

•

Selected Examples  
of  
CONIC DUALITY AT WORK

- Robust Linear Optimization
- Synthesis of Linear Controllers
- Matrix Cube Theorem

**A. Nemirovski**

Arkadi.Nemirovski@isye.gatech.edu

# Linear Optimization Problem, its Data and Structure

## ♣ Linear Optimization problem:

$$\min_x \{c^T x + d : Ax \leq b\} \quad (\text{LO})$$

- $x \in \mathbf{R}^n$ : vector of *decision variables*,
- $c \in \mathbf{R}^n$  and  $d \in \mathbf{R}$  form the *objective*,
- $A$ : an  $m \times n$  *constraint matrix*,
- $b \in \mathbf{R}^m$ : *right hand side*.

♠ Problem's structure: its sizes  $m, n$ .

♠ Problem's data:  $(c, d, A, b)$ .

## Data Uncertainty

♣ The data of typical real world LOs are partially *uncertain* — not known exactly when the problem is being solved.

♠ Sources of data uncertainty:

- Prediction errors. Some of data entries (future demands, returns, etc.) do not exist when the problem is solved and hence are replaced with their forecasts.

- Measurement errors: Some of the data (parameters of technological devices and processes, contents associated with raw materials, etc.) cannot be measured exactly, and their true values drift around the measured “nominal” values.
- Implementation errors: Some of the decision variables (planned intensities of technological processes, parameters of physical devices we are designing, etc.) cannot be implemented exactly as computed. *The implementation errors are equivalent to artificial data uncertainties.*

Indeed, the impact of implementation errors

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

on the validity of the constraint

$$a_{i1}x_1 + \dots + a_{in}x_n \leq b_i$$

is *as if* there were no implementation errors, but the data of the constraint was subject to perturbations

$$a_{ij} \mapsto (1 + \epsilon_j)a_{ij}, \quad b_i \mapsto b_i - \sum_j a_{ij}\delta_j.$$

## Data Uncertainty: Traditional Treatment and Dangers

♣ Traditionally,

♠ “small” (fractions of percents) data uncertainty is just ignored, the problem is solved “as it is” – with the nominal data, and the resulting nominal optimal solution is forwarded to the end user;

♠ “large” data uncertainty is assigned with a probability distribution and is treated via Stochastic Programming techniques.

♠ Fact: *in many cases, even small data uncertainty can make the nominal solution heavily infeasible and thus practically meaningless.*

## ♣ Example: Antenna Design

♠ [Physics:] *Directional density of energy transmitted by an monochromatic antenna placed at the origin is proportional to  $|D(\delta)|^2$ , where the antenna's diagram  $D(\delta)$  is a complex-valued function of 3-D direction (unit 3-D vector)  $\delta$ .*

♠ [Physics:] *For an antenna array — a complex antenna comprised of a number of antenna elements, the diagram is*

$$D(\delta) = \sum_j x_j D_j(\delta) \quad (*)$$

- $D_j(\cdot)$ : diagrams of elements
- $x_j$ : complex *weights* – design parameters responsible for how the elements in the array are invoked.

♠ Antenna Design problem: *Given diagrams*

$$D_1(\cdot), \dots, D_k(\cdot)$$

*and a target diagram  $D_*(\cdot)$ , find the weights  $x_i \in \mathbf{C}$  such that the synthesized diagram  $(*)$  is as close as possible to the target diagram  $D_*(\cdot)$ .*

♥ When  $D_j(\cdot)$ ,  $D_*(\cdot)$ , same as the weights, are real and the “closeness” is quantified by the uniform norm on a finite grid  $\Gamma$  of directions, Antenna Design becomes the LO problem

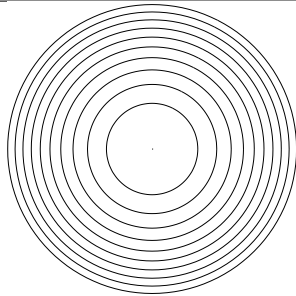
$$\min_{x \in \mathbf{R}^n, \tau} \left\{ \tau : -\tau \leq D_*(\delta) - \sum_j x_j D_j(\delta) \leq \tau \quad \forall \delta \in \Gamma \right\}.$$

♠ **Example:** Consider planar antenna array comprised of 10 elements (circle surrounded by 9 rings of equal areas) in the plane XY (Earth's surface"), and our goal is to send most of the energy "up," along the 12° cone around the Z-axis:

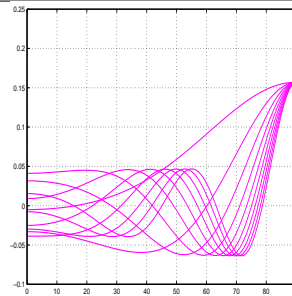
- Diagram of a ring  $\{z = 0, a \leq \sqrt{x^2 + y^2} \leq b\}$ :

$$D_{a,b}(\theta) = \frac{1}{2} \int_a^b \left[ \int_0^{2\pi} r \cos(2\pi r \lambda^{-1} \cos(\theta) \cos(\phi)) d\phi \right] dr,$$

- $\theta$ : altitude angle
- $\lambda$ : wavelength



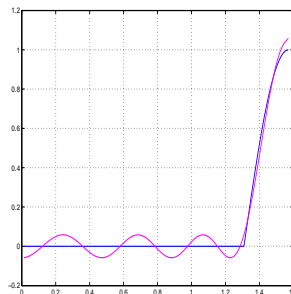
10 antenna elements,  
equal areas, outer radius 1 m



Diagrams of the elements  
vs the altitude angle  $\theta$ ,  $\lambda = 50$  cm

- **Nominal design problem:**

$$\tau_* = \min_{x \in \mathbf{R}^{10}, \tau} \left\{ \tau : -\tau \leq D_*(\theta_i) - \sum_{j=1}^{10} x_j D_j(\theta_i) \leq \tau, 1 \leq i \leq 240, \theta_i = \frac{i\pi}{480} \right\}$$



Target (blue) and nominal  
optimal (magenta) diagrams,

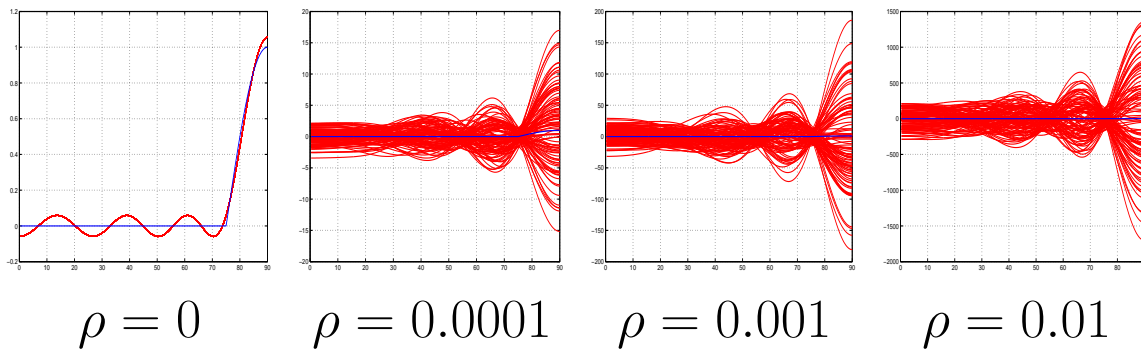
$$\tau_* = 0.0589$$

**But:** The design variables are characteristics of physical devices and as such they cannot be implemented exactly as computed.

What happens when there are implementation errors:

$$x_j^{\text{fact}} = (1 + \xi_j)x_j^{\text{comp}}, \quad \xi_j \sim \text{Uniform}[-\rho, \rho]$$

with small  $\rho$ ?



“Dream and reality,” nominal optimal design: **samples of 100 actual diagrams (red)** for different uncertainty levels. **Blue: the target diagram**

	Dream	Reality								
	$\rho = 0$	$\rho = 0.0001$			$\rho = 0.001$			$\rho = 0.01$		
	value	min	mean	max	min	mean	max	min	mean	max
$\ \cdot\ _{\infty}$ -distance to target	<b>0.059</b>	1.280	<b>5.671</b>	14.04	11.42	<b>56.84</b>	176.6	39.25	<b>506.5</b>	1484
energy concentration	<b>85.1%</b>	0.5%	<b>16.4%</b>	51.0%	0.1%	<b>16.5%</b>	48.3%	0.5%	<b>14.9%</b>	47.1%

Quality of nominal antenna design: dream and reality. Data over 100 samples of actuation errors per each uncertainty level  $\rho$ .

♠ **Conclusion:** *Nominal optimal design is completely meaningless...*

## ♣ Example: NETLIB Case Study.

♠ **NETLIB:** a collection of LO problems for testing LO algorithms.

♠ **Constraint # 372** of the NETLIB problem PILOT4:

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

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

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

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

♠ **Note:** The coefficients in  $a$ , except for the coefficient **1** at  $x_{880}$ , are “ugly reals” like -15.79081 or -84.644257. Ugly coefficients characterize certain technological devices and processes; as such *they could hardly be known to high accuracy* and coincide with the “true” data within accuracy of 3-4 digits, not more.

Question: Assuming that the ugly entries in  $a$  are 0.1%-accurate approximations of the true data  $\tilde{a}$ , what is the effect of this uncertainty on the validity of the “true” constraint  $\tilde{a}^T x \geq b$  as evaluated at  $x^*$ ?



## Answer:

- The minimum, over all 0.1% perturbations  $a \mapsto \tilde{a}$  of ugly entries in  $a$ , value of  $\tilde{a}^T x^* - b$ , is  $< -104.9$ , that is, *with 0.1% perturbations of ugly coefficients, the violation of the constraint as evaluated at the nominal solution can be as large as 450% of the right hand side!*
- With independent *random* 0.1%-perturbations of ugly coefficients,
  - the violation of the constraint *at average* is as large as **125%** of the right hand side;
  - the probability of violating the constraint by *at least* 150% of the right hand side is as large as **0.18**.
- ♣ Among 90 NETLIB problems, perturbing ugly coefficients by just **0.01%** results in violating some of the constraints, as evaluated at nominal optimal solutions,
  - by more than 50% – in 13 problems,
  - by more than 100% – in 6 problems.
  - by **210,000%** – in PILOT4.

♣ **Conclusion:** *In applications of LO, there exists a real need of a technique capable of detecting cases when data uncertainty can heavily affect the quality of the nominal solution, and in these cases to generate a “reliable” solution, one that is immunized against uncertainty.*

**Robust Optimization** is aimed at satisfying the above need.

## Uncertain Linear Optimization Problems

♣ **Definition:** *An uncertain LO problem is a collection*

$$\left\{ \min_x \{c^T x + d : Ax \leq b\} \right\}_{(c,d,A,b) \in \mathcal{U}} \quad (LO_{\mathcal{U}})$$

*of LO problems (instances)  $\min_x \{c^T x + d : Ax \leq b\}$  of common structure (i.e., with common numbers  $m$  of constraints and  $n$  of variables) with the data varying in a given uncertainty set  $\mathcal{U} \subset \mathbf{R}^{(m+1) \times (n+1)}$ .*

♠ Usually we assume that the uncertainty set is parameterized, in an affine fashion, by *perturbation vector*  $\zeta$  varying in a given *perturbation set*  $\mathcal{Z}$ :

$$\mathcal{U} = \left\{ \left[ \begin{array}{c|c} c^T & d \\ \hline A & b \end{array} \right] = \underbrace{\left[ \begin{array}{c|c} c_0^T & d_0 \\ \hline A_0 & b_0 \end{array} \right]}_{\substack{\text{nominal} \\ \text{data } D_0}} + \sum_{\ell=1}^L \zeta_{\ell} \underbrace{\left[ \begin{array}{c|c} c_{\ell}^T & d_{\ell} \\ \hline A_{\ell} & b_{\ell} \end{array} \right]}_{\substack{\text{basic} \\ \text{shifts } D_{\ell}}} : \zeta \in \mathcal{Z} \subset \mathbf{R}^L \right\}.$$

**Example:** When speaking about PILOT4, we tacitly used the following model of uncertainty:

Uncertainty affects only the ‘ugly’ coefficients  $\{a_{ij} : (i, j) \in \mathcal{J}\}$  in the constraint matrix, and every one of them is allowed to run, independently of all other coefficients, through the interval

$$[a_{ij}^n - \rho_{ij}|a_{ij}^n|, a_{ij}^n + \rho_{ij}|a_{ij}^n|]$$

- $a_{ij}^n$ : nominal values of the data
- $\rho_{ij}$ : perturbation levels (which in the experiment were set to  $\rho = 0.001$ ).

- **Perturbation set:** The box

$$\{\zeta = \{\zeta_{ij}\}_{(i,j) \in \mathcal{J}} : -\rho_{ij} \leq \zeta_{ij} \leq \rho_{ij}\}$$

- **Parameterization of the data by perturbation vector:**

$$\left[ \begin{array}{c|c} c^T & d \\ \hline A & b \end{array} \right] = \left[ \begin{array}{c|c} [c^n]^T & d^n \\ \hline A^n & b^n \end{array} \right] + \sum_{(i,j) \in \mathcal{J}} \zeta_{ij} \left[ \begin{array}{c|c} \hline e_i e_j^T & \hline \hline \end{array} \right]$$

$$\begin{aligned}
& \left\{ \min_x \{c^T x + d : Ax \leq b\} \right\}_{(c,d,A,b) \in \mathcal{U}} \\
\mathcal{U} = & \left\{ \underbrace{\begin{bmatrix} c_0^T & d_0 \\ A_0 & b_0 \end{bmatrix}}_{\text{nominal data } D_0} + \sum_{\ell=1}^L \zeta_\ell \underbrace{\begin{bmatrix} c_\ell^T & d_\ell \\ A_\ell & b_\ell \end{bmatrix}}_{\text{basic shifts } D_\ell} : \zeta \in \mathcal{Z} \subset \mathbf{R}^L \right\}. \quad (\mathbf{LO}_{\mathcal{U}})
\end{aligned}$$

♣ There is no universally defined notion of a “solution to a *family* of optimization problems,” like  $(\mathbf{LO}_{\mathcal{U}})$ .

Consider “decision environment” as follows:

A.1. All decision variables in  $(\mathbf{LO}_{\mathcal{U}})$  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.”

A.2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set  $\mathcal{U}$ .

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

$$\left\{ \min_x \{c^T x + d : Ax \leq b\} \right\}_{(c,d,A,b) \in \mathcal{U}} \quad (\mathbf{LO}_{\mathcal{U}})$$

♣ In the above decision environment, the only meaningful candidate solutions to  $(\mathbf{LO}_{\mathcal{U}})$  are the *robust feasible* ones.

Definition:  $x \in \mathbf{R}^n$  is called a *robust feasible solution* to  $(\mathbf{LO}_{\mathcal{U}})$ , if  $x$  is feasible for all instances:

$$Ax \leq b \quad \forall (c, d, A, b) \in \mathcal{U}.$$

Indeed, by [A.1](#) a meaningful candidate solution should be independent of the data, i.e., it should be just a fixed vector  $x$ . By [A.2-3](#), it should satisfy the constraints, whatever be a realization of the data from  $\mathcal{U}$ .

♠ Acting in the same “worst-case-oriented” fashion, it makes sense to quantify the quality of a candidate solution  $x$  by the *guaranteed* (the worst, over the data from  $\mathcal{U}$ ) value of the objective:

$$\sup \{c^T x + d : (c, d, A, b) \in \mathcal{U}\}$$

$$\left\{ \min_x \{c^T x + d : Ax \leq b\} \right\}_{(c,d,A,b) \in \mathcal{U}} \quad (\mathbf{LO}_{\mathcal{U}})$$

♠ Now we can associate with  $(\mathbf{LO}_{\mathcal{U}})$  the problem of finding the best, *in terms of the guaranteed value of the objective*, among the *robust feasible solutions*:

$$\min_{t,x} \{t : c^T x + d \leq t, Ax \leq b \forall (c, d, A, b) \in \mathcal{U}\} \quad (\mathbf{RC})$$

This is called the *Robust Counterpart* of  $(\mathbf{LO}_{\mathcal{U}})$ .

Note: Passing from LOs of the form

$$\min_x \{c^T x + d : Ax \leq b\}$$

to their equivalents

$$\min_{t,x} \{t : c^T x + d \leq t, Ax \leq b\}$$

we always may assume that the objective is certain, and the RC respects this equivalence.

$\Rightarrow$  We lose nothing by assuming the objective in  $(\mathbf{LO}_{\mathcal{U}})$  certain, in which case we can think of  $\mathcal{U}$  as of the set in the space  $\mathbf{R}^{m \times (n+1)}$  of the  $[A, b]$ -data, and the RC reads

$$\min_x \{c^T x : Ax \leq b \forall [A, b] \in \mathcal{U}\}. \quad (\mathbf{RC})$$

$$\left\{ \min_x \{c^T x : Ax \leq b\} \right\}_{(A,b) \in \mathcal{U}} \quad (\mathbf{LO}_{\mathcal{U}})$$

↓

$$\min_x \{c^T x : Ax \leq b \ \forall [A, b] \in \mathcal{U}\} \quad (\mathbf{RC})$$

♣ **Fact I:** The RC of uncertain LO with certain objective is a purely *constraint-wise* construction: when building the RC, we replace every constraint  $a_i^T x \leq b_i$  of the instances with its RC

$$a_i^T x \leq b_i \ \forall [a_i^T, b_i] \in \mathcal{U}_i \quad (\mathbf{RC}_i)$$

where  $\mathcal{U}_i$  is the projection of the uncertainty set  $\mathcal{U}$  on the space of data  $[a_i^T, b_i]$  of  $i$ -th constraint.

♣ **Fact II:** The RC remains intact when extending the uncertainty set  $\mathcal{U}$  to its closed convex hull.

When  $(\mathbf{LO}_{\mathcal{U}})$  has certain objective, the RC remains intact when extending  $\mathcal{U}$  to the direct product of *closed convex hulls* of  $\mathcal{U}_i$ . Thus, the transformation

$$\mathcal{U} \mapsto \mathcal{U}^+ = [\text{cl Conv}(\mathcal{U}_1)] \times \dots \times [\text{cl Conv}(\mathcal{U}_m)]$$

keeps the RC intact.

♠ From now on, we always assume uncertainty set  $\mathcal{U}$  convex, and perturbation set  $\mathcal{Z}$  – convex and closed.

$$\left\{ \min_x \{c^T x : Ax \leq b\} \right\}_{[A,b] \in \mathcal{U}} \quad (\mathbf{LO}_{\mathcal{U}})$$



$$\min_x \{c^T x : Ax \leq b \forall [A,b] \in \mathcal{U}\} \quad (\mathbf{RC})$$

♣ The central questions associated with the concept of RC are:

**A.** *What is the “computational status” of the RC? When is it possible to process the RC efficiently?*

— to be addressed in-depth below.

**B.** *How to come-up with meaningful uncertainty sets?*

— modeling issue **NOT** to be addressed in this talk.



$$\min_x \{c^T x : Ax \leq b \forall [A, b] \in \mathcal{U}\} \quad (\text{RC})$$

♣ Potentially bad news: The RC is a *semi-infinite* optimization problem (finitely many variables, infinitely many constraints) and as such can be computationally intractable.

Example: Consider the “nearly linear” semi-infinite constraint

$$\begin{aligned} & \|Px - p\|_1 \leq 1, \quad \forall [P, p] \in \mathcal{U} \\ & \mathcal{U} = \{[P_*, p] : p = B\zeta, \|\zeta\|_2 \leq 1\} \end{aligned}$$

To check whether  $x = 0$  is robust feasible is, in general, NP-hard!

$$\min_x \{c^T x : Ax \leq b \forall [A, b] \in \mathcal{U}\} \quad (\text{RC})$$

♣ Good news: The RC of an uncertain LO problem is computationally tractable, provided the uncertainty set  $\mathcal{U}$  is so.

**Explanation, I:** The RC can be written down as the optimization problem

$$\min_x \{c^T x : f_i(x) \leq 0, i = 1, \dots, m\}$$
$$f_i(x) = \sup_{[A,b] \in \mathcal{U}} [a_i^T x - b_i]$$

- The functions  $f_i(x)$  are convex (due to their origin) and efficiently computable (as maxima of affine functions over computationally tractable convex sets).
- Thus, the RC is a Convex Programming program with efficiently computable objective and constraints, and problems of this type are efficiently solvable.

♣ Recalling that the RC is a “constraint-wise” construction, all we need is to reformulate in a tractable form a *single* semi-infinite constraint

$$\begin{aligned} \forall \alpha = [a; b] \in \{\alpha^n + \mathcal{A}\zeta : \zeta \in \mathcal{Z}\} \subset \mathbf{R}^{n+1} : \\ \alpha^T[x; 1] \equiv a^T x + b \leq 0. \end{aligned} \quad (*)$$

♠ Consider several instructive cases when tractable reformulation of (\*) is easy – does not require any theory.

1. **Scenario uncertainty**  $\mathcal{Z} = \text{Conv}\{\zeta^1, \dots, \zeta^N\}$ . Setting  $\alpha^j = \alpha^n + \mathcal{A}\zeta^j$ ,  $1 \leq j \leq N$ , we get

$$(*) \Leftrightarrow \{\alpha^j[x; 1] \leq 0, 1 \leq j \leq N\}$$

2.  **$\|\cdot\|_p$ -uncertainty**  $\mathcal{Z} = \{\zeta \in \mathbf{R}^L : \|\zeta\|_p \leq 1\}$ ,

$$\|\zeta\|_p = \begin{cases} \left(\sum_{\ell=1}^L |\zeta_\ell|^p\right)^{1/p}, & 1 \leq p < \infty \\ \max_\ell |\zeta_\ell|, & p = \infty \end{cases}$$

We have

$$(*) \Leftrightarrow [\alpha^n]^T[x; 1] + \|\mathcal{A}^T[x; 1]\|_{p_*} \leq 0, \quad \frac{1}{p} + \frac{1}{p_*} = 1$$

In particular, *Interval uncertainty*

$$\{\alpha : (\alpha^n)_j - \delta_j \leq \alpha_j \leq \alpha_j^n + \delta_j, 1 \leq j \leq n+1\}$$

leads to

$$(*) \Leftrightarrow [\alpha^n]^T[x; 1] + \sum_{i=1}^n \delta_i |x_i| + \delta_{n+1} \leq 0$$

## General Well-Structured Case

**Definition:** Let us say that a set  $\mathcal{X} \subset \mathbf{R}^N$  is *well-structured*, if it admits a *well-structured representation* — a representation of the form

$$\mathcal{X} = \left\{ x \in \mathbf{R}^N : \exists u \in \mathbf{R}^M : \begin{cases} A_0x + B_0u + c_0 = 0 \\ A_1x + B_1u + c_1 \in \mathbf{K}_1 \\ \dots \\ A_Kx + B_Ku + c_K \in \mathbf{K}_K \end{cases} \right\},$$

where  $\mathbf{K}_k$ , for every  $k \leq K$ , is a simple cone, specifically,  
 — either *nonnegative orthant*  $\mathbf{R}_+^m = \{x \in \mathbf{R}^m : x \geq 0\}$ ,  $m = m_k$ ,  
 — or a *Lorentz cone*  $\mathbf{L}^m = \{x \in \mathbf{R}^m : x_m \geq \sqrt{x_1^2 + \dots + x_{m-1}^2}\}$ ,  $m = m_k$ ,  
 — or a *Semidefinite cone*  $\mathbf{S}_+^m$  — the cone of positive semidefinite matrices in the space  $\mathbf{S}^m$  of real symmetric  $m \times m$  matrices,  $m = m_k$ .

**Example 1:** The set  $\mathcal{X} = \{x \in \mathbf{R}^N : \|x\|_1 \leq 1\}$  admits *polyhedral representation*

$$\begin{aligned} \mathcal{X} &= \{x \in \mathbf{R}^N : \exists u \in \mathbf{R}^N : -u_i \leq x_i \leq u_i, \sum_i u_i \leq 1\} \\ &= \left\{ x \in \mathbf{R}^n : \exists u \in \mathbf{R}^N : A_1x + B_1u + c_1 \equiv \begin{bmatrix} u_1 - x_1 \\ u_1 + x_1 \\ \vdots \\ u_N - x_N \\ u_N + x_N \\ 1 - \sum_i u_i \end{bmatrix} \in \mathbf{R}_+^{2N+1} \right\} \end{aligned}$$

**Example 2:** The set  $\mathcal{X} = \{x \in \mathbf{R}_+^4 : x_1x_2x_3x_4 \geq 1\}$  admits *conic quadratic representation*

$$\begin{aligned} \mathcal{X} &= \left\{ x \in \mathbf{R}_+^4 : \exists u \in \mathbf{R}^3 : \begin{cases} 0 \leq u_1 \leq \sqrt{x_1x_2} \\ 0 \leq u_2 \leq \sqrt{x_3x_4} \\ 1 \leq u_3 \leq \sqrt{u_1u_2} \end{cases} \right\} \\ &= \left\{ x \in \mathbf{R}^n : \exists u \in \mathbf{R}^3 : \begin{cases} [x_1; x_2; x_3; x_4; u_1; u_2; u_3 - 1] \in \mathbf{R}_+^7 \\ [2u_1; x_1 - x_2; x_1 + x_2] \in \mathbf{L}^3 \\ [2u_2; x_3 - x_4; x_3 + x_4] \in \mathbf{L}^3 \\ [2u_3; u_1 - u_2; u_1 + u_2] \in \mathbf{L}^3 \end{cases} \right\} \end{aligned}$$

**Example 3:** The set  $\mathcal{X}$  of  $m \times n$  matrices  $X$  with nuclear norm (sum of singular values)  $\leq 1$  admits *semidefinite representation*

$$\mathcal{X} = \left\{ X \in \mathbf{R}^{m \times n} : \exists u = (U \in \mathbf{S}^m, V \in \mathbf{S}^n) : \begin{aligned} &\text{Tr}(U) + \text{Tr}(V) \leq 2 \\ &\begin{bmatrix} U & X \\ X^T & V \end{bmatrix} \succeq 0 \end{aligned} \right\}.$$

$$\mathcal{X} = \left\{ x \in \mathbf{R}^N : \exists u \in \mathbf{R}^M : \begin{cases} A_0x + B_0u + c_0 = 0 \\ A_1x + B_1u + c_1 \in \mathbf{K}_1 \\ \dots \\ A_Kx + B_ku + c_K \in \mathbf{K}_K \end{cases} \right\}, \quad (*)$$

♣ **Good news** on **w**ell-**s**tructured **r**epresentations:

• Computational tractability: *Minimizing a linear objective over a set given by (\*) reduces to solving a well-structured conic program*

$$\min_{x,u} \left\{ c^T x : \begin{cases} A_0x + B_0u + c_0 = 0 \\ A_1x + B_1u + c_1 \in \mathbf{K}_1 \\ \dots \\ A_Kx + B_ku + c_K \in \mathbf{K}_K \end{cases} \right\},$$

*and thus can be done in a theoretically (and to some extent — also practically) efficient manner by polynomial time interior point algorithms.*

• Extremely powerful expressive abilities: *w.-s.r.'s admit a simple fully algorithmic calculus which makes it easy to build a w.-s.r. for the result of a convexity-preserving operation with convex sets (like taking intersections, direct sums, affine images, inverse affine images, polars, etc.) via w.-s.r.'s of the operands.*

**As a result**, *for all practical purposes, all computationally tractable convex sets arising in Optimization admit explicit w.-s.r.'s.*



♣ Coming back to Robust Counterpart of a scalar linear inequality, we want to answer the following

Question: *How to describe in a tractable form the fact that a vector  $x$  satisfies the semi-infinite constraint*

$$\forall \alpha = [a; b] \in \{\alpha^n + \mathcal{A}\zeta : \zeta \in \mathcal{Z}\} \subset \mathbf{R}^{n+1} : \quad (*)$$

$$\alpha^T [x; 1] \equiv a^T x + b \leq 0.$$

♠ We are about to give a nice answer in the case when the perturbation set  $\mathcal{Z}$  is *well structured*.

• Substituting  $\alpha = \alpha^n + \mathcal{A}\zeta$  into  $\alpha^T [x; 1] \leq 0$ , we arrive at the inequality

$$\zeta^T \underbrace{\mathcal{A}^T [x; 1]}_{-p(x)} + \underbrace{[\alpha^n]^T [x; 1]}_{q(x)} \leq 0,$$

or, which is the same,

$$(p(x))^T \zeta \geq q(x) \quad (!)$$

Note that  $p(x)$  and  $q(x)$  are affine in  $x$  and that  $x$  satisfies (\*) if and only if (!) holds true for every  $\zeta \in \mathcal{Z}$ , that is, if and only if the relation (!), considered as a scalar linear inequality in variables  $\zeta$ , is a consequence of the constraints defining  $\mathcal{Z}$ .

Thus, our goal is to understand when a scalar linear inequality is a consequence of the system of constraints defining a well-structured set.

The answer is given by *Conic Duality*.



## Excursion to Conic Duality

♣ Situation: We are given a system of linear equations and conic constraints

$$\mathcal{A}_0 w - b_0 = 0 \ \& \ \mathcal{A}_i w - b_i \in \mathbf{K}_i, \ 1 \leq i \leq I \quad (C)$$

- $w$ : variable from Euclidean space  $E$  with inner product  $\langle \cdot, \cdot \rangle$
- $\mathbf{K}_i$ : closed convex pointed cones with nonempty interiors in Euclidean spaces  $E_i$  with inner products  $\langle \cdot, \cdot \rangle_i$
- $w \mapsto \mathcal{A}_i w - b_i : E \rightarrow E_i$ : affine mappings

♠ Case of primary interest: Every  $\mathbf{K}_i$  is

- or *nonnegative orthant*  $\mathbf{R}_+^{m_i} = \{y \in \mathbf{R}^{m_i} : y \geq 0\}$ ,  
( $E_i = \mathbf{R}^{m_i}$ ,  $\langle y, z \rangle_i = y^T z$ ),
- or *Lorentz cone*  $\mathbf{L}_+^{m_i} = \{y \in \mathbf{R}^{m_i} : y_{m_i} \geq \|[y_1; \dots; y_{m_i-1}]\|_2\}$   
( $E_i = \mathbf{R}^{m_i}$ ,  $\langle y, z \rangle_i = y^T z$ ),
- or *semidefinite cone*  $\mathbf{S}_+^{m_i}$  ( $E_i = \mathbf{S}^{m_i}$  is the space of  $m_i \times m_i$  symmetric matrices with the Frobenius inner product  $\langle y, z \rangle_i = \text{Tr}(yz) = \sum_{i,j} y_{ij} z_{ij}$ )

♠ Goal: To understand when a scalar linear inequality

$$\langle p, w \rangle \geq q \quad (*)$$

is a consequence of (C), meaning that (\*) is satisfied at every solution to (C).

♣ Given a system

$$\mathcal{A}_0 w - b_0 = 0 \ \& \ \mathcal{A}_i w - b_i \in \mathbf{K}_i, \ 1 \leq i \leq I \quad (C)$$

- $w$ : variable from Euclidean space  $(E, \langle \cdot, \cdot \rangle)$
- $\mathbf{K}_i$ : closed convex pointed cones with nonempty interiors in Euclidean spaces  $(E_i, \langle \cdot, \cdot \rangle_i)$

of linear equations and conic constraints, we want to understand when a scalar linear inequality

$$\langle p, w \rangle \geq q \quad (*)$$

is a consequence of the system.

♣ Linear aggregation:

- Choose **Lagrange multipliers**  $y_i, 0 \leq i \leq I$ , such that for  $i \geq 1$ ,  $y_i$  belongs to the cone dual to  $\mathbf{K}_i$ :

$$\forall i > 0 : y_i \in \mathbf{K}_i^* := \{y : \langle y, u \rangle_i \geq 0 \ \forall u \in \mathbf{K}_i\}$$

- Since  $y_i \in \mathbf{K}_i^*$ , the conic constraint  $\mathcal{A}_i w - b_i \in \mathbf{K}_i$  implies that  $\langle y_i, [\mathcal{A}_i w - b_i] \rangle_i \geq 0$ , that is,

$$(C) \Rightarrow \langle \mathcal{A}_i^* y_i, w \rangle \geq \langle b_i, y_i \rangle_i \quad (+)$$

$$\left[ \begin{array}{l} \mathcal{A}^* : F \rightarrow E \text{ is the conjugate of } \mathcal{A} : E \rightarrow F : \\ \langle \mathcal{A}^* y, x \rangle_E \equiv \langle y, \mathcal{A} x \rangle_F \end{array} \right]$$

Equations  $\mathcal{A}_0 w - b_0 = 0$  imply (+) when  $i = 0$ .

- Summing up inequalities (+) over  $i = 0, 1, \dots, I$ , we arrive at the following conclusion:

♠ Whenever  $y_i, 0 \leq i \leq I$  are such that  $y_i \in \mathbf{K}_i^*$  when  $i \geq 1$ , and

$$\sum_{i=0}^I \mathcal{A}_i^* y_i = p \ \text{and} \ \sum_{i=0}^I \langle b_i, y_i \rangle_i \geq q$$

the scalar linear inequality (\*) is a consequence of (C).

$$\mathcal{A}_0 w - b_0 = 0 \ \& \ \mathcal{A}_i w - b_i \in \mathbf{K}_i, \ 1 \leq i \leq I \quad (C)$$

- $w$ : variable from Euclidean space  $(E, \langle \cdot, \cdot \rangle)$
- $\mathbf{K}_i$ : closed convex pointed cones with nonempty interiors in Euclidean spaces  $(E_i, \langle \cdot, \cdot \rangle_i)$

$$\langle p, w \rangle \geq q \quad (*)$$

♣ Conic Farkas Lemma: *Let (C) be strictly feasible:*

$$\exists \bar{w} : \mathcal{A}_0 \bar{w} - b_0 = 0 \ \& \ \mathcal{A}_i \bar{w} - b_i \in \text{int} \mathbf{K}_i, \ 1 \leq i \leq I$$

*Then (\*) is a consequence of (C) if and only if (\*) can be obtained from (C) by linear aggregation, i.e. if and only if*

$$\exists y_0, \{y_i \in \mathbf{K}_i^*\}_{i=1}^I : \sum_{i=0}^I \mathcal{A}_i^* y_i = p \ \text{and} \ \sum_{i=0}^I \langle b_i, y_i \rangle_i \geq q.$$

*When all  $\mathbf{K}_i$  are nonnegative orthants, the statement remains true when replacing strict feasibility with feasibility.*

♣ Corollary [Conic Duality]: Consider a conic problem

$$\text{Opt}(P) = \min_x \{ \langle c, w \rangle : \mathcal{A}_0 w = b_0 \ \& \ \mathcal{A}_i w - b_i \in \mathbf{K}_i, 1 \leq i \leq I \} \quad (P)$$

along with its dual problem

$$\text{Opt}(D) = \max_{y_1, \dots, y_I} \left\{ \sum_{i=0}^I \langle b_i, y_i \rangle_i : y_i \in \mathbf{K}_i^* \ \forall i \geq 1, \sum_{i=0}^I \mathcal{A}_i^* y_i = c \right\} \quad (D)$$

Weak Duality: One always has  $\text{Opt}(P) \geq \text{Opt}(D)$ .

Strong Duality: Assuming that (P) is strictly feasible and below bounded, (D) is solvable, and

$$\text{Opt}(P) = \text{Opt}(D).$$

When all  $\mathbf{K}_i$  are nonnegative orthants, Strong Duality holds when strict feasibility of (P) is replaced with feasibility.

Proof of Strong Duality: Apply Conic Farkas Lemma to the inequality

$$\langle c, x \rangle \geq \text{Opt}(P)$$

which clearly is the consequence of the system of constraints in (P).

♠ Note: Nonnegative orthants/Lorentz/Semidefinite cones are self-dual.

$\Rightarrow$  When all  $\mathbf{K}_i$  are of nonnegative orthants, or Lorentz cones, or semidefinite cones, the constraints  $y_i \in \mathbf{K}_i^*$  in (D) take the form  $y_i \in \mathbf{K}_i$ .

♣ Conic Farkas Lemma is really deep already in the case when all conic inequalities in question are scalar:  $K_i = \mathbf{R}_+$  for all  $i$ . In this case it becomes the usual **Farkas Lemma**: *A linear inequality*

$$p^T x \geq q \quad (*)$$

is a consequence of *solvable system of linear inequalities*

$$Ax \geq b$$

if and only if there exists  $y \geq 0$  such that

$$A^T y = p \ \& \ b^T y \geq q,$$

or, which is the same, if and only if the target inequality (\*) can be obtained by linear aggregation (taking weighted sum with nonnegative weights) of inequalities from the system and the identically true inequality

$$0^T x \geq -1.$$

**While simply-looking, this fact is really deep.** Indeed, consider the following derivation

$$\left\{ \begin{array}{l} -1 \leq x \leq 1 \\ -1 \leq y \leq 1 \end{array} \right\} \rightarrow x^2 \leq 1, y^2 \leq 1 \Rightarrow x^2 + y^2 \leq 2$$

$$\Rightarrow x + y = 1 \cdot x + 1 \cdot y \leq \sqrt{1^2 + 1^2} \sqrt{x^2 + y^2} \leq \sqrt{2} \sqrt{2} = 2$$

[we have used Cauchy Inequality]

In this chain of implications, the starting point is a solvable system of linear inequalities, and the conclusion “ $x + y \leq 2$ ” is a linear inequality as well; however, the steps use “highly nonlinear” arguments. According to Farkas Lemma, every chain of this type, whatever long and with whatever complicated steps, can be replaced with linear aggregation.

*A statement with that huge “predictive power” indeed must be deep!*

## Back to Robust Linear Optimization

♣ Now we are ready to answer the question

When the scalar linear inequality in variables  $\zeta$

$$(p(x))^T \zeta \geq q(x) \quad (!)$$

with  $p(x), q(x)$  affine in  $x$  holds true for all  $\zeta \in \mathcal{Z}$ , where  $\mathcal{Z}$  is well structured:

$$\mathcal{Z} = \left\{ \zeta \in \mathbf{R}^L : \exists u \in \mathbf{R}^M : \begin{cases} A_0 \zeta + B_0 u + c_0 = 0 \\ A_1 \zeta + B_1 u + c_1 \in \mathbf{K}_1 \\ \dots \\ A_K \zeta + B_K u + c_K \in \mathbf{K}_K \end{cases} \right\}, \quad (*)$$

$\mathbf{K}_i$ : Nonnegative orthants/Lorentz/Semidefinite cones.

♠ Assume that either all  $\mathbf{K}_k$  are nonnegative orthants, or  $(*)$  is strictly feasible. We ask when the scalar linear inequality

$$(p(x))^T \zeta + 0^T u \geq q(x)$$

in variables  $w = [\zeta; u]$  holds true for all  $[\zeta; u]$  satisfying the constraints in  $(*)$ . By conic Farkas Lemma, this is so if and only if there exist  $y_0, y_1, \dots, y_K$  such that

$$\begin{aligned} A_0^T y_0 + \sum_{k=1}^K A_k^T y_k &= p(x), \\ B_0^T y_0 + \sum_{k=1}^K B_k^T y_k &= 0, \\ -y_0^T c_0 - \sum_{k=1}^K y_k^T c_k &\geq q(x) \\ y_k &\in \mathbf{K}_k^* = \mathbf{K}_k, \quad 1 \leq k \leq K \end{aligned} \quad (**)$$

Since  $p(x)$  and  $q(x)$  are affine in  $x$ ,  $(**)$  is a system of linear equations and conic constraints in variables  $x, y_0, y_1, \dots, y_K$

♠ We conclude that The set  $\mathcal{X}$  of those  $x$  for which  $(!)$  holds true for all  $\zeta \in \mathcal{Z}$  is well-structured:

$$\mathcal{X} = \{x : \exists y_0, y_1, \dots, y_k : (x, y_0, \dots, y_k) \text{ satisfy } (**)\}$$