeq 0, \quad \forall f \in \mathcal{U} \right\}$$

f_N

Uncertainty set of load vector:

$$\mathcal{U} = \{ f = f_N + B\zeta \mid \|\zeta\|_2 \leq \rho \}$$

We need the RC of the LMI:

$$A_\zeta = \left[\begin{array}{c|c} 2\tau & \zeta^T B^T + f_N^T \\ \hline f_N + B\zeta & A(t) \end{array} \right] \succeq 0, \ \forall \zeta : \|\zeta\|_2 \leq \rho \,. \tag{5}$$

Now

$$\mathcal{A}_\zeta = \mathcal{A}^N(\tau, t) + [L^T \zeta R + R^T \zeta L]$$

where

$$L = (0_n, B^T) \quad R = (1, 0_n^T)$$

$$A^N(\tau, t) = \begin{bmatrix} 2\tau & f_N^T \\ f_N & A(t) \end{bmatrix}$$

$L^T \zeta R = \begin{pmatrix} 0 & 0 \\ B\zeta & 0 \end{pmatrix}$ $\quad R^T \zeta L = \begin{pmatrix} 0 & \zeta^T B^T \\ 0 & 0 \end{pmatrix}$

$R^T R = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

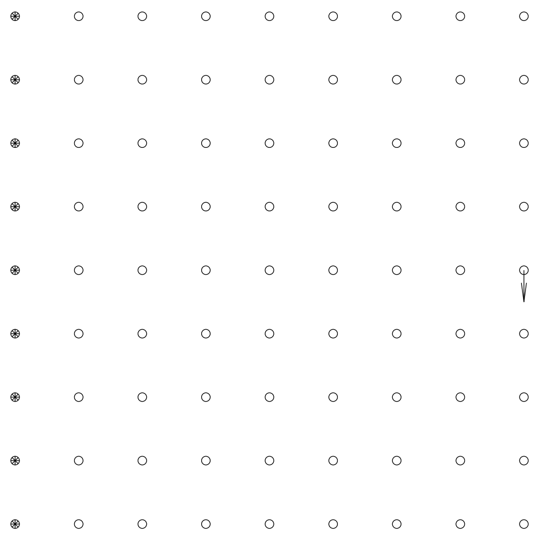
By Theorem 1, (5) is equivalent to the single LMI

$$\left[\begin{array}{c|c} \lambda I & \rho L \\ \hline \rho L^T & \mathcal{A}^N - \lambda R^T R \end{array}\right] \succeq 0 \tag{6}$$

which here becomes

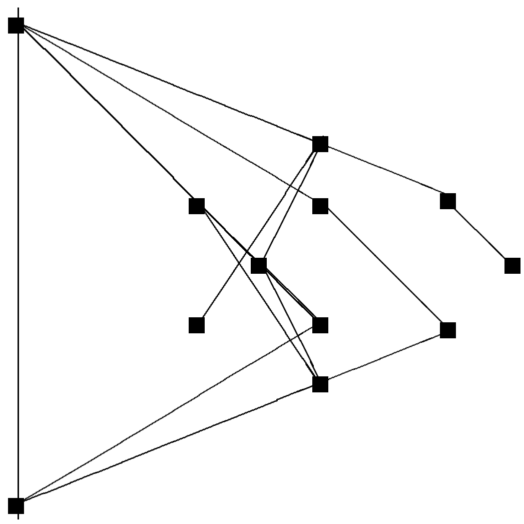
$$\left[\begin{array}{c|cc} \lambda I & 0 & \rho B^T \\ \hline 0 & 2\tau - \lambda & f_N^T \\ \rho B & f_N & A(t) \end{array}\right] \succeq 0 \tag{7}$$

Example: **Assume we are designing a planar truss – a cantilever; the 9×9 nodal structure and the only load of interest f^* are as shown on the picture:**



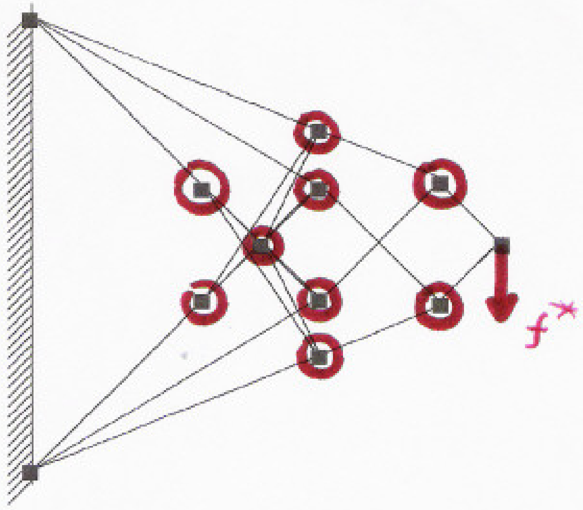
9×9 ground structure and the load of interest

The optimal single-load design yields a nice truss as follows:



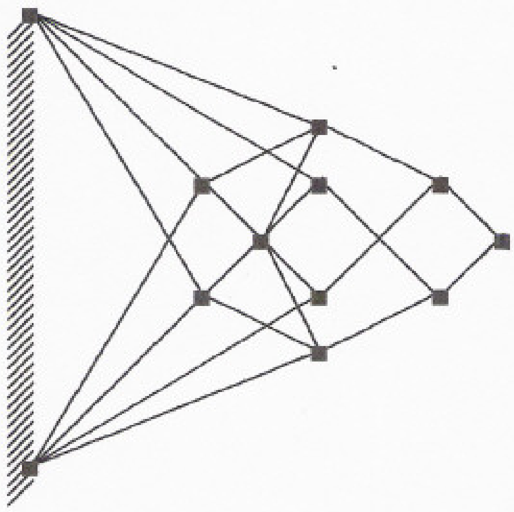
Optimal cantilever (single-load design)
the compliance is 1.000

• **Passing from the single-load to the robust design, we modify the result as follows:**



Optimal cantilever (single-load design)

$\Downarrow$



"Robust" cantilever

Compliances	**Design**	
	Single-load	**Robust**
Compliance w.r.t. f^*	**1.000**	**1.0024**
max **compliance w.r.t. loads** f**:** $\| f \| \leq 0.1 \| f^* \|$	**32000**	**1.03**

A Guide to Deriving Robust Counterparts

Consider the uncertain constraint

$$f(a,x) \leq 0\,, \tag{5}$$

where $x \in I\!R^n$ is the optimization variable, $f(.,x)$ is concave for all $x \in I\!R^n$, and $a \in I\!R^m$ is an uncertain vector, which is only known to reside in a set U. The robust counterpart of (5) is then

$$(RC) \qquad\qquad f(a,x) \leq 0\,, \quad \forall\, a \in U\,, \tag{6}$$

where the uncertainty set U is modeled as follows:

$$U = \{a = a^0 + A\zeta \mid \zeta \in Z \subset I\!R^L\}\,.$$

Here $a^0 \in I\!R^m$ is the so-called "nominal value", the matrix A is given column wise: $A = (a^1 a^2, \ldots, a^L) \in I\!R^{m \times L}$, ζ is called the vector of "primitive uncertainties", and Z is a given nonempty, convex and compact set, with $0 \in ri(Z)$.

With this formulation, it is required to determine the value of x before the actual realization of a is available ("here and now" decisions).

Definition 1 The nominal vector a^0 is called *regular* if $a^0 \in ri(\text{dom}\, f(.,x))$, $\forall\, x$.

Note that when a^0 is regular and since $0 \in ri(Z)$, then the following holds:

$$ri(U) \cap ri(\text{dom}\, f)(.,x)) \neq \emptyset\,,\ \forall\, x\,. \tag{7}$$

The robust inequality (RC) can be rewritten as

$$\max_{a \in U} f(a,x) \le 0\,. \tag{8}$$

A General Principle

The dual of the optimization problem in (8) has the general form

$$\min\{g(b,x) \mid b \in Z(x)\}\,. \tag{9}$$

Under suitable convexity and regularity conditions on $f(.,x)$ and U (such as (7)) strong duality holds between the maximization problem in (8) and (9); hence, x is robust feasible if and only if

$$\min\{g(b,x) \mid b \in Z(x)\} \le 0\,. \tag{10}$$

So finally, x is robust feasible for (6) *if and only if* x and b solve the system

$$\begin{cases} g(b,x) \le 0 \\ b \in Z(x)\,. \end{cases} \tag{11}$$

In case strong duality does not hold, we still have (by weak duality) that (10) implies (9). Hence, whenever x and some b solve (11), then x satisfies (6), i.e., it is robust feasible.

Fenchel Duality

The primal problem is given as follows:

$$(P) \qquad \inf\{f(x) - g(x) | x \in \text{dom}(f) \cap \text{dom}(g)\}\,.$$

The Fenchel dual of (P) is given by:

$$(D) \qquad \sup\{g_*(y) - f^*(y) | y \in \text{dom}(g_*) \cap \text{dom}(f^*)\}\,.$$

The Fenchel duality theorem is stated next.

Theorem 5 If $ri(\text{dom}(f)) \cap ri(\text{dom}(g)) \neq \emptyset$ then the optimal values of (P) and (D) are equal and the maximal value of (D) is attained.

If $ri(\text{dom}(g_*)) \cap ri(\text{dom} f^*)) \neq \emptyset$ then the optimal values of (P) and (D) are equal and the minimum value of (P) is attained. $\square$

Note that since $f^{**} = f$ and $g_{**} = g$, we have that the dual of (D) is (P).

The next basic result gives an equivalent reformulation for (RC) which can be used extensively to derive tractable RCs.

Theorem 2 Let a^0 be regular. Then the vector $x \in I\!R^n$ satisfies (RC) *if and only if* $x \in I\!R^n$, $v \in I\!R^m$ satisfy the single inequality

$$(FRC) \qquad (a^0)^T v + \delta^*(A^T v|Z) - f_*(v,x) \leq 0\,, \tag{12}$$

in which the support function δ^* and the partial concave conjugate function f_* are defined in (3) and (1), respectively.

Proof Using the definition of indicator functions (2) and using Fenchel duality, we have

$$F(x) := \max_{a \in U} f(a,x) \tag{13}$$

$$= \max_{a \in R^m} \{f(a,x) - \delta(a|U)\} \tag{14}$$

$$= \min_{v \in R^m} \{\delta^*(v|U) - f_*(v,x)\,, \tag{15}$$

(strong duality holds by regularity) where

$$f_*(v,x) = \inf_{a \in I\!R^m} \{a^T v - f(a,x)\}\,,$$

and

$$\delta^*(v|U) = \sup_{\zeta \in Z} \{a^T v \mid a = a_0 + A\zeta\} \tag{16}$$

$$= (a^0)^T v + \sup_{\zeta \in Z} v^T A\zeta \tag{17}$$

$$= (a^0)^T v + \delta^*(A^T v|Z)\,. \tag{18}$$

By (18) and (15), $F(x) \leq 0$ is exactly condition (12).

Remarks

1. In (FRC), the computation involving f are completely independent from those involving Z.

2. To derive (FRC) we did not assume $f(a,x)$ to be convex in x. However, if $f(a,x)$ is convex in x, $\forall\, a \in U$, then $f_*(v,x)$ is concave in (v,x), so then (FRC) is a convex inequality in (v,x).

3. It is interesting to observe that robustifying a nonlinear constraint may have a “convexification effect”. This is illustrated in the next nominal constraint which is nonconvex, but whose robust counterpart is convex. We consider the following robust counterpart:

$$f(a,x) := \sum_{i=1}^{m} a_i f_i(x) \le b \quad \forall\, a \in U\,,$$

where $U = \{a \in I\!R^m | \|a - a^0\|_\infty \le \rho\}$, and let $\rho + a_i^0 \ge 0$, $i = 1, \ldots, m$. We assume that $f_i(x)$, $i = 1, \ldots, m$, are convex and $f_i(x) \ge 0$, $\forall\, x$. Suppose (some of) the nominal values a_i^0 are negative, which means that the nominal inequality

$$\sum_{i=1}^{m} a_i^0 f_i(x) \le b$$

may not be convex. However, in this case the (FRC)

$$\sum_{i=1}^{m} (a_i^0 + \rho) f_i(x) \le b$$

is indeed a convex inequality.

Conjugate Functions, Support Functions and Fenchel Duality

We outline some basic results on conjugate functions, support functions and Fenchel duality.

We start with some well-known results on conjugate functions. First, note that f^* is closed convex, and g_* is closed concave; moreover, $f^{**} = f$ and $g_{**} = g$. It is well-known that for $a > 0$

$$(af)^*(y) = af^*\left(\frac{y}{a}\right)$$

and for $\tilde{f}(x) = f(ax)$, $a > 0$, we have $\qquad \tilde{f}^*(y) = f^*\left(\frac{y}{a}\right)$

and for $\tilde{f}(x) = f(x - a)$ we have $\qquad \tilde{f}^*(y) = f^*(y) + ay\,.$

We frequently use the following sum-rules for conjugate functions.

Lemma 1 Assume that f_i, $i = 1, \ldots, m$, are convex, and the intersection of the relative interiors of the domains of f_i, $i = 1, \ldots, m$, is nonempty, i.e., $\cap_{i=1}^m ri(\text{dom } f_i) \neq \emptyset$. Then

$$\left(\sum_{i=1}^m f_i\right)^*(s) = \inf_{\{v^i\}_{i=1}^m} \left\{ \sum_{i=1}^m f_i^*(v^i) \,\middle|\, \sum_{i=1}^m v_i = s \right\},,$$

and the inf is attained for some v_i, $i = 1, \ldots, m$.

In particular, let $S_1, \ldots, S_k$ be closed convex sets, such that $\cap_i ri(S_i) \neq \emptyset$, and let $S = \cap_{i=1}^k S_i$. Then

$$\delta^*(y|S) = \min \left\{ \sum_{i=1}^k \delta^*(v^i|S_i) \,\middle|\, \sum_{i=1}^k v^i = y \right\}.$$

$\square$

We now state three results which are used to derive tractable robust counterparts. The first lemma relates the conjugate of the adjoint function ($f^{\diamondsuit}_{(x)} = xf\left(\frac{1}{x}\right)$, $x > 0$) to the conjugate of the original function. Note that $f^{\diamondsuit}(x)$ is convex if $f(x)$ is convex. The next proposition can be used in cases where f^* is not available in closed form, but $(f^{\diamondsuit})^*$ is available as such.

Lemma 2 For the conjugate of a function $f : I\!R_+ \longrightarrow I\!R$ and the conjugate of its adjoint $f^{\diamondsuit}$, we have $\quad f^*(s) = \inf\{y \in I\!R : (f^{\diamondsuit})^*(-y) \leq -s\}$. $\square$

The next proposition can be used in cases where f^{-1} is not available in closed form, but $(f^{-1})^*$ is available as such.

Lemma 3 Let $f : I\!R \longrightarrow I\!R$ be strictly increasing and concave. Then, for all $y > 0$

$$(f^{-1})^*(y) = -yf_*\left(\frac{1}{y}\right) = -(f_*)^{\diamondsuit}(y)\,. \qquad \square$$

The next proposition gives a useful result related to the conjugate of a function after linear transformations.

Lemma 4 Let A be a linear transformation from $I\!R^n$ to $I\!R^m$. Assume there exists an x such that $Ax \in ri(\text{dom}\ g)$. Then, for each convex function g on $I\!R^m$, one has

$$(gA)^*(z) = \inf_y\{g^*(y) \mid A^T y = z\}\,.$$

where for each z the infimum is attained, and where the function gA is defined by

$$(gA)(x) = g(Ax)\,. \qquad \square$$

$f(t)$	$f^*(s)$ (domain)
t	$0\ (s = 1)$
t^2	$s^2/4\ (s \in I\!R)$
$\lvert t\rvert^p/p\ (p > 1)$	$\lvert s\rvert^q/q\ (s \in I\!R)$
$-t^p/p\ (t \geq 0,\ 0 < p < 1)$	$-(-s)^q/q\ (s \leq 0)$
$-\log t\ (t > 0)$	$-1 - \log(-s)\ (s < 0)$
$t \log t\ (t > 0)$	$e^{s-1}\ (s \in I\!R)$
e^t	$\begin{cases} s\log s - s\ (s > 0) \\ 0\ (s = 0) \end{cases}$
$\log(1 + e^t)$	$\begin{cases} s\log s + (1-s)\log(1-s)\ (0 < s < 1) \\ 0\ (s = 0, 1) \end{cases}$
$\sqrt{1 + t^2}$	$-\sqrt{1 - s^2}\ (-1 \leq s \leq 1)$

Some examples for f, with conjugate f^*. The parameters p and q are related as follows: $1/p + 1/q = 1$.

Example (Transformed uncertainty region). Suppose $f(\tilde{a}, x) = \tilde{a}^T x - \beta$ and the uncertainty region $\tilde{U}$ is defined as follows:

$$\tilde{U} = \left(\tilde{a} \mid h(\tilde{a}) = h_1(\tilde{a}_1), \ldots, h_m(\tilde{a}_m))^T \in U\right\} ,$$

where $h_i(\cdot)$ is convex for each i, and moreover we assume that h_i^{-1} exists for all i. By substituting $\tilde{a} = h^{-1}(a)$, we obtain for (RC)

$$h^{-1}(a)^T x \leq \beta \quad \forall\, a \in U \,.$$

The corresponding (FRC) becomes:

$$(a^0)^T v + \delta^*(A^T v|Z) + \sum_{i=1}^{n} x_i \left((h_i)^{-1}\right)_* (v_i/x_i) \leq \beta \,.$$

Using the result in the conjugate of h^{-1}, we finally get:

$$(a^0)^T v + \delta^*(A^T v|Z) + \sum_{i=1}^{n} v_i (h_i)^*(x_i/v_i) \leq \beta \,.$$

The result shows that even if we cannot compute a closed form for h_i^{-1}, we can still construct the robust counterpart. As an example, take $h_i(t) = -(t + \log t)$. There is no closed form for h_i^{-1}, but there is an explicit formula for the conjugate of h_i:

$$h_i^*(s) = -1 - \log(-(s+1)) \,. \qquad \square$$

Finally, we give some examples in which $f(a, x)$ cannot be written as $f(a)^T g(x)$, but still $f_*(v, x)$ can be computed.

Nonconcave uncertainty

So far, it was assumed that $f(a,x)$ is concave in a for each x. If this assumption does not hold, one may try to reformulate the (RC) problem to regain convexity. We describe different ways for such a reformulation:

Reparametrizing and computing convex hull. Suppose $f(a,x)$ can be written as $f(a)^T g(x)$, where $f(a)$ is not necessarily concave and/or $g(x)$ may attain positive and negative values. Let us, for ease of notation, also assume that $m = L$, $A = 1$, and $a^0 = 0$. By the parametrization $b = f(a)$, (RC) becomes

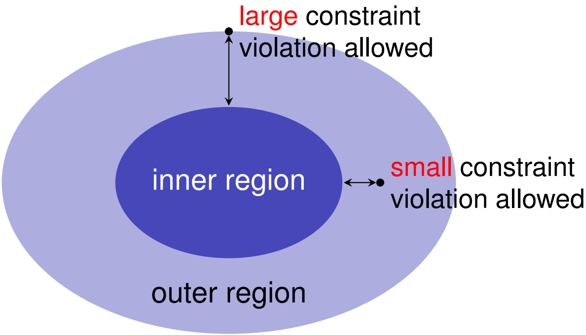
$$b^T g(x) \leq 0 \quad \forall\, (a,b) \in \bar{U}\,,$$

where $\bar{U} = \{(a,b) \mid a \in U,\ b = f(a)\}$. Since the left-hand side of this constraint is linear in the uncertain parameter b, we may replace $\bar{U}$ by conv$(\bar{U})$. Hence, if we can compute conv$(\bar{U})$, we can apply Theorem 2.

A well-known example is quadratic uncertainty and an ellipsoidal uncertainty region. For this case, it has been proved that conv$(\bar{U})$ can be formulated as an LMI. The resulting robust counterpart is therefore a system of LMIs.

Globalized Robust Optimization

Idea Globalized Robust Optimization



- Inner uncertainty region: full feasibility as in classical RO
- Allow restricted constraint violations for parameters in the outer uncertainty region
- Allowed violation depends on distance to the inner region

Globalized Robust Counterpart

$$f(a,x) \le \min_{\tilde{a} \in U_1} \phi(a,\tilde{a}) \qquad \forall a \in U_2 \tag{GRC}$$

Uncertainty regions

- U_1: inner uncertainty region (convex and compact)
- U_2: outer uncertainty region (convex)
- Z_1 and Z_2 corresponding primitive uncertainty regions

Violation measure

$\phi(a,\tilde{a})$: "distance" between a and $\tilde{a}$

- nonnegative and jointly convex in both arguments
- $\phi(a,a) = 0$ for all $a \in \mathbb{R}^m$
- Examples:
  - Norm based: $\phi(a,\tilde{a}) = \theta \cdot \|a - \tilde{a}\|$ or $\phi(a,\tilde{a}) = \theta \cdot \|a - \tilde{a}\|^2$
  - Phi-divergence distances

Example: Simple Case

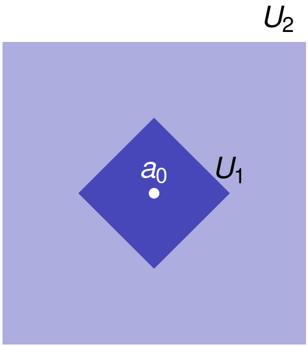
**Linear constraint,
violation linear in distance**

$$a^T x - b \leq \min_{\tilde{a} \in U_1} \theta \cdot \|a - \tilde{a}\|_2 \qquad \forall a \in U_2$$

with

$$U_i = \left\{ a = a_0 + A\zeta \,\middle|\, \zeta \in Z_i \right\} \quad i = 1, 2$$
$$Z_1 = \left\{ \zeta \in \mathbb{R}^L \,\middle|\, \|\zeta\|_1 \leq \rho_1 \right\}$$
$$Z_2 = \left\{ \zeta \in \mathbb{R}^L \,\middle|\, \|\zeta\|_\infty \leq \rho_2 \right\}$$

where $0 < \rho_1 < \rho_2$.



Example: Simple Case (continued)

Original GRC

$$a^T x - b \leq \min_{\tilde{a} \in U_1} \theta \cdot \|a - \tilde{a}\|_2 \qquad \forall a \in U_2$$

GRC theorem

$$a_0^T(v+w) + \delta^*(A^T v \mid Z_1) + \delta^*(A^T w \mid Z_2) - f_*(v+w, x) + \phi^*(v; -v) \leq 0$$

Tractable GRC

$$\begin{cases} a_0^T(v+w) + \rho_1 \|A^T v\|_\infty + \rho_2 \|A^T w\|_1 - b + 0 \leq 0 \\ v + w = x \\ \|v\|_2 \leq \theta \end{cases}$$

Theorem 1 *Let $f(.,x)$ be a concave function in $I\!R^m$ for all $x \in I\!R^n$, and $\phi : I\!R^m \times I\!R^m \to I\!R$ a convex and nonnegative function for which $f(a,a) = 0$ for all $a \in I\!R^m$. Let the set $Z_1 \subset I\!R^L$ be nonempty, convex, and compact with $0 \in \mathrm{ri}(Z_1)$, let Z_2 be a convex set such that $Z_1 \subset Z_2$, and let the sets U_1 and U_2 be defined by*

$$U_i = \{a = a_0 + A\zeta \mid \zeta \in Z_i\}, \qquad i = 1, 2,$$

where $a_0 \in I\!R^m$ and $A \in I\!R^{m\times L}$.

Then the vector $x \in I\!R^n$ satisfies the GRC

$$f(a,x) \le \min_{a' \in U_1} \phi(a,a') \qquad \forall\, a \in U_2, \tag{1}$$

if and only if there exist $v, w \in I\!R^m$ that satisfy the single inequality

$$a_0^T(v+w) + \delta^*(A^T v \mid Z_1) + \delta^*(A^T w \mid Z_2) - f_*(v+w,x) + \phi^*(v;-v) \le 0\,. \tag{2}$$

Simplification of the Robust Counterparts for Conic Quadratic Optimization

Inhomogeneous S-lemma

Lemma 1 *Let A, D be symmetric matrices of the same size, and let the quadratic form $z^T Az + 2b^T z + c$ be strictly positive at some point. Then the implication*

$$z^T Az + 2b^T z + c \geq 0 \Rightarrow z^T Dz + 2e^T z + f \geq 0$$

holds true if and only if

$$\exists \lambda \geq 0 : \left(\begin{array}{c|c} D - \lambda A & e - \lambda b \\ \hline e^T - \lambda b^T & f - \lambda c \end{array} \right) \succeq 0. \tag{3}$$

$\square$

However, although theoretically speaking (6) is tractable, computationally speaking it is not!

Aim: to show that the RC is a conic quadratic programming problem.

Core problem: find a tractable robust counterpart for:

$$a^T Da + 2a^T d \leq \gamma \quad \forall\, a : a^T Aa \leq \rho^2.$$

Simultaneously diagonalizability

Definition 2 *Real symmetric matrices A and B are called simultaneously diagonalizable if there exists a nonsingular matrix S such that both S^TAS and S^TBS are diagonal.*

The following theorem, proved by Uhlig (1972), gives a sufficient condition for simultaneously diagonalizability.

Theorem 3 *Real symmetric matrices A and B can be simultaneously diagonalized if $\{x|x^TAx = 0\} \cap \{x|x^TBx = 0\} = \{0\}$.*

Note that if one of the matrices A and B is positive definite, then these two matrices can be simultaneously diagonalized.

How to calculate S?

Suppose B is positive definite, then:

1. compute the Cholesky factorization $B = GG^T$
2. compute $C = G^{-1}AG^{-T}$
3. use symmetric QR-algorithm to compute the Schur decomposition $Q^TCQ = \operatorname{diag}(\alpha_1, \cdots, \alpha_n)$
4. set $S = G^{-T}Q$.

We now have $S^TBS = I$ and $S^TAS = \operatorname{diag}(\alpha_1, \cdots, \alpha_n)$, i.e., matrices A and B can be diagonalized by S.

Nonconvex quadratic problems (with one constraint)

We consider the following optimization problem:

$$(P) \quad \begin{cases} \min & \frac{1}{2} z^T D z + e^T z \\ \text{s.t.} & \frac{1}{2} z^T A z + b^T z + c \le 0, \end{cases}$$

where $D, A \in \mathbb{R}^{n\times n}$ are symmetric, $z, b, e \in \mathbb{R}^n$, and $c \in \mathbb{R}$. We assume that A and D can be simultaneously diagonalized: $\exists$ nonsingular S such that

$$S^T A S = \operatorname{diag}(\alpha_1, \cdots, \alpha_n) \quad \text{and} \quad S^T D S = \operatorname{diag}(\delta_1, \cdots, \delta_n).$$

Using the change of variables $z = Sx$ and change of parameters $\beta = S^T b$, $\epsilon = S^T e$, and by setting $y_i = \frac{1}{2} x_i^2$, we can rewrite problem (P) as follows:

$$(P_1) \quad \begin{cases} \min & \delta^T y + \epsilon^T x \\ \text{s.t.} & \alpha^T y + \beta^T x + c \le 0 \\ & \frac{1}{2} x_i^2 - y_i = 0, \quad \forall i. \end{cases}$$

We consider the following convex relaxation of (P_1):

$$(P_2) \quad \begin{cases} \min & \delta^T y + \epsilon^T x \\ \text{s.t.} & \alpha^T y + \beta^T x + c \le 0 \\ & \frac{1}{2} x_i^2 - y_i \le 0 \quad \forall i. \end{cases}$$

The following theorem shows the equivalence of (P_1) and (P_2).

Theorem 4 *Assume that there exists a strictly feasible solution to (P). Then, if there exists an optimal solution to (P_2) then there exists an optimal solution to (P_1).*

Hence, the (probably nonconvex) quadratic problem (P) can be solved by solving a convex quadratic optimization problem (P_2).

Example of nonconvex quadratic problems

.

$$\min -\frac{1}{2}z_1^2 - \frac{1}{2}z_2^2 - z_2 \quad \text{s.t.} \quad z_1^2 + \frac{1}{2}z_2^2 + z_2 \leq 1.$$

The corresponding problem (P_2) is:

$$\min -y_1 - y_2 - x_2 \quad \text{s.t.} \quad 2y_1 + y_2 + x_2 \leq 1, \quad \frac{1}{2}x_1^2 - y_1 \leq 0 \quad \frac{1}{2}x_2^2 - y_2 \leq 0.$$

The optimal objective value of this problem is -1, and the optimal solution $x_1^* = 0, x_2^* = 0, y_1^* = 0, y_2^* = 1$, and the KKT multipliers $u = 1, \mu_1 = 1, \mu_2 = 0$. This solution clearly does not satisfy $y_i^* = \frac{1}{2}(x_i^*)^2$.

However, such a solution is given by $\bar{x}_1 = 0, \bar{x}_2 = -1 \pm \sqrt{3}, \bar{y}_1 = 0, \bar{y}_2 = 2 \mp \sqrt{3}$, with objective value -1.

Dual problem of (P_2)

One can even simplify problem $(P_2)$by taking the dual.

Theorem 6 *Suppose there exists a strictly feasible solution to* (P_2)*. Then the objective values of* (P_2) *and the following dual problem are equal:*

$$(D_2) \quad \left\{ \begin{array}{cc} \max_{v \in \mathbb{R}} & -\sum_i \frac{(v\beta_i + \epsilon_i)^2}{2(\delta_i + v\alpha_i)} + cv \\ \textit{s.t.} & \delta_i + v\alpha_i \geq 0, \quad \forall i \\ & v \geq 0. \end{array} \right.$$

Note: the objective of (D_2) is concave in v.

Sharpening the S-lemma in the case of SDG

We start with the inhomogeneous version of the fundamental S-lemma.

Lemma 7 *Let A, D be symmetric matrices of the same size, and let the quadratic form $z^TAz + 2b^Tz + c$ be strictly positive at some point. Then the implication*

$$z^TAz + 2b^Tz + c \geq 0 \Rightarrow z^TDz + 2e^Tz + f \geq 0$$

holds true if and only if

$$\exists \lambda \geq 0 : \left(\begin{array}{c|c} D - \lambda A & e - \lambda b \\ \hline e^T - \lambda b^T & f - \lambda c \end{array} \right) \succeq 0. \qquad (10)$$

$\square$

In case that the matrices A and D are simultaneously diagonalizable we can sharpen the S-lemma, i.e., the LMI can be replaced by a simple convex constraint.

Lemma 8 *Let A, D be symmetric matrices of the same size and simultaneously diagonalizable by S into $\operatorname{diag}(\alpha_1, \ldots, \alpha_n)$ and $\operatorname{diag}(\delta_1, \ldots, \delta_n)$, respectively. Let the quadratic form $z^T Az + 2b^T z + c$ be strictly positive at some point. Then the implication*

$$z^T Az + 2b^T z + c \geq 0 \Rightarrow z^T Dz + 2e^T z + f \geq 0$$

holds true if and only if there exist $v \in \mathbb{R}$ such

$$\begin{cases} -\sum_i \frac{(v\beta_i - \epsilon_i)^2}{\delta_i - v\alpha_i} - cv + f \geq 0 \\ \delta_i - v\alpha_i \geq 0 \\ v \geq 0, \end{cases}$$

in which $\beta = S^T b$ and $\epsilon = S^T e$.

Applications - General

- There are also many examples of optimization problems that are not conic quadratic, but can be reformulated as such.
- See the excellent paper by Lobo et al. (2009).
- Examples: max of norms problems, logarithmic Chebychev approximation, quadratic/linear fractional problems.
- The robust counterpart of the original problem and the reformulated conic quadratic problem are not always equivalent.

  E.g., for the sum of norms problem mentioned the reformulation does not allow you to use implementation error.

Conclusions

- In case of SDG, the S-lemma can be sharpened, and be extended to the case of three quadratic forms.
- In case of SDG, a convex quadratic constraint with ellipsoidal uncertainty can be transformed into a conic quadratic constraint.
- In case of SDG, a conic quadratic constraint with ellipsoidal uncertainty can be transformed into a ’nearly’ conic quadratic constraint, with convex level sets.
- This has many applications!

Robust Optimization and Chance Constraints

Chance Constraints

$$p(w) \equiv \text{Prob}\left\{ w_0 + \sum_{\ell=1}^{d} z_\ell w_\ell \geq 0 \right\} \geq 1 - \epsilon \qquad \text{(C)}$$

- In general, (C) can be difficult to process:
  - The feasible set X of (C) can be nonconvex, which makes it problematic to optimize under the constraint.
  - Even when convex, X can be "computationally intractable":

    Let $z \sim$ Uniform$([0.1]^d)$. In this case, X is convex (Lagoa et al., 2005); however, *unless $P = NP$, there is no algorithm capable to compute $p(w)$ within accuracy δ in time polynomial in the size of the (rational) data w and in* $\ln(1/\delta)$ (L. Khachiyan, 1989).
- When (C) is difficult to process "as it is", one can look for a *safe tractable approximation of (C) — a computationally tractable convex set U_ϵ such that* $U_\epsilon \subset X \equiv \{w : p(w) \geq \epsilon\}$.

Probabilistic Guarantees via RO

$$f_0(x) + \sum_{l=1}^{d} z_l f_l(x) \leq 0\,. \qquad (1)$$

Assumption

$z_1, z_2, \ldots, z_d$ independent rv's

$z_l \sim \mathbf{P}_l \in \mathcal{P}_l$ (compact all prob. dist. in $\mathcal{P}_l$ has common support $= [-1, 1]$).

Definition A vector x satisfying, for a given $0 < z < 1$:

$$\Pr\{f_0(x) + \Sigma z_l f_l(x) \leq 0\} \geq 1 - \epsilon \quad \text{(chance constraint)} \qquad (2)$$

provides a *safe approximation* of (1).

Challenge Find uncertainty set for z, U_ϵ s.t. the Robust Counterpart of (1):

$$f_0(x) + \Sigma z_l f_l(x) \leq 0,\ \forall\, z \in U_\epsilon \qquad (3)$$

is a safe approximation of (1), i.e., every x satisfying (3) satisfies the CC (2).

Theorem

$$\begin{aligned} & U_\epsilon = B \cap (M + E_\epsilon) \\ & B = \{u \in I\!R^d \mid \|u\|_\infty \le 1\} \\ \text{where} \quad & M = \{u | \mu_l^- \le u_l \le \mu_l^+,\ l = 1, \ldots, d\} \\ & E = \{u | \Sigma u_l^2 / \sigma_l^2 \le 2 \log(1/\epsilon)\} \end{aligned} \tag{4}$$

μ_l^-, μ_l^+ and σ_l are such that

$$A_l(y) \le \max(\mu_l^- y, \mu_l^+ y) + \frac{\sigma_l^2}{2}\, y_l^2, \quad \forall\, l = 1, \ldots, d$$

where

$$A_l(y) = \max_{P_l \in \mathcal{P}_l} \log \left(\int \exp(ys) dP_l(s) \right) .$$

Values of $\mu_l^{\pm}$, σ_l are explicitly known for various families $\mathcal{P}_l$, e.g.

$$\left.\begin{array}{c}\mathrm{supp}(\mathcal{P})\subset[-1,1]\\ \text{P unimodal and symmetric}\end{array}\right\}\begin{array}{c}\mu_l^{\pm}=0\\ \sigma_l=\sqrt{1/6}\end{array}$$

$$\left.\begin{array}{c}\mathrm{supp}(\mathcal{P})\subset[-1,1]\\ \text{P unimodal}\end{array}\right\}\begin{array}{c}\mu_l^{-}=-1/2,\mu_l^{+}=1/2\\ \sigma_l=\sqrt{1/24}\end{array}$$

Moreover, for the LP case ($f_l(y)$ affine) the RC of (3), with U as in the Theorem, is conic quadratic or LP.

$$(1) \qquad a(\zeta)^T x \le b(\zeta)$$

$$a(\zeta) = a^0 + \sum_{\ell=1}^{L} \zeta_\ell a^\ell, \quad b(\zeta) = b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell$$

$$\zeta_1, \ldots, \zeta_L \ \text{ i.i.d.}, \ \ E(\zeta_\ell) = 0\,, \quad |\zeta_\ell| \le 1$$

$$(\mathrm{CC})_\varepsilon \ \boxed{\mathrm{Prob}_\zeta \left(a(\zeta)^T x \le b(\zeta)\right) \ge 1 - \varepsilon}$$

Let $\mathcal{U}_\Omega = \left\{\zeta \in \mathbb{R}^L \mid \|\zeta\|_2 \le \Omega\right\}$.

Consider the RC of (1) w.r.t. $\mathcal{U}_\Omega$:

$$a(\zeta)^T x \le b(\zeta) \quad \forall\, \zeta \in \mathcal{U}_\Omega$$

which we already know is equivalent to

$$(\mathrm{RC})_\Omega \ \boxed{(a^0)^T x + \Omega \sqrt{\textstyle\sum_{\ell=1}^{L} \left((a^\ell)^T x - b_\ell\right)^2} \le b^0}$$

Theorem 1 *If x solves $(\mathrm{RC})_\Omega$ with $\Omega \ge \sqrt{2\log(1/\varepsilon)}$, then x solves $(\mathrm{CC})_\varepsilon$*

$$\mathrm{OR:} \quad \left\{ \begin{array}{l} x \ \textit{solves}\ (\mathrm{RC})_\Omega \quad \textit{then } x \textit{ solves} \\ (\mathrm{CC})_\varepsilon \quad \textit{with} \quad \varepsilon < e^{-\Omega^2/2} \end{array} \right.$$

e.g., $\ \Omega = 7.44,\ 1 - \varepsilon = 1 - 10^{-12}$.

Discussion What if we ignore the stochastic information and just use $|\zeta_\ell| \leq 1$? In this case, the CC coincide with the RC of the linear eq. with uncertainty

$$\mathcal{U}_{\mathrm{Box}} = \{\zeta \mid |\zeta_i| \leq 1, \quad \ell = 1, \ldots, L\}$$

which is here

$$\sum \left|(a^\ell)^T x - b^\ell\right| \leq b^0 - (a^o)^T x \tag{3}$$

Hence

$$x \;\; \text{feasible for (3)} \;\; \Rightarrow CC \text{ is feasible with prob. 1.} \tag{4}$$

When using the stochastic information, the RC is

$$\Omega \sqrt{\sum \left[(a^\ell)^T x - b^\ell\right]^2} \quad \leq b^0 - (a^0)^T x \tag{5}$$

which corresponds to $\mathcal{U}_{\mathrm{Ball}}$, and we have

$$x \;\; \text{feasible for (4)} \;\; \Rightarrow CC \text{ is feasible with prob. } 1 - \exp(-\Omega^2/2) \tag{6}$$

(For $\Omega = 7.44 \quad 1 - \exp(-\Omega^2/2) = 1 - 10^{-12}$ so (4) and (6) are indistinguishable!)

But, for large L

$$\frac{\mathrm{Vol}(\mathcal{U}_{\mathrm{Box}})}{\mathrm{Vol}(\mathcal{U}_{\mathrm{Ball}})} \longrightarrow \infty \quad \text{superexponentially}$$

(ratio is > 1 starting with $L = 237$).

For small L, we could use

$$\mathcal{U} = \mathcal{U}_{\mathrm{Box}} \cap \mathcal{U}_{\mathrm{Ball}} = \left\{ \zeta \in \mathbb{R}^L \mid \|\zeta\|_\infty \leq 1, \ \|\zeta\|_2 \leq \Omega \right\}$$

Here the RC is

$$\left\{ \begin{array}{ll} \sum |z_\ell| + \Omega \sqrt{\sum w_\ell^2} & \leq b^0 - (a^0)^T x \\ z_\ell + w_\ell = b^\ell = (a^\ell)^T x & \end{array} \right. \tag{7}$$

and we have

$$\begin{array}{lcl} \text{If } x \text{ is a component of a feasible} & & CC \text{ is feasible with} \\ \text{solution } (x, z, w) \text{ of (7)} & \Longrightarrow & \text{prob. } 1 - \exp(-\Omega^2/2) \end{array} \tag{8}$$

Discussion With $\mathcal{U} = \mathcal{U}_{\rm Box} \cap \mathcal{U}_{\rm Ball}$, we get always (for all L) less conservative RC (incomparable so for $\Omega > 7$ compared to $\mathcal{U}_{\rm Box}$) and yet guarantee the CC with prob. essen. 1.

Striking phenomenon Consider the special case of P in Case I:

$$Pr(\zeta_i = \pm 1) = 1/2\,.$$

Note that here

$$\|\zeta\|_2^2 = L\,;$$

hence, if $L > \Omega^2$, the set $\mathcal{U}$ **does not contain even a single realization of** ζ, and yet the CC holds with high probability.

Conclusion The "immunization power" of the RC (7) cannot be explained by the fact that the underlying perturbation set $\mathcal{U}$ contains "nearly all" realizations of the random perturbation vector.

Illustration: Single-Period Portfolio Selection

There are 200 assets. Asset #200 ("money in the bank") has yearly return $r_{200} = 1.05$ and zero variability. The yearly returns r_ℓ, $\ell = 1, \ldots, 199$ of the remaining assets are independent random variables taking values in the segments $[\mu_\ell - \sigma_\ell, \mu_\ell + \sigma_\ell]$ with expected values μ_ℓ; here

$$\mu_\ell = 1.05 + 0.3\frac{(200-\ell)}{199},\ \sigma_\ell = 0.05 + 0.6\frac{(200-\ell)}{199},\ \ell = 1, \ldots, 199\,.$$

The goal is to distribute \$1 between the assets in order to maximize the return of the resulting portfolio, the required risk level being $\varepsilon = 0.5\%$.

We want to solve the uncertain LO problem

$$\max_{y,t}\left\{t : \sum_{t=1}^{199} r_\ell y_\ell + r_{200}y_{200} - t \geq 0, \sum_{\ell=0}^{200} y_\ell = 1,\ y_\ell \geq 0\,\forall\,\ell\right\},$$

where y_ℓ is the capital to be invested into asset $\#\ell$.

The uncertain data are the returns r_ℓ, $\ell = 1, \dots, 199$; their natural parameterization is

$$r_\ell = \mu_\ell + \sigma_\ell \zeta_\ell \,,$$

where ζ_ℓ, $\ell = 1, \dots, 199$, are independent random perturbations with zero mean varying in the segments $[-1, 1]$. Setting $x = [y; -t] \in \mathbb{R}^{201}$, the problem becomes

$$\left\{ \begin{array}{lll} \text{minimize} & x_{201} & \\ \text{subject to} & & \\ (a) & \left[a^0 + \sum_{\ell=1}^{199} \zeta_\ell a^\ell\right]^T x - \left[b^0 + \sum_{\ell=1}^{199} \zeta_\ell b^\ell\right] \le 0 & \\ (b) & \sum_{j=1}^{200} x_\ell = 1 & \\ (c) & x_\ell \ge 0,\ \ell = 1, \dots, 200 & \end{array} \right. \qquad (4)$$

where

$$\begin{array}{rcl} a^0 & = & [-\mu_1; -\mu_2; \dots; -\mu_{199}; -r_{200}; -1]; a^\ell = \sigma_\ell \cdot [0_{\ell-1,1}; 1; 0_{201-\ell,1}],\ \ell = 1, \dots, 199; \\ b^\ell & = & 0,\ \ell = 0, 1, \dots, 199\,. \end{array}$$

The only uncertain constraint in the problem is the linear inequality (a). We consider 3 perturbation sets along with the associated robust counterparts of problem (4).

1. *Box RC* which ignores the information on the stochastic nature of the perturbations affecting the uncertain inequality and uses the only fact that these perturbations vary in $[-1, 1]$. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\}$;
2. *Ball-Box* with the safety parameter $\Omega = \sqrt{2\ln(1/\varepsilon)} = 3.255$, which ensures that the optimal solution of the associated RC (a CQ prob.) satisfies (a) with probability at least $1 - \varepsilon = 0.995$. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\}, \ \|\zeta\|_2 \leq 3.255\}$;

Results

Box RC. The associated RC is the LP

$$\max_{y,t}\left\{t:\ \begin{array}{l}\sum\limits_{\ell=1}^{199}(\mu_\ell-\sigma_\ell)y_\ell+1.05y_{200}\geq t\\ \sum\limits_{\ell=1}^{200}y_\ell=1,\ y\geq 0\end{array}\right\};$$

as it should be expected, this is nothing but the instance of our uncertain problem corresponding to the worst possible values $r_\ell=\mu_\ell-\sigma_\ell,\ \ \ell=1,...,199$, of the uncertain returns. Since these values are less than the guaranteed return for money, the robust optimal solution prescribes to keep our initial capital in the bank with guaranteed yearly return 1.05.

Ball-Box RC. The associated RC is the conic quadratic problem

$$\max_{y,z,w,t}\left\{ t: \begin{array}{l} \sum_{\ell=1}^{199}(\mu_\ell y_\ell + 1.05y_{200} - \sum_{\ell=1}^{199}|z_\ell| - 3.255\sqrt{\sum_{\ell=1}^{199} w_\ell^2} \geq t \\ z_\ell + w_\ell = y_\ell, \quad \ell = 1,\dots,199,\ \sum_{\ell=1}^{200} y_\ell = 1,\ y \geq 0 \end{array} \right\}.$$

The robust optimal value is 1.1200, meaning 12.0% profit with risk as low as $\varepsilon = 0.5\%$.

Robust Solution of Uncertain Optimization Problems under Ambiguous Stochastic Data

Aharon Ben-Tal

Technion—Israel Institute of Technology

Joint work with K. Postek, D. den Hertog and B. Melenberg

Deterministic optimization problem (z known parameters)

$$\begin{array}{ll} & \min_{x\in X} \; f(x,z) \\ s.t. & \\ & \quad g(x,z) \le 0\,. \end{array}$$

What if z is uncertain?

Case I z is only known to reside in a bounded set U (called ***uncertainty set***).

Approach: Robust optimization

$$\left\{ \begin{array}{ll} & \min_{x\in X} \; \max_{z\in U} F(x,z) \\ s.t. & \\ & \quad \max_{z\in U} g(x,z) \le 0 \end{array} \right.$$

For tractability it is needed that f and g are convex in x and *concave in z*.

Case II z is a random vector with known distribution function $\mathbb{P}$.

Approach

$$\begin{array}{ll} & \min_{x\in X} E_{\mathbb{P}} f(x,z) \\ s.t. & \\ & \text{Probability}_{\mathbb{P}}\{g(x,z) \le 0\} \ge 1-\epsilon \end{array}$$

Difficulties: Obj. function involves multiple integration. Constraint is usually NP-hard (nonconvex).

Case III z is stochastic but its distribution $\mathbb{P}$ is not known exactly.

It is only known to belong to a family $\mathcal{P}$ of distributions.

Approach

$$\begin{array}{ll} & \min_{x\in X} \max_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}} f(x,z) \\ s.t. & \\ & \max_{\mathbb{P}\in\mathcal{P}} \textbf{Prob}_{\mathbb{P}}\{g(x,z)\le 0\} \ge 1-\epsilon \end{array}$$

This is **Robust Optimization under Stochastic Ambiguity**.

Ambiguous constraints

Consider a constraint

$$f(\mathbf{x}, \mathbf{z}) \leq 0,$$

where

- $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$ is the decision vector
- $\mathbf{z} \in \mathbb{R}^{n_\mathbf{z}}$ is an uncertain parameter vector
- $f(\cdot, \mathbf{z})$ is convex for all $\mathbf{z}$.

Assume that $\mathbf{z}$ has a stochastic nature:

- follows a probability distribution $\mathbb{P}$
- $\mathbb{P}$ belongs to an *ambiguity set* $\mathcal{P}$, based on some partial information

Distributionally Robust Optimization

Two types of constraints:

- worst-case expected feasibility constraints:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \leq 0, \qquad \text{(WC-EF)}$$

Distributionally Robust Optimization

Two types of constraints:

- worst-case expected feasibility constraints:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x},\mathbf{z}) \leq 0, \tag{WC-EF}$$

- worst-case chance constraints:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(f(\mathbf{x},\mathbf{z}) > 0\right) \leq \epsilon. \tag{WC-CC}$$

(WC-EF) is used to construct *safe approximations* of (WC-CC).

Ambiguity set $\mathcal{P}$

Ambiguity set $\mathcal{P}$ should be such that it is possible to obtain good, computationally tractable upper bounds on

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$$

Most frequently, $\mathcal{P}$ consists of $\mathbb{P}$ with known:

- mean
- (co)variance matrix
- possibly, higher order moment information

Major works: Scarf (1958), Dupačová (1977), Birge and Wets (1987), Birge and Dulá (1991), Gallego (1992), Gallego, Ryan & Simchi-Levi (2001), Delage and Ye (2010), Wiesemann et al. (2014) and many others...

Scarf's (1958) newsvendor problem

Scarf considered a newsvendor problem with a single product such that

- c is the selling price per unit
- $z \geq 0$ is the uncertain demand with known mean μ and variance σ^2
- x is the number of items chosen by the newsvendor to purchase

The objective is to maximize the worst-case expected profit:

$$\max_{x} \sup_{\mathbb{P} \in \mathcal{P}} \min\{x, z\} - cx$$

where

$$\mathcal{P}_z = \{\mathbb{P} : \quad \mathbb{E}_{\mathbb{P}} z = \mu, \quad \mathbb{E}_{\mathbb{P}} (z - \mu)^2 = \sigma^2\}.$$

Scarf's result

In our framework the problem is given by

$$\begin{aligned} \max \quad & t \\ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(x,z) \leq -t \\ & x \geq 0, \end{aligned}$$

where $f(x,z) = cx - \min\{x,z\}$.
The optimal value is equal to:

$$t^* = \mu\left((1-c) + \sqrt{\sigma^2 c(1-c)/\mu^2}\right)^+ .$$

However, no closed-form *tight* bound for the worst-case expectation of a general convex $f(x,\cdot)$ under mean-variance information!

Forgotten result of Ben-Tal and Hochman (1972)

An exact upper bound when the dispersion measure is the mean absolute deviation (MAD).

Theorem

Assume that a one-dimensional random variable z has support included in $[a, b]$ and its mean and mean absolute deviation are μ and d:

$$\mathcal{P} = \{\mathbb{P} : \quad supp(z) \subseteq [a, b], \quad \mathbb{E}_{\mathbb{P}} z = \mu, \quad \mathbb{E}_{\mathbb{P}} |z - \mu| = d\}.$$

Then, for any convex function $g : \mathbb{R} \to \mathbb{R}$ it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} g(z) = p_1 g(a) + p_2 g(\mu) + p_3 g(b),$$

where $p_1 = \frac{d}{2(\mu - a)}, \quad p_3 = \frac{d}{2(b - \mu)}, \quad p_2 = 1 - p_1 - p_3.$

Generalization to multiple dimensions

The result of Ben-Tal and Hochman (1972) generalizes to multidimensional $\mathbf{z}$ with independent components.

$$\mathcal{P} = \{\mathbb{P} : \text{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_{\mathbb{P}} z_i = \mu_i, \quad \mathbb{E}_{\mathbb{P}} |z_i - \mu_i| = d_i, \quad z_i \perp z_j\}.$$

Independence implies that the worst-case distribution is a product of the per-component worst-case distributions.

For each convex $g(\cdot)$ it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} g(\mathbf{z}) = \sum_{\alpha \in \{1,2,3\}^{n_{\mathbf{z}}}} \left(\prod_{i=1}^{n_{\mathbf{z}}} p^i_{\alpha_i} \right) g(\tau^1_{\alpha_1}, \ldots, \tau^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}})$$

where $p^i_{\alpha_i}$ and $\tau^i_{\alpha_i}$ depend only on a_i, b_i, μ_i, and d_i (not on $g(\cdot)$).

Lower bound result

Ben-Tal and Hochman (1972) provide also an exact formula for the lower bound on the expectation if additionally, it is known that $\mathbb{P}(z \geq \mu) = \beta$:

$$\mathcal{P}_\beta = \{\mathbb{P} : \quad \mathbb{P} \in \mathcal{P}, \quad \mathbb{P}(z \geq \mu) = \beta\}.$$

Then, for any convex function $g : \mathbb{R} \to \mathbb{R}$ it holds that

$$\inf_{\mathbb{P} \in \mathcal{P}_\beta} \mathbb{E}_\mathbb{P} g(z) = \beta g\left(\mu + \frac{d}{2\beta}\right) + (1-\beta) g\left(\mu - \frac{d}{2(1-\beta)}\right).$$

Application to (WC-EF)

Setting

Consider (WC-EF):

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x},\mathbf{z}) \leq 0,$$

where $f(\mathbf{x},\cdot)$ is convex. Then, (WC-EF) is equivalent to:

$$g^U(\mathbf{x}) \leq 0$$

where

$$g^U(\mathbf{x}) = \sum_{\alpha\in\{1,2,3\}^{n_\mathbf{z}}} \left(\prod_{i=1}^{n_\mathbf{z}} p^i_{\alpha_i} \right) f(\mathbf{x}, \tau^1_{\alpha_1}, \ldots, \tau^{n_\mathbf{z}}_{\alpha_{n_\mathbf{z}}})$$

with fixed $\tau^i_{\alpha_i} \in \{a_i, \mu_i, b_i\}$, $p^i_{\alpha_i}$. Note that $g^U(\mathbf{x})$ is convex.

What can be gained by the lower bound result?

If for each z_i we also know $\beta_i = \mathbb{P}(z_i \geq \mu_i)$ then we know not only the

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$$

but also

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$$

by a generalization of the lower-bound result to multiple dimensions.

Hence, an entire, tight interval for $\mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$ is provided!

If β_i unknown, assume several values and evaluate the lower bound for each.

Specific applications - convex functions

Constraints that are convex in the uncertain parameter $\mathbf{z}$ are generally intractable in the classical (worst-case oriented) RO.

This can occur, e.g., when:

- implementation error $\mathbf{x} \mapsto \mathbf{x} + \mathbf{z}$ is present: $f(\mathbf{x}, \mathbf{z}) = g(\mathbf{x} + \mathbf{z})$
- linear decision rules $\mathbf{x} = v + V\mathbf{z}$ are applied: $f(\mathbf{x}, \mathbf{z}) = g(v + V\mathbf{z})$
- the objective has a sum-of-max form: $f(\mathbf{x}, \mathbf{z}) = \sum_i \max\{l_1(\mathbf{x}, \mathbf{z}), \ldots, l_{I_i}(\mathbf{x}, \mathbf{z})\}$, where $l_j(\cdot, \cdot)$ are bilinear

Inventory management (Ben-Tal et al. (2004))

Problem characteristics:

- 6 periods to manage the inventory
- involves purchase, holding and shortage costs
- the uncertain parameter is the deviation of product demand $\mathbf{z}_t$ from its forecast value μ_t
- decision variables are ordering decisions $\mathbf{x}_t(\mathbf{z}^t)$ (linear decisions of observed past demand)

We assume independence of the deviations z_t with mean 0 and known MAD.

Inventory management (Ben-Tal et al. (2004))

Research question: What is the difference in average-case performnce between

- RO solutions - solutions that minimize the worst-case outcome
- WCE solutions - solutions that minimize the worst-case expectation of the outcome?

Inventory management (Ben-Tal et al. (2004))

Research question: What is the difference in average-case performnce between

- RO solutions - solutions that minimize the worst-case outcome
- WCE solutions - solutions that minimize the worst-case expectation of the outcome?

Objective type	β	Total costs	
		RO	WCE
Worst-case value	-	1950	2384
Expectation range	0.25	[1255,1280]	[1004,1049]
Expectation range	0.5	[1223,1280]	[970,1049]
Expectation range	0.75	[1230,1280]	[994,1049]

We assume three possible values of β (skewness of the product demand).

Inventory management (Ben-Tal et al. (2004))

Objective type	β	Total costs	
		RO	WCE
Worst-case value	-	1950	2384
Expectation range	0.25	[1255,1280]	[1004,1049]
Expectation range	0.5	[**1223**, **1280**]	[**970**, **1049**]
Expectation range	0.75	[1230,1280]	[994,1049]

- WCE solution performs worse on the worst-case basis (as expected)
- WCE solution yields strictly better average-case performance than the robust solution - for $\beta = 0.5$ the entire WCE interval lies below the RO interval
- even more so, the WCE solutions are better by one standard deviation than the RO solutions

Average-case enhancement of RO solutions

Suppose that $f(\mathbf{x}, \mathbf{z})$ is RO tractable (concave or special-case convex in $\mathbf{z}$) and we minimize $\sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z})$.

Often, there are multiple worst-case optimal $\mathbf{x}$ (Iancu and Trichakis (2013)). Answer? Choose the one with the best average-case performance.

1. Solve the RO problem minimizing the worst-case objective:

$$\min_{\mathbf{x}} \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z})$$

   Denote optimal value by $\bar{t}$.

2. Solve the second-stage enhancing problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \\ \text{s.t.} \quad & \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z}) \leq \bar{t} \end{aligned}$$

Inventory management (Ben-Tal et al. (2004))

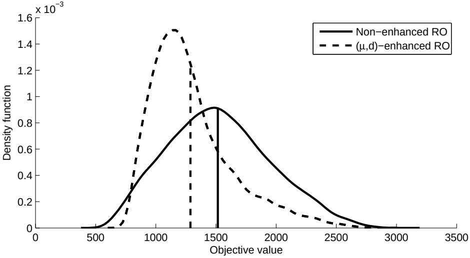
Research question: What is the difference in simulation performnce between

- non-enhanced RO solutions
- enhanced RO solutions?

Demand in the simulation is sampled uniformly from the uncertainty set.

We study the distribution of the total costs obtained in the simulation study by each of the solutions.

Inventory management (Ben-Tal et al. (2004))



The enhanced solution is better because:

- both solutions yield the same worst-case value (3149)
- the worst-case mean objective value brought by the enhanced solution is 1284 vs 1502 of the non-enhanced solution (down by 15%)
- the histogram of total costs obtained by the enhanced solution is visibly more skewed to the left (smaller total costs more probable)

Application to (WC-CC)

Setting

We assume w.l.o.g. that $\text{supp}(z_i) \in [-1, 1]$, $\mathbb{E}_{\mathbb{P}} z_i = 0$ and $\mathbb{E}_{\mathbb{P}} |z_i - 0| = d$.

Consider the (WC-CC):

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\mathbf{a}^T(\mathbf{z})\mathbf{x} > b(\mathbf{z})\right) \leq \epsilon,$$

where

$$[\mathbf{a}(\mathbf{z}); b(\mathbf{z})] = [\mathbf{a}^0; b^0] + \sum_{i=1}^{n_\mathbf{z}} z_i [\mathbf{a}^i; b^i].$$

Safe approximations

As such, (WC-CC) is intractable and we need a *safe approximation* - a computationally tractable set $\mathcal{S}$ of deterministic constraints such that

$$\mathbf{x} \text{ feasible for } \mathcal{S} \Rightarrow \mathbf{x} \text{ feasible for (WC-CC)}$$

How to construct safe approximations?
The crucial step is a construction of an upper bound on the moment generating function (MGF) of $\mathbf{z}$ (Ben-Tal et al. (2009)):

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T\mathbf{z}).$$

.

Recall: For each convex $g(\cdot)$ it holds that

$$\sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}} g(z) = \sum_{\alpha\in\{1.2.3\}^n} \left(\prod_{i=1}^n p^i_{\alpha_i}\right) g\left(\tau^1_{\alpha^1}, \ldots, \tau^{n_z}_{\alpha_{n_z}}\right)$$

where $p^i_{\alpha_i}$ and $\tau^i_{\alpha_i}$ depend only on a_i, b_i, μ_i, and d_i (not on $g(\cdot)$).

This formula has 3^n terms!

However:

$$\begin{aligned}
&\sup_{\mathbb{P}\in P} \log\left(E_{\mathbb{P}} \exp(w^T z)\right) = \sup_{\mathbb{P}\in\mathcal{P}} \log\left(E_{\mathbb{P}}\left(e^{w_1 z_1+\cdots+w_n z_n}\right)\right) \\
&= \sup_{\mathbb{P}\in\mathcal{P}} \log\left(E_{\mathbb{P}} \prod_{i=1}^n e^{w_i z_i}\right) = \text{ due to } z_i\text{'s being independent} \\
&= \sup_{\mathbb{P}\in\mathcal{P}} \log\left(\prod_{i=1}^n E\, e^{w_i z_i}\right) = \sup_{\mathbb{P}\in\mathcal{P}} \sum_{i=1}^n \left(\log E\, e^{w_i z_i}\right).
\end{aligned}$$

So here we need to apply the (B-H) upper (lower) bound separately to each on the n one-variable convex functions $E\, e^{w_i z_i}$!

MGF with our distributional assumptions

We know exactly the worst-case value of the MGF (not just an upper bound):

$$
\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T\mathbf{z}) &= \prod_{i=1}^{n_{\mathbf{z}}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(w_i z_i) \\
&= \prod_{i=1}^{n_{\mathbf{z}}} \left(\frac{d}{2}\exp(-w_i) + 1 - d + \frac{d}{2}\exp(w_i) \right) \\
&= \prod_{i=1}^{n_{\mathbf{z}}} \left(d\cosh(w_i) + 1 - d \right)
\end{aligned}
$$

Using this fact, we are able to construct three safe approximations of increasing tightness and increasing complexity.

An example of a safe approximation

Theorem

Let

$$[\mathbf{a}(\mathbf{z}); b(\mathbf{z})] = [\mathbf{a}^0; b^0] + \sum_{i=1}^{n_\mathbf{z}} z_i [\mathbf{a}^i; b^i].$$

If there exists $\alpha > 0$ *such that* $(\mathbf{x}, \alpha)$ *satisfies the constraint*

$$(\mathbf{a}^0)^T \mathbf{x} - b_0 + \alpha \log \left(\sum_{i=1}^{n_\mathbf{z}} \left(d_i \cosh \left(\frac{(\mathbf{a}^i)^T \mathbf{x} - b^i}{\alpha} \right) + 1 - d_i \right) \right) + \alpha \log(1/\epsilon) \leq 0,$$

then $\mathbf{x}$ *satisfies the (WC-CC):* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{a}^T(\mathbf{z})\mathbf{x} > b(\mathbf{z}) \right) \leq \epsilon.$

The approximating constraint is convex in $(\mathbf{x}, \alpha)$!

Antenna array (Ben-Tal and Nemirovski (2002))

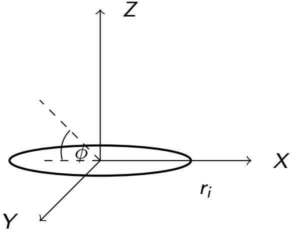
1. We consider an optimization problem with 40 circular antennas.
2. Each antenna has its diagram $D_i(\phi)$ - a plot of intensity of signal sent to different directions.
3. The diagram of the set of 40 antennas is the sum of their diagrams .

$$D(\phi) = \sum_{i=1}^{n} x_i D_i(\phi)$$

4. To the i-th antenna we can send a different amount of power x_i.
5. **Objective:** Set the x_i's in such a way that the diagram has the desired shape.

Application - antenna array optimization

Consider a circular antenna:



Energy sent in angle ϕ is characterized by *diagram*

Diagram of a single antenna:

$$D_i(\phi) = \frac{1}{2}\int_0^{2\pi} \cos\left(\frac{2\pi i}{40}\cos(\phi)\cos(\theta)\right) d\theta$$

Diagram of n antennas

$$D(\phi) = \sum_{i=1}^{n} x_i D_i(\phi)$$

x_i - power assigned to antenna i

Objective: construct $D(\phi)$ as close as possible to the desired $D^*(\phi)$ using the antennas available.

Antenna array (Ben-Tal and Nemirovski (2002))

Problem conditions:

- for $77^\circ < \phi \leq 90^\circ$ the diagram is nearly uniform:

$$0.9 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 77^\circ < \phi \leq 90^\circ$$

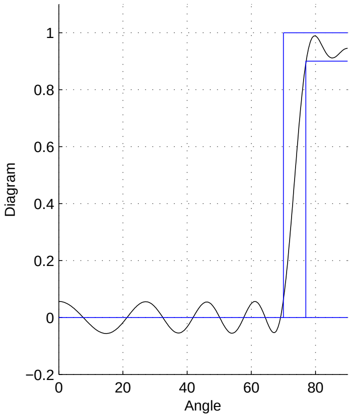
- for $70^\circ < \phi \leq 77^\circ$ the diagram is bounded:

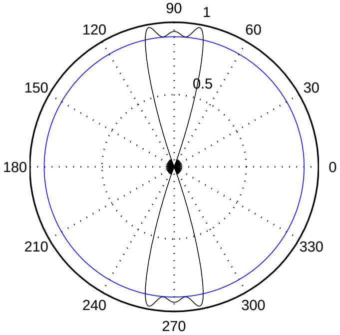
$$-1 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 70^\circ < \phi \leq 77^\circ$$

- we minimize the maximum absolute diagram value over $0^\circ < \phi \leq 70^\circ$:

$$\min \max_{0^\circ < \phi \leq 70^\circ} \left| \sum_{i=1}^{n} x_i D_i(\phi) \right|$$

Desired diagram graphically





Optimization problem to be solved

$$
\begin{aligned}
\min \quad & \tau \\
\text{s.t.} \quad & -\tau \le \sum_{i=1}^{n} x_i D_i(\phi) \le \tau, \quad 0 \le \phi \le 70^\circ \\
& -1 \le \sum_{i=1}^{n} x_i D_i(\phi) \le 1, \quad 70^\circ \le \phi \le 77^\circ \\
& 0.9 \le \sum_{i=1}^{n} x_i D_i(\phi) \le 1, \quad 77^\circ \le \phi \le 90^\circ
\end{aligned}
$$

Typically, decisions x_i suffer from implementation error z_i:

$$x_i \mapsto \widetilde{x}_i = (1 + z_i) x_i$$

We want each constraint to hold with probability at least $1 - \epsilon$!

Implementation error

Typically, decisions x_i suffer from implementation error z_i:

$$x_i \mapsto \widetilde{x}_i = (1 + \rho z_i) x_i$$

We want each constraint to hold with probability at least $1 - \epsilon$ for all $\mathbb{P} \in \mathcal{P}$, for example:

$$\mathbb{P}\left(\sum_{i=1}^{n} x_i (1 + \rho z_i) D_i(\phi) \leq 1 \right) \geq 1 - \epsilon, \quad 77^\circ < \phi \leq 90^\circ, \quad \forall \mathbb{P} \in \mathcal{P}$$

Two solutions:

- nominal: no implementation error
- robust: $\rho = 0.001$ and $\epsilon = 0.001$.

Implementation error and chance constraints

Assumptions on implementation errors $z_1, \ldots, z_n$:

- independence: z_i independent from z_j for $i \neq j$
- support: $-1 \leq z_i \leq 1$ for all i
- zero mean: $\mathbb{E}_{\mathbb{P}} z_i = 0$ for all i
- mean absolute deviation: $\mathbb{E}_{\mathbb{P}} |z_i - \mu_i| = d_i$ for all i

Hence the ambiguity set:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P} : \ z_i \in [-1, 1], \ \mathbb{E}_{\mathbb{P}} z_i = 0, \ \mathbb{E}_{\mathbb{P}} |z_i - \mu_i| = d_i, \ z_i \perp z_j\}$$

Safe approximations of chance constraints are typically constructed by constructing upper bounds on *moment-generating functions* $M(\mathbf{w})$:

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} M(\mathbf{w}) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp\left(\mathbf{w}^T \mathbf{z}\right)$$

Ben-Tal and Hochman (1972)

Assume $f : \mathbb{R} \mapsto \mathbb{R}$ is convex and z follows distribution $\mathbb{P}$ belonging to ambiguity set $\mathcal{P}$ such that

$$\mathcal{P} = \left\{ \mathbb{P} : \text{supp}(z) = [a^-, a^+], \quad \mathbb{E}_{\mathbb{P}} z = \mu, \quad \mathbb{E}_{\mathbb{P}} |z - \mu| = d \right\}$$

Then it holds that:

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(z_i) = p_1 f(a^-) + p_2 f(\mu) + p_3 f(a^+)$$

where $p_1 = \frac{d}{2(a^+ - \mu)}, \quad p_3 = \frac{d}{2(\mu - a^-)}, \quad p_2 = 1 - p_1 - p_3.$

Using this to the MGF problem with our assumptions on $z_1, \ldots, z_n$ we have

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp\left(\mathbf{w}^T \mathbf{z}\right) = \prod_{i=1}^{n} \left(d_i \cosh(w_i) + 1 - d_i\right)$$

How does it apply to antenna implementation error

We need $\mu_i^-, \mu_i^+, \sigma_i$ such that:

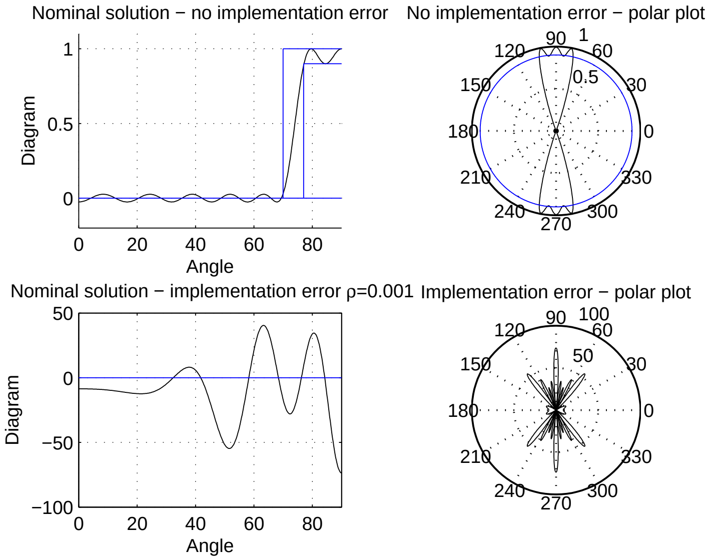
$$d_i \cosh(t) + 1 - d_i \leq \exp\left(\max\{\mu_i^+ t, \mu_i^- t\} + \frac{1}{2}\sigma_i^2 t\right)$$

for all $t \in \mathbb{R}$.
We easily find the right values $\mu_i^- = \mu_i^+ = 0$ and

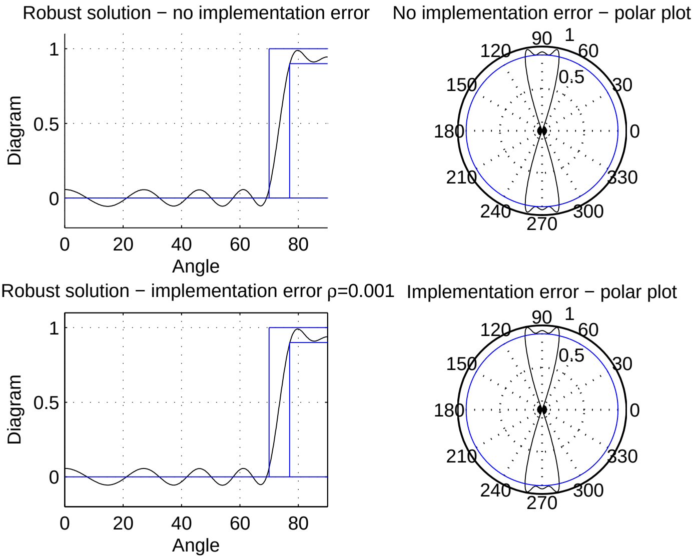
$$\sigma_i = \sup_{t \in \mathbb{R}} \sqrt{\frac{2\log\left(d_i \cosh(t) + 1 - d_i\right)}{t^2}}$$

to satisfy this requirement.

Nominal solution - dream and reality



Robust solution - dream and reality



Paper

Postek, K., Ben-Tal, A., Den Hertog, D., & Melenberg, B. (2015).

Exact robust counterparts of ambiguous stochastic constraints under mean and dispersion information.

- Operations Research 2018

Recovery of signals from noisy outputs

$\boxed{\text{The Estimation Problem}}$

$$y = Hx + w$$

Given y, find an estimator $\hat{x}$, which is as "close" as possible to x.

w random vector

$E(w) = 0, \;\; \mathrm{cov}(w) = C$ positive definite

CLASSICAL METHODS are based on minimizing data error $\|y - Hx\|$.

CLASSICAL APPROACH (Gauss,...)

Closeness measured by (standardized) data error

$$\|C^{-1/2}(y - H\hat{x})\|_2$$

Least Squares Estimator

$$\hat{x}_{LS} = \arg\min_x \|C^{-1/2}(y - Hx)\|_2$$

convex optimization

SOLUTION (H full column rank)

$$\boxed{\hat{x}_{LS} = (H^T C^{-1} H)^{-1} H^T C^{-1} y}$$

a *linear estimator*

$$\hat{x} = Gy$$

CLASSICAL MODIFICATION (Tikhonov,...)

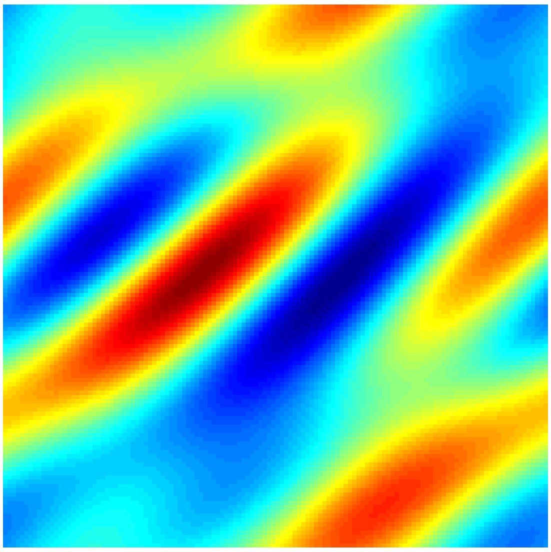
$$\boxed{\hat{x}_T = \arg\min_x \left\{ \|C^{-1/2}(y - Hx)\|^2 + \lambda \|x\|^2 \right\}}$$

still convex optimization

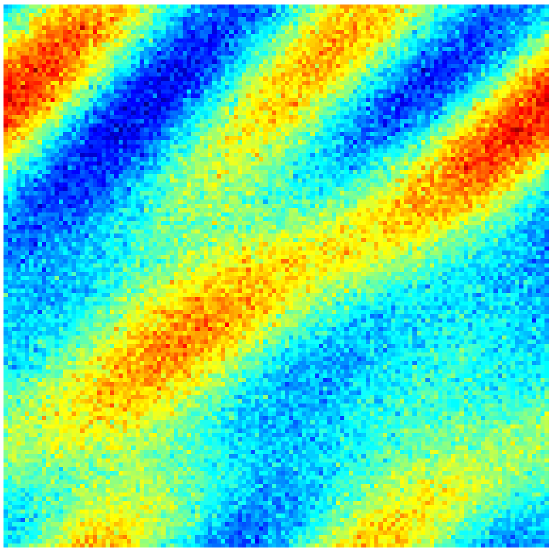
SOLUTION

$$\hat{x}_T = (H^T C^{-1} H + \lambda I)^{-1} H^T C^{-1} y$$

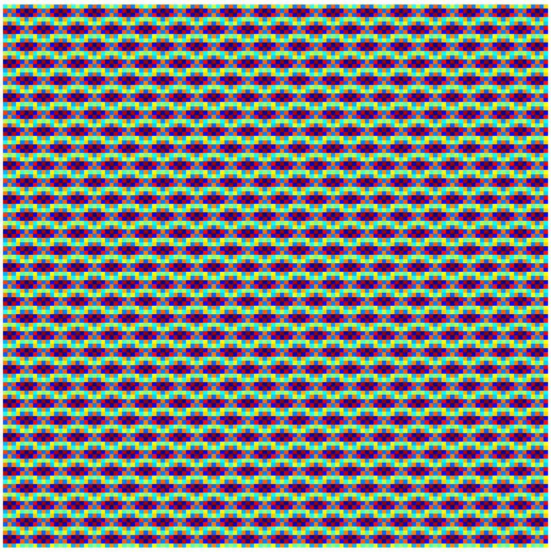
also a linear estimator.



True signal



Observations



LS

MSE estimator

$$\min_{\hat{x}} \; E\|x-\hat{x}\|^2$$

With a *linear estimator* $\hat{x} = Gy$ problem becomes

$$\min_{G} \{ \underbrace{x^T(I-GH)^T(I-GH)x}_{\text{bias}} + \underbrace{Tr(GCG)}_{\text{variance}} \}$$

but x unknown!

"Solution": minimal variance unbiased estimator
$GH = I$

Solution: Same as $\hat{x}_{L_S} \cdots$

Our approach: minmax MSE linear estimator:
$\hat{x} = Gy$, where:

$$\min_{G} \max_{\|x\|_T \le L} \{ x^T(I-GH)^T(I-GH)x + Tr(GCG) \}$$

$$\min_G \{L^2 \lambda \max(T^{-1/2}(I - GH)^T (I - GH) T^{-1/2})$$
$$+ Tr(GCG^T)\} \tag{7}$$

$$\Updownarrow$$

$$\boxed{\begin{array}{ll} \min\limits_{s.t} & L^2\lambda + t \\ & T^{-1/2}(I - GH)^T (I - GH) T^{-1/2} \leq \lambda I \\ & Tr(GCG^T) \leq t \end{array}} \tag{8}$$

Not an SDP ... yet.

Schur's complement:

$$T^{-1/2}(I - GH)^T(I - GH)T^{-1/2} \preceq \lambda I$$

$$\Leftrightarrow \underbrace{\begin{pmatrix} \lambda I & T^{-1/2}(I - GH)^T \\ (I - GH)T^{-1/2} & I \end{pmatrix} \succeq 0}_{LMI}$$

$$Tr(GCG^T) \le t \Leftrightarrow \underbrace{\begin{pmatrix} t & g^T \\ g & I \end{pmatrix}}_{LMI} \succeq 0$$

where $g = \text{vec}\,(GC^{1/2})$.

Theorem I: Original MinMax MSE problem (1) is equivalent to the SDP problem:

$$\boxed{\begin{array}{l} \min\limits_{s.t.} L^2\lambda + t \\ \begin{pmatrix} \lambda I & T^{-1/2}(I - GH)^T \\ (I - GH)T^{-1/2} & I \end{pmatrix} \succeq 0 \\ \begin{pmatrix} t & g^T \\ g & I \end{pmatrix} \succeq 0 \end{array}}$$

Theorem II: For the special case $T = I$, SDP can be solved explicitly. The optimal MMX MSE estimator is

$$\hat{x}_{\mathrm{mmx}} = \alpha \underbrace{(H^T C^{-1} H)^{-1} H^T C^{-1} y}_{\hat{x}_{LS}}$$

where $$\alpha = \frac{L^2}{L^2 + Tr((H^T C^{-1} H)^{-1})}$$

$\boxed{\text{Proof Structure}}$

(I) Establish the structure of the optimal solution

$$\boxed{G = VDV^T(H^TC^{-1}H)^{-1}H^TC^{-1}}$$

where V is the orthogonal matrix diagonalizing $H^TC^{-1}H$, i.e.,

$$H^TC^{-1}H = V\Sigma V^*$$
$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$$

This is obtained by optimality condition. Using this, we end up with an equivalent problem in variable (matrix) D (Problem B below).

(II) Show that $\exists$ an optimal matrix D which is diagonal.

(III) Find the diagonal elements of D.

The First Part of the Proof

The optimization problem

$$(A) \qquad \boxed{\begin{array}{l} \min\limits_{s.t.} \quad L^2\lambda + Tr(GCG^T) \\ \begin{pmatrix} \lambda I & (I - GH)^T \\ I - GH & I \end{pmatrix} \succeq 0 \end{array}}$$

Form the Lagrangian:

$$\begin{aligned} L(G,\lambda,U) &= L^2\lambda + Tr(GCG^T) \\ &- \quad Tr\left\{ \begin{pmatrix} U_1 & U_2^T \\ U_2 & U_3 \end{pmatrix} \begin{pmatrix} \lambda I & (I - GH)^T \\ I - GH & I \end{pmatrix} \right\} \\ &= \quad L^2\lambda + Tr(GCG^T) - \lambda Tr(U_1) - 2Tr(U_2(I - GH)) \\ &- \quad Tr(U_3)\,, \end{aligned}$$

$$U \quad := \quad \begin{pmatrix} U_1 & U_2^T \\ U_2 & U_3 \end{pmatrix} \succeq 0$$

Differentiating the Lagrangian with respect to G:

$$\frac{\partial L}{\partial G} = 0 \quad \Leftrightarrow \quad \boxed{G = U_2 H^T C^{-1}}$$

$$\Leftrightarrow \quad GH = U_2 (H^T C^{-1} H)$$

$$\Rightarrow \quad U_2 = (GH)(H^T C^{-1} H)^{-1}$$

change of variables: $D = V^T (GH) V \qquad (V^T V = I)$

$$\Leftrightarrow V D V^T = GH$$

$$\Rightarrow \boxed{U_2 = V D V^T (H^T C^{-1} H)^{-1}}$$

$$\boxed{G = V D V^T (H^T C^{-1} H)^{-1} H^T C^{-1}}$$

In particular, if the orthogonal matrix V is chosen as the matrix which diagonalizes $H^T C^{-1} H$, i.e.

$$(H^T C^{-1} H) = V \underbrace{\mathrm{diag}(\sigma_1, \ldots, \sigma_n)}_{\Sigma} V^T$$

then our problem (A), after substituting G becomes

(B) $$\boxed{\begin{array}{c}\min_{D,\lambda} \; L^2\lambda + Tr(D^T D\Sigma^{-1}) \\ (I-D)^T(I-D) \preccurlyeq \lambda I\end{array}}$$

Second part of the proof ("optimal D can be chosen diagonal").

Let $\mathcal{J}_n$ be the set of 2^n matrices which are $n \times n$, diagonal, with the entries in the diagonal being $+1$ or -1.

Claim If D^* is an optimal solution of (B), then so is

$$JD^*J, \quad \forall\ J \in \mathcal{J}_n$$

Proof

$$\begin{array}{rcl} Tr[(JDJ)^T(JDJ)\Sigma^{-1} & = & Tr(D^T D\Sigma^{-1})] \\ (I-JDJ)^T(I-JDJ) \preceq \lambda I & \Leftrightarrow & (I-D)^T(I-D) \end{array}$$

Conclusion Since (B) is a convex problem $\Rightarrow$ its optimal solution set is convex, so if D^* is an optimal solution, so is

$$\frac{1}{2^n} \sum_{J\in\mathcal{J}_n} (JD^*J)$$

$$D = \begin{bmatrix} a, & b \\ c, & d \end{bmatrix}$$

$$\mathcal{J}_2 = \Bigg\{ \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{J_1}, \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}}_{J_2}, \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{J_3} \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}}_{J_4}$$

$$J_1 D J_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad J_2 D J_2 = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$$

$$J_3 D J_3 = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix} \qquad J_4 D J_4 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\frac{1}{4} \sum_{i=1}^{4} J_i D J_i = \frac{1}{4} \begin{bmatrix} 4a & 0 \\ 0 & 4d \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$$

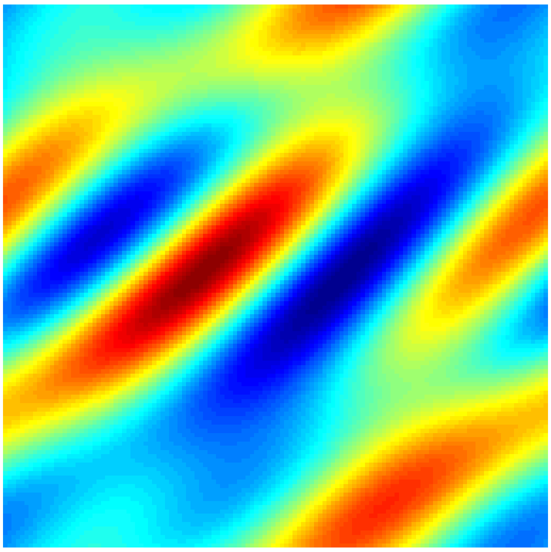
General result

$$\frac{1}{2^n} \sum_{J \in \mathcal{J}_n} J D J = \text{diag } D$$

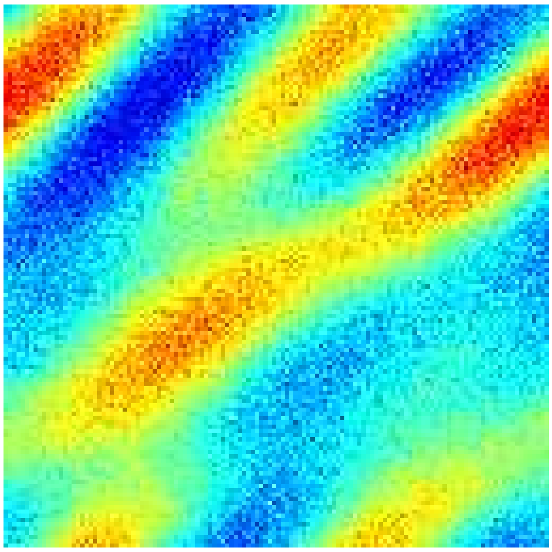
Part 3 of the proof with $D = \text{diag}(d_1, \ldots, d_n)$ problem (A) $\Leftrightarrow$ (B) reduces to

$$\boxed{\begin{array}{ll} \min\limits_{d_i, \lambda} & L^2 \lambda + \sum (d_i^2/\sigma_i) \\ \text{s.t.} & (1 - d_i)^2 \leq \lambda, \quad \forall\, i \end{array}}$$

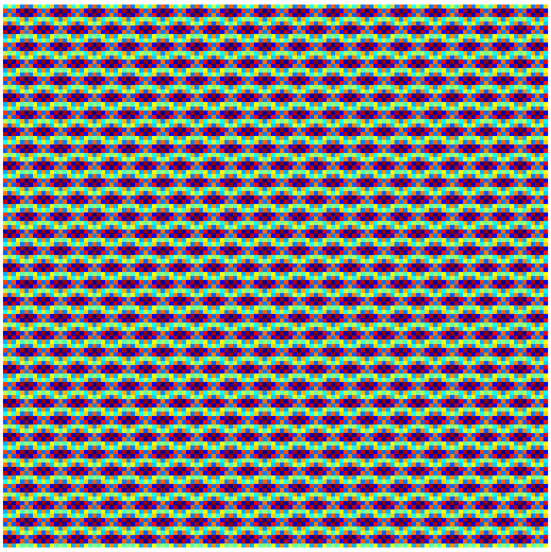
This problem can be solved analytically, which gives the final result claimed in Theorem II.



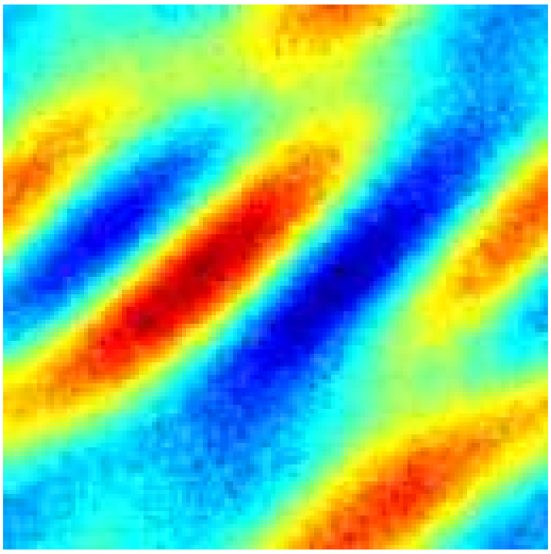
True signal



Observations



LS



Minmax Use

Adjustable Robust Optimization

Challenges: Adjustable Robust Optimization

♣ **Aside of applications of the RO methodology in various subject areas, an important venue of the RO-related research is** *extending the RO methodology beyond the scope of the RC approach as presented so far.*

The most important in this respect is, we believe, passing to *Adjustable Robust Optimization, where the decision variables are allowed to "adjust themselves"*, **to come extent,** *to the true values of the uncertain data.*

♠ **One of the central assumptions which led us to the notion of Robust Counterpart reads:**

> **A.1.** All decision variables in uncertain problem represent "here and now" decisions; they should be assigned specific numerical values as a result of solving the problem *before* the actual data "reveals itself."

While being adequate to many decision making situations, A.1 is not a "universal truth."

♠ **In some cases,** *not all decision variables represent "here and now" decisions.* **In dynamical decision making some of the variables represent "wait and see" decisions and as such can depend on the portion of the true data which "reveals itself" before the moment when the decision is being made.**

Example: In an inventory affected by uncertain demand, there are no reasons to specify all replenishment orders in advance; the true time to specify the replenishment order of period t is the beginning of this period, and thus we can allow this order to depend on the actual demands in periods $1, ..., t-1$.

♠ **Usually,** *not all decision variables represent actual decisions;* **there exist also "analysis" (or slack) variables which do not represent decisions at all and are used to convert the problem into a desired form, e.g., one of a LO problem.** *Since the analysis variables do not represent actual decisions, why not to allow them to depend on the entire true data?*

Adjustable Robust Optimization

$$
\text{(NLP)} \quad \left[\begin{array}{l} \min_x \; f(x) \\ \qquad\quad g(x, \boldsymbol{z}) \le 0 \end{array} \right.
$$

$\boldsymbol{z}$ uncertain parameter vector, $z \in \mathcal{U} \leftarrow$ uncertainty set.

E.g., in the LP case

$$
\text{(LP)} \quad \left[\begin{array}{l} \min \; c^T(x) \\ \qquad Ax - b \le 0 \le 0 \end{array} \right.
$$

$$
\boldsymbol{z} = (c, A, b).
$$

Robust counterpart of (NLP):

$$
\text{(RC)} \quad \left[\begin{array}{l} \min \; f(x) \\ \qquad g(x, z) \le 0 \quad \forall \, z \in \mathcal{U} \, . \end{array} \right.
$$

Implicitly, it is understood that optimal x has to be found *prior* to the realization of the data z.

- In many practical cases, part of the decision variables can be determined after (part of) the data vector z is realized (e.g., multi-stage decision problems).
- Adjustable/nonadjustable variables.

Uncertain LP with Adjustable Variables

$$(\mathcal{LP}_z) \quad \left\{ \begin{array}{l} \min_{u,v}\ c^T u \\ \qquad\qquad Uu + Vv \le b \end{array} \right\}_{(U,V,b)\in Z}$$

$u =$ vector of nonadjustable variables
("here and now" decisions)

$v =$ vector of adjustable variables
("wait and see" decisions)

V is called the **recourse matrix**.

The (usual) **RC** of $(\mathcal{LP}_z)$:

$$(\mathrm{RC}) \quad \left\{ \begin{array}{l} \min_{u,v}\ c^T u \\ \qquad\qquad Uu + Vv \le b\,,\ \forall\,(U,V,b)\in Z \end{array} \right.$$

The new *adjustable* RC (**ARC**) is defined

$$(\mathrm{ARC}) \quad \left\{ \begin{array}{l} \min_{u}\ c^T u \\ \qquad\qquad \forall\,(U,V,b)\in Z, \quad \exists\, v : \\ \qquad\qquad Uu + Vv \le b \end{array} \right.$$

ARC is less conservative.

Example (RC versus ARC)

Consider the following simple example:

$$\min_{u,v}\left\{-u : \begin{array}{rcl} (1-2\xi)u+v & \geq & 0 \\ \xi u - v & \geq & 0 \\ u & \leq 1 & \end{array}\right\}_{0\leq\xi\leq 1}$$

The RC and the ARC are, respectively, the problems:

$$\min_{u}\left\{-u : \begin{array}{c} \exists\, v\ \forall\,(\xi\in[0,1]) : \\ (1-2\xi)u+v\geq 0, \\ \xi u - v \geq 0,\ u\leq 1 \end{array}\right\} \qquad \text{(RC)}$$

$$\min_{u}\left\{-u : \begin{array}{c} \forall\,(\xi\in[0,1])\ \exists\, v : \\ (1-2\xi)u+v\geq 0, \\ \xi u - v \geq 0,\ u\leq 1 \end{array}\right\} \qquad \text{(ARC)}$$

Let us solve the RC:

$$\xi = 1 \Rightarrow \left\{\begin{array}{r} -u+v\geq 0 \\ u-v\geq 0 \end{array}\right. ; \quad \xi = 0 \Rightarrow \left\{\begin{array}{r} u+v\geq 0 \\ -v\geq 0 \end{array}\right. .$$

The only u, v that satisfy these inequalities are $u = v = 0$; therefore the only feasible solution to RC is $u = 0$, and the **optimal value is 0**. When solving ARC, *for every* $u \in [0, 1]$, taking $v = \xi u$, we see that u is feasible. Thus, the optimal solution to ARC is $u = 1$, and the *optimal value is* -1. Hence, the ARC of our simple problem is completely different from its RC, i.e., the ARC is *essential*.

Tractability Status of the ARC

Consider an uncertain Linear Programming problem

$$(\mathcal{LP}_z) \quad \left\{ \begin{array}{l} \min_u c^T u \\ s.t. \quad Uu + \boldsymbol{V}v \le b \end{array} \right\}_{[U,V,b]\in Z}$$

Theorem 1 *In the case when (fixed recourse)*

$$\mathcal{Z} = \mathrm{Conv}\left\{[U_1, \boldsymbol{V}, b_1], \ldots, U_N, \boldsymbol{V}, b_N]\right\},$$

the ARC of $\mathcal{LP}_z$ is equivalent to the usual LP problem

$$\min_{u,v_1,\ldots,v_N} \left\{c^T u : U_\ell u + V v_\ell \ge b_\ell,\ \ell = 1, \ldots, N\right\}.$$

Corollary 1 *In the case when*

1) *uncertainty set $\mathcal{Z}$ is a polytope given as a convex hull of a finite set of scenarios,*

2) *V is certain (fixed recourse).*

The Adjustable Robust Counterpart of the $\mathcal{LP}_z$ is an explicit LP program and thus is computationally tractable.

However, if only one of the properties 1) and 2) takes place, the ARC is not necessarily computationally tractable.

Computationally Intractable ARC's

Theorem 2 *In the case of an uncertainty set defined as a convex hull of finitely many scenarios with varying (from scenario to scenario) matrices $\boldsymbol{V_i}$:*

$$\mathcal{Z} = \mathrm{Conv}\left\{[U_1, V_1, b_1], \ldots, [U_N, V_N, b_N]\right\} ,$$

the ARC of $\mathcal{LP}_z$ can be computationally intractable.

Theorem 3 *In the case when $\mathcal{Z}$ is a general-type polytope given by a list of linear inequalities, the ARC of $\mathcal{LP}_z$ is a computationally intractable problem even when the coefficients of the analysis variables are certain.*

- What can be done when the ARC is computationally intractable?

  A natural solution in these cases would be to switch from "exact" ARC's to approximate ones.

Approximated Adjustable Robust Counterpart

When passing from an uncertain problem to its adjustable Robust Counterpart, we allow the analysis variables v to tune themselves to the true data ζ. Now, let us impose a restriction on *how* the analysis variables can be tuned to the data; specifically, assume that for u given v is *allowed to be an affine function of the data*:

$$\text{(LDR)} \qquad v = w + W\zeta\,.$$

With this approach, we are interested in the decision variables u that can be extended, by properly chosen w, W, to a solution of the infinite system of inequalities

$$U\boldsymbol{u} + V(\boldsymbol{w} + \boldsymbol{W}\zeta) \leq b,\ \forall\,\zeta = [U, V, b] \in \mathcal{Z}$$

in variables $\boldsymbol{u}$, $\boldsymbol{w}$, $\boldsymbol{W}$, and the approximate ARC of $\mathcal{LP}_z$ becomes the optimization program

$$\text{(AARC)} \qquad \min_{u,w,W} \left\{ \begin{array}{c} c^T u : \forall\,\zeta = [U, V, b] \in \mathcal{Z} \\ Uu + V(w + W\zeta) \leq b \end{array} \right\}.$$

Note that the AARC is "in-between" the usual RC of problem $\mathcal{P}_z$ and the Adjustable RC of the problem (AARC is less conservative than RC; solution of AARC is a "policy"); to get the RC, one should set to zero the variable W. Since the AARC seems simpler than the general ARC of $\mathcal{LP}_z$, there are chances to arrive at computationally tractable "approximate" ARC in the cases where the ARC is intractable.

Let the uncertainty set $\mathcal{Z}$ be given in the parametric form

$$\mathcal{Z} = \left\{ [U, V, b] = [U^0, V^0, b^0] + \sum_{\ell=1}^{L} \xi_\ell [U^\ell, V^\ell, b^\ell] : \xi \in \mathcal{X} \right\},$$

using the LDR:

$$v = v(\xi) = v^0 + \sum_\ell \xi_\ell v^\ell.$$

The **AARC of an uncertain LP** becomes:

$$\left\{ \begin{array}{l} \min\limits_{u, v^0, v^1, \ldots, v^L} c^T u \\ \left[U^0 + \Sigma\, \xi_\ell\, U^\ell\right] u + [V^0 + \Sigma\, \xi_\ell\, V^\ell][v^0 + \Sigma\, \xi_\ell\, v^\ell] \\ \qquad\qquad\qquad \leq [b^0 + \Sigma\, \xi_\ell\, b^\ell], \;\; \forall \xi \in \mathcal{X}. \end{array} \right.$$

For which perturbation set $\mathcal{X}$ is this SIP tractable?

Theorem 4 *Consider an uncertain LP with fixed recourse and a cone-representable perturbation set:*

$$\mathcal{X} = \{\xi \mid \exists \omega : A\xi + B\omega - d \in \mathcal{K}\} \subset R^L.$$

When the cone $\mathcal{K}$ is

1. *a nonnegative orthant R^n_+*
2. *a (finite) direct product of Lorentz cones*

$$L^m = \left\{ x \in R^m : x_m \geq \sqrt{x_1^2 + \cdots + x_{m-1}^2} \right\}$$

3. *a semidefinite cone (cone of positive semidefinite symmetric matrices),*

then: the AARC of uncertain LP is an explicit Linear, Conic Quadratic or Semidefinite Programming problem, respectively, of sizes polynomial in those of the description of the perturbation set and of the parametrization mapping.

Bad News: In the non-fixed resource case, the ARC can become NP-hard.

Remedy ??

Tight Approximation on AARC

Recall the "perturbation-based" model of AARC:

$$\min_{u,v^0,v^1,\ldots,v^L} c^T u$$
$$\left[U^0 + \Sigma\,\xi_\ell\,U^\ell\right] u + [V^0 + \Sigma\,\xi_\ell\,V^\ell][v^0 + \Sigma\,\xi_\ell\,v^\ell]$$
$$\geq [b^0 + \Sigma\,\xi_\ell\,b^\ell],\ \forall\,\xi \in \mathcal{X}.$$

with a $\cap$ ellipsoids perturbation set:

$$\mathcal{X} = \mathcal{X}_\rho \equiv \{\xi \mid \xi^T S_\ell\,\xi \leq \rho^2,\ \ell = 1,\ldots,L\},$$

with $\rho > 0$, $S_\ell \succeq 0$, $\Sigma\,X_\ell \succ 0$.

Note that this ellipsoidal uncertainty allows for a wide variety of symmetric (with respect to the origin) and convex perturbation sets. For example,

- setting $\ell = 1$, the perturbation set is an ellipsoid centered at the origin;
- setting $L = \dim\xi$ and $\xi^T S_\ell \xi = a_\ell^{-2}\xi_\ell^2$, $\ell = 1,\ldots,L$, we get, as the perturbation set, the box $\mathcal{X} = \{|\xi_\ell| \leq \rho a_\ell,\ \ell = 1,\ldots,L\}$ centered at the origin;
- choosing as S_ℓ dyadic matrices $g_\ell g_\ell^T$, we can get, as the perturbation set (centered at the origin) polytope $\mathcal{X} = \{\xi : |g_\ell^T\xi| \leq \rho,\ \ell = 1,\ldots,L\}$.

Theorem 5 *Consider an uncertain LP with an ellipsoidal perturbation set:*

$$\mathcal{X} = \mathcal{X}_\rho \equiv \left\{\xi \mid \xi^T S_\ell \xi \leq \rho^2,\ \ell = 1, \ldots, L\right\},$$

where $\rho > 0$, $S_\ell \succeq 0$, $\Sigma\, S_\ell \succ 0$ *along with the*

Semidefinite program

$$\text{(SDP)} \quad \min_{\lambda^1,\ldots,\lambda^m,x=[u,v^0,v^1,\ldots,v^L]} c^T u$$
$$\text{s.t.} \quad \left(\begin{array}{c|c} \Gamma_i(x) - \rho^{-2}\Sigma\lambda^i_\ell S_\ell & \beta_i(x) \\ \hline \beta_i^T(x) & \alpha_i(x) - \Sigma\lambda^i_\ell(x) \end{array}\right) \succeq 0,$$
$$\lambda^i \geq 0,\ i = 1, \ldots, m,$$

where $x \equiv [u, v^0, v^1, \ldots, v^L]$ *and*

- $\alpha_i(x) \equiv [U_i^0 u + V_i^0 v^0 - b_i^0]$
- $\beta_i^\ell(x) \equiv \frac{[U_i^\ell u + V_i^0 v^\ell + V_i^\ell v^0 - b_i^\ell]}{2}$, $\ell = 1, \ldots, L$
- $\Gamma_i^{(\ell,k)}(x) \equiv \frac{V_i^k v^\ell + V_i^\ell v^k}{2}$, $\ell, k = 1, \ldots, L$.

Then: Problem (SDP) is a "conservative approximation" to the AARC: whenever x *can be extended, by some* $\lambda^1, \ldots, \lambda^m$*, to a feasible solution of this semidefinite program,* x *is feasible for the AARC.*

What is the level of conservativeness?

$$
\begin{array}{lll}
\text{(SDP)} & \min\limits_{\lambda^1,\ldots,\lambda^m,x=[u,v^0,v^1,\ldots,v^L]} & c^T u \\
 & s.t. & \left(\begin{array}{c|c} \Gamma_i(x)-\boldsymbol{\rho}^{-2}\Sigma\lambda^i_\ell S_\ell & \beta_i(x) \\ \hline \beta_i^T(x) & \alpha_i(x)-\Sigma\lambda^i_\ell(x) \end{array}\right) \succeq 0, \\
 & & \lambda^i \geq 0,\ i=1,\ldots,m,
\end{array}
$$

Theorem 6 *In the case of simple ellipsoidal uncertainty (i.e. $L=1$), problem (SDP)* **is exactly equivalent** *to the AARC.*

(Proof based on S-lemma)

Theorem 7 *In the case of $L>1$, (SDP) is a tight approximation of the AARC. Specifically, the projection on the x-space of the feasible set of (SDP)* **is contained** *in the feasible set of the AARC, the perturbation level being ρ, and* **contains** *the feasible set of the AARC, the perturbation level being $\Omega\rho$, where*

$$
\boxed{\Omega = 0(1)\ln L}
$$

In particular, the optimal value in (SDP) is in-between the optimal values of the AARC's corresponding to the perturbation levels ρ and $\Omega\rho$.

Proof based on Approx. S-Lemma

Application: Inventory Model

Consider a single product inventory system, which is comprised of a warehouse and of I factories. The planning horizon is T periods. At a period t:

- d_t is the demand for the product. All the demands must be satisfied;
- $v(t)$ is the amount of the product in the warehouse at the beginning of the period ($v(1)$ is given);
- $p_i(t)$ – the i'th order of the period – is the amount of the product that will be produced during the period by factory i and used to satisfy the demand of the period (and, perhaps, to replenish the warehouse); these are the decision variables.
- $P_i(t)$ is the maximal prod. capacity of factory i;
- $c_i(t)$ is the cost of producing one unit of the product at factory i.

Other parameters of the problem are:

- $V_{\min}$ is the minimal level of inventory that must be at the warehouse at each given moment;
- $V_{\max}$ is the maximal storing capacity of the warehouse;
- Q_i is the total maximal production capacity of the i'th factory throughout the planning horizon.

Our goal is to minimize the total production cost over all factories and the entire planning horizon. The Linear Programming problem modeling this is the following:

$$\min_{p_i(t), v(t), F} F$$

$$\sum_{t=1}^{T} \sum_{i=1}^{I} c_i(t) p_i(t) \le F$$

$$0 \le p_i(t) \le P_i(t), \qquad 1 \le i \le I,\ 1 \le t \le T$$

$$\sum_{t=1}^{T} p_i(t) \le Q(i), \qquad 1 \le i \le I$$

$$v(t+1) = v(t) + \sum_{i=1}^{I} p_i(t) - d_t, \quad 1 \le t \le T$$

$$V_{\min} \le v(t) \le V_{\max}, \quad 2 \le t \le T+1\,.$$

We can eliminate the v variables:

$$\min_{p_i(t), F} F$$

$$\sum_{t=1}^{T} \sum_{i=1}^{I} c_i(t) p_i(t) \le F$$

$$0 \le p_i(t) \le P_i(t), \qquad 1 \le i \le I,\ 1 \le t \le T$$

$$\sum_{t=1}^{T} p_i(t) \le Q(i), \qquad 1 \le i \le I$$

$$V_{\min} \le v(1) + \sum_{s=1}^{t} \sum_{i=1}^{I} p_i(s) - \sum_{s=1}^{t} d_s \le V_{\max},\ 1 \le t \le T\,.$$

The decision on orders $p_i(t)$ is made at the beginning of period t, and we are allowed to make these decisions on the basis of demands d_r, $r \in I_t$, where I_t is a given subset of the segment $\{1, \ldots, t\}$. The only uncertain data in the problem are the demands d_t, and all we know is that

$$d_t \in [d_t^* - \theta d_t^*, d_t^* + \theta d_t^*] , \quad t = 1, \ldots, T ,$$

with given positive θ and d_t^*. Applying the AARC methodology, we restrict our decision-making policy with *affine decision rules*

$$p_i(t) = \pi_{i,t}^0 + \sum_{r \in I_t} \pi_{i,t}^r d_r ,$$

where the coefficients $\pi_{i,t}^r$ are our new decision variables. With this approach, we get an AARC problem which is itself an LP.

Illustrative Example: The Data

There are 3 factories for production of umbrellas, and one warehouse. The decisions concerning production are made every two weeks, and we are planning production for one year. Thus, the time horizon is **24 periods**. The nominal demand d_t^* is seasonal, reaching its maximum in winter, and behaves according to the following function:

$$d_t^* = 1000\left(1 - \frac{1}{2}\sin\left(\frac{\pi(t-1)}{12}\right)\right), \quad t = 1, \ldots, 24.$$

We assume that the uncertainty level is θ, i.e., $d_t \in [(1-\theta)d_t^*, (1+\theta)d_t^*]$.

The production costs per unit of the product depend on the factory and on time, and follow the same seasonal pattern as the demand, i.e., rise in winter and fall in summer. Cost per unit for a factory i at a period t is given by:

$$c_i(t) = \alpha_i \left(1 - \frac{1}{2}\sin\left(\frac{\pi(t-1)}{12}\right)\right), \quad t = 1, \ldots, 24$$

$$\alpha_1 = 1, \quad \alpha_2 = 1.5, \quad \alpha_3 = 2\,.$$

The maximal production capacity of each one of the factories at each two-week period is $\boldsymbol{P_i(t) = 567}$ units, and the *total production capacity* of each one of the factories for a year is $\boldsymbol{Q_i = 13600}$. The inventory at the warehouse should not be less than 500 ($V_{\min}$) units, and cannot exceed 2000 ($V_{\max}$) units.

The initial inventory at warehouse is $v(1) = 1250$.

The Experiments

In every one of the experiments, the corresponding management policy was tested against a given number (100) of simulations; in every one of the simulations, the actual demand d_t of period t was drawn at random, according to the ***uniform*** distribution, from the segment $[(1-\theta)d_t^*, (1+\theta)d_t^*]$, where θ was the "uncertainty level" characteristic for the experiment. The demands of distinct periods were independent of each other.

We conducted two series of experiments.

1. We checked the ***influence of*** the demand uncertainty θ on the total production costs corresponding to the management policy yielded by the optimal solution to the AARC. We ***compared*** this cost ***to*** the "Perfect Hindsight" (PH) one, i.e., the cost we would have paid in the case when all the demands were known to us in advance and we were using the corresponding optimal management policy as given by the optimal solution of certain LP.

The Results

- The influence of the uncertainty level on the management cost.

Here we tested the robustly adaptive management policy *with the standard information basis* against different levels of uncertainty, specifically, the levels of 20%, 10% and 5%.

	AARC cost	PH cost	
Uncertainty	Mean	Mean	"price of robustness"
5%	24615	23782	3.5%
10%	25211	23607	5.9%
20%	26860	23735	13%

Std. in range 100–780.

The Results

- The influence of the "information basis" on the management cost: **Robust Counterpart**

An interesting question is what is the uncertainty level which still allows for a priori decision. To get an answer, we solved the RC of our uncertain problem for the uncertainty levels 20%, 10% and 5% and all the time got ***infeasible RC's***. Only at the uncertainty level as small as 2.5%, the RC becomes feasible and yields the management costs as follows:

	RC cost	PH cost	
Uncertainty	Mean	Mean	price of robustness
2.5%	25287	23842	6%
		AARC	$\to$ 2%

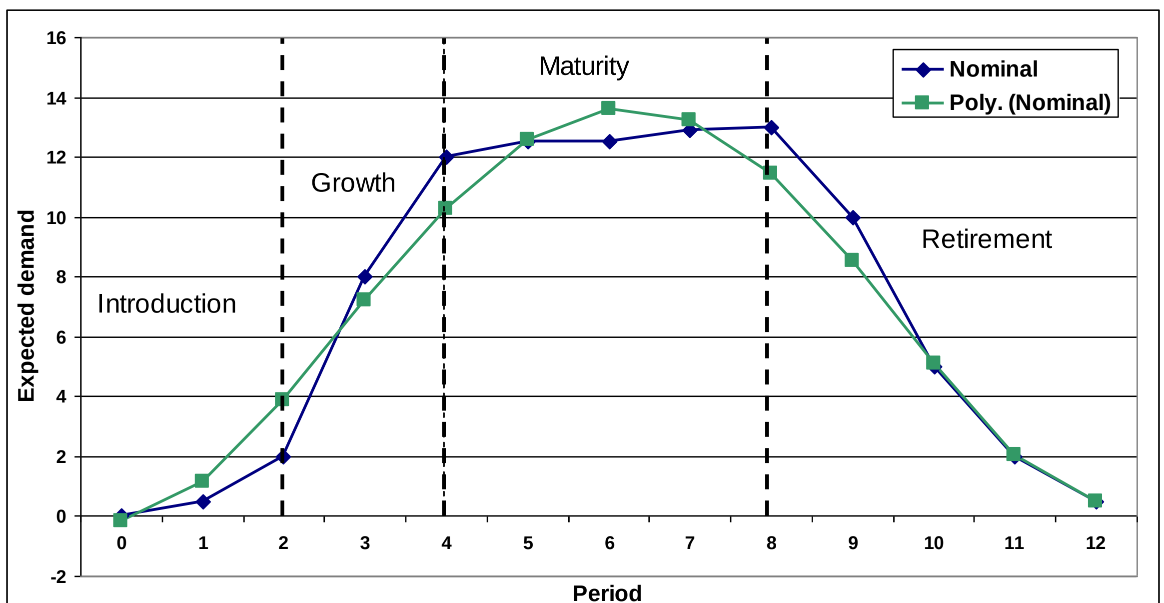
Note that even at this unrealistically small uncertainty level, the price of robustness for the policy yielded by the RC is larger by 6% than the PH cost (while for the AARC this difference is just 2%).

The Price of Robustness?

Example: Robust Optimization of Multi-Period Production Planning under Uncertainty

Problem definition

- Single-product production planning problem
  - Finite horizon T
  - Uncertain demand
  - Discrete periods
- Expected (nominal) demand follows a typical life-cycle pattern: $\bar{d}(t) = -0.17 + 0.25t + 1.23t^2 - 0.18t^3 + 0.004t^4 + 0.0002t^5$



Problem definition (cont.)

- In each period $t=1,\dots,T$:
  - Holding and shortage costs incurred for each unit of surplus or shortage, respectively
  - Income is realized through sales
- In each period $t=1,\dots,T-1$:
  - Unsatisfied demand is backlogged
- In period T :
  - Salvage value is gained for surplus units
  - Unsatisfied demand is lost

Mathematical model

T	planning horizon		
q_t	production quantity in units	c_t	production cost per unit
d_t	demand in units	m_t	selling price per unit
$\bar{d}_t$	nominal demand in units	s_T	salvage value per unit
I_t	inventory level	h_t	holding cost per unit
k	initial inventory (non-negative)	p_t	shortage cost per unit

$$q^t = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_t \end{pmatrix} \qquad d^t = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_t \end{pmatrix}$$

- The model:

$$\underset{q\geq 0}{Max}\left\{ \begin{array}{l} F(q,d) = \overbrace{\sum_{t=1}^{T} m_t d_t - m_T \max(0,-I_T)}^{\text{Sales}} + \overset{\text{Salvage}}{s_T \max(0, I_T)} + \\ -\sum_{t=1}^{T}\Big[\underset{\text{Production}}{c_t q_t} + \underset{\text{Holding}}{h_t \max(0, I_t)} + \underset{\text{Shortage}}{p_t \max(0,-I_t)}\Big] \end{array} \right\} \quad (1)$$

- Denote

$$I_t = I_t\left(q^t, d^t\right) = \sum_{\tau=1}^{t}\left(q_\tau - d_\tau\right) + I_0; \quad I_t^+ = \max\left(0, I_t\right); \quad I_t^- = \max\left(0, -I_t\right)$$

LP model

- The piecewise linear model (1) can be written as the following LP model:

$$\underset{q,\ I^-}{Max}\ F$$

$$s.t.$$

$$\sum_{t=1}^{T} m_t d_t - m_T I_T^- - \sum_{t=1}^{T}\left[c_t q_t + p_t I_t^- + \overline{h_t}\left(I_t^- + I_0 + \sum_{i=1}^{t}(q_i - d_i)\right)\right] \ge F$$

$$\left.\begin{array}{l} I_t^- + I_0 + \sum_{i=1}^{t}(q_i - d_i) \ge 0 \\ I_t^-, q_t \ge 0 \end{array}\right\} \quad t = 1, \ldots, T \qquad (2)$$

Affinely Adjustable Robust Counterpart (AARC)

- Two types of decision variables: *adjustable* and *non-adjustable*
- *Non-adjustable* should be determined before the uncertain data is revealed
- *Adjustable* can depend on past realizations of the uncertain data
- In the AARC the dependence of the *adjustable* variables on past data is a linear decision rule (LDR) as follows:

$$\begin{aligned} q_t &= \alpha_t^0 + \sum_{r \in H_{1t}} \alpha_t^r d_r \\ I_t^- &= \beta_t^0 + \sum_{r \in H_{2t}} \beta_t^r d_r \end{aligned} \qquad (4)$$

The price of robustness

- One of the limitations of the RO methodology is its conservative optimal solutions. Its guaranteed objective value might be inferior as compared with the nominal solution (NOM)
- The difference between the nominal and the robust objective values is called the price of robustness (POR)

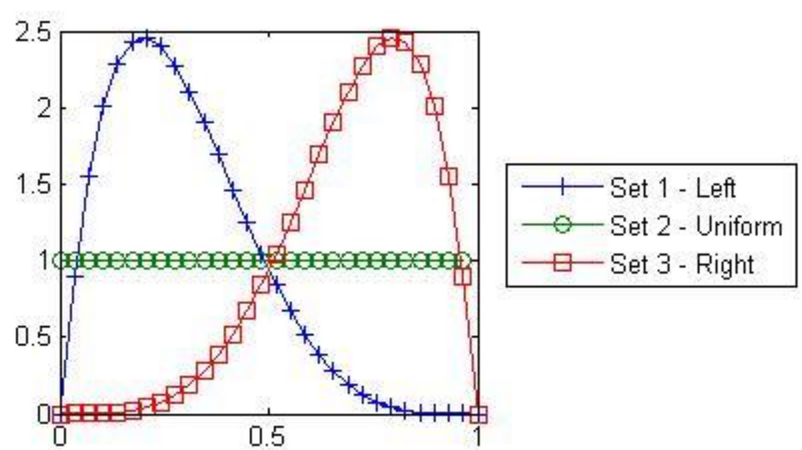
The actual price of robustness

- In practice it is rare that the demand takes the worst case scenario values $\Rightarrow$ POR is inappropriate
- We base the comparison on L demand simulations:
  - $AP(q_A) = \frac{1}{L}\sum_{l=1}^{L} F(q_A, d^l)$ - the average "actual" profit according to the policy of method A
  - Define the actual price of robustness (APOR) for method A as the difference between $AP(q_{NOM})$ and $AP(q_A)$

Simulations

- For several uncertainty levels ρ three sets of $L=100$ demand vectors were generated
- Each vector consists of $T=12$ entries
- The demand vectors entries of each set were generated from a Beta distribution with specific shape parameters supported by the uncertainty set for the demand

Set number	Set description	Beta distribution shape parameters α	β
1	Left	2	5
2	Uniform	1	1
3	Right	5	2



The actual price of robustness (APOR) for the AARC method

Doesn't depend on the uncertainty level

		Uncertainty level (in %)		
		2	14	30
Optimal solution	$F(q_{NOM}, \bar{d})$	98.71	98.71	98.71
	$F(q_{AARC})$	95.3	74.82	46.92
	POR	3.4	23.9	51.79
Set 1 - Left	$AP(q_{NOM})$	95.72	77.74	53.77
	$AP(q_{AARC})$	96.87	85.83	71.97
	$APOR$	-1.16	-8.09	-18.20
Set 2 - Uniform	$AP(q_{NOM})$	97.05	87.03	73.69
	$AP(q_{AARC})$	97.94	93.29	87.23
	$APOR$	-0.89	-6.26	-13.54
Set 3 - Right	$AP(q_{NOM})$	94.38	87.03	73.69
	$AP(q_{AARC})$	99.06	101.15	101.59
	$APOR$	-4.69	-32.81	-27.90

***$AP(q_{NOM})$* does depend on the uncertainty level**

Robust Solutions of Uncertainly Affected

Linear Dynamic Systems

♣ **Generic application:** *Affine control of uncertainty-affected Linear Dynamical Systems.*

♠ **Consider Linear Time-Varying Dynamical system**

$$\begin{array}{rcl} x_{t+1} &=& A_t x_t + B_t u_t + R_t d_t \\ y_t &=& C_t x_t \\ x_0 &=& z \end{array} \qquad (S)$$

- x_t: **state;** • u_t: **control** • y_t: **output;**
- d_t: **uncertain input;** • **z: initial state**

to be controlled over finite time horizon $t = 0, 1, ..., T$**.**

♠ **Assume that a "desired behaviour" of the system is given by a system of convex inclusions**

$$D_i w - b_i \in \mathcal{Q}_i, \; i = 1, ..., m$$

on the state-control trajectory

$$w = (x_0, x_1, ..., x_{T+1}, u_0, u_1, ..., u_T),$$

and the goal of the control is to minimize a given linear objective $f(w)$**.**

$$\begin{array}{rcl} x_{t+1} &=& A_t x_t + B_t u_t + R_t d_t \\ y_t &=& C_t x_t \\ x_0 &=& z \end{array} \qquad \textbf{(S)}$$

♠ **Restricting ourselves with affine output-based control laws**

$$u_t = \xi_{t0} + \sum_{\tau=0}^{t} \Xi_{t\tau} y_\tau, \qquad (*)$$

the problem of interest is

(!) *Find an affine control law $(*)$ which ensures that the resulting state-control trajectory w satisfies the system of convex inclusions*

$$D_i w - b_i \in \mathcal{Q}_i,\ \ i = 1, ..., m$$

and minimizes, under this restriction, a given linear objective $f(w)$.

Dynamics (S) makes w a known function of inputs $d = (d_0, d_1, ..., d_T)$, the initial state z and the parameters ξ of the control law $(*)$:

$$w = W(\xi; d, z).$$

Consequently, (!) is the optimization problem

$$\min_{\xi} \left\{ f(W(\xi; d, z)) : D_i W(\xi; d, z) - b_i \in \mathcal{Q}_i,\ i = 1, ..., m \right\} \qquad (\mathbf{U})$$

$$\boxed{\begin{array}{ll} \textbf{open loop dynamics:} & \left\{\begin{array}{rcl} x_{t+1} &=& A_t x_t + B_t u_t + R_t d_t \\ y_t &=& C_t x_t \\ x_0 &=& z \end{array}\right. \\ \textbf{control law:} & u_t = \xi_{t0} + \sum\limits_{\tau=0}^{t} \Xi_{t\tau} y_\tau \end{array}}$$

$$\Downarrow$$

$$\boxed{w := (u_0, ..., u_T, x_0, ..., x_{T+1}) = W(\xi; d, z)}$$

$$\Downarrow$$

$$\boxed{\min_{\xi} \{f(W(\xi; d, z)) : D_i W(\xi; d, z) - b_i \in \mathcal{Q}_i,\ i = 1, ..., m\} \qquad \textbf{(U)}}$$

Note: **Due to presence of uncertain input trajectory d and possible uncertainty in the initial state, (U) is an uncertain problem.**

Difficulty: **While linearity of the dynamics and the control law make $W(\xi; d, z)$ linear in (d, z), the dependence of $W(\cdot, \cdot)$ on the parameters $\xi = \{\xi_{t0}, \Xi_{t\tau}\}_{0 \le \tau \le t \le T}$ of the control law is highly nonlinear**

$\Rightarrow$ (U) is *not* a bi-affine problem, which makes inapplicable the theory we have developed. In fact, (U) seems to be intractable already when there is no uncertainty in d, z!

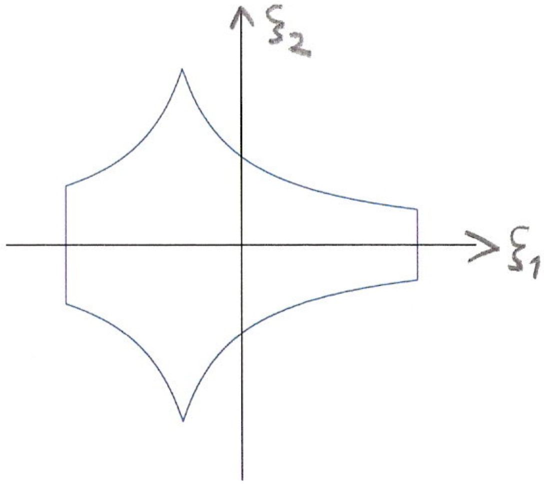
♣ **Example:**

$$\begin{array}{rcl} x_0 & = & 0 \\ x_{t+1} & = & x_t + u_t + d_t \\ y_t & = & x_t \\ \hline u_t & = & \xi_t y_t \end{array}$$

- **Due to non-affine dependence $u_2 = \xi_2((1+\xi_1)d_0+d_1)$ on d_1, d_2, the design specification**

  Whenever the inputs satisfy $|d_t| \le 1$, $t = 0,1$, the controls should satisfy $|u_t| \le 3$, $t = 0,1,2$

specifies a *non-convex* **domain in the plane (ξ_1, ξ_2) of control parameters:**



$u_0 = 0 \quad u_1 = \xi_1 d_0$

$u_2 = \xi_2 [(1+\xi_1) d_0 + d_1]$

$\begin{cases} |u_1| \le 3, \ |u_2| \le 3 \\ \forall \ |d_t| \le 1 \end{cases}$

X**-axis:** ξ_1, Y**-axis:** ξ_2

Remedy: suitable re-parameterization of affine control laws.

♣ Consider a closed loop system along with its *model*:

closed loop system:	model:
$x_{t+1} = A_t x_t + B_t u_t + R_t d_t$	$\widehat{x}_{t+1} = A_t \widehat{x}_t + B_t u_t$
$y_t = C_t x_t$	$\widehat{y}_t = C_t \widehat{x}_t$
$x_0 = z$	$\widehat{x}_0 = 0$
$u_t = U_t(y_0, ..., y_t)$	

♠ Observation: We can run the model in an on-line fashion, so that at time t, before the decision on u_t should be made, we have in our disposal *purified outputs*

$$v_t = y_t - \widehat{y}_t.$$

♠ Fact I [Equivalence]: Every transformation $(d, z) \mapsto w$ which can be obtained from an affine control law based on outputs:

$$u_t = \xi_{t0} + \sum_{\tau=0}^{t} \Xi_{t\tau} y_\tau \qquad (*)$$

can be obtained from an affine control law based on purified outputs:

$$u_t = \eta_{t0} + \sum_{\tau=0}^{t} H_{t\tau} v_\tau \qquad (**)$$

and vice versa.

system:	**model:**
$$\begin{array}{rcl} x_{t+1} &=& A_t x_t + B_t u_t + R_t d_t \\ y_t &=& C_t x_t \\ x_0 &=& z \end{array}$$	$$\begin{array}{rcl} \widehat{x}_{t+1} &=& A_t \widehat{x}_t + B_t u_t \\ \widehat{y}_t &=& C_t \widehat{x}_t \\ \widehat{x}_0 &=& 0 \end{array}$$

(S)

control law:

$$\begin{array}{rcl} v_t &=& y_t - \widehat{y}_t \\ u_t &=& \eta_{t0} + \sum\limits_{\tau=0}^{t} H_{t\tau} v_\tau \end{array} \qquad (**)$$

♠ **Fact II [bi-affinity]: The state-control trajectory** $w = W(\eta; d, z)$ **of (S) is affine in** (d, z) **when the parameters** $\eta = \{\eta_{t0}, H_{t\tau}\}_{0\leq\tau\leq t\leq T}$ **of the control law** $(**)$ **are fixed,** *and is affine in* η *when* (d, z) *is fixed.*

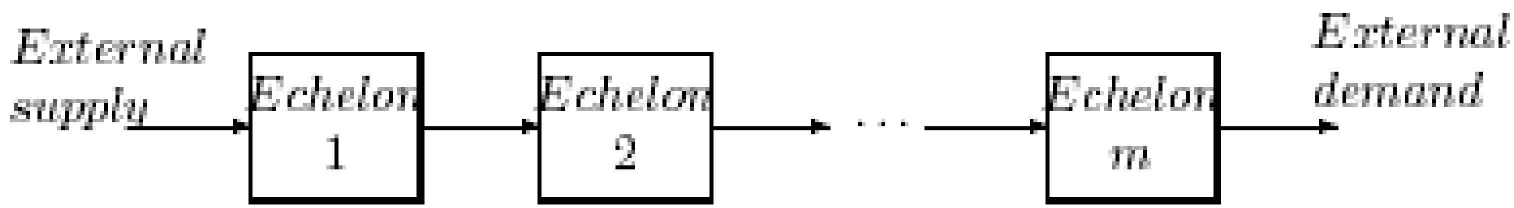
♠ **Corollary:** *With parameterization* $(**)$ *of affine control laws, the problem*

> *Find an affine control law* $(*)$ *which ensures that the resulting state-control trajectory* w *satisfies the system of convex inclusions*
>
> $$D_i w - b_i \in \mathcal{Q}_i,\ i = 1, ..., m$$
>
> *and minimizes, under this restriction, a given linear objective* $f(w)$.

becomes an uncertain bi-affine optimization problem and as such can be processed via the CRC approach.

In particular, in the case when $\mathcal{Q}_i$ *are one-dimensional, the CRC of the problem is computationally tractable, provided that the normal range* $\mathcal{U}$ *of* (d, z) *and the associated cone* $\mathcal{L}$ *are so. If* $\mathcal{U}$, $\mathcal{L}$ *and the norms used to measure distances are polyhedral, CRC is just an explicit LP program.*

Supply chain control – GRC implementation for control problems



- x_t^j=amount echelon j orders from j-1 at the beginning of period t
- Y_t^j =inventory level in echelon *j* at the end of period *t*
- z^j =initial inventory level at echelon *j*
- d_t =external demand at period *t*

- $T^L(j) = l(j) + M(j-1) + L(j)$ the delay between the time an order is placed and <u>received</u> in echelon *j*.
- $T^M(j) = l(j+1) + M(j)$
  the delay between the time an order is placed and <u>shipped</u> from echelon

Supply chain control – GRC implementation for control problems

- Main objective : minimizing cost
- Sub objective: stabilizing the system
- Problem Characteristics:
  - Finite horizon
  - Multi echelon
  - Delays
  - Backlogging
  - Demand must be satisfied and is uncertain

$$
\begin{aligned}
&\min_{y,x} \sum_{j,t} [c_t^j x_t^j + w_t^j] \\
&s.t. \\
&\qquad \begin{array}{rcl}
y_t^j & = & y_{t-1}^j + x_{t-T^L(j)}^j - x_{t-T^M(j)}^{j+1} \\
y_t^m & = & y_{t-1}^m + x_{t-T^L(m)}^m - d_{t-T^M(m))}
\end{array} \\
&\qquad \left.\begin{array}{rcl}
w_t^j & \geq & h_t^j y_t^j \\
w_t^j & \geq & -p_t^j y_t^j \\
y_t^j & \geq & \underline{a}^j \\
y_t^j & \leq & \overline{a}^j \\
x_t^j & \leq & b^j \\
x_t^j & \geq & 0 \\
w_t^j & \geq & 0 \\
y_0^j & = & z^j
\end{array}\right\} \forall j \in \{1, ..., m\}
\end{aligned}
$$

Supply chain control

- Eliminating the equalities recursively will give us a LP problem of the form we discussed.
- How do we control this system? What are the consequences of different types of control?

LV control

- Lets assume we take the control suggested by Love [Love,1979] using target inventory: $x_t^j = x_{t-1}^{j+1} + \frac{1}{2}(\Upsilon^j - y_{t-1}^j) \ \forall j \in 1, \ldots, m$
- Resulting in the LP problem:

$$
\begin{array}{l}
\min\limits_{x,\tilde{x},y,w,\pi} \sum\limits_{j,t}[c_t^j x_t^j + w_t^j] \\
s.t. \\
\left.\begin{array}{rcl}
y_t^j & = & y_{t-1}^j + x_{t-T^L(j)}^j - x_{t-T^M(j)}^{j+1} \ \forall j \in \{1,..,m-1\} \\
y_t^m & = & y_{t-1}^m + x_{t-T^L(m)}^m - d_{t-T^M(m)} \\
& & \left.\begin{array}{rcl}
w_t^j & \geq & h_t^j y_t^j \\
w_t^j & \geq & -p_t^j y_t^j \\
\tilde{x}_t^j & = & x_{t-1}^{j+1} + \frac{1}{2}(\Upsilon^j - y_{t-1}^j) \\
\tilde{x}_t^j & \leq & \mathrm{M}\pi_t^j \\
x_t^j & \geq & 0 \\
x_t^j & \geq & \tilde{x}_t^j \\
x_t^j & \leq & \mathrm{M}\pi_t^j \\
x_t^j & \leq & \mathrm{M}(1-\pi_t^j) + \tilde{x}_t^j \\
w_t^j & \geq & 0 \\
y_0^j & = & z^j \\
\pi_t^j & \in & \{0,1\}
\end{array}\right\} \forall j \in \{1,\ldots,m\}
\end{array}\right\} \forall t \in \{1,\ldots,n\}
\end{array}
$$

ILV control

- We can further improve this method by making the reference inventory a decision variable rather than a constant.

GRC control

- Applying the GRC to the AARC problem assuming

$$D = \mathcal{D} + \mathcal{L}_{\mathcal{D}}, \qquad Z = \mathcal{Z} + \mathcal{L}_{\mathcal{Z}}$$

$\mathcal{D}$ and $\mathcal{Z}$ are of the form of a multi-dimensional box

Robust Counterpart (AARC)

$$y_t^j = y_t^j(d, z) \text{ affine function}$$
$$x_t^j = x_t^j(d, z)$$

constraints must hold $\forall\ d \in D, z \in Z$

CRC A typical constraint is of the form

$$F_i(d, z) \in K_i \quad (K_i = \text{interval})$$

Its CRC version being

$$(*) \qquad \operatorname{dist}\left(F_i(d, z), K_i\right) \le \sum_{t=1}^{n} \alpha_{it} \operatorname{dist}\left(d_t, D_t\right) + \sum_{j=1}^{m} \alpha_{ij} \operatorname{dist}\left(z_j, Z_j\right)$$

When D_t, Z_j are polyhedral (*) can be converted to linear inequalities.

The purified outputs corresponding to the dynamic system (1) – (3) are here

$$v_t^j = \begin{cases} z_0^m - \sum_{t=1}^{t-T^M(m)} d_z & \text{if } j = m \\ z_0^j & j < m \end{cases}$$

The affine control law is here

$$x_t^j = \eta_0^{x,t,j} + \sum_{l=1}^{m} \sum_{\tau=1}^{T} \eta_{\tau l}^{x,t,j} v_\tau^l$$

where $\eta_{\tau l}^{x,t,j} = 0 \qquad \forall \tau \geq t$ (non anticipativity)

Example

- [Love, 1979], Oscillating demand:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
d_t	6	6	6	6	6	6	6	6	7	8	9	10	9	8	7	6	5	4	5	6

- Horizon: n=20
- Echelons: m=3
- Cost: c=2, p=3,h=1
- Initial inventory: z=12
- Lead time: L=2
- No other delays

Inventory Behavior – "amplification of oscillation"