Lecture 1

Robust Linear Optimization: Motivation, Concepts, Tractability

In this lecture, we introduce the concept of the uncertain Linear Optimization problem and its Robust Counterpart, and study the computational issues associated with the emerging optimization problems.

1.1 Data Uncertainty in Linear Optimization

1.1.1 Linear Optimization Problem, its data and structure

The Linear Optimization (LO) *problem* is

$$\min_x \left\{ c^T x + d : Ax \leq b \right\}; \tag{1.1.1}$$

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

Clearly, the constant term d in the objective, while affecting the optimal value, does not affect the optimal solution, this is why it is traditionally skipped. When treating the LO problems with *uncertain data* there are good reasons not to neglect this constant term.

When speaking about optimization (or whatever other) problems, we usually distinguish between problems's *structure* and problems's *data*. When asked "what is the data of the LO problem (1.1.1)," everybody will give the same answer: "the data of the problem are the collection (c, d, A, b)." . As about the structure of (1.1.1), it, given the form in which we write the problem down, is specified by the number m of constraints and the number n of variables.

Usually not all constraints of an LO program, as it arises in applications, are of the form $a^T x \leq$ const; there can be linear "$\geq$" inequalities and linear equalities as well. Clearly, the constraints of the latter two types can be represented equivalently by linear "$\leq$" inequalities, and we will assume henceforth that these are the only constraints in the problem.

1.1.2 Data uncertainty: sources

Typically, the data of real world LOs is not known exactly when the problem is being solved. The most common reasons for data uncertainty are as follows:

- Some of data entries (future demands, returns, etc.) do not exist when the problem is solved and hence are replaced with their forecasts. These data entries are thus subject to *prediction errors*;
- Some of the data (parameters of technological devices/processes, contents associated with raw materials, etc.) cannot be measured exactly, and their true values drift around the measured "nominal" values; these data are subject to *measurement errors*;
- Some of the decision variables (intensities with which we intend to use various technological processes, parameters of physical devices we are designing, etc.) cannot be implemented exactly as computed. The resulting *implementation errors* are equivalent to appropriate artificial data uncertainties.

  Indeed, the contribution of a particular decision variable x_j to the left hand side of constraint i is the product $a_{ij}x_j$. A typical implementation error can be modeled as $x_j \mapsto (1+\xi_j)x_j + \eta_j$, where ξ_j is the multiplicative, and η_j is the additive component of the error. The effect of this error is *as if* there were no implementation error at all, but the coefficient a_{ij} got the multiplicative perturbation: $a_{ij} \mapsto a_{ij}(1+\xi_j)$, and the right hand side b_i of the constraint got the additive perturbation $b_i \mapsto b_i - \eta_j a_{ij}$.

1.1.3 Data uncertainty: dangers

In the traditional LO methodology, a small data uncertainty (say, 0.1% or less) is just ignored; the problem is solved *as if* the given ("nominal") data were exact, and the resulting *nominal* optimal solution is what is recommended for use, in hope that small data uncertainties will not affect significantly the feasibility and optimality properties of this solution, or that small adjustments of the nominal solution will be sufficient to make it feasible. In fact these hopes are not necessarily justified, and sometimes even small data uncertainty deserves significant attention. We are about to present two instructive examples of this type.

Motivating example I: Synthesis of Antenna Arrays.

Consider a monochromatic transmitting antenna placed at the origin. Physics says that

1. The directional distribution of energy sent by the antenna can be described in terms of *antenna's diagram* which is a complex-valued function $D(\delta)$ of a 3D direction δ. The directional distribution of energy sent by the antenna is proportional to $|D(\delta)|^2$.
2. When the antenna is comprised of several antenna elements with diagrams $D_1(\delta)$,..., $D_k(\delta)$, the diagram of the antenna is just the sum of the diagrams of the elements.

In a typical Antenna Design problem, we are given several antenna elements with diagrams $D_1(\delta)$,...,$D_n(\delta)$ and are allowed to multiply these diagrams by complex *weights* x_i (which in reality corresponds to modifying the output powers and shifting the phases of the elements). As a result, we can obtain, as a diagram of the array, any function of the form

$$D(\delta) = \sum_{j=1}^{n} x_j D_j(\delta),$$

and our goal is to find the weights x_j which result in a diagram as close as possible, in a prescribed sense, to a given "target diagram" $D_*(\delta)$.

Example 1.1 Antenna Design Consider a planar antenna comprised of a central circle and 9 concentric rings of the same area as the circle (figure 1.1.a) in the XY-plane ("Earth's surface"). Let the wavelength be $\lambda = 50$cm, and the outer radius of the outer ring be 1 m (twice the wavelength).

One can easily see that the diagram of a ring $\{a \leq r \leq b\}$ in the plane XY (r is the distance from a point to the origin) as a function of a 3-dimensional direction δ depends on the altitude (the angle θ between the direction and the plane) only. The resulting function of θ turns out to be *real-valued*, and its analytic expression is

$$D_{a,b}(\theta) = \frac{1}{2}\int\limits_a^b \left[\int\limits_0^{2\pi} r\cos\left(2\pi r\lambda^{-1}\cos(\theta)\cos(\phi)\right)d\phi\right]dr.$$

Fig. 1.1.b represents the diagrams of our 10 rings for $\lambda = 50$cm.

Assume that our goal is to design an array with a real-valued diagram which should be axial symmetric with respect to the Z-axis and should be "concentrated" in the cone $\pi/2 \geq \theta \geq \pi/2 - \pi/12$. In other words, our target diagram is a real-valued function $D_*(\theta)$ of the altitude θ with $D_*(\theta) = 0$ for $0 \leq \theta \leq \pi/2 - \pi/12$ and $D_*(\theta)$ somehow approaching 1 as θ approaches $\pi/2$. The target diagram $D_*(\theta)$ used in this example is given in figure 1.1.c (blue).

Let us measure the discrepancy between a synthesized diagram and the target one by the Tschebyshev distance, taken along the equidistant 240-point grid of altitudes, i.e., by the quantity

$$\tau = \max_{i=1,\dots,240}\left|D_*(\theta_i) - \sum_{j=1}^{10} x_j \underbrace{D_{r_{j-1},r_j}(\theta_i)}_{D_j(\theta_i)}\right|, \quad \theta_i = \frac{i\pi}{480}.$$

Our design problem is simplified considerably by the fact that the diagrams of our "building blocks" and the target diagram are real-valued; thus, we need no complex numbers, and the problem we should finally solve is

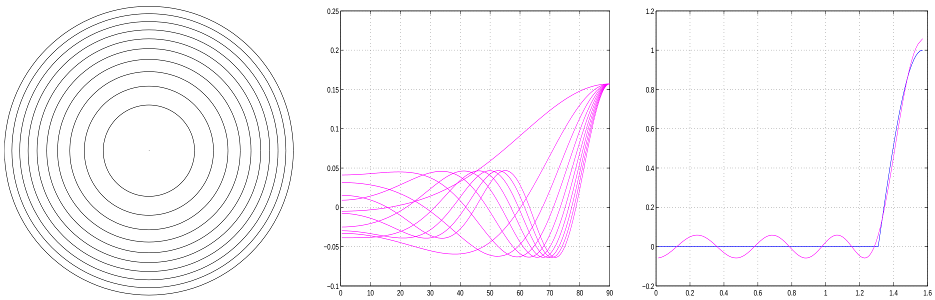
$$\min_{\tau\in\mathbb{R},x\in\mathbb{R}^{10}}\left\{\tau : -\tau \leq D_*(\theta_i) - \sum_{j=1}^{10} x_j D_j(\theta_i) \leq \tau,\ i = 1,\dots,240\right\}. \tag{1.1.2}$$

This is a simple LP program; its optimal solution x^* results in the diagram depicted at figure 1.1.c (magenta). The uniform distance between the actual and the target diagrams is ≈ 0.0589 (recall that the target diagram varies from 0 to 1).

Now recall that our design variables are characteristics of certain physical devices. In reality, of course, we cannot tune the devices to have precisely the optimal characteristics x_j^*; the best we may hope for is that the actual characteristics $x_j^{\rm fct}$ will coincide with the desired values x_j^* within a small margin ρ, say, $\rho = 0.1\%$ (this is a fairly high accuracy for a physical device):

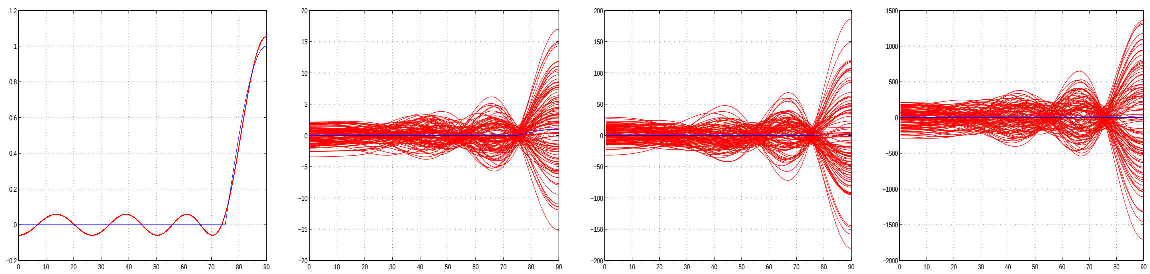
$$x_j^{\rm fct} = (1+\xi_j)x_j^*,\ |\xi_j| \leq \rho = 0.001.$$

It is natural to assume that the *actuation errors* ξ_j are random with the mean value equal to 0; it is perhaps not a great sin to assume that these errors are independent of each other. Note that as it was already explained, the consequences of our actuation errors are *as if* there were no actuation errors at all, but the coefficients $D_j(\theta_i)$ of variables x_j in (1.1.2) were subject to perturbations $D_j(\theta_i) \mapsto (1+\xi_j)D_j(\theta_i)$.



(a) (b) (c)

Figure 1.1: Synthesis of antennae array.
(a): 10 array elements of equal areas in the XY-plane; the outer radius of the largest ring is 1m, the wavelength is 50cm.
(b): "building blocks" — the diagrams of the rings as functions of the altitude angle θ.
(c): the target diagram (blue) and the synthesized diagram (magenta).



$\rho = 0$ $\rho = 0.0001$ $\rho = 0.001$ $\rho = 0.01$

Figure 1.2: "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	0.059	1.280	5.671	14.04	11.42	56.84	176.6	39.25	506.5	1484
energy concentration	85.1%	0.5%	16.4%	51.0%	0.1%	16.5%	48.3%	0.5%	14.9%	47.1%

Table 1.1: Quality of nominal antenna design: dream and reality. Data over 100 samples of actuation errors per each uncertainty level ρ.

Since the actual weights differ from their desired values x_j^*, the actual (random) diagram of our array of antennae will differ from the "nominal" one we see on figure 1.1.c. How large could be the difference? Looking at figure 1.2, we see that the difference can be dramatic. The diagrams corresponding to $\rho > 0$ are not even the worst case: given ρ, we just have taken as $\{\xi_j\}_{j=1}^{10}$ 100 samples of 10 independent numbers distributed uniformly in $[-\rho, \rho]$ and have plotted the diagrams corresponding to $x_j = (1 + \xi_j)x_j^*$. Pay attention not only to the shape, but also to the scale (table 1.1): the target diagram varies from 0 to 1, and the nominal diagram (the one corresponding to the exact optimal x_j) differs from the target by no more than by 0.0589 (this is the optimal value in the "nominal" problem (1.1.2)). The data in table 1.1 show that when $\rho = 0.001$, the typical $\| \cdot \|_\infty$ distance between the actual diagram and the target one is by 3 (!) orders of magnitude larger. Another meaningful way, also presented in table 1.1, to understand what is the quality of our design is via *energy concentration* – the fraction of the total emitted energy which "goes up," that is, is emitted along the spatial angle of directions forming angle at most $\pi/12$ with the Z-axis. For the nominal design, *the dream* (i.e., with no actuation errors) energy concentration is as high as 85% – quite respectable, given that the spatial angle in question forms just 3.41% of the entire hemisphere. This high concentration, however, exists only in our imagination, since actuation errors of magnitude ρ as low as 0.01% reduce the average energy concentration (which, same as the diagram itself, now becomes random) to just 16%; the lower 10% quantile of this random quantity is as small as 2.2% – 1.5 times less than the fraction (3.4%) which the "going up" directions form among all directions. The bottom line is that "un reality" our nominal optimal design is completely meaningless.

Motivating example II: `NETLIB` Case Study

`NETLIB` includes about 100 not very large LOs, mostly of real-world origin, used as the standard benchmark for LO solvers. In the study to be described, we used this collection in order to understand how "stable" are the feasibility properties of the standard – "nominal" – optimal solutions with respect to small uncertainty in the data. To motivate the methodology of this "case study", here is the constraint # 372 of the problem `PILOT4` from `NETLIB`:

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

The related *nonzero* coordinates in the optimal solution x^* of the problem, as reported by CPLEX (one of the best commercial LP solvers), are as follows:

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

The indicated optimal solution makes (C) an equality within machine precision.

Observe that most of the coefficients in (C) are "ugly reals" like -15.79081 or -84.644257. We have all reasons to believe that coefficients of this type characterize certain technological devices/processes, and as such *they could hardly be known to high accuracy*. It is quite natural to assume that the "ugly coefficients" are in fact uncertain – they coincide with the "true" values of the corresponding data within accuracy of 3-4 digits, not more. The only exception is the coefficient 1 of x_{880} – it perhaps reflects the structure of the underlying model and is therefore exact – "certain".

Assuming that the uncertain entries of a are, say, 0.1%-accurate approximations of unknown entries of the "true" vector of coefficients $\widetilde{a}$, we looked what would be the effect of this uncertainty on the validity of the "true" constraint $\widetilde{a}^T x \geq b$ at x^*. Here is what we have found:

• The minimum (over all vectors of coefficients $\widetilde{a}$ compatible with our "0.1%-uncertainty hypothesis") value of $\widetilde{a}^T x^* - b$, is < -104.9; in other words, the violation of the constraint can be as large as 450% of the right hand side!

• Treating the above worst-case violation as "too pessimistic" (why should the true values of all uncertain coefficients differ from the values indicated in (C) in the "most dangerous" way?), consider a more realistic measure of violation. Specifically, assume that the true values of the uncertain coefficients in (C) are obtained from the "nominal values" (those shown in (C)) by random perturbations $a_j \mapsto \widetilde{a}_j = (1 + \xi_j)a_j$ with independent and, say, uniformly distributed on $[-0.001, 0.001]$ "relative perturbations" ξ_j. What will be a "typical" relative violation

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

of the "true" (now random) constraint $\widetilde{a}^T x \geq b$ at x^*? The answer is nearly as bad as for the worst scenario:

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

Table 2.1. 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 "clearly uncertain" data coefficients can make the "nominal" optimal solution x^* heavily infeasible and thus – practically meaningless.*

A "case study" reported in [8] shows that the phenomenon we have just described is not an exception – in 13 of 90 *NETLIB* Linear Programming problems considered in this study, already 0.01%-perturbations of "ugly" coefficients result in violations of some constraints as evaluated at the nominal optimal solutions by more than 50%. In 6 of these 13 problems the magnitude of constraint violations was over 100%, and in PILOT4 — "the champion" — it was as large as 210,000%, that is, 7 orders of magnitude larger than the relative perturbations in the data.

The conclusion is as follows:

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.

We are about to introduce the *Robust Counterpart* approach to uncertain LO problems aimed at coping with data uncertainty.

1.2 Uncertain Linear Problems and their Robust Counterparts

1.2.1 Uncertain LO problem

Definition 1.1 *An uncertain Linear Optimization problem is a collection*

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

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

We always assume that the uncertainty set is parameterized, in an affine fashion, by *perturbation vector* ζ varying in a given *perturbation set* $\mathcal{Z}$:

$$\mathcal{U} = \left\{ \left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = \underbrace{\left[\begin{array}{c|c} c_0^T & d_0 \\ \hline A_0 & b_0 \end{array}\right]}_{\substack{\text{nominal} \\ \text{data } D_0}} + \sum_{\ell=1}^{L} \zeta_\ell \underbrace{\left[\begin{array}{c|c} c_\ell^T & d_\ell \\ \hline A_\ell & b_\ell \end{array}\right]}_{\substack{\text{basic} \\ \text{shifts } D_\ell}} : \zeta \in \mathcal{Z} \subset \mathbb{R}^L \right\}. \qquad (1.2.1)$$

For example, when speaking about `PILOT4`, we, for the sake of simplicity, tacitly assumed uncertainty only in the constraint matrix, specifically, as follows: every coefficient a_{ij} is allowed to vary, independently of all other coefficients, in the interval $[a^{\rm n}_{ij} - \rho_{ij}|a^{\rm n}_{ij}|, a^{\rm n}_{ij} + \rho_{ij}|a^{\rm n}_{ij}|]$, where $a^{\rm n}_{ij}$ is the nominal value of the coefficient — the one in the data file of the problem as presented in NETLIB, and ρ_{ij} is the perturbation level, which in our experiment was set to 0.001 for all "ugly" coefficients $a^{\rm n}_{ij}$ and was set to 0 for "nice" coefficients, like the coefficient 1 at $\mathtt{x}_{880}$. Geometrically, the corresponding perturbation set is just a box

$$\zeta \in \mathcal{Z} = \{\zeta = \{\zeta_{ij} \in [-1,1]\}_{i,j: a^{\rm n}_{ij} \text{ is ugly}}\},$$

and the parameterization of the a_{ij}-data by the perturbation vector is

$$a_{ij} = \begin{cases} a^{\rm n}_{ij}(1+\zeta_{ij}), & a^{\rm n}_{ij} \text{ is ugly} \\ a^{\rm n}_{ij}, & \text{otherwise} \end{cases}$$

Remark 1.1 *If the perturbation set* $\mathcal{Z}$ *in* (1.2.1) *itself is represented as the image of another set* $\widehat{\mathcal{Z}}$ *under affine mapping* $\xi \mapsto \zeta = p + P\xi$, *then we can pass from perturbations* ζ *to perturbations*

ξ:

$$\begin{array}{rcl}\mathcal{U} &=& \left\{\left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = D_0 + \sum\limits_{\ell=1}^{L} \zeta_\ell D_\ell : \zeta \in \mathcal{Z}\right\} \\ &=& \left\{\left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = D_0 + \sum\limits_{\ell=1}^{L} [p_\ell + \sum\limits_{k=1}^{K} P_{\ell k}\xi_k] D_\ell : \xi \in \widehat{\mathcal{Z}}\right\} \\ &=& \left\{\left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = \underbrace{\left[D_0 + \sum\limits_{\ell=1}^{L} p_\ell D_\ell\right]}_{\widehat{D}_0} + \sum\limits_{k=1}^{K} \xi_k \underbrace{\left[\sum\limits_{\ell=1}^{L} P_{\ell k} D_\ell\right]}_{\widehat{D}_k} : \xi \in \widehat{\mathcal{Z}}\right\}.\end{array}$$

It follows that when speaking about perturbation sets with simple geometry (parallelotopes, ellipsoids, etc.), we can normalize these sets to be "standard." For example, a parallelotope is by definition an affine image of a unit box $\{\xi \in \mathbb{R}^k : -1 \leq \xi_j \leq 1, j = 1, ..., k\}$, which gives us the possibility to work with the unit box instead of a general parallelotope. Similarly, an ellipsoid is by definition the image of a unit Euclidean ball $\{\xi \in \mathbb{R}^k : \|x\|_2^2 \equiv x^T x \leq 1\}$ under affine mapping, so that we can work with the standard ball instead of the ellipsoid, etc. We will use this normalization whenever possible.

1.2.2 Robust Counterpart of Uncertain LO

Note that a *family* of optimization problems like ($\mathrm{LO}_\mathcal{U}$), in contrast to a single optimization problem, is not associated by itself with the concepts of feasible/optimal solution and optimal value. How to define these concepts depends on the underlying "decision environment." Here we focus on an environment with the following characteristics:

A.1. All decision variables in ($\mathrm{LO}_\mathcal{U}$) 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 $\mathcal{U}$ given by (1.2.1).

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

Note that A.1 – A.3 are *assumptions* on our decision environment (in fact, the strongest ones within the methodology we are presenting); while being meaningful, these assumptions in no sense are automatically valid; In the mean time, we shall consider relaxed versions of these assumptions and consequences of these relaxations.

Assumptions A.1 — A.3 determine, essentially in a unique fashion, what are the meaningful, "immunized against uncertainty," feasible solutions to the uncertain problem ($\mathrm{LO}_\mathcal{U}$). By A.1, these should be fixed vectors; by A.2 and A.3, they should be *robust feasible*, that is, they should satisfy all the constraints, whatever the realization of the data from the uncertainty set. We have arrived at the following definition.

Definition 1.2 *A vector $x \in \mathbb{R}^n$ is a robust feasible solution to* ($\mathrm{LO}_\mathcal{U}$)*, if it satisfies all realizations of the constraints from the uncertainty set, that is,*

$$Ax \leq b \quad \forall (c, d, A, b) \in \mathcal{U}. \tag{1.2.2}$$

As for the objective value to be associated with a robust feasible) solution, assumptions A.1 — A.3 do not prescribe it in a unique fashion. However, "the spirit" of these worst-case-oriented assumptions leads naturally to the following definition:

Definition 1.3 *Given a candidate solution x, the robust value $\widehat{c}(x)$ of the objective in* ($\mathrm{LO}_{\mathcal{U}}$) *at x is the largest value of the "true" objective $c^T x + d$ over all realizations of the data from the uncertainty set:*

$$\widehat{c}(x) = \sup_{(c,d,A,b)\in\mathcal{U}} [c^T x + d]. \tag{1.2.3}$$

After we agree what are meaningful candidate solutions to the uncertain problem ($\mathrm{LO}_{\mathcal{U}}$) and how to quantify their quality, we can seek the best robust value of the objective among all robust feasible solutions to the problem. This brings us to the central concept of our methodology, *Robust Counterpart* of an uncertain optimization problem, which is defined as follows:

Definition 1.4 *The Robust Counterpart of the uncertain LO problem* ($\mathrm{LO}_{\mathcal{U}}$) *is the optimization problem*

$$\min_x \left\{ \widehat{c}(x) = \sup_{(c,d,A,b)\in\mathcal{U}} [c^T x + d] : Ax \le b \ \forall (c,d,A,b) \in \mathcal{U} \right\} \tag{1.2.4}$$

of minimizing the robust value of the objective over all robust feasible solutions to the uncertain problem.

An optimal solution to the Robust Counterpart is called a robust optimal solution to ($\mathrm{LO}_{\mathcal{U}}$), *and the optimal value of the Robust Counterpart is called the robust optimal value of* ($\mathrm{LO}_{\mathcal{U}}$).

In a nutshell, the robust optimal solution is simply "the best uncertainty-immunized" solution we can associate with our uncertain problem.

1.2.3 More on Robust Counterparts

We start with several useful observations.

A. The Robust Counterpart (1.2.4) of $\mathrm{LO}_{\mathcal{U}}$ can be rewritten equivalently as the problem

$$\min_{x,t} \left\{ t : \begin{array}{rcl} c^T x - t & \le & -d \\ Ax & \le & b \end{array} \right\} \forall (c,d,A,b) \in \mathcal{U} \right\}. \tag{1.2.5}$$

Note that we can arrive at this problem in another fashion: we first introduce the extra variable t and rewrite instances of our uncertain problem ($\mathrm{LO}_{\mathcal{U}}$) equivalently as

$$\min_{x,t} \left\{ t : \begin{array}{rcl} c^T x - t & \le & -d \\ Ax & \le & b \end{array} \right\},$$

thus arriving at an equivalent to ($\mathrm{LO}_{\mathcal{U}}$) uncertain problem in variables x, t with the objective t that is not affected by uncertainty at all. The RC of the reformulated problem is exactly (1.2.5). We see that

> *An uncertain LO problem can always be reformulated as an uncertain LO problem with certain objective. The Robust Counterpart of the reformulated problem has the same objective as this problem and is equivalent to the RC of the original uncertain problem.*

As a consequence, *we lose nothing when restricting ourselves with uncertain LO programs with certain objectives* and we shall frequently use this option in the future.

> We see now why the constant term d in the objective of (1.1.1) should not be neglected, or, more exactly, should not be neglected if it is uncertain. When d is certain, we can account for it by the shift $t \mapsto t - d$ in the slack variable t which affects only the optimal value, but not the optimal solution to the Robust Counterpart (1.2.5). When d is uncertain, there is no "universal" way to eliminate d without affecting the optimal solution to the Robust Counterpart (where d plays the same role as the right hand sides of the original constraints).

B. Assuming that (LO$_{\mathcal{U}}$) is with certain objective, the Robust Counterpart of the problem is

$$\min_x \left\{c^T x + d : Ax \leq b,\, \forall (A, b) \in \mathcal{U}\right\} \tag{1.2.6}$$

(note that the uncertainty set is now a set in the space of the constraint data $[A, b]$). We see that

The Robust Counterpart of an uncertain LO problem with a certain objective is a purely "constraint-wise" construction: to get RC, we act as follows:

- *preserve the original certain objective as it is, and*
- *replace every one of the original constraints*

$$(Ax)_i \leq b_i \Leftrightarrow a_i^T x \leq b_i \tag{C$_i$}$$

(a_i^T is i-th row in A) with its Robust Counterpart

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

where $\mathcal{U}_i$ is the projection of $\mathcal{U}$ on the space of data of i-th constraint:

$$\mathcal{U}_i = \{[a_i; b_i] : [A, b] \in \mathcal{U}\}.$$

In particular,

The RC of an uncertain LO problem with a certain objective remains intact when the original uncertainty set $\mathcal{U}$ is extended to the direct product

$$\widehat{\mathcal{U}} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_m$$

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

Example 1.2 The RC of the system of uncertain constraints

$$\{x_1 \geq \zeta_1,\, x_2 \geq \zeta_2\} \tag{1.2.7}$$

with $\zeta \in \mathcal{U} := \{\zeta_1 + \zeta_2 \leq 1, \zeta_1, \zeta_2 \geq 0\}$ is the infinite system of constraints

$$x_1 \geq \zeta_1,\, x_1 \geq \zeta_2 \;\; \forall \zeta \in \mathcal{U};$$

on variables x_1, x_2. The latter system is clearly equivalent to the pair of constraints

$$x_1 \geq \max_{\zeta \in \mathcal{U}} \zeta_1 = 1,\;\; x_2 \geq \max_{\zeta \in \mathcal{U}} \zeta_2 = 1. \tag{1.2.8}$$

The projections of $\mathcal{U}$ to the spaces of data of the two uncertain constraints (1.2.7) are the segments $\mathcal{U}_1 = \{\zeta_1 : 0 \leq \zeta_1 \leq 1\}$, $\mathcal{U}_2 = \{\zeta_2 : 0 \leq \zeta_2 \leq 1\}$, and the RC of (1.2.7) w.r.t. the uncertainty set $\widehat{\mathcal{U}} = \mathcal{U}_1 \times \mathcal{U}_2 = \{\zeta \in \mathbb{R}^2 : 0 \leq \zeta_1, \zeta_2 \leq 1\}$ clearly is (1.2.8).

> The conclusion we have arrived at seems to be counter-intuitive: it says that it is immaterial whether the perturbations of data in different constraints are or are not linked to each other, while intuition says that such a link should be important. We shall see later (lecture 5) that this intuition is valid when a more advanced concept of *Adjustable* Robust Counterpart is considered.

C. If x is a robust feasible solution of (C_i), then x remains robust feasible when we extend the uncertainty set $\mathcal{U}_i$ to its convex hull $\mathrm{Conv}(\mathcal{U}_i)$. Indeed, if $[\bar{a}_i; \bar{b}_i] \in \mathrm{Conv}(\mathcal{U}_i)$, then

$$[\bar{a}_i; \bar{b}_i] = \sum_{j=1}^{J} \lambda_j [a_i^j; b_i^j],$$

with appropriately chosen $[a_i^j; b_i^j] \in \mathcal{U}_i$, $\lambda_j \geq 0$ such that $\sum\limits_j \lambda_j = 1$. We now have

$$\bar{a}_i^T x = \sum_{j=1}^{J} \lambda_j [a_i^j]^T x \leq \sum_j \lambda_j b_i^j = \bar{b}_i,$$

where the inequality is given by the fact that x is feasible for $\mathrm{RC}(\mathrm{C}_i)$ and $[a_i^j; b_i^j] \in \mathcal{U}_i$. We see that $\bar{a}_i^T x \leq \bar{b}_i$ for all $[\bar{a}_i; \bar{b}_i] \in \mathrm{Conv}(\mathcal{U}_i)$, QED.

By similar reasons, the set of robust feasible solutions to (C_i) remains intact when we extend $\mathcal{U}_i$ to the closure of this set. Combining these observations with **B**, we arrive at the following conclusion:

> *The Robust Counterpart of an uncertain LO problem with a certain objective remains intact when we extend the sets $\mathcal{U}_i$ of uncertain data of respective constraints to their closed convex hulls, and extend $\mathcal{U}$ to the direct product of the resulting sets.*
>
> *In other words, we lose nothing when assuming from the very beginning that the sets $\mathcal{U}_i$ of uncertain data of the constraints are closed and convex, and $\mathcal{U}$ is the direct product of these sets.*

In terms of the parameterization (1.2.1) of the uncertainty sets, the latter conclusion means that

> *When speaking about the Robust Counterpart of an uncertain LO problem with a certain objective, we lose nothing when assuming that the set $\mathcal{U}_i$ of uncertain data of i-th constraint is given as*
>
> $$\mathcal{U}_i = \left\{ [a_i; b_i] = [a_i^0; b_i^0] + \sum_{\ell=1}^{L_i} \zeta_\ell [a_i^\ell; b_i^\ell] : \zeta \in \mathcal{Z}_i \right\}, \tag{1.2.9}$$
>
> *with a closed and convex perturbation set $\mathcal{Z}_i$.*

D. An important modeling issue. In the usual — with certain data — Linear Optimization, constraints can be modeled in various equivalent forms. For example, we can write:

$$\begin{array}{ll} (a) & a_1 x_1 + a_2 x_2 \leq a_3 \\ (b) & a_4 x_1 + a_5 x_2 = a_6 \\ (c) & x_1 \geq 0, x_2 \geq 0 \end{array} \tag{1.2.10}$$

or, equivalently,

$$\begin{array}{ll} (a) & a_1x_1 + a_2x_2 \le a_3 \\ (b.1) & a_4x_1 + a_5x_2 \le a_6 \\ (b.2) & -a_5x_1 - a_5x_2 \le -a_6 \\ (c) & x_1 \ge 0, x_2 \ge 0. \end{array} \tag{1.2.11}$$

Or, equivalently, by adding a slack variable s,

$$\begin{array}{ll} (a) & a_1x_1 + a_2x_2 + s = a_3 \\ (b) & a_4x_1 + a_5x_2 = a_6 \\ (c) & x_1 \ge 0, x_2 \ge 0, s \ge 0. \end{array} \tag{1.2.12}$$

However, when (part of) the data $a_1, ..., a_6$ become *uncertain*, not all of these equivalences remain valid: the RCs of our now uncertainty-affected systems of constraints are not equivalent to each other. Indeed, denoting the uncertainty set by $\mathcal{U}$, the RCs read, respectively,

$$\left.\begin{array}{ll} (a) & a_1x_1 + a_2x_2 \le a_3 \\ (b) & a_4x_1 + a_5x_2 = a_6 \\ (c) & x_1 \ge 0, x_2 \ge 0 \end{array}\right\} \forall a = [a_1; ...; a_6] \in \mathcal{U}. \tag{1.2.13}$$

$$\left.\begin{array}{ll} (a) & a_1x_1 + a_2x_2 \le a_3 \\ (b.1) & a_4x_1 + a_5x_2 \le a_6 \\ (b.2) & -a_5x_1 - a_5x_2 \le -a_6 \\ (c) & x_1 \ge 0, x_2 \ge 0 \end{array}\right\} \forall a = [a_1; ...; a_6] \in \mathcal{U}. \tag{1.2.14}$$

$$\left.\begin{array}{ll} (a) & a_1x_1 + a_2x_2 + s = a_3 \\ (b) & a_4x_1 + a_5x_2 = a_6 \\ (c) & x_1 \ge 0, x_2 \ge 0, s \ge 0 \end{array}\right\} \forall a = [a_1; ...; a_6] \in \mathcal{U}. \tag{1.2.15}$$

It is immediately seen that while the first and the second RCs are equivalent to each other,[1] they are *not* equivalent to the third RC. The latter RC is more conservative than the first two, meaning that whenever (x_1, x_2) can be extended, by a properly chosen s, to a feasible solution of (1.2.15), (x_1, x_2) is feasible for (1.2.13)≡(1.2.14) (this is evident), but not necessarily vice versa. In fact, the gap between (1.2.15) and (1.2.13)≡(1.2.14) can be quite large. To illustrate the latter claim, consider the case where the uncertainty set is

$$\mathcal{U} = \{a = a_\zeta := [1 + \zeta; 2 + \zeta; 4 - \zeta; 4 + \zeta; 5 - \zeta; 9] : -\rho \le \zeta \le \rho\},$$

where ζ is the data perturbation. In this situation, $x_1 = 1$, $x_2 = 1$ is a feasible solution to (1.2.13)≡(1.2.14), provided that the uncertainty level ρ is $\le 1/3$:

$$(1 + \zeta) \cdot 1 + (2 + \zeta) \cdot 1 \le 4 - \zeta \, \forall (\zeta : |\zeta| \le \rho \le 1/3) \, \&(4 + \zeta) \cdot 1 + (5 - \zeta) \cdot 1 = 9 \; \forall \zeta.$$

At the same time, when $\rho > 0$, our solution $(x_1 = 1, x_2 = 1)$ cannot be extended to a feasible solution of (1.2.15), since the latter system of constraints is infeasible and remains so even after eliminating the equality (1.2.15.b).

Indeed, in order for x_1, x_2, s to satisfy (1.2.15.a) for all $a \in \mathcal{U}$, we should have

$$x_1 + 2x_2 + s + \zeta[x_1 + x_2] = 4 - \zeta \;\; \forall (\zeta : |\zeta| \le \rho);$$

when $\rho > 0$, we therefore should have $x_1 + x_2 = -1$, which contradicts (1.2.15.c)

[1] Clearly, this always is the case when an equality constraint, certain or uncertain alike, is replaced with a pair of opposite inequalities.

The origin of the outlined phenomenon is clear. Evidently the inequality $a_1x_1+a_2x_2 \leq a_3$, where all a_i and x_i are fixed reals, holds true if and only if we can "certify" the inequality by pointing out a real $s \geq 0$ such that $a_1x_1 + a_2x_2 + s = a_3$. When the data a_1, a_2, a_3 become uncertain, the restriction on (x_1, x_2) to be robust feasible for the uncertain inequality $a_1x_1 + a_2x_2 \leq a_3$ for all $a \in \mathcal{U}$ reads, "in terms of certificate," as

$$\forall a \in \mathcal{U}\, \exists s \geq 0 : a_1x_1 + a_2x_2 + s = a_3,$$

that is, the certificate s should be allowed to depend on the true data. In contrast to this, in (1.2.15) we require from both the decision variables x *and the slack variable ("the certificate")* s to be independent of the true data, which is by far too conservative.

What can be learned from the above examples is that when modeling an uncertain LO problem one should avoid whenever possible converting inequality constraints into equality ones, unless all the data in the constraints in question are certain. Aside from avoiding slack variables,[2] this means that restrictions like "total expenditure cannot exceed the budget," or "supply should be at least the demand," which in LO problems with certain data can harmlessly be modeled by equalities, in the case of uncertain data should be modeled by inequalities. This is in full accordance with common sense saying, e.g., that when the demand is uncertain and its satisfaction is a must, it would be unwise to forbid surplus in supply. Sometimes a good for the RO methodology modeling requires eliminating "state variables" — those which are readily given by variables representing actual decisions — via the corresponding "state equations." For example, time dynamics of an inventory is given in the simplest case by the state equations

$$\begin{aligned} &x_0 = c \\ &x_{t+1} = x_t + q_t - d_t,\ t = 0, 1, ..., T, \end{aligned}$$

where x_t is the inventory level at time t, d_t is the (uncertain) demand in period $[t, t+1)$, and variables q_t represent actual decisions – replenishment orders at instants $t = 0, 1, ..., T$. A wise approach to the RO processing of such an inventory problem would be to eliminate the state variables x_t by setting

$$x_t = c + \sum_{\tau=1}^{t-1} q_\tau,\ t = 0, 1, 2, ..., T+1,$$

and to get rid of the state equations. As a result, typical restrictions on state variables (like "x_t should stay within given bounds" or "total holding cost should not exceed a given bound") will become uncertainty-affected inequality constraints on the actual decisions q_t, and we can process the resulting inequality-constrained uncertain LO problem via its RC.[3]

1.2.4 What is Ahead

After introducing the concept of the Robust Counterpart of an uncertain LO problem, we confront two major questions:

1. What is the "computational status" of the RC? When is it possible to process the RC efficiently?

2. How to come-up with meaningful uncertainty sets?

[2]Note that slack variables do not represent actual decisions; thus, their presence in an LO model contradicts assumption A.1, and thus can lead to too conservative, or even infeasible, RCs.

[3]For more advanced robust modeling of uncertainty-affected multi-stage inventory, see lecture 5.

The first of these questions, to be addressed in depth in section 1.3, is a "structural" one: what should be the structure of the uncertainty set in order to make the RC computationally tractable? Note that the RC as given by (1.2.5) or (1.2.6) is a *semi-infinite* LO program, that is, an optimization program with simple linear objective and *infinitely many* linear constraints. In principle, such a problem can be "computationally intractable" — NP-hard.

Example 1.3 Consider an uncertain "essentially linear" constraint

$$\{\|Px - p\|_1 \leq 1\}_{[P;p]\in\mathcal{U}}\,, \tag{1.2.16}$$

where $\|z\|_1 = \sum\limits_j |z_j|$, and assume that the matrix P is certain, while the vector p is uncertain and is parameterized by perturbations from the unit box:

$$p \in \{p = B\zeta : \|\zeta\|_\infty \leq 1\}\,,$$

where $\|\zeta\|_\infty = \max\limits_\ell |\zeta_\ell|$ and B is a given positive semidefinite matrix. To check whether $x = 0$ is robust feasible is exactly the same as to verify whether $\|B\zeta\|_1 \leq 1$ whenever $\|\zeta\|_\infty \leq 1$; or, due to the evident relation $\|u\|_1 = \max\limits_{\|\eta\|_\infty \leq 1} \eta^T u$, the same as to check whether $\max\limits_{\eta,\zeta}\left\{\eta^T B\zeta : \|\eta\|_\infty \leq 1, \|\zeta\|_\infty \leq 1\right\} \leq 1$. The maximum of the bilinear form $\eta^T B\zeta$ with positive semidefinite B over η, ζ varying in a convex symmetric neighborhood of the origin is always achieved when $\eta = \zeta$ (you may check this by using the polarization identity $\eta^T B\zeta = \frac{1}{4}(\eta+\zeta)^T B(\eta+\zeta) - \frac{1}{4}(\eta-\zeta)^T B(\eta-\zeta)$). Thus, to check whether $x = 0$ is robust feasible for (1.2.16) is the same as to check whether the maximum of a given nonnegative quadratic form $\zeta^T B\zeta$ over the unit box is ≤ 1. The latter problem is known to be NP-hard,[4] and therefore so is the problem of checking robust feasibility for (1.2.16).

The second of the above is a modeling question, and as such, goes beyond the scope of purely theoretical considerations. However, theory, as we shall see in section 2.1, contributes significantly to this modeling issue.

1.3 Tractability of Robust Counterparts

In this section, we investigate the "computational status" of the RC of uncertain LO problem. The situation here turns out to be as good as it could be: we shall see, essentially, that *the RC of the uncertain LO problem with uncertainty set $\mathcal{U}$ is computationally tractable whenever the convex uncertainty set $\mathcal{U}$ itself is computationally tractable.* The latter means that we know in advance the affine hull of $\mathcal{U}$, a point from the relative interior of $\mathcal{U}$, and we have access to an efficient *membership oracle* that, given on input a point u, reports whether $u \in \mathcal{U}$. This can be reformulated as a precise mathematical statement; however, we will prove a slightly restricted version of this statement that does not require long excursions into complexity theory.

1.3.1 The Strategy

Our strategy will be as follows. First, we restrict ourselves to uncertain LO problems with a certain objective — we remember from item **A** in Section 1.2.3 that we lose nothing by this restriction. Second, all we need is a "computationally tractable" representation of the RC of a *single* uncertain linear constraint, that is, an equivalent representation of the RC by an explicit

[4] In fact, it is NP-hard to compute the maximum of a nonnegative quadratic form over the unit box with inaccuracy less than 4% [55].

(and "short") system of efficiently verifiable convex inequalities. Given such representations for the RCs of every one of the constraints of our uncertain problem and putting them together (cf. item **B** in Section 1.2.3), we reformulate the RC of the problem as the problem of minimizing the original linear objective under a finite (and short) system of explicit convex constraints, and thus — as a computationally tractable problem.

To proceed, we should explain first what does it mean to represent a constraint by a system of convex inequalities. Everyone understands that the system of 4 constraints on 2 variables,

$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1, \tag{1.3.1}$$

represents the nonlinear inequality

$$|x_1| + |x_2| \leq 1 \tag{1.3.2}$$

in the sense that both (1.3.2) and (1.3.1) define the same feasible set. Well, what about the claim that the system of 5 linear inequalities

$$-u_1 \leq x_1 \leq u_1, -u_2 \leq x_2 \leq u_2, u_1 + u_2 \leq 1 \tag{1.3.3}$$

represents the same set as (1.3.2)? Here again everyone will agree with the claim, although we cannot justify the claim in the former fashion, since the feasible sets of (1.3.2) and (1.3.3) live in different spaces and therefore cannot be equal to each other!

What actually is meant when speaking about "equivalent representations of problems/constraints" in Optimization can be formalized as follows:

Definition 1.5 *A set $X^+ \subset \mathbb{R}^n_x \times \mathbb{R}^k_u$ is said to represent a set $X \subset \mathbb{R}^n_x$, if the projection of X^+ onto the space of x-variables is exactly X, i.e., $x \in X$ if and only if there exists $u \in \mathbb{R}^k_u$ such that $(x, u) \in X^+$:*

$$X = \left\{x : \exists u : (x, u) \in X^+\right\}.$$

A system of constraints $\mathcal{S}^+$ in variables $x \in \mathbb{R}^n_x$, $u \in \mathbb{R}^k_u$ is said to represent a system of constraints $\mathcal{S}$ in variables $x \in \mathbb{R}^n_x$, if the feasible set of the former system represents the feasible set of the latter one.

With this definition, it is clear that the system (1.3.3) indeed represents the constraint (1.3.2), and, more generally, that the system of $2n + 1$ linear inequalities

$$-u_j \leq x_j \leq u_j,\ j = 1, ..., n, \sum_j u_j \leq 1$$

in variables x, u represents the constraint

$$\sum_j |x_j| \leq 1.$$

To understand how powerful this representation is, note that to represent the same constraint in the style of (1.3.1), that is, without extra variables, it would take as much as 2^n linear inequalities.

Coming back to the general case, assume that we are given an optimization problem

$$\min_x \left\{f(x) \text{ s.t. } x \text{ satisfies } \mathcal{S}_i,\ i = 1, ..., m\right\}, \tag{P}$$

where $\mathcal{S}_i$ are systems of constraints in variables x, and that we have in our disposal systems $\mathcal{S}_i^+$ of constraints in variables x, v^i which represent the systems $\mathcal{S}_i$. Clearly, the problem

$$\min_{x,v^1,...,v^m} \left\{f(x) \text{ s.t. } (x, v^i) \text{ satisfies } \mathcal{S}_i^+,\ i = 1, ..., m\right\} \tag{P$^+$}$$

is equivalent to (P): the x component of every feasible solution to (P$^+$) is feasible for (P) with the same value of the objective, and the optimal values in the problems are equal to each other, so that the x component of an ϵ-optimal (in terms of the objective) feasible solution to (P$^+$) is an ϵ-optimal feasible solution to (P). We shall say that (P$^+$) represents equivalently the original problem (P). What is important here, is that a representation can possess desired properties that are absent in the original problem. For example, an appropriate representation can convert the problem of the form $\min_x\{\|Px - p\|_1 : Ax \le b\}$ with n variables, m linear constraints, and k-dimensional vector p, into an LO problem with $n + k$ variables and $m + 2k + 1$ linear inequality constraints, etc. Our goal now is to build a representation capable of expressing equivalently a semi-infinite linear constraint (specifically, the robust counterpart of an uncertain linear inequality) as a finite system of explicit convex constraints, with the ultimate goal to use these representations in order to convert the RC of an uncertain LO problem into an explicit (and as such, computationally tractable) convex program.

The outlined strategy allows us to focus on a *single* uncertainty-affected linear inequality — a family

$$\left\{a^T x \le b\right\}_{[a;b]\in\mathcal{U}}, \tag{1.3.4}$$

of linear inequalities with the data varying in the uncertainty set

$$\mathcal{U} = \left\{[a; b] = [a^0; b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell] : \zeta \in \mathcal{Z}\right\} \tag{1.3.5}$$

— and on "tractable representation" of the RC

$$a^T x \le b \quad \forall \left([a; b] = [a^0; b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell] : \zeta \in \mathcal{Z}\right) \tag{1.3.6}$$

of this uncertain inequality.

By reasons indicated in item **C** of Section 1.2.3, we assume from now on that the associated perturbation set $\mathcal{Z}$ is convex.

1.3.2 Tractable Representation of (1.3.6): Simple Cases

We start with the cases where the desired representation can be found by "bare hands," specifically, the cases of *interval* and *simple ellipsoidal* uncertainty.

Example 1.4 Consider the case of *interval uncertainty*, where $\mathcal{Z}$ in (1.3.6) is a box. W.l.o.g. we can normalize the situation by assuming that

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

In this case, (1.3.6) reads

$$\begin{array}{lll} & [a^0]^T x + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell & \forall (\zeta : \|\zeta\|_\infty \le 1) \\ \Leftrightarrow & \sum_{\ell=1}^{L} \zeta_\ell [[a^\ell]^T x - b^\ell] \le b^0 - [a^0]^T x & \forall (\zeta : |\zeta_\ell| \le 1,\ \ell = 1, ..., L) \\ \Leftrightarrow & \max_{-1\le\zeta_\ell\le1} \left[\sum_{\ell=1}^{L} \zeta_\ell [[a^\ell]^T x - b^\ell]\right] \le b^0 - [a^0]^T x & \end{array}$$

The concluding maximum in the chain is clearly $\sum\limits_{\ell=1}^{L} |[a^\ell]^T x - b^\ell|$, and we arrive at the representation of (1.3.6) by the explicit convex constraint

$$[a^0]^T x + \sum_{\ell=1}^{L} |[a^\ell]^T x - b^\ell| \leq b^0, \tag{1.3.7}$$

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

$$\left\{ \begin{array}{l} -u_\ell \leq [a^\ell]^T x - b^\ell \leq u_\ell,\ \ell = 1, ..., L, \\ [a^0]^T x + \sum\limits_{\ell=1}^{L} u_\ell \leq b^0. \end{array} \right. \tag{1.3.8}$$

Example 1.5 Consider the case of *ellipsoidal uncertainty* where $\mathcal{Z}$ in (1.3.6) is an ellipsoid. W.l.o.g. we can normalize the situation by assuming that $\mathcal{Z}$ is merely the ball of radius Ω centered at the origin:

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

In this case, (1.3.6) reads

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

and we arrive at the representation of (1.3.6) by the explicit convex constraint ("conic quadratic inequality")

$$[a^0]^T x + \Omega \sqrt{\sum_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2} \leq b^0. \tag{1.3.9}$$

1.3.3 Tractable Representation of (1.3.6): General Case

Now consider a rather general case when the perturbation set $\mathcal{Z}$ in (1.3.6) is given by a *conic representation* (cf. section A.2.4 in Appendix):

$$\mathcal{Z} = \left\{ \zeta \in \mathbb{R}^L : \exists u \in \mathbb{R}^K : P\zeta + Qu + p \in \mathbf{K} \right\}, \tag{1.3.10}$$

where $\mathbf{K}$ is a closed convex pointed cone in $\mathbb{R}^N$ with a nonempty interior, P, Q are given matrices and p is a given vector.

In the case when $\mathbf{K}$ is *not* a polyhedral cone, assume that this representation is strictly feasible:

$$\exists (\bar{\zeta}, \bar{u}) : P\bar{\zeta} + Q\bar{u} + p \in \mathrm{int} K. \tag{1.3.11}$$

In fact, in the sequel we would lose nothing by further restricting K to be a *canonical cone* – (finite) direct product of "simple" cones $K^1, ..., K^S$:

$$K = K^1 \times ... \times K^S, \tag{1.3.12}$$

where every K^s is

- either the nonnegative orthant $\mathbb{R}^n_+ = \{x = [x_1; ...; x_n] \in \mathbb{R}^n : x_i \geq 0 \,\forall i\}$,
- or the Lorentz cone $\mathbf{L}^n = \{x = [x_1; ...; x_n] \in \mathbb{R}^n : x_n \geq \sqrt{\sum_{i=1}^{n-1} x_i^2}\}$,
- or the semidefinite cone $\mathbf{S}^n_+$. This cone "lives" in the space $\mathbf{S}^n$ of real symmetric $n \times n$ matrices equipped with the Frobenius inner product $\langle A, B\rangle = \mathrm{Tr}(AB) = \mathrm{Tr}(AB^T) = \sum_{i,j} A_{ij}B_{ij}$; the cone itself is comprised of all positive semidefinite symmetric $n \times n$ matrices.

As a matter of fact,

- the family $\mathcal{F}$ of all convex sets admitting conic representations involving canonical cones is extremely nice – it is closed w.r.t. all basic operations preserving convexity, like taking finite intersections, arithmetic sums, images and inverse images under affine mappings, etc. Moreover, conic representation of the result of such an operation is readily given by conic representation of the operands; see section A.2.4 for the corresponding "calculus." As a result, handling convex sets from the family in question is fully algorithmic and computationally efficient;
- the family $\mathcal{F}$ is extremely wide: as a matter of fact, for all practical purposes one can think of $\mathcal{F}$ as of the family of *all* computationally tractable convex sets arising in applications.

Theorem 1.1 *Let the perturbation set $\mathcal{Z}$ be given by* (1.3.10)*, and in the case of non-polyhedral* $\mathbf{K}$*, let also* (1.3.11) *take place. Then the semi-infinite constraint* (1.3.6) *can be represented by the following system of conic inequalities in variables $x \in \mathbb{R}^n, y \in \mathbb{R}^N$:*

$$\begin{array}{l} p^T y + [a^0]^T x \leq b^0, \\ Q^T y = 0, \\ (P^T y)_\ell + [a^\ell]^T x = b^\ell, \ell = 1, ..., L, \\ y \in \mathbf{K}_*, \end{array} \tag{1.3.13}$$

where $\mathbf{K}_* = \{y : y^T z \geq 0 \,\forall z \in \mathbf{K}\}$ is the cone dual to $\mathbf{K}$.

Proof. We have

$$\begin{array}{cl} & x \text{ is feasible for } (1.3.6) \\ \Leftrightarrow & \sup\limits_{\zeta \in \mathcal{Z}} \Big\{ \underbrace{[a^0]^T x - b^0}_{d[x]} + \sum\limits_{\ell=1}^{L} \zeta_\ell \underbrace{\Big[[a^\ell]^T x - b^\ell\Big]}_{c_\ell[x]} \Big\} \leq 0 \\ \Leftrightarrow & \sup\limits_{\zeta \in \mathcal{Z}} \left\{ c^T[x]\zeta + d[x] \right\} \leq 0 \\ \Leftrightarrow & \sup\limits_{\zeta \in \mathcal{Z}} c^T[x]\zeta \leq -d[x] \\ \Leftrightarrow & \max\limits_{\zeta, v} \left\{ c^T[x]\zeta : P\zeta + Qv + p \in \mathbf{K} \right\} \leq -d[x]. \end{array}$$

The concluding relation says that x is feasible for (1.3.6) if and only if the optimal value in the conic program

$$\max_{\zeta, v} \left\{ c^T[x]\zeta : P\zeta + Qv + p \in \mathbf{K} \right\} \tag{CP}$$

is $\leq -d[x]$. Assume, first, that (1.3.11) takes place. Then (CP) is strictly feasible, and therefore, applying the Conic Duality Theorem (Theorem A.1), the optimal value in (CP) is $\leq -d[x]$ if and only if the optimal value in the conic dual to the (CP) problem

$$\min_y \left\{ p^T y : Q^T y = 0, P^T y = -c[x], y \in \mathbf{K}_* \right\}, \tag{CD}$$

is attained and is $\leq -d[x]$. Now assume that $\mathbf{K}$ is a polyhedral cone. In this case the usual LO Duality Theorem, (which does not require the validity of (1.3.11)), yields exactly the same conclusion: the optimal value in (CP) is $\leq -d[x]$ if and only if the optimal value in (CD) is achieved and is $\leq -d[x]$. In other words, under the premise of the Theorem, x is feasible for (1.3.6) if and only if (CD) has a feasible solution y with $p^T y \leq -d[x]$. □

Observing that nonnegative orthants, Lorentz and Semidefinite cones are self-dual, and thus their finite direct products, i.e., canonical cones, are self-dual as well,[5] we derive from Theorem 1.1 the following corollary:

Corollary 1.1 *Let the nonempty perturbation set in* (1.3.6) *be:*
(i) *polyhedral, i.e., given by* (1.3.10) *with a nonnegative orthant* $\mathbb{R}^N_+$ *in the role of* $\mathbf{K}$*, or*
(ii) *conic quadratic representable, i.e., given by* (1.3.10) *with a direct product* $\mathbf{L}^{k_1} \times ... \times \mathbf{L}^{k_m}$ *of Lorentz cones* $\mathbf{L}^k = \{x \in \mathbb{R}^k : x_k \geq \sqrt{x_1^2 + ... + x_{k-1}^2}\}$ *in the role of* $\mathbf{K}$*, or*
(iii) *semidefinite representable, i.e., given by* (1.3.10) *with the positive semidefinite cone* $\mathbf{S}^k_+$ *in the role of* $\mathbf{K}$*.*
In the cases of (ii), (iii) *assume in addition that* (1.3.11) *holds true. Then the Robust Counterpart* (1.3.6) *of the uncertain linear inequality* (1.3.4) — (1.3.5) *with the perturbation set* $\mathcal{Z}$ *admits equivalent reformulation as an explicit system of*
— *linear inequalities, in the case of* (i),
— *conic quadratic inequalities, in the case of* (ii),
— *linear matrix inequalities, in the case of* (iii).
In all cases, the size of the reformulation is polynomial in the number of variables in (1.3.6) *and the size of the conic description of* $\mathcal{Z}$*, while the data of the reformulation is readily given by the data describing, via* (1.3.10)*, the perturbation set* $\mathcal{Z}$*.*

Remark 1.2 A. Usually, the cone $\mathbf{K}$ participating in (1.3.10) is the direct product of simpler cones $\mathbf{K}^1, ..., \mathbf{K}^S$, so that representation (1.3.10) takes the form

$$\mathcal{Z} = \{\zeta : \exists u^1, ..., u^S : P_s\zeta + Q_s u^s + p_s \in \mathbf{K}^s,\ s = 1, ..., S\}. \tag{1.3.14}$$

In this case, (1.3.13) becomes the system of conic constraints in variables $x, y^1, ..., y^S$ as follows:

$$\begin{array}{l} \sum\limits_{s=1}^{S} p_s^T y^s + [a^0]^T x \leq b^0, \\ Q_s^T y^s = 0,\ s = 1, ..., S, \\ \sum\limits_{s=1}^{S} (P_s^T y^s)_\ell + [a^\ell]^T x = b^\ell, \ell = 1, ..., L, \\ y^s \in \mathbf{K}^s_*, s = 1, ..., S, \end{array} \tag{1.3.15}$$

where K^s_* is the cone dual to K^s.
B. Uncertainty sets given by LMIs seem “exotic”; however, they can arise under quite realistic circumstances, see section 1.5.

[5]Since the cone dual to a direct product of cones K^s clearly is the direct product of cones K^s_* dual to K^s.

Examples

We are about to apply Theorem 1.1 to build tractable reformulations of the semi-infinite inequality (1.3.6) in two particular cases. While at a first glance no natural "uncertainty models" lead to the "strange" perturbation sets we are about to consider, it will become clear later that these sets are of significant importance — they allow one to model *random* uncertainty.

Example 1.6 *$\mathcal{Z}$ is the intersection of concentric co-axial box and ellipsoid, specifically,*

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : -1 \leq \zeta_\ell \leq 1, \ell \leq L, \sqrt{\sum_{\ell=1}^{L} \zeta_\ell^2/\sigma_\ell^2} \leq \Omega\}, \tag{1.3.16}$$

where $\sigma_\ell > 0$ and $\Omega > 0$ are given parameters.

Here representation (1.3.14) becomes

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : P_1\zeta + p_1 \in \mathbf{K}^1, P_2\zeta + p_2 \in \mathbf{K}^2\},$$

where

• $P_1\zeta \equiv [\zeta; 0]$, $p_1 = [0_{L\times 1}; 1]$ and $\mathbf{K}^1 = \{(z,t) \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_\infty\}$, whence $\mathbf{K}^1_* = \{(z,t) \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_1\}$;

• $P_2\zeta = [\Sigma^{-1}\zeta; 0]$ with $\Sigma = \text{Diag}\{\sigma_1, ..., \sigma_L\}$, $p_2 = [0_{L\times 1}; \Omega]$ and $\mathbf{K}^2$ is the Lorentz cone of the dimension $L+1$ (whence $\mathbf{K}^2_* = \mathbf{K}^2$)

Setting $y^1 = [\eta_1; \tau_1]$, $y^2 = [\eta_2; \tau_2]$ with one-dimensional τ_1, τ_2 and L-dimensional η_1, η_2, (1.3.15) becomes the following system of constraints in variables τ, η, x:

$$\begin{array}{llcl}
(a) & \tau_1 + \Omega\tau_2 + [a^0]^T x & \leq & b^0, \\
(b) & (\eta_1 + \Sigma^{-1}\eta_2)_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L, \\
(c) & \|\eta_1\|_1 & \leq & \tau_1 \quad [\Leftrightarrow [\eta_1; \tau_1] \in \mathbf{K}^1_*], \\
(d) & \|\eta_2\|_2 & \leq & \tau_2 \quad [\Leftrightarrow [\eta_2; \tau_2] \in \mathbf{K}^2_*].
\end{array}$$

We can eliminate from this system the variables τ_1, τ_2 — for every feasible solution to the system, we have $\tau_1 \geq \bar{\tau}_1 \equiv \|\eta_1\|_1$, $\tau_2 \geq \bar{\tau}_2 \equiv \|\eta_2\|_2$, and the solution obtained when replacing τ_1, τ_2 with $\bar{\tau}_1$, $\bar{\tau}_2$ still is feasible. The reduced system in variables x, $z = \eta_1$, $w = \Sigma^{-1}\eta_2$ reads

$$\begin{array}{rcl}
\sum_{\ell=1}^{L} |z_\ell| + \Omega\sqrt{\sum_\ell \sigma_\ell^2 w_\ell^2} + [a^0]^T x & \leq & b^0, \\
z_\ell + w_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L,
\end{array} \tag{1.3.17}$$

which is also a representation of (1.3.6), (1.3.16).

Example 1.7 [*"budgeted uncertainty"*] Consider the case where $\mathcal{Z}$ is the intersection of $\|\cdot\|_\infty$- and $\|\cdot\|_1$-balls, specifically,

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \leq 1,\ \|\zeta\|_1 \leq \gamma\}, \tag{1.3.18}$$

where γ, $1 \leq \gamma \leq L$, is a given "uncertainty budget."

Here representation (1.3.14) becomes

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : P_1\zeta + p_1 \in \mathbf{K}^1, P_2\zeta + p_2 \in \mathbf{K}^2\},$$

where

• $P_1\zeta \equiv [\zeta; 0]$, $p_1 = [0_{L\times 1}; 1]$ and $\mathbf{K}^1 = \{[z; t] \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_\infty\}$, whence $\mathbf{K}^1_* = \{[z; t] \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_1\}$;

• $P_2\zeta = [\zeta; 0]$, $p_2 = [0_{L\times 1}; \gamma]$ and $\mathbf{K}^2 = \mathbf{K}^1_* = \{[z; t] \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_1\}$, whence $\mathbf{K}^2_* = \mathbf{K}^1$.

Setting $y^1 = [z; \tau_1]$, $y^2 = [w; \tau_2]$ with one-dimensional τ and L-dimensional z, w, system (1.3.15) becomes the following system of constraints in variables τ_1, τ_2, z, w, x:

$$\begin{array}{llcl}(a) & \tau_1 + \gamma\tau_2 + [a^0]^T x & \leq & b^0,\\ (b) & (z+w)_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L,\\ (c) & \|z\|_1 & \leq & \tau_1 \quad [\Leftrightarrow [\eta_1; \tau_1] \in \mathbf{K}^1_*],\\ (d) & \|w\|_\infty & \leq & \tau_2 \quad [\Leftrightarrow [\eta_2; \tau_2] \in \mathbf{K}^2_*].\end{array}$$

Same as in Example 1.6, we can eliminate the τ-variables, arriving at a representation of (1.3.6), (1.3.18) by the following system of constraints in variables x, z, w:

$$\begin{array}{rcl}\sum_{\ell=1}^{L} |z_\ell| + \gamma \max_\ell |w_\ell| + [a^0]^T x & \leq & b^0,\\ z_\ell + w_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L,\end{array} \tag{1.3.19}$$

which can be further converted into the system of linear inequalities in z, w and additional variables.

1.4 How it Works: Motivating Examples Revisited

In this section, we outline the results of Robust Optimization methodology as applied to our "motivating examples."

1.4.1 Robust Synthesis of Antenna Arrays

In the situation of the Antenna Design problem (1.1.2), the "physical" uncertainty comes from the actuation errors $x_j \mapsto (1+\xi_j)x_j$; as we have already explained, these errors can be modeled equivalently by the perturbations $D_j(\theta_i) \mapsto D_{ij} = (1+\xi_j)D_j(\theta_i)$ in the coefficients of x_j. Assuming that the errors ξ_j are bounded by a given *uncertainty level* ρ, and that this is the only a priori information on the actuation errors, we end up with the uncertain LO problem

$$\left\{ \min_{x,\tau} \left\{ \tau : -\tau \leq \sum_{j=1}^{J=10} D_{ij}x_j - D_*(\theta_i) \leq \tau, 1 \leq i \leq I = 240 \right\} : |D_{ij} - D_j(\theta_i)| \leq \rho |D_j(\theta_i)| \right\}.$$

The Robust Counterpart of the problem is the semi-infinite LO program

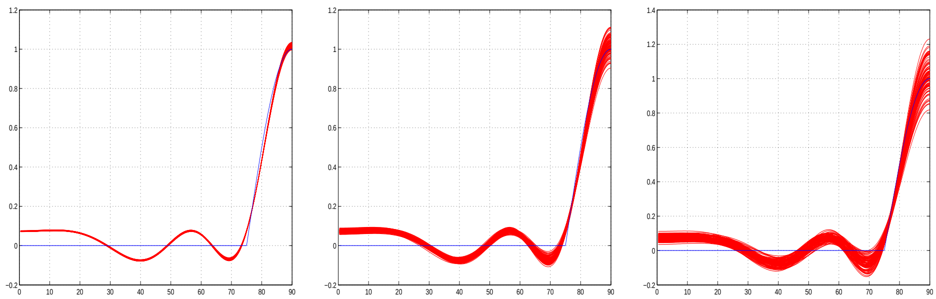
$$\min_{x,\tau} \left\{ \tau : -\tau \leq \sum_j D_{ij}x_j \leq \tau, 1 \leq i \leq I\ \forall D_{ij} \in [\underline{G}_{ij}, \overline{G}_{ij}] \right\} \tag{1.4.1}$$

with $\underline{G}_{ij} = G_j(\theta_i) - \rho|G_j(\theta_i)|$, $\overline{G}_{ij} = G_j(\theta_i) + \rho|G_j(\theta_i)|$. The generic form of this semi-infinite LO is

$$\min_y \left\{ c^T y : Ay \leq b \forall [A,b] : [\underline{A}, \underline{b}] \leq [A, b] \leq [\overline{A}, \overline{b}] \right\} \tag{1.4.2}$$

where $\leq$ for matrices is understood entrywise and $[\underline{A}, \underline{b}] \leq [\overline{A}, \overline{b}]$ are two given matrices. This is a very special case of polyhedral uncertainty set, so that our theory says that the RC is equivalent to an explicit LO program. In fact we can point out (one of) LO reformulation of the Robust Counterpart without reference to any theory: it is immediately seen that (1.4.2) is equivalent to the LO program

$$\min_{y,z} \left\{ c^T y : \underline{A}z + \overline{A}(y+z) \leq \underline{b}, z \geq 0, y+z \geq 0 \right\}. \tag{1.4.3}$$



$\rho = 0.01$ $\rho = 0.05$ $\rho = 0.1$

Figure 1.3: "Dream and reality," robust optimal design: samples of 100 actual diagrams (red) for different uncertainty levels. Blue: the target diagram.

	Reality								
	$\rho = 0.01$			$\rho = 0.05$			$\rho = 0.1$		
	min	mean	max	min	mean	max	min	mean	max
$\\|\cdot\\|_\infty$-distance to target	0.075	0.078	0.081	0.077	0.088	0.114	0.082	0.113	0.216
energy concentration	70.3%	72.3%	73.8%	63.6%	71.6%6	79.3%	52.2%	70.8%	87.5%

Table 1.2: Quality of robust antenna design. Data over 100 samples of actuation errors per each uncertainty level ρ.
For comparison: for nominal design, with the uncertainty level as small as $\rho = 0.001$, the average $\|\cdot\|_\infty$-distance of the actual diagram to target is as large as 56.8, and the expected energy concentration is as low as 16.5%.

Solving (1.4.1) for the uncertainty level $\rho = 0.01$, we end up with the robust optimal value 0.0815, which, while being by 39% worse than the nominal optimal value 0.0589 (which, as we have seen, exists only in our imagination and says nothing about the actual performance of the nominal optimal design), still is reasonably small. Note that the robust optimal value, in sharp contrast with the nominally optimal one, does say something meaningful about the actual performance of the underlying *robust* design. In our experiments, we have tested the robust optimal design associated with the uncertainty level $\rho = 0.01$ versus actuation errors of this and larger magnitudes. The results are presented on figure 1.3 and in table 1.2. Comparing these figure and table with their "nominal design" counterparts, we see that the robust design is incomparably better than the nominal one.

`NETLIB` Case Study

The corresponding uncertainty model ("ugly coefficients a_{ij} in the constraint matrix independently of each other vary in the segments $[a^{\rm n}_{ij} - \rho|a^{\rm n}_{ij}|, a^{\rm n}_{ij} + \rho|a^{\rm n}_{ij}|]$, $\rho > 0$ being the uncertainty level) clearly yields the RCs of the generic form (1.4.2). As explained above, these RCs can be straightforwardly converted to explicit LO programs which are of nearly the same sizes and sparsity as the instances of the uncertain LPs in question. It turns out that at the uncertainty level 0.1% ($\rho = 0.001$), all these RCs are feasible, that is, we can immunize the solutions against this uncertainty. Surprisingly, this immunization is "nearly costless" – the robust optimal values of all 90 NETLIB LOs considered in [8] remain within 1% margin of the nominal optimal values. For further details, including what happens at larger uncertainty levels, see [8].

1.5 Non-Affine Perturbations

In the first reading this section can be skipped.

So far we have assumed that the uncertain data of an uncertain LO problem are *affinely* parameterized by a perturbation vector ζ varying in a closed convex set $\mathcal{Z}$. We have seen that this assumption, combined with the assumption that $\mathcal{Z}$ is computationally tractable, implies tractability of the RC. What happens when the perturbations enter the uncertain data in a nonlinear fashion? Assume w.l.o.g. that every entry a in the uncertain data is of the form

$$a = \sum_{k=1}^{K} c_k^a f_k(\zeta),$$

where c_k^a are given coefficients (depending on the data entry in question) and $f_1(\zeta), ..., f_K(\zeta)$ are certain basic functions, perhaps non-affine, defined on the perturbation set $\mathcal{Z}$. Assuming w.l.o.g. that the objective is certain, we still can define the RC of our uncertain problem as the problem of minimizing the original objective over the set of robust feasible solutions, those which remain feasible for all values of the data coming from $\zeta \in \mathcal{Z}$, but what about the tractability of this RC? An immediate observation is that the case of nonlinearly perturbed data can be immediately reduced to the one where the data are affinely perturbed. To this end, it suffices to pass from the original perturbation vector ζ to the new vector

$$\widehat{\zeta}[\zeta] = [\zeta_1; ...; \zeta_L; f_1(\zeta); ...; f_K(\zeta)].$$

As a result, the uncertain data become *affine* functions of the new perturbation vector $\widehat{\zeta}$ which now runs through the image $\widetilde{\mathcal{Z}} = \widehat{\zeta}[\mathcal{Z}]$ of the original uncertainty set $\mathcal{Z}$ under the mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$. As we know, in the case of affine data perturbations the RC remains intact when replacing a given perturbation set with its closed convex hull. Thus, we can think about our uncertain LO problem as an affinely perturbed problem where the perturbation vector is $\widehat{\zeta}$, and this vector runs through the closed convex set $\widehat{\mathcal{Z}} = \mathrm{cl}\,\mathrm{Conv}(\widehat{\zeta}[\mathcal{Z}])$. We see that formally speaking, the case of general-type perturbations can be reduced to the one of affine perturbations. This, unfortunately, does not mean that non-affine perturbations do not cause difficulties. Indeed, in order to end up with a computationally tractable RC, we need more than affinity of perturbations and convexity of the perturbation set — we need this set to be computationally tractable. And the set $\widehat{\mathcal{Z}} = \mathrm{cl}\,\mathrm{Conv}(\widehat{\zeta}[\mathcal{Z}])$ may fail to satisfy this requirement even when both $\mathcal{Z}$ and the *nonlinear* mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$ are simple, e.g., when $\mathcal{Z}$ is a box and $\widehat{\zeta} = [\zeta; \{\zeta_\ell \zeta_r\}_{\ell,r=1}^L]$, (i.e., when the uncertain data are quadratically perturbed by the original perturbations ζ).

We are about to present two generic cases where the difficulty just outlined does not occur (for justification and more examples, see section 5.3.2).

Ellipsoidal perturbation set $\mathcal{Z}$, quadratic perturbations. Here $\mathcal{Z}$ is an ellipsoid, and the basic functions f_k are the constant, the coordinates of ζ and the pairwise products of these coordinates. This means that the uncertain data entries are quadratic functions of the perturbations. W.l.o.g. we can assume that the ellipsoid $\mathcal{Z}$ is centered at the origin: $\mathcal{Z} = \{\zeta : \|Q\zeta\|_2 \leq 1\}$, where $\mathrm{Ker}Q = \{0\}$. In this case, representing $\widehat{\zeta}[\zeta]$ as the matrix $\left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right]$, we have the following semidefinite representation of $\widehat{\mathcal{Z}} = \mathrm{cl}\,\mathrm{Conv}(\widehat{\zeta}[\mathcal{Z}])$:

$$\widehat{\mathcal{Z}} = \left\{ \left[\begin{array}{c|c} & w^T \\ \hline w & W \end{array}\right] : \left[\begin{array}{c|c} 1 & w^T \\ \hline w & W \end{array}\right] \succeq 0, \mathrm{Tr}(QWQ^T) \leq 1 \right\}$$

(for proof, see Lemma 5.4).

Separable polynomial perturbations. Here the structure of perturbations is as follows: ζ runs through the box $\mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \leq 1\}$, and the uncertain data entries are of the form

$$a = p_1^a(\zeta_1) + ... + p_L^a(\zeta_L),$$

where $p_\ell^a(s)$ are given algebraic polynomials of degrees not exceeding d; in other words, the basic functions can be split into L groups, the functions of ℓ-th group being $1 = \zeta_\ell^0, \zeta_\ell, \zeta_\ell^2, ..., \zeta_\ell^d$. Consequently, the function $\widehat{\zeta}[\zeta]$ is given by

$$\widehat{\zeta}[\zeta] = [[1; \zeta_1; \zeta_1^2; ...; \zeta_1^d]; ...; [1; \zeta_L; \zeta_L^2; ...; \zeta_L^d]].$$

Setting $P = \{\widehat{s} = [1; s; s^2; ...; s^d] : -1 \leq s \leq 1\}$, we conclude that $\widetilde{\mathcal{Z}} = \widehat{\zeta}[\mathcal{Z}]$ can be identified with the set $P^L = \underbrace{P \times ... \times P}_{L}$, so that $\widehat{\widetilde{\mathcal{Z}}}$ is nothing but the set $\underbrace{\mathcal{P} \times ... \times \mathcal{P}}_{L}$, where $\mathcal{P} = \mathrm{Conv}(P)$. It remains to note that the set $\mathcal{P}$ admits an explicit semidefinite representation, see Lemma 5.2.

1.6 Exercises

Exercise 1.1 Prove the fact stated in the beginning of section 1.4.1:

(!) The RC of an uncertain LO problem with certain objective and simple interval uncertainty in the constraints — the uncertain problem

$$\mathcal{P} = \left\{ \min_x \left\{ c^T x : Ax \leq b \right\}, [\underline{A}, \underline{b}] \leq [A, b] \leq [\overline{A}, \overline{b}] \right\}$$

is equivalent to the explicit LO program

$$\min_{u,v} \left\{ c^T x : \overline{A}u - \underline{A}v \leq \underline{b}, u \geq 0, v \geq 0, u - v = x \right\} \tag{1.6.1}$$

Exercise 1.2 Represent the RCs of every one of the uncertain linear constraints given below:

$$\begin{array}{ll} a^T x \leq b, [a; b] \in \mathcal{U} = \{[a; b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \|\zeta\|_p \leq \rho\} & \\ \qquad\qquad\qquad\qquad [p \in [1, \infty]] & (a) \\ a^T x \leq b, [a; b] \in \mathcal{U} = \{[a; b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \|\zeta\|_p \leq \rho, \zeta \geq 0\} & \\ \qquad\qquad\qquad\qquad [p \in [1, \infty]] & (b) \\ a^T x \leq b, [a; b] \in \mathcal{U} = \{[a; b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \|\zeta\|_p \leq \rho\} & \\ \qquad\qquad\qquad\qquad [p \in (0, 1)] & (c) \end{array}$$

as explicit convex constraints.

Exercise 1.3 Represent in tractable form the RC of uncertain linear constraint

$$a^T x \leq b$$

with $\cap$-ellipsoidal uncertainty set

$$\mathcal{U} = \{[a,b] = [a^{\rm n}; b^{\rm n}] + P\zeta : \zeta^T Q_j \zeta \leq \rho^2,\ 1 \leq j \leq J\},$$

where $Q_j \succeq 0$ and $\sum_j Q_j \succ 0$.

The goal of subsequent exercises is to find out whether there is a "gap" between feasibility/optimality properties of *instances* of an uncertain LO problem

$$\mathcal{P} = \left\{ \min_x \left\{ c^T x : Ax \leq b \right\} : [A,b] \in \mathcal{U} \right\}$$

and similar properties of its RC

$$\mathrm{Opt} = \min_x \left\{ c^T x : Ax \leq b\ \forall [A,b] \in \mathcal{U} \right\}. \qquad \text{(RC)}$$

Specifically, we want to answer the following questions:

- Is it possible that every instance of $\mathcal{P}$ is feasible, while (RC) is not so?
- Is it possible that (RC) is feasible, but its optimal value is worse than those of all instances?
- Under which natural conditions feasibility of (RC) is equivalent to feasibility of all instances, and the robust optimal value is the maximum of optimal values of instances.

Exercise 1.4 Consider two uncertain LO problems

$$\begin{array}{rl}
\mathcal{P}_1 = & \left\{ \min\limits_x \{-x_1 - x_2 : 0 \leq x_1 \leq b_1, 0 \leq x_2 \leq b_2, x_1 + x_2 \geq p\} : b \in \mathcal{U}_1 \right\}, \\
& \mathcal{U}_1 = \{b : 1 \geq b_1 \geq 1/3, 1 \geq b_2 \geq 1/3, b_1 + b_2 \geq 1\}, \\
\mathcal{P}_2 = & \left\{ \min\limits_x \{-x_1 - x_2 : 0 \leq x_1 \leq b_1, 0 \leq x_2 \leq b_2, x_1 + x_2 \geq p\} : b \in \mathcal{U}_2 \right\}, \\
& \mathcal{U}_2 = \{b : 1 \geq b_1 \geq 1/3, 1 \geq b_2 \geq 1/3\}.
\end{array}$$

(p is a parameter).

1. Build the RC's of the problems.
2. Set $p = 3/4$. Is there a gap between feasibility properties of the instances of $\mathcal{P}_1$ and those of the RC of $\mathcal{P}_1$? Is there a similar gap in the case of $\mathcal{P}_2$?
3. Set $p = 2/3$. Is there a gap between the largest of the optimal values of instances of $\mathcal{P}_1$ and the optimal value of the RC? Is there a similar gap in the case of $\mathcal{P}_2$?

The results of Exercise 1.4 demonstrate that there could be a huge gap between feasibility/optimality properties of the RC and those of instances. We are about to demonstrate that this phenomenon does *not* occur in the case of a "constraint-wise" uncertainty.

Definition 1.6 *Consider an uncertain LO problem with certain objective*

$$\mathcal{P} = \left\{ \min_x \left\{ c^T x : Ax \le b \right\} : [A, b] \in \mathcal{U} \right\}$$

with convex compact uncertainty set $\mathcal{U}$, and let $\mathcal{U}_i$ be the projection of $\mathcal{U}$ on the set of data of i-th constraint:

$$\mathcal{U}_i = \left\{ [a_i^T, b_i] : \exists [A, b] \in \mathcal{U} \text{ such that } [a_i^T, b_i] \text{ is } i\text{-th row in } [A, b] \right\}.$$

Clearly, $\mathcal{U} \subset \mathcal{U}^+ = \prod_i \mathcal{U}_i$. We call uncertainty constraint-wise, if $\mathcal{U} = \mathcal{U}^+$, and call the uncertain problem, obtained from $\mathcal{P}$ by extending the original uncertainty set $\mathcal{U}$ the constraint-wise envelope of $\mathcal{P}$.

Note that by claim on p. 10, when passing from uncertain LO problem to its constraint-wise envelope, the RC remains intact.

Exercise 1.5 1. Consider the uncertain problems $\mathcal{P}_1$ and $\mathcal{P}_2$ from Exercise 1.4. Which one of them, if any, has constraint-wise uncertainty? Which one of them, if any, is the constraint-wise envelope of the other problem?

2. * Let $\mathcal{P}$ be an uncertain LO program with constraint-wise uncertainty such that the feasible sets of all instances belong to a given in advance convex compact set X (e.g., all instances share common system of certain box constraints). Prove that in this case there is no gap between feasibility/optimality properties of instances and those of the RC: the RC if feasible if and only if all instances are so, and in this case the optimal value of the RC is equal to the maximum, over the instances, of the optimal values of the instances.