Introduction to Robust Optimization

Part 1

Giuseppe C. Calafiore

Dipartimento di Automatica e Informatica
Politecnico di Torino

IPAM, UCLA Sept. 2010

Outline

Abstract

- We give an introductory overview of the basics of Robust (convex) Optimization (RO): a methodology aimed at immunizing optimization problems against *uncertainty* in the data.

- In Part 1 of the talk, we provide the fundamental definitions and tools, and then introduce RO by discussing robust counterparts of linear programming (LP) problems affected by deterministic data uncertainty. Besides worst-case immunization, we shall also discuss probabilistic immunization, when data uncertainty is described by a stochastic model.

- In Part 2, we extend the RO methodology from LP to second order conic programs (SOCP) and semidefinite programs (SDP) affected by deterministic uncertainty, focusing on tractability and approximation issues.

Sources

This presentation is mainly based on the material found in the book:

- *Robust Optimization*, by A. Ben Tal, L. El Ghaoui, A. Nemirovski; Princeton University Press, 2009

  to which the audience is referred to also as a comprehensive source of pointers to the recent literature.

The part on probability constrained linear programs is based on the paper:

- G. Calafiore, L. El Ghaoui, ‘Distributionally Robust Chance-Constrained Linear Programs with Applications.” Journal of Optimization Theory and Applications (Springer), Vol. 130, n. 1, pp. 1-22, 2006.

Outline

Motivation

- The optimal solution of an optimization problem depends on the *data* describing the problem. For example, the optimal solution x^* of an LP

$$\min_x \quad c^\top x \quad \text{s.t.}$$
$$Ax \leq b$$

obviously depends on A, b.

- In turn, the data comes from physical description of the problem. E.g., A, b may depend on geometry, physical characteristics (weight, density), forecast data, load scenarios, etc. These data are thus almost always affected by *uncertainty*, coming from measurement errors, mechanical inaccuracies, variability in forecasts (e.g. fluctuations in supply/demand levels), etc.
- When typical, nominal, data are used to solve the optimization problem, there is in general no guarantee that the optimal solution x^*, based on these nominal data, will still be optimal, or even feasible, when implemented in practice, on "true" data.
- RO aims at finding solutions that are **resilient** to pre-specified ranges of variation in the input parameters.

A General Model

- $x \in \mathbb{R}^n$: the decision variable
- $\zeta_i \in \mathbb{R}^k$: the *i*-th uncertain parameter
- $f_0(x) : \mathbb{R}^n \to \mathbb{R}$: the (convex) objective function
- $f_i(x, \zeta_i) : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$: the *i*-th *uncertain* constraint function. f_i is convex in the first variable
- $\mathcal{U}_i \subseteq \mathbb{R}^k$: the set where ζ_i lives

Definition (Robust Optimization Problem (ROP))

$$\begin{aligned} \min_x \quad & f_0(x) \quad \text{s.t.:} \\ & f_i(x, \zeta_i) \leq 0, \quad \forall \zeta_i \in \mathcal{U}_i, \quad i = 1, \ldots, m. \end{aligned}$$

A General Model

Definition (Robust Feasible Solution)

$x \in \mathbb{R}^n$ is a *robust feasible solution* for the ROP if $x \in \mathcal{X}$, where

$$\mathcal{X} = \{x : f_i(x, \zeta_i) \le 0, \forall \zeta_i \in \mathcal{U}_i, \quad i = 1, \ldots, m\}$$

Remarks:

- $\mathcal{X}$ is *convex* (intersection of convex sets);
- ROP is a convex optimization problem;
- nevertheless, if $\mathcal{U}_i$ are continuous sets, ROP entails an infinite number of constraints (semi-infinite optimization problem) and can be very hard to solve.

A Simple LP

Consider the following robust version of a simple LP with one uncertain constraint:

$$\min_{x} \quad x_2 \quad \text{s.t.:}$$

$$x_1 \le 1, \; -x_1 \le 1$$

$$x_2 \le 1$$

$$a(\zeta)^\top x \le 1, \; \forall \zeta \in \mathcal{U},$$

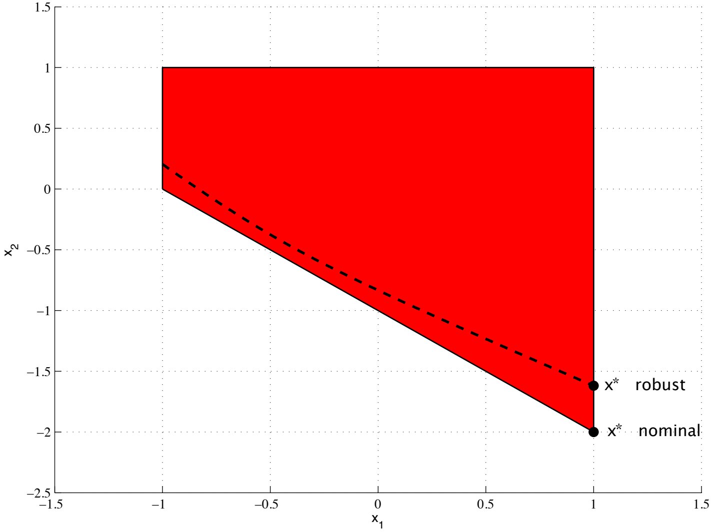
where $a(\zeta) = \bar{a} + \rho\zeta$, being

$$\bar{a} = [-1 \; -1], \quad \mathcal{U} = \{\zeta : \|\zeta\| \le 1\}, \quad \rho = 0.2.$$

Nominal feasible region is in Red (figure). Note that:

- Nominal optimal solution is unfeasible for the robust problem;
- Robustification may change the "nature" of the problem. Namely, robust problem is no longer an LP.

A Simple LP



Outline

Robust LP

We next focus on Linear Progams.

Discuss two cases with easily representable and tractable robust counterparts:

- LP with interval and polyhedral uncertainty;
- LP with ellipsoidal uncertainty.

Robust LP with Interval Uncertainty

Consider a robust LP

$$\begin{array}{lll} \min\limits_x & c^\top x & \text{s.t.} \\ & A(\zeta)x \leq b & \forall \zeta \in \mathcal{U}, \end{array}$$

where $A(\zeta) = \bar{A} + \zeta$, and $\mathcal{U}$ is a box:

$$\mathcal{U} = \{\zeta : |\zeta| \leq Z\},$$

where Z is an $m \times n$ matrix of nonnegative elements. Clearly,

$$\begin{aligned} A(\zeta)x \leq b \, \forall \zeta \in \mathcal{U} \quad & \Leftrightarrow \quad \bar{A}x + \zeta x \leq b \, \forall \zeta : |\zeta| \leq Z \\ & \Leftrightarrow \quad \bar{A}x + Z|x| \leq b. \end{aligned}$$

Robust LP with Interval Uncertainty

Let $\mathcal{X} = \{x : \bar{A}x + Z|x| \leq b\}$, and let

$$\mathcal{X}^+ = \{(x, u) : |x| \leq u; \ \bar{A}x + Zu \leq b\}.$$

Note that $x \in \mathcal{X} \Leftrightarrow \exists u$: $(x, u) \in \mathcal{X}^+$.
Indeed, if $x \in \mathcal{X}$, just take $u = |x|$ and we see that $(x, u) \in \mathcal{X}^+$.
Conversely, if $(x, u) \in \mathcal{X}^+$, then $\bar{A}x + Z|x| \leq \bar{A}x + Zu \leq b$, thus $x \in \mathcal{X}$.

The set $\mathcal{X}^+$ is said to be a *lifting* of $\mathcal{X}$, meaning that the projection of $\mathcal{X}^+$ onto the space of x variables coincides with $\mathcal{X}$. Lifting is a standard "trick" in robust optimization.

Robust LP with Interval Uncertainty

Thus, robust LP with interval uncertainty is representable by a standard LP in (x, u):

$$
\begin{aligned}
\min_{x,u} \quad & c^\top x \quad \text{s.t.} \\
& -u \leq x \leq u \\
& Ax + Zu \leq b.
\end{aligned}
$$

Robust LP with Polyhedral Uncertainty

Interval uncertainty is a special case of polyhedral uncertainty. Consider a robust LP

$$\min_x \quad c^\top x \quad \text{s.t.}$$

$$a_i^\top(\zeta_i)x \leq b_i, \ \forall \zeta_i \in \mathcal{U}_i, \ i = 1, \ldots, m,$$

where $a_i(\zeta_i) = \bar{a}_i + \zeta_i$, and ζ_i lies in a polyhedron:

$$\mathcal{U}_i = \{\zeta_i : D_i\zeta_i \leq d_i\}.$$

Robust LP with Polyhedral Uncertainty

For the *i*-th robust constraint, we have

$$a_i^\top(\zeta_i)x \leq b_i,\ \forall \zeta_i \in \mathcal{U}_i \quad \Leftrightarrow \quad \bar{a}_i^\top x + \max_{\zeta_i \in \mathcal{U}_i} \zeta_i^\top x \leq b_i$$

$$\Leftrightarrow \quad \max_{\zeta_i : D_i \zeta_i \leq d_i} \zeta_i^\top x \leq b_i - \bar{a}_i^\top x.$$

Next recall LP duality:

$$\begin{array}{l} \max\ \zeta_i^\top x \text{ s.t.:} \\ D_i \zeta_i \leq d_i \end{array} \quad \Leftrightarrow \quad \begin{array}{l} \min\ u_i^\top d_i \text{ s.t.:} \\ u_i^\top D_i = x \\ u_i \geq 0 \end{array}$$

and note that:

$$\min_{u_i^\top D_i = x, u_i \geq 0} u_i^\top d_i \leq t \ \Leftrightarrow\ u_i^\top d_i \leq t \text{ for some } u_i : u_i^\top D_i = x, u_i \geq 0.$$

Robust LP with Polyhedral Uncertainty

Therefore, the i-th robust constraint is represented in lifted space as:

$$\max_{\zeta_i : D_i \zeta_i \leq d_i} \zeta_i^\top x \leq b_i - \bar{a}_i^\top x$$

$$\Updownarrow$$

$$\begin{aligned} u_i^\top d_i &\leq b_i - \bar{a}_i^\top x \\ u_i^\top D_i &= x \\ u_i &\geq 0 \end{aligned}$$

And the overall robust polyhedral LP is representable again as a standard LP in $(x, u_1, \ldots, u_m)$ variables:

$$\min_{x, u_1, \ldots, u_m} c^\top x \quad \text{s.t.:}$$

$$\begin{aligned} \bar{a}_i^\top x + u_i^\top d_i &\leq b_i, \; i = 1, \ldots, m \\ u_i^\top D_i &= x, \; i = 1, \ldots, m \\ u_i &\geq 0, \; i = 1, \ldots, m. \end{aligned}$$

Robust LP with Ellipsoidal Uncertainty

A second case which is easily converted into an explicit and exact representation is that of affine ellipsoidal uncertainty:
Consider the robust LP

$$\min_x \quad c^\top x \quad \text{s.t.}$$

$$a_i^\top(\zeta_i)x \leq b_i, \ \forall \zeta_i \in \mathcal{U}_i, \ i = 1, \ldots, m,$$

where $a_i(\zeta_i) = \bar{a}_i + E_i\zeta_i$, and ζ_i lies in the Euclidean unit ball:

$$\mathcal{U}_i = \{\zeta_i : \|\zeta_i\| \leq 1\}.$$

In this case, each robust constraint is equivalent to:

$$\max_{\|\zeta_i\| \leq 1} \zeta_i^\top E_i^\top x \leq b_i - \bar{a}_i^\top x,$$

where the max on the left is achieved at $\zeta_i = E_i^\top x / \|E_i^\top x\|$.

Robust LP with Ellipsoidal Uncertainty

Thus, the i-th robust constraint is equivalent to

$$\|E_i^\top x\| \leq b_i - \bar{a}_i^\top x,$$

and the robust LP with ellipsoidal uncertainty reduces to the following explicit program:

$$\min_x \quad c^\top x \quad \text{s.t.}$$
$$\|E_i^\top x\| \leq b_i - \bar{a}_i^\top x, \;\; i = 1, \ldots, m.$$

Remark: This is no longer an LP! It is a (convex) second order cone program (SOCP). Still solvable very efficiently.

Parenthesis: Cones & co.

- A nonempty subset K of a Euclidean space $(E, \langle \cdot, \cdot \rangle_E)$ is called a **cone**, if whenever $t_1, t_2 \geq 0$ and $x_1, x_2 \in K$, we have $t_1 x_1 + t_2 x_2 \in K$.
- A cone K is called *regular*, if it is closed, possesses a nonempty interior and does not contain lines (*pointed*).
- If K is a cone, then the set

$$K^* = \{e \in E : \langle e, h \rangle_E \geq 0 \, \forall h \in K\}$$

is said the cone *dual* to K.

Examples

- Nonnegative orthant: $\mathbb{R}^n{}_+ = \{x \in \mathbb{R}^n : x \geq 0\}$
- Lorentz cone: $L^n = \{x \in \mathbb{R}^n : x_n \geq \sqrt{\sum_{j=1}^{n-1} x_j^2}\}$
- Semidefinite cone: $S_+^n = \{A \in \mathbb{R}^{n,n} : A = A^\top, A \succeq 0\}$.

Conic optimization

A *conic problem* is an optimization problem of the form:

$$\mathrm{Opt}(P) = \min_x \left\{ \langle c, x \rangle_E : \begin{array}{l} A_i x - b_i \in K_i,\ i = 1, \dots, m, \\ Ax = b \end{array} \right\} \qquad (P)$$

where

- $(E, \langle \cdot, \cdot \rangle_E)$ is a Euclidean space of *decision vectors* x and $c \in E$ is the *objective*;
- A_i, $1 \le i \le m$, are linear maps from E into Euclidean spaces $(F_i, \langle \cdot, \cdot \rangle_{F_i})$, $b_i \in F_i$ and $K_i \subset F_i$ are regular cones;
- A is a linear mapping from E into a Euclidean space $(F, \langle \cdot, \cdot \rangle_F)$ and $b \in F$.

Examples

- *Linear Programming:* conic problems associated with nonnegative orthants $\mathbb{R}^m{}_+$: $\min_x\{c^\top x : Ax - b \geq 0\}$;
- *Second Order Cone Programming:* conic problems associated with cones which are *finite direct products* of Lorentz cones:

$$\min_x \left\{c^\top x : [A_1; \ldots; A_m]x - [b_1; \ldots; b_m] \in L^{k_1} \times \ldots \times L^{k_m}\right\}$$

  where A_i are $k_i \times \dim x$ matrices and $b_i \in \mathbb{R}^{k_i}$;
- *Semidefinite Programming:* conic problems associated with cones which are *finite direct products* of Semidefinite cones:

$$\min_x \left\{c^\top x : A_i^0 + \sum_{j=1}^{\dim x} x_j A_i^j \succeq 0,\ 1 \leq i \leq m\right\},$$

  where A_i^j are symmetric matrices of appropriate sizes.

Conic duality

- The origin of conic duality is the desire to find a systematic way to *bound from below* the optimal value in a conic (primal) problem (P).
- The dual of (P) is the problem of maximizing a lower bound on the objective of (P):

$$\mathrm{Opt}(D) = \max_{z,\{y_i\}} \left\{ \langle z, b\rangle_F + \sum_i \langle y_i, b_i\rangle_{F_i} : \begin{array}{l} y_i \in K_i^*,\ 1 \le i \le m, \\ A^* z + \sum_i A_i^* y_i = c \end{array} \right\} \quad (D)$$

Conic duality theorem

Consider a primal-dual pair of conic problems (P), (D):

- (i) [Weak Duality] One has $\mathrm{Opt}(D) \leq \mathrm{Opt}(P)$.
- (ii) [Symmetry] The duality is symmetric: (D) is a conic problem, and the problem dual to (D) is (equivalent to) (P).
- (iii) [Strong Duality] If one of the problems (P), (D) is strictly feasible and bounded, then the other problem is solvable, and $\mathrm{Opt}(P) = \mathrm{Opt}(D)$.

  If both the problems are strictly feasible, then both are solvable with equal optimal values.

Conic representation of sets

- Let $\mathcal{K}$ be a family of regular cones. A set $Y \in \mathbb{R}^n$ is called $\mathcal{K}$-representable if it can be expressed in the form:

$$Y = \{y \in \mathbb{R}^n : \exists u \in \mathbb{R}^m : Ay + Bu - b \in K\},$$

  where $K \in \mathcal{K}$ and A, B, b are matrices and vectors of appropriate dimensions.

- Geometrically, a $\mathcal{K}$-representation of Y is the representation of Y as the *projection* on the space of y-variables of the set $Y_+ = \{[y; u] : Ax + Bu - b \in K\}$.

- given a $\mathcal{K}$-rep. of the feasible domain Y, we can rewrite an optimization program over Y as a conic program involving a cone from the family $\mathcal{K}$:

$$\min_{x=[y;u]} \left\{c^\top x := d^\top y : Ay + Bu - b \in K\right\}.$$

Conic representation of sets

- When $\mathcal{K}$ is the family of all nonnegative orthants, a $\mathcal{K}$-rep. of Y allows to rewrite the optimization program over Y as a Linear Program;
- When $\mathcal{K}$ is the family of all finite direct products of Lorentz cones, a $\mathcal{K}$-rep. of Y allows to rewrite the optimization program over Y as Conic Quadratic program;
- When $\mathcal{K}$ is the family of all finite direct products of positive semidefinite cones, a $\mathcal{K}$-rep. of Y allows to rewrite the optimization program over Y as a Semidefinite Program.

Note that a $\mathcal{K}$-representable set is always convex.

Conic representation of sets
An Example

- Consider the optimization problem

$$\min_{y} \sum_{i=1}^{m} |a_i^\top y - b_i|$$

Is this an LP?

- The answer is that it can be represented as an LP, via a conic representation.
- First use "epigraphic" form

$$\min_{t,y} \ t : \sum_{i=1}^{m} |a_i^\top y - b_i| \leq t$$

- Then use conic representation:

$$\{(y,t) : \sum_{i=1}^{m} |a_i^\top y - b_i| \leq t\} = \left\{ (y,t) : \exists u : \begin{array}{l} t - \sum_{i=1}^{m} u_i \geq 0 \\ u_i - a_i^\top y + b_i \geq 0 \\ u_i + a_i^\top y - b_i \geq 0 \end{array} \right\}$$

- Therefore the problem is represented in standard LP form as:

$$\min_{y,t,u} \quad t \quad \text{s.t.:}$$

$$t - \sum_{i=1}^{m} u_i \geq 0$$

$$u_i - a_i^\top y + b_i \geq 0$$

$$u_i + a_i^\top y - b_i \geq 0.$$

A quite general tractable robust LP

- Consider a family of uncertainty-affected linear inequalities:

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

- With the data varying in the uncertainty set

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

- Where $\mathcal{Z}$ has a conic representation:

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

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

In the case when K is *not* polyhedral, assume that the representation is strictly feasible: $\exists(\bar{\zeta}, \bar{u}) : P\bar{\zeta} + Q\bar{u} + p \in \operatorname{int} K$.

A quite general tractable robust LP

Theorem

The semi-infinite constraint $\{a^\top x \leq b\}$, $\forall [a; b] \in \mathcal{U}$ can be represented by the following system of conic inequalities in variables $x \in \mathbb{R}^n$, $y \in \mathbb{R}^N$:

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

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

A quite general tractable robust LP

Corollary

Let the nonempty $\mathcal{Z}$ be:

(i) polyhedral (i.e., K is the nonnegative orthant $\mathbb{R}^N_+$), or

(ii) conic quadratic representable (i.e., K is a direct product of Lorentz cones $L^k = \{x \in \mathbb{R}^k : x_k \geq \sqrt{x_1^2 + \ldots + x_{k-1}^2}\}$), or

(iii) semidefinite representable (i.e., K is the positive semidefinite cone S^k_+)

(in the cases of (ii), (iii) assume in addition that $\mathcal{Z}$ has nonempty interior.)

Then the robust linear inequality $\{a^\top x \leq b\}, \forall [a; b] \in \mathcal{U}$ 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).

Outline

Probability Constrained Linear Programs

- So far we worked with so-called "uncertain-but-bounded," deterministic data model: one only knows the domain $\mathcal{U}$ of the admissible data;
- In many engineering situations, however, one has further information on the data uncertainty, besides its "domain";
- For instance, some stochastic model of the uncertainty may be available.
- The stochastic model may be complete, i.e. one knows the exact probability distribution of the uncertainty, or it can be itself uncertain (or *ambiguous*), e.g. one only knows some moments of the distribution, or only knows that the distribution belongs to a given family of distributions.
- We next discuss some introductory issues on linear programs with stochastic data.

Example: LP with Gaussian data

$$\min_x \; c^\top x \qquad \text{s.t.}$$

$$\text{Prob}\{a_i^\top x \le b_i\} \ge 1 - \epsilon, \quad i = 1, \ldots, m$$

where $a_i \sim N(\bar{a}_i, \Sigma_i)$, $\Sigma_i \succ 0$, and $\epsilon \in (0, 1)$ is a probabilistic level measuring the acceptable risk of constraint violation.
How to solve such problem? Is it convex?
Answer:

$$\text{Prob}\{a_i^\top x \le b_i\} = \text{Prob}\left\{ \frac{(a_i - \bar{a}_i)^\top x}{\sigma(x)} \le \frac{b_i - \bar{a}_i^\top x}{\sigma(x)} \right\}$$

where $\sigma^2(x) = x^\top \Sigma_i x$. Now, $\frac{(a_i - \bar{a}_i)^\top x}{\sigma(x)} \sim N(0, 1)$ and hence

$$\text{Prob}\{a_i^\top x \le b_i\} = \Psi\left(\frac{b_i - \bar{a}_i^\top x}{\sigma(x)} \right)$$

where Ψ is the standard Gaussian cumulative distribution function.

Example: LP with Gaussian data

Therefore, we have that

$$\mathrm{Prob}\{a_i^\top x \leq b_i\} \geq 1-\epsilon \quad \Leftrightarrow \Psi\left(\frac{b_i - \bar{a}_i^\top x}{\sigma(x)}\right) \geq 1-\epsilon$$

i.e. iff

$$\Psi^{-1}(1-\epsilon)\sqrt{x^\top \Sigma_i x} + \bar{a}_i^\top x - b_i \leq 0$$

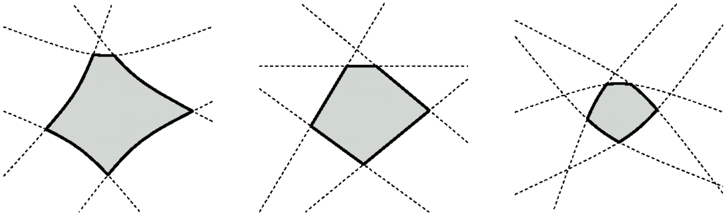
This is a convex (second order cone) constraint on x, whenever $\epsilon \leq 0.5$

$$\Downarrow$$

Probability constrained LP is efficiently solvable in this case.

Example: LP with Gaussian data

$$\{x \mid \mathbf{Prob}(a_i^T x \le b_i) \ge \eta,\ \ i = 1, \ldots, m\}$$



$\eta = 10\%$ $\eta = 50\%$ $\eta = 90\%$

LP with random data uncertainty

- We consider uncertain linear programs

$$\min_{x \in \mathbb{R}^n} \quad c^\top x \quad \text{subject to:}$$
$$a_i^\top x + b_i \leq 0, \quad i = 1, \ldots, m$$

where $x \in \mathbb{R}^n$ is the decision variable, and data $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, are random, i.e.

$$d_i \doteq \begin{bmatrix} a_i \\ b_i \end{bmatrix}, \quad i = 1, \ldots, m$$

are independent $(n+1)$-dimensional random vectors.

- In the chance-constrained approach, we fix *risk levels* $\epsilon_i \in (0, 1)$, and enforce the constraints in probability

$$\min_{x \in \mathbb{R}^n} \quad c^\top x \quad \text{subject to:}$$
$$\mathbb{P}\{a_i^\top x + b_i \leq 0\} \geq 1 - \epsilon_i, \quad i = 1, \ldots, m.$$

Outline

Depending on the distribution of (a, b), the set

$$\mathcal{X} \doteq \{x : \mathbb{P}\{a^\top x + b \leq 0\} \geq 1 - \epsilon\}$$

may be non-convex, hence optimization problem 'difficult,' or even when $\mathcal{X}$ is convex, it can be hard to express it explicitly.

- ★ Give explicit convex characterization of $\mathcal{X}$, for distributions that are radially symmetric with respect to the Euclidean norm.
- ★ Discuss several results on the distributionally robust constraint
  $$\inf_{(a,b)\sim\mathcal{F}} \mathbb{P}\{a^\top x + b \leq 0\} \geq 1 - \epsilon$$
  for certain families $\mathcal{F}$ of distributions.
- ★ Discuss the 'uniformity principle' for chance constraints.
- ★ Example.

Notation

- Under the assumption that the m constraints are independent, we may concentrate w.l.o.g. on a single generic constraint

$$\mathbb{P}\{a^\top x + b \leq 0\} \geq 1 - \epsilon, \quad \epsilon \in (0, 1)$$

- Define

$$d^\top \doteq \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^{n+1}, \quad a \in \mathbb{R}^n,\ b \in \mathbb{R}$$

$$\begin{aligned} \hat{d} &\doteq E\{d^\top\} = E\{[a^\top \ \ b]\} \doteq [\hat{a}^\top \ \ \hat{b}]; \\ \Gamma &\doteq \mathrm{var}\{d\} = \mathrm{var}\{[a^\top \ \ b]\} \doteq \begin{bmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{12}^\top & \gamma_{22} \end{bmatrix} \succeq 0. \end{aligned}$$

Notation

- We set
$$\tilde{x} \doteq \begin{bmatrix} x \\ 1 \end{bmatrix} \in \mathbb{R}^{n+1}, \quad \varphi(x) \doteq d^\top \tilde{x}$$
and
$$\begin{aligned} \hat{\varphi}(x) &\doteq E\{\varphi(x)\} = \hat{d}^\top \tilde{x}, \\ \sigma^2(x) &\doteq \mathrm{var}\{\varphi(x)\} = \tilde{x}^\top \Gamma \tilde{x}. \end{aligned}$$
- Define also the normalized random variable
$$\bar{\varphi}(x) = \frac{\varphi(x) - \hat{\varphi}(x)}{\sigma(x)}$$
such that our chance constraint is equivalently rewritten as
$$\mathbb{P}\{\varphi(x) \leq 0\} = \mathbb{P}\left\{\bar{\varphi}(x) \leq -\frac{\hat{\varphi}(x)}{\sigma(x)}\right\} \geq 1 - \epsilon.$$

Chance Constraints for Radial Distributions

For a significant class of probability distributions on d, the chance constraint can be *explicitly* expressed as a deterministic convex constraint on x.

We next introduce the class of multivariate distributions of interest.

Definition. A random vector $d \in \mathbb{R}^{n+1}$ has a Q-radial distribution with *defining function* $g(\cdot)$, if $d - E\{d\} = Q\omega$, where $Q \in \mathbb{R}^{n+1,n+1}$ is a fixed, positive definite matrix, and ω is a random vector having probability density f_ω that only depends on the Euclidean norm of ω, i.e. $f_\omega(\omega) = g(\|\omega\|)$. The function $g(\cdot)$ that defines the radial shape of the density is named the 'defining function' of d. $\star$

N.B.: Matrix Q is related to the covariance $\mathrm{var}\{d\} = \Gamma$, by

$$Q = \nu\Gamma^{1/2}, \quad \nu \doteq \left(V_{n+1}\int_0^\infty r^{n+2}g(r)dr\right)^{-1/2}.$$

where V_{n+1} denotes the volume of the unit ball in $\mathbb{R}^{n+1}$

Chance Constraints for Radial Distributions

Theorem. For any $\epsilon \in (0, 0.5]$, the chance constraint

$$\mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon,$$

where d has Q-radial distribution with defining function $g(\cdot)$ and covariance Γ, is equivalent to the convex second order cone constraint

$$\kappa_{\epsilon,r}\sigma(x) + \hat{\varphi}(x) \leq 0,$$

where $\kappa_{\epsilon,r} = \nu\Psi^{-1}(1-\epsilon)$, being Ψ the cumulative probability function of the density

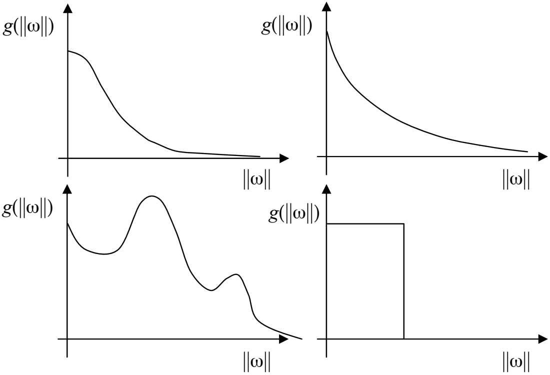
$$f_{\bar{\varphi}(x)/\nu}(\xi) = S_n \int_0^\infty g(\sqrt{\rho^2 + \xi^2})\rho^{n-1} d\rho,$$

$$\nu \doteq \left(V_{n+1} \int_0^\infty r^{n+2} g(r) dr \right)^{-1/2},$$

where S_n denotes the surface of the unit ball in $\mathbb{R}^n$.

Chance Constraints for Radial Distributions

Examples of defining functions for radial distributions.



Example: Gaussian density

- The Gaussian distribution $\mathcal{N}(\hat{d}, \Gamma)$ is Q-radial with $Q = \Gamma^{1/2}$, $\nu = 1$ and defining function

$$g(r) = \frac{1}{(2\pi)^{(n+1)/2}} \exp(-r^2/2)$$

- Consequently, $f_{\bar{\varphi}(x)}$ is the Gaussian density function

$$f_{\bar{\varphi}(x)}(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-\xi^2/2),$$

  and Ψ is the standard Gaussian cumulative probability function $\Psi(\xi) = \Psi_G(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} \exp(-t^2/2)dt$.
- For $\epsilon \in (0, 0.5]$, the safety parameter κ_ϵ results

$$\kappa_\epsilon = \kappa_{\epsilon,G} = \Psi_G^{-1}(1-\epsilon).$$

- We here recover a classical result (see e.g. Prékopa).

Uniform Distribution on Ellipsoidal Support

Lemma. Let $d - \hat{d} \in \mathbb{R}^{n+1}$ be uniformly distributed in the ellipsoid

$$\mathcal{E} = \{\xi = Qz : \|z\| \leq 1\},$$

where $Q \doteq \nu \Gamma^{1/2}$, Γ is positive definite, and $\nu \doteq \sqrt{n+3}$. Then, for any $\epsilon \in (0, 0.5]$, the chance constraint

$$\mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon,$$

is equivalent to the convex second order cone constraint

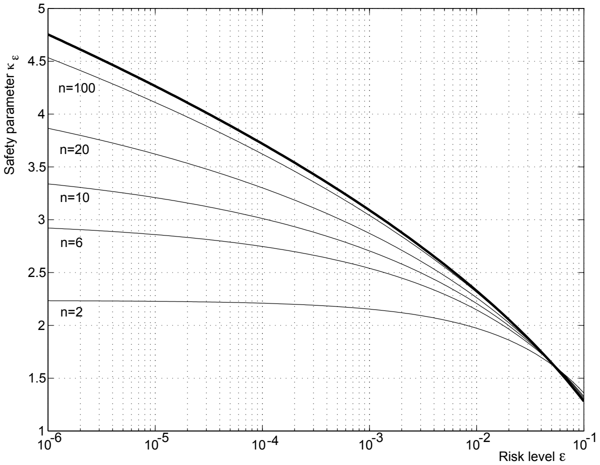
$$\kappa_{\epsilon,u} \sigma(x) + \hat{\varphi}(x) \leq 0,$$

where

$$\kappa_{\epsilon,u} = \nu \sqrt{\Psi^{-1}_{\mathrm{Beta}}(1 - 2\epsilon)},$$

being $\Psi_{\mathrm{Beta}}(\cdot)$ the cumulative distribution of a $\mathrm{Beta}\left(\frac{1}{2}; \frac{n}{2} + 1\right)$ probability density. ▸ Beta distribution

Comparison of Safety Parameters



Comparison between $\kappa_{\epsilon,G}$ (thick line) and $\kappa_{\epsilon,u}$ (light lines), for $\epsilon \in [10^{-6}, 10^{-1}]$, and various values of *n*.

Distributional Robustness

We next discuss the 'distributionally robust' chance constraint

$$\inf_{d \sim \mathcal{F}} \mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon$$

for the following distribution classes:

- $\mathcal{F}$ is the class $(\hat{d}, \Gamma)$ of all distributions of mean $\hat{d}$ and covariance Γ;
- $\mathcal{F}$ is the class $(\hat{d}, L)_{\mathcal{I}}$ of all distributions having mean $\hat{d}$, such that the elements of d are independently distributed in bounded intervals;
- $\mathcal{F}$ is a class of radially symmetric non-increasing densities (RSNID).

Distributional Robustness

Distributions with given mean and covariance.

Theorem. For any $\epsilon \in (0,1)$, the distributionally robust chance constraint

$$\inf_{d \sim (\hat{d}, \Gamma)} \mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon$$

with $\Gamma \succ 0$, is equivalent to the convex second order cone constraint

$$\kappa_\epsilon \sigma(x) + \hat{\varphi}(x) \leq 0,$$

with

$$\kappa_\epsilon = \sqrt{\frac{1-\epsilon}{\epsilon}}$$

⋆

Proof is based on a classical result of Marshall and Olkin.

Distributional Robustness

Distributions with given mean and covariance + central symmetry.

Let $(\hat{d}, \Gamma)_{\mathcal{S}}$ denote the family of symmetric distributions having mean $\hat{d}$ and covariance Γ. Symmetric meaning that the measure μ is such that $\mu(A) = \mu(-A)$, for all Borel sets $A \subseteq \mathbb{R}^{n+1}$.

Lemma. For any $\epsilon \in (0, 0.5]$, the symmetric distributionally robust chance constraint

$$\inf_{d \sim (\hat{d}, \Gamma)_{\mathcal{S}}} \mathbb{P}\{d^\top \tilde{x} \le 0\} \ge 1 - \epsilon$$

with $\Gamma \succ 0$, holds if

$$\kappa_\epsilon \sigma(x) + \hat{\varphi}(x) \le 0,$$

with

$$\kappa_\epsilon = \sqrt{\frac{1}{2\epsilon}}$$

$\star$

Proof is based on Chebychev mean-variance inequality for symmetric distributions (I. Popescu, 2002).

Distributional Robustness

Random data in independent intervals.

- Assume d of the form

$$d_i = \hat{d}_i + \delta_i, \quad i = 1, \ldots, n+1$$

  where $\delta \in \mathbb{R}^{n+1}$ is a zero-mean random vector of independent elements which are bounded in intervals

$$\delta_i \in [\ell_i^-, \ell_i^+], \quad \ell_i^+ \geq 0 \geq \ell_i^-$$

- Let us denote with $(\hat{d}, L)_{\mathcal{I}}$ the family of distributions on d satisfying the above definition, where L is a diagonal matrix containing the interval widths

$$L \doteq \mathrm{diag}(\ell_1^+ - \ell_1^-, \ldots, \ell_{n+1}^+ - \ell_{n+1}^-).$$

Distributional Robustness

Random data in independent intervals.

The following result holds.
Lemma. For any $\epsilon \in (0, 1)$, the distributionally robust chance constraint

$$\inf_{d \sim (\hat{d}, L)_{\mathcal{I}}} \mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon$$

holds if

$$\sqrt{\frac{1}{2} \ln \frac{1}{\epsilon}} \|L\tilde{x}\| + \hat{\varphi}(x) \leq 0.$$

$\star$

Distributional Robustness

Outline of Proof:

Let $d^\top \tilde{x} = \hat{d}^\top \tilde{x} + \sum_{i=1}^{n+1} \xi_i,$, where
$\xi_i \doteq x_i \delta_i, \quad i = 1, \ldots, n; \quad \xi_{n+1} \doteq \delta_{n+1}.$
Recall that $\varphi(x) = d^\top \tilde{x}$ and $\hat{\varphi}(x) = \hat{d}^\top \tilde{x}$, and hence

$$\mathbb{P}\{\varphi(x) \leq 0\} = \mathbb{P}\left\{\sum_{i=1}^{n+1} \xi_i \leq -\hat{\varphi}(x)\right\}.$$

Now, ξ_i's are zero-mean, independent and bounded in intervals of width $|x_i|(\ell_i^+ - \ell_i^-)$. Therefore, applying Hoeffding's tail probability inequality we obtain that, if $\hat{\varphi}(x) \leq 0$ then

$$\mathbb{P}\{\varphi(x) \leq 0\} \geq 1 - \exp\left(\frac{-2\hat{\varphi}^2(x)}{(\ell_{n+1}^+ - \ell_{n+1}^-)^2 + \sum_{i=1}^n x_i^2(\ell_i^+ - \ell_i^-)^2}\right)$$

from which the statement easily follows.

Distributional Robustness

Radially symmetric non-increasing densities (RSNID)

- Consider the sets

$$\begin{aligned}\mathcal{H}(\hat{d}, P) &\doteq \{d = \hat{d} + P\omega : \|\omega\|_\infty \le 1\} \\ \mathcal{E}(\hat{d}, Q) &\doteq \{d = \hat{d} + Q\omega : \|\omega\| \le 1\},\end{aligned}$$

where $P = \mathrm{diag}(p_1, \ldots, p_{n+1}) \succ 0$, $Q \succ 0$.

- The classes of densities of interest are defined as follows.

Definition (RSNID). A random vector $d \in \mathbb{R}^{n+1}$ has a probability distribution within the class $\mathcal{F}_\mathcal{H}$ (resp. $\mathcal{F}_\mathcal{E}$) if

$$d - E\{d\} = P\omega \quad \text{resp. } d - E\{d\} = Q\omega$$

where ω is a random vector having probability density f_ω such that

$$f_\omega(\omega) = \begin{cases} g(\|\omega\|_\infty), & \text{for } \|\omega\|_\infty \le 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{resp. } f_\omega(\omega) = \begin{cases} g(\|\omega\|), & \text{for } \|\omega\| \le 1 \\ 0 & \text{otherwise,} \end{cases}$$

and where $g(\cdot)$ is a non-increasing function. ⋆

Distributional Robustness

Radially symmetric non-increasing densities (RSNID)

- Notice that the uniform distribution on ellipsoidal support discussed previously belongs to the class $\mathcal{F}_{\mathcal{E}}$.
- Also, the uniform distribution on the orthotope $\mathcal{H}(\hat{d}, P)$ belongs to the class $\mathcal{F}_{\mathcal{H}}$.
- The following proposition, based on the 'uniformity principle' (Barmish & Lagoa, 1997) states an important property of these uniform distributions, namely, that they are actually the worst-case distributions in the considered classes.

Distributional Robustness

Radially symmetric non-increasing densities (RSNID)

Uniformity Principle for RSNID.
For any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint

$$\inf_{d \sim \mathcal{F}_{\mathcal{H}}} \mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon$$

is equivalent to the chance constraint

$$\mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad d \sim U(\mathcal{H}(\hat{d}, P))$$

where $U(\mathcal{H}(\hat{d}, P))$ is the uniform distribution over $\mathcal{H}(\hat{d}, P)$.
Similarly, for any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint

$$\inf_{d \sim \mathcal{F}_{\mathcal{E}}} \mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon$$

is equivalent to the chance constraint

$$\mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad d \sim U(\mathcal{E}(\hat{d}, Q))$$

where $U(\mathcal{E}(\hat{d}, Q))$ is the uniform distribution over $\mathcal{E}(\hat{d}, Q)$.

Distributional Robustness

Radially symmetric non-increasing densities (RSNID)

- As a result of previous Proposition, we have that a distributionally robust constraint over the family $\mathcal{F}_{\mathcal{E}}$ is equivalent to a probability constraint involving uniform density over ellipsoidal support, which in turn is converted into an explicit convex second order cone constraint.
- A distributionally robust constraint over the family $\mathcal{F}_{\mathcal{H}}$ is instead equivalent to a probability constraint involving uniform density over the orthotope $\mathcal{H}(\hat{d}, P)$.
- Next Lemma provides an explicit sufficient condition for enforcement of the robust constraint over $\mathcal{F}_{\mathcal{H}}$.

Distributional Robustness

Radially symmetric non-increasing densities (RSNID)

Lemma. For any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint

$$\inf_{d \sim \mathcal{F}_{\mathcal{H}}} \mathbb{P}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon$$

holds if

$$\sqrt{\frac{1}{6} \ln \frac{1}{\epsilon}} \, \|2P\tilde{x}\| + \hat{\varphi}(x) \leq 0.$$

$\star$

▸ Skip proof

Distributional Robustness

Outline of Proof (1/4):

We start by establishing a simple auxiliary result:

Let ξ be a zero-mean random variable uniformly distributed in the interval $[-c, c]$, $c \geq 0$, then for any $\lambda \geq 0$ it holds that

$$\ln E\{e^{\lambda\xi}\} \leq \frac{\lambda^2 c^2}{6}.$$

The above fact is proved as follows: Compute in closed form

$$E\{e^{\lambda\xi}\} = \frac{\sinh(\lambda c)}{\lambda c}$$

and consider the function $\psi(z) \doteq \ln \frac{\sinh(z)}{z}$, $z \doteq \lambda c$, extended by continuity to $\psi(0) = 0$. Then, we have $\psi'(0) = 0$ and $\psi''(0) = 1/3$, and moreover $\psi''(z) \leq 1/3$, $\forall z$. Therefore, by Taylor expansion with Lagrange remainder we have that for some $\theta \in [0, z]$

$$\psi(z) = \psi(0) + z\psi'(0) + \frac{z^2}{2}\psi''(\theta) \leq \frac{z^2}{6},$$

from which our preliminary statement follows.

Distributional Robustness

Outline of Proof (2/4):

Now, we write $d = \hat{d} + P\omega$ (we recall that $P = \mathrm{diag}(p_1, \ldots, n+1) \succ 0$), and hence $d^\top \tilde{x} = \hat{d}^\top \tilde{x} + \sum \tilde{x}_i p_i \omega_i$, and

$$\mathbb{P}\{d^\top \tilde{x} \le 0\} = \mathbb{P}\left\{\sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x)\right\}$$

where $\xi_i = \tilde{x}_i p_i \omega_i$, $\hat{\varphi}(x) = \hat{d}^\top \tilde{x}$.
Then, we observe that, by the uniformity principle, the infimum of the probability is attained when the ω_i's are uniformly distributed in $[-1, 1]$. Therefore, in the worst-case the ξ_i's are zero-mean, independent, and uniformly distributed in intervals $|x_i|[-p_i, p_i]$, for $i = 1, \ldots, n$, and $[-p_{n+1}, p_{n+1}]$ respectively.

Distributional Robustness

Outline of Proof (3/4):

By the Chernoff bounding method applied to the Markov probability inequality, we next have that for $\hat{\varphi}(x) \leq 0$ and any $\lambda \geq 0$

$$\mathbb{P}\left\{\sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x)\right\} \leq \frac{E\left\{e^{\lambda \sum_{i=1}^{n+1} \xi_i}\right\}}{e^{-\lambda\hat{\varphi}(x)}} = \frac{\prod_{i=1}^{n+1} E\left\{e^{\lambda\xi_i}\right\}}{e^{-\lambda\hat{\varphi}(x)}}.$$

By the preliminary result, we further have

$$E\left\{e^{\lambda\xi_i}\right\} \leq e^{\frac{(\lambda p_i x_i)^2}{6}}, \quad i = 1, \ldots, n, \text{ and } E\left\{e^{\lambda\xi_{n+1}}\right\} \leq e^{\frac{(\lambda p_{n+1})^2}{6}}$$

and hence

$$\mathbb{P}\left\{\sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x)\right\} \leq e^{\lambda^2 \|2P\tilde{x}\|^2/24 + \lambda\hat{\varphi}(x)} \leq e^{-6\frac{\hat{\varphi}^2(x)}{\|2P\tilde{x}\|^2}}$$

where the last inequality obtains selecting $\lambda \geq 0$ so to minimize the bound, which results in

$$\lambda = \frac{-12\hat{\varphi}(x)}{\|2P\tilde{x}\|^2}.$$

Distributional Robustness

Outline of Proof (4/4):

Finally, probability

$$\mathbb{P}\left\{\sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x)\right\}$$

is smaller than $\epsilon \in (0, 0.5]$ if $\hat{\varphi}(x) \leq 0$ and

$$\|2P\tilde{x}\|^2 \ln \frac{1}{\epsilon} \leq 6\hat{\varphi}^2(x),$$

which can be compactly rewritten as the convex second order cone constraint

$$\sqrt{\frac{1}{6} \ln \frac{1}{\epsilon}} \|2P\tilde{x}\| + \hat{\varphi}(x) \leq 0,$$

thus proving the claim.

Example: Portfolio Optimization

- Consider a portfolio $p^- \in \mathbb{R}^n$ consisting of n assets (which may include cash), where $p_i^- \geq 0$, $i = 1, \ldots, n$ represent the dollar value of the current holdings in each asset.
- In the so-called single-stage portfolio problem, the investor should decide an adjustment $x \in \mathbb{R}^n$ of the portfolio, where x_i denotes the dollar amount transacted in asset i ($x_i > 0$ for buying, and $x_i \leq 0$ for selling), and after the transaction the adjusted portfolio $p = p^- + x$ is held for a fixed amount of time.
- In the classical Markowitz framework, the investor goal is to maximize the total expected wealth at the end of the investment period, while satisfying a set of constraints on the portfolio, which may include bounds on the amounts held in each single asset, as well as limits on the exposure to risk.

Example: Portfolio Optimization

- A standard approach to the problem is to assume that the *return* r_i of asset i over the considered period is a random variable, and that the expected value $\hat{r}_i$ and covariance terms of the returns are known.
- We denote with $r \doteq [r_1 \ \cdots \ r_n]^\top$ the vector of returns, with $\hat{r}$ its expected value, and with Σ its covariance, and consider the family $(\hat{r}, \Sigma)$ of all possible distributions on the returns, compatible with the given mean and covariance.
- Now, defining $R \doteq \mathrm{diag}(r_1, \ldots, r_n)$, we have that the portfolio at the end of the investment period is the random vector

$$p^+ = R(p^- + x),$$

while the total wealth is $\mathbf{1}^\top p^+$, where $\mathbf{1}$ denotes a vector of ones.

Example: Portfolio Optimization

- We here consider the following simplified portfolio optimization problem:

$$\max_x \quad E\{\mathbf{1}^\top p^+\} \text{ subject to} \tag{1}$$

$$\mathbf{1}^\top x + \tau(x) \leq 0 \tag{2}$$

$$p^- + x \geq 0 \tag{3}$$

$$\sup_{r\sim(\hat{r},\Sigma)} \mathbb{P}\{\mathbf{1}^\top p^+ \leq w_{low}\} \leq \epsilon \tag{4}$$

  where (2) is a budget constraint taking into account all transaction costs $\tau(x)$, (3) is a no-shortselling constraint, and (4) is a shortfall risk constraint. In words, this latter constraint imposes a small probability ϵ on the event that the end-of-period wealth be lower than an undesired level w_{low}.
- Notice that we do not assume a specific distribution for the returns, but rather impose the probability constraint robustly with respect to all probability distributions on the returns that have the specified mean and covariances.

Example: Portfolio Optimization

- In our schematic example, the transaction costs are assumed to be proportional to the transacted amounts, i.e.

$$\tau(x) \doteq \alpha \sum_{i=1}^{n} |x_i| = \alpha \|x\|_1,$$

where $\alpha \geq 0$ is the fixed unit transaction cost.

- Now, the distributionally robust constraint (4) holds if

$$\kappa \|\Sigma^{1/2}(p^- + x)\| - \hat{r}^\top (p^- + x) + w_{low} \leq 0,$$

with $\kappa = \sqrt{(1-\epsilon)/\epsilon}$. Therefore, the stochastic problem (1)–(4) is converted into the explicit convex problem

$$\begin{aligned} \max_x \quad & \hat{r}^\top (p^- + x) \text{ subject to} \\ & \mathbf{1}^\top x + \alpha \|x\|_1 \leq 0 \\ & p^- + x \geq 0 \\ & \kappa \|\Sigma^{1/2}(p^- + x)\| - \hat{r}^\top (p^- + x) + w_{low} \leq 0. \end{aligned}$$

Example: Portfolio Optimization

- For the purpose of the example, we considered a portfolio holding period of 20 days, and five assets from the S&P 500 basket (tickers: AOL, CSCO, DELL, EQR, TXN), plus cash, i.e. $n = 6$. Cash is assumed to have unit return and zero covariance (riskless asset).
- We (crudely) estimated the 20-day average returns and covariances for the assets from historical data (closing prices from 2002-05-14 to 2003-05-13, using 0.98 forgetting factor), obtaining

$$\hat{r}^\top = \begin{bmatrix} 1.2018 & 1.2197 & 1.1744 & 1.0698 & 1.3776 & 1 \end{bmatrix},$$

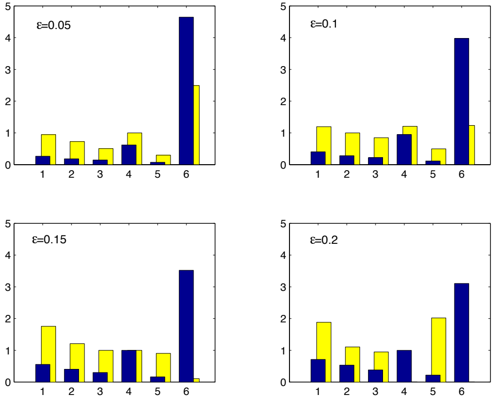
and

$$\Sigma = \begin{bmatrix} 0.5014 & 0.1839 & 0.1471 & 0.0616 & 0.2354 & 0 \\ 0.1839 & 0.4817 & 0.2350 & 0.0891 & 0.4921 & 0 \\ 0.1471 & 0.2350 & 0.4868 & 0.0594 & 0.3275 & 0 \\ 0.0616 & 0.0891 & 0.0594 & 0.0733 & 0.1587 & 0 \\ 0.2354 & 0.4921 & 0.3275 & 0.1587 & 1.8530 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Example: Portfolio Optimization

- Assuming transaction cost $\alpha = 1\%$, initial holdings $p^{-} = [1\,1\,1\,1\,1\,1]^{\top}$ and $w_{low} = 4$, we obtained the updated portfolios shown in the next figure, for different values of risk level ϵ.
- Notice that the use of a distributionally robust probability constraint leads to very 'cautious' choices of the portfolios (strong bias towards riskless asset).
- More 'aggressive' portfolios are instead obtained if we assume that the random returns obey to a Gaussian distribution, in which case the constant κ in (5) is set to $\kappa = \Psi_G^{-1}(1-\epsilon)$, where Ψ_G is the standard Gaussian cumulative function.

Example: Portfolio Optimization



Composition of optimal portfolios for different values of risk level ϵ. The abscissae report the asset type (1=AOL, 2=CSCO, 3=DELL, 4=EQR, 5=TXN, 6=cash). The blue bars show the composition of the distributionally robust portfolios, while the yellow ones refer to the Gaussian case.

End of Part 1

Thank you for your attention!

Beta distribution

The density function of a $\text{Beta}_{(a,b)}$ random variable is

$$\beta(x; a, b) = B^{-1}(a, b)x^{a-1}(1 - x)^{b-1}, \quad x \in [0, 1],$$

where $B(a, b)$ is the normalization constant.

Hoeffding's Inequality

- Let $x_1, \ldots, x_n$ be independent, zero-mean random variables bounded in itervals: $x_i \in [a_i, b_i]$, $i = 1, \ldots, n$.
- Let $S_n = x_1 + x_2 + \cdots + x_n$
- Then,

$$\mathbb{P}\{S_n \geq t\} \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Chernoff bounding

- **Markov's inequality**

  If S is any random variable and $a > 0$, then

$$\mathbb{P}\{|S| \geq a\} \leq \frac{\mathrm{E}(|S|)}{a}.$$

- The Chernoff bound for a random variable S, which is the sum of n independent random variables $x_1, \ldots, x_n$, is obtained by applying the Markov's inequality on $\exp(tS)$, for some well-chosen value of t.

- For any $t > 0$,

$$\mathbb{P}\{S \geq a\} = \mathbb{P}\left\{\mathrm{e}^{tS} \geq \mathrm{e}^{ta}\right\} \leq \frac{\mathbf{E}\left[\mathrm{e}^{tS}\right]}{\mathrm{e}^{ta}} = \frac{\prod_i E[\mathrm{e}^{tx_i}]}{\mathrm{e}^{ta}}.$$

Conic duality

- Primal:
$$\mathrm{Opt}(P) = \min_x \left\{ \langle c, x\rangle_E : \begin{array}{l} A_i x - b_i \in K_i,\ i = 1, \ldots, m, \\ Ax = b \end{array} \right\} \qquad (P)$$
- Let $y_i \in K_i^*$ and $z \in F$. By the definition of the dual cone, for every x feasible for (P) we have
$$\langle A_i^* y_i, x\rangle_E - \langle y_i, b_i\rangle_{F_i} \equiv \langle y_i, Ax_i - b_i\rangle_{F_i} \geq 0,\ 1 \leq i \leq m,$$
and of course $\langle A^* z, x\rangle_E - \langle z, b\rangle_F = \langle z, Ax - b\rangle_F = 0$.
- Summing up the resulting inequalities, we get
$$\langle A^* z + \sum_i A_i^* y_i, x\rangle_E \geq \langle z, b\rangle_F + \sum_i \langle y_i, b_i\rangle_{F_i}. \qquad (C)$$
- By its origin, this scalar linear inequality on x is a consequence of the constraints of (P), that is, it is valid for all feasible solutions x to (P).
- It may happen that the left hand side in this inequality is, identically in $x \in E$, equal to the objective $\langle c, x\rangle_E$; this happens if and only if
$$A^* z + \sum_i A_i^* y_i = c.$$
Whenever it is the case, the right hand side of (C) is a valid lower bound on the optimal value in (P). The dual problem is nothing but the problem of maximizing this lower bound:
$$\mathrm{Opt}(D) = \max_{z, \{y_i\}} \left\{ \langle z, b\rangle_F + \sum_i \langle y_i, b_i\rangle_{F_i} : \begin{array}{l} y_i \in K_i^*,\ 1 \leq i \leq m, \\ A^* z + \sum_i A_i^* y_i = c \end{array} \right\}.$$