Robust Convex Programming: Randomized Solutions and Confidence Levels

Giuseppe Calafiore

Dipartimento di Automatica e Informatica

Politecnico di Torino, Italy

Joint work with M.C. Campi, Università di Brescia.

Goal

- Introduce a new concept of solution for robust convex programs (RCP), in a probabilistic setting.
- Show that RCPs can be efficiently ‘solved’, within this framework.
- Analyze issues of complexity and reliability for the computed solution.
- Discuss some numerical examples.

Introduction

- Many engineering problems can be cast as optimization problems subject to convex constraints that need to be satisfied for all possible values of some unknown-but-bounded parameters.

- The usual solution approach is to transform the original semi-infinite optimization problem into a standard convex optimization one, by means of relaxation techniques.

- The feasible set of the relaxed problem is in general an inner approximation of the original feasible set, and therefore an upper bound on the actual optimal solution is obtained.

Introduction

- In this talk, we present an alternative randomized approach: by randomly sampling the uncertain parameters, we substitute the original infinite constraint set with a finite set of N constraints, and solve the resulting problem.

- Using statistical learning techniques, we provide an explicit bound on the measure (probability or volume) of the original constraints that are possibly violated by the randomized solution. This volume rapidly decreases to zero as N is increased.

Robust Convex Optimization

$$\text{RCP:} \quad \min_{x \in \mathbb{R}^n} c^T x \ \text{ subject to } x \in \mathcal{X} \cap \Omega,$$

- $\mathcal{X}$ is a closed convex subset in $\mathbb{R}^n$ with non-empty interior

- $\Omega \subset \mathbb{R}^n$ can be expressed as the (usually infinite) intersection of convex sets

$$\Omega \doteq \bigcap_{\delta \in \Delta \subset \mathbb{R}^\ell} \{x : \ f(x, \delta) \leq 0\}$$

- $f(x, \delta) : \mathcal{X} \times \Delta \to \mathbb{R}^p$ is convex in x, and the inequality is to be intended element-wise

- RCP is still convex

Robust Convex Optimization

Examples:

- **Robust Linear Programs**

$$\min_{x \in \mathbb{R}^n} \quad c^T x \quad \text{subject to}$$
$$A(\delta)x \leq b(\delta), \ \forall \delta \in \Delta,$$

where $A(\delta) \in \mathbb{R}^{m,n}$.

- **Robust Semidefinite Programs**

$$\min_{x \in \mathbb{R}^n} \quad c^T x \quad \text{subject to}$$
$$F_0(\delta) + \sum_{i=1}^{n} x_i F_i(\delta) \prec 0, \ \forall \delta \in \Delta, \ F_i = F_i^T.$$

Robust Convex Optimization

Solution Techniques:

- General RCPs are numerically hard to solve
- In some (rare) cases, RCPs are reducible to standard convex programs, and hence solved exactly
- In some cases, RCPs admit efficiently computable convex relaxations (i.e. an upper bound on the optimal objective is minimized)
- For general uncertainty structures entering non-linearly the data: no efficient method (or potentially conservative methods).

An Alternative

The probabilistic setting.

- Assume that a probability density (pdf) $p_\delta(\delta)$ is defined over the support Δ.

- Let $x \in \mathcal{X}$ be a candidate solution, and define

  The *probability of violation* of x

  $$V(x) \doteq \text{Prob}\{\delta \in \Delta : \; f(x,\delta) > 0\}.$$

  For instance, if the uniform density is assumed on Δ, then $V(x)$ measures the volume of ‘bad’ parameters δ such that the constraint $f(x,\delta) \leq 0$ is violated.

- Our goal is to devise an algorithm that returns with high probability a candidate solution having a small associated probability of violation.

The Probabilistic Setting

ϵ-level solutions.

Definition. Let $\epsilon \in [0, 1]$. We say that $x \in \mathcal{X}$ is an ϵ-level robustly feasible solution for RCP, with respect to the pdf $p_\delta(\delta)$, if $V(x) \leq \epsilon$.

Randomized RCP.

N iid samples $\delta^{(1)}, \ldots, \delta^{(N)}$ are drawn according to $p_\delta(\delta)$. Consider the convex optimization problem

$$\widetilde{\text{RCP}}_N : \quad \min_{x \in \mathbb{R}^n} c^T x \quad \text{subject to} \quad \begin{array}{l} x \in \mathcal{X} \\ f(x, \delta^{(i)}) \leq 0, \ i = 1, \ldots, N \end{array}$$

$\widehat{x}_N$: a corresponding optimal solution.

The Probabilistic Setting

- $\widehat{x}_N$ is the outcome of a random experiment (it depends on the extraction $\delta^{(1)}, \ldots, \delta^{(N)}$), i.e. it is a random variable

- By definition, a randomized algorithm returns a 'solution' not always, but with high probability

- Next key theorem states that, for given $\epsilon > 0$, and pre-specified success probability, if a sufficient number of samples is used, then $\widetilde{\mathrm{RCP}}_N$ returns an ϵ-level optimal solution.

Main Result

Theorem.

Let level $\epsilon \in [0,1]$ and $\beta \in [0,1]$ be given, and let

$$N \geq 1 + \frac{n}{\epsilon\beta}.$$

Then, with probability greater than $1 - \beta$, the randomized optimization problem $\widetilde{\text{RCP}}_N$ returns an optimal solution $\widehat{x}_N$ which is ϵ-level robustly feasible for RCP.

Preliminaries and sketch of proof

Support constraints.

- Consider a convex problem $\mathcal{P}$ in $\vartheta \in \mathbb{R}^d$

$$\mathcal{P}: \quad \min\ s(\vartheta) \quad \text{subject to} \quad \vartheta \in \mathcal{X}_i,\ i = 1, \ldots, m,$$

  where $s(\vartheta)$ is a linear objective, and $\mathcal{X}_i$, $i = 1, \ldots, m$ are closed convex sets.

- Let $\mathcal{P}_k$, $k = 1, \ldots, m$ be obtained from $\mathcal{P}$, removing the k-th constraint

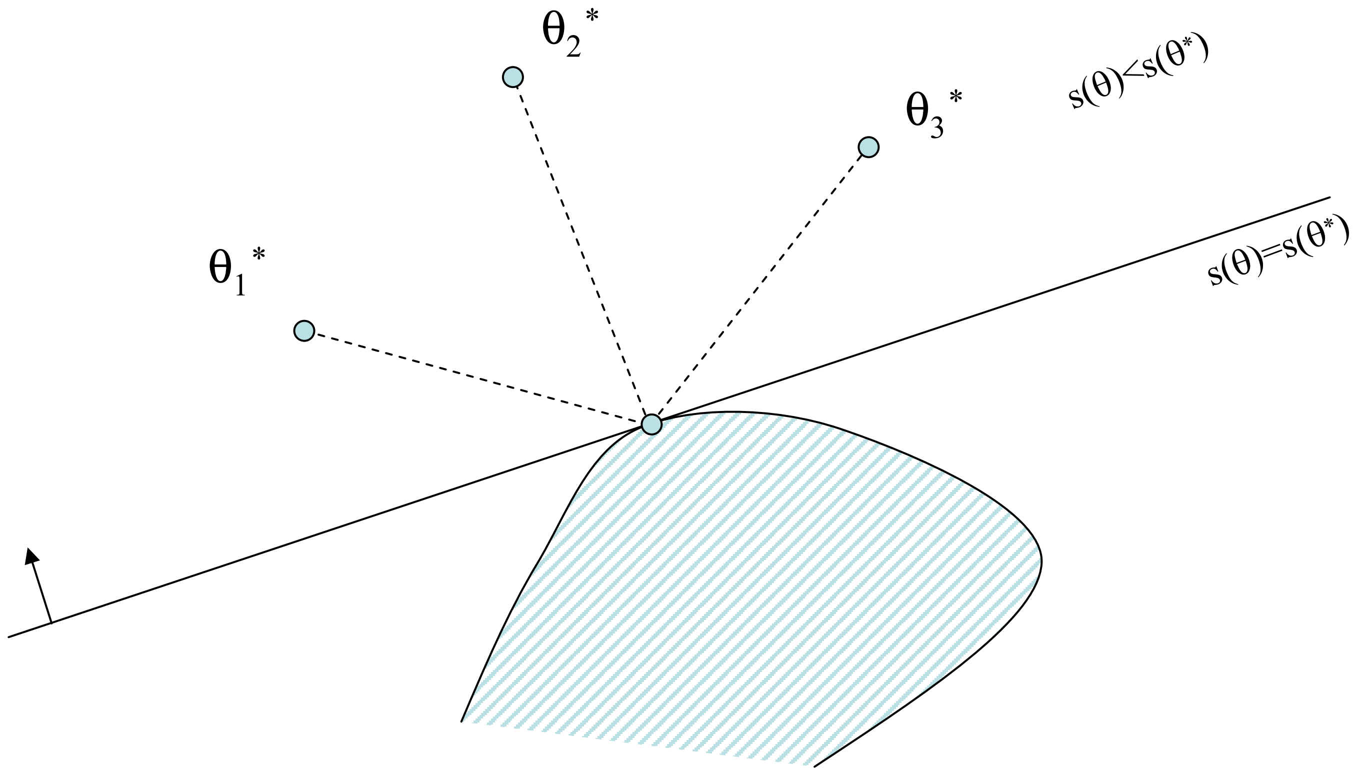
$$\mathcal{P}_k: \quad \min\ s(\vartheta) \quad \text{subject to} \quad \vartheta \in \mathcal{X}_i,\ i = 1, \ldots, k-1, k+1, \ldots, m.$$

  Let ϑ^* be optimal for $\mathcal{P}$, and ϑ_k^* be optimal for $\mathcal{P}_k$.

- We say that the k-th constraint $\mathcal{X}_k$ is a *support constraint* for $\mathcal{P}$, if $\mathcal{P}_k$ has an optimal solution ϑ_k^* such that $s(\vartheta_k^*) < s(\vartheta^*)$.

Theorem.

The number of support constraints for problem $\mathcal{P}$ is at most d.



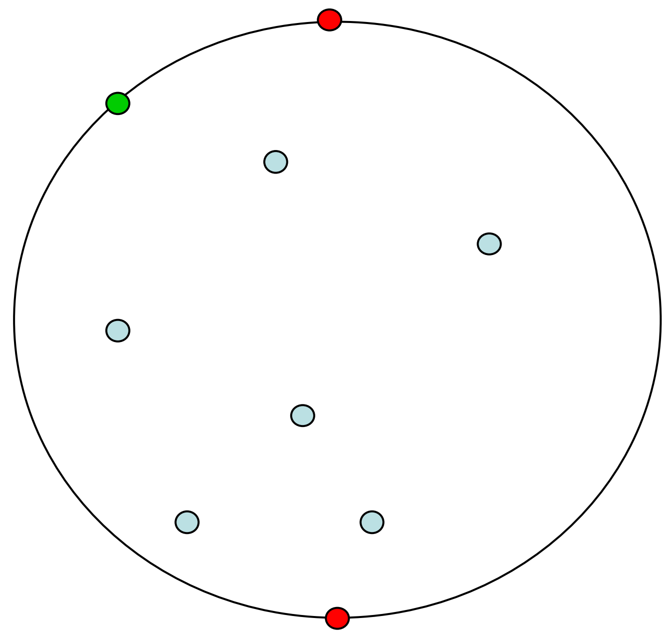
θ_2^*
$s(\theta)<s(\theta^*)$
θ_3^*
θ_1^*
$s(\theta)=s(\theta^*)$

Intuitive Example

Minimum radius sphere containing given points

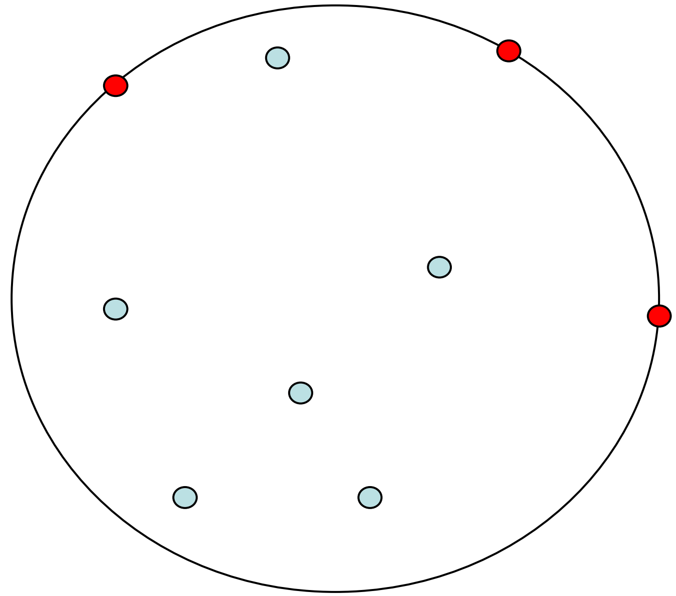
Case a)

- # of vars.: d=3
- # of constraints: m=9
- # of active constraints: 3
- # of support constraints: 2



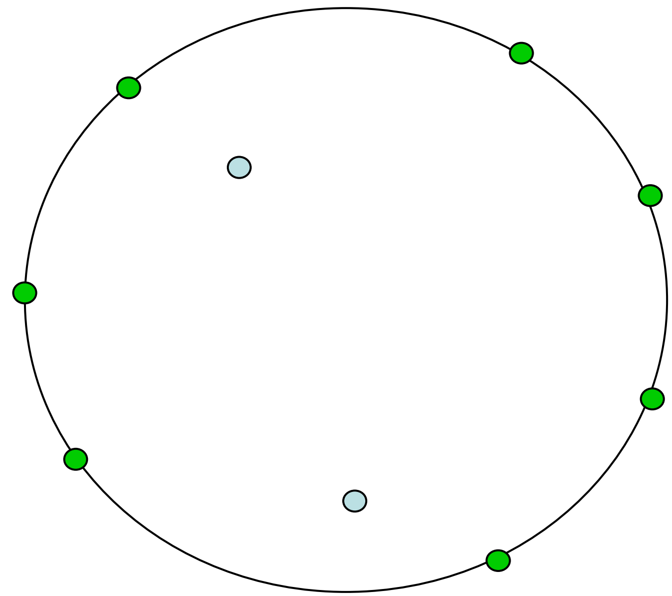
Case b)

# of vars.: d=3
- # of constraints: m=9
- # of active constraints: 3
- # of support constraints: 3



Case c)

# of vars.: d=3
- # of constraints: m=9
- # of active constraints: 7
- # of support constraints: 0



Idea of proof of main result

- Use leave-one-out estimation technique using $N+1$ samples $z^{(1)}, \dots, z^{(N+1)}$ extracted according to $p_\delta(\delta)$

- $V(\widehat{x}_N)$ is a r.v. belonging to $[0,1]$. Let

$$\begin{aligned} \bar{V}_N &\doteq E_{p_\delta^N}[V(\widehat{x}_N)], \\ \hat{\bar{V}}_N &\doteq \frac{1}{N+1}\sum_{k=1}^{N+1}\begin{cases} 1, & \text{if } f(\widehat{x}_N^k, z^{(k)}) > 0 \\ 0, & \text{otherwise.}\end{cases} \end{aligned}$$

- Since there are at most n points $\widehat{x}_N^k$ such that the constraints are violated, it follows that

$$\hat{\bar{V}}_N \leq \frac{n}{N+1}$$

- We prove that $\bar{V}_N = E_{p_\delta^{N+1}}[\hat{\bar{V}}_N]$, and therefore

$$\bar{V}_N \leq \frac{n}{N+1}$$

- The result follows using a standard bound involving expected value and tail probability

A-priori and a-posteriori probabilities

A-priori assessments.

- *Before* running the optimization, it is guaranteed by main Theorem that if $N \geq 1+n/\epsilon\beta$ samples are drawn, the solution of the randomized program will be ϵ-level robustly feasible, with probability greater than $1-\beta$.

- However, the a-priori levels ϵ, β are generally chosen not too small, for technological reasons related to limitations on the number of constraints that one specific optimization software can deal with

A-priori and a-posteriori probabilities

A-posteriori assessments.

- *After* a candidate solution $\widehat{x}_N$ is computed, one can assess the probability of feasibility using for instance Monte-Carlo techniques

- Generate a new batch of $\tilde{N}$ independent random samples of $\delta \in \Delta$, and simply construct the *empirical probability* of constraint violation, say $\widehat{V}_{\tilde{N}}(\widehat{x}_N)$.

  Then, the classical Chernoff inequality guarantees that

  $$|\widehat{V}_{\tilde{N}}(\widehat{x}_N) - V(\widehat{x}_N)| \leq \tilde{\epsilon}$$

  holds with confidence greater than $1 - \tilde{\beta}$, provided that

  $$\tilde{N} \geq \frac{\log 2/\tilde{\beta}}{2\tilde{\epsilon}^2}$$

  test samples are drawn.

Numerical Examples

Robust linear programs (1/3).

- For comparison purposes, consider a robust LP that admits an exact solution

- In particular, assume that each row $a_i^T(\delta)$ of $A(\delta)$ belongs to an ellipsoid, i.e.

$$a_i(\delta) = \hat{a}_i + E_i\delta_i, \; \|\delta_i\| \leq 1, \; i = 1, \ldots, m,$$

  $\delta = [\delta_1^T \cdots \delta_m^T]^T \in \mathbb{R}^{mn}$.

- Notice that the constraint $a_i^T(\delta)x \leq b_i$ holds for all $\delta \in \Delta$ if and only if

$$\max_{\|\delta_i\| \leq 1} \hat{a}_i^T x + \delta_i^T E_i x \leq b_i,$$

  which in turn holds, if and only if $\hat{a}_i^T x + \|E_i x\| \leq b_i$.

Numerical Examples

Robust linear programs (2/3).

- Therefore, this robust linear program has an *exact* reformulation as the second order cone program (SOCP)

$$\min_{x \in \mathbb{R}^n} \quad c^T x \quad \text{subject to}$$

$$\widehat{a}_i^T x + \|E_i x\| \leq b_i, \ i = 1, \ldots, m.$$

- To pursue the randomized approach, we assume that each vector δ_i is uniformly distributed over the ball $\|\delta_i\| \leq 1$, and for fixed ϵ, β we determine N according to our bound, and draw N samples $\delta^{(i)}, \ldots, \delta^{(N)}$.

- The randomized counterpart of the problem is therefore the linear program

$$\min_{x \in \mathbb{R}^n} \quad c^T x \quad \text{subject to}$$

$$A(\delta^{(i)}) x \leq b, \ i = 1, \ldots, N.$$

Numerical Examples

Robust linear programs (3/3).

- Consider the following numerical data

$$A(\delta) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} + 0.2 \begin{bmatrix} \delta_1^T \\ \delta_2^T \\ \delta_3^T \\ \delta_4^T \end{bmatrix}, \ \|\delta_i\| \leq 1, \ i = 1, \ldots, 4$$

  and $b = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}^T$, $c = -\begin{bmatrix} 1 & 1 \end{bmatrix}$.

- The exact robust solution is

$$x^* = [0.7795 \ 0.7795]^T$$

- For the randomized counterpart, we selected probabilistic levels $\epsilon = \beta = 0.01$, which requires $N = 20,001$ randomized constraints.

- The resulting linear program was readily solved on a PC using Matlab LP, yielding the solution

$$\widehat{x}_N = [0.7796 \ 0.7798]^T$$

Numerical Examples

Robust least-squares (1/5).

- We considered an example of robust polynomial interpolation borrowed from El Ghaoui & Lebret, SIMAX 18(4), 1997.

- For given integer $n \geq 1$, we seek a polynomial of degree $n-1$, $p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ that interpolates m given points (a_i, y_i), $i = 1, \ldots, m$, that is

$$p(a_i) = y_i, \; i = 1, \ldots, m.$$

- If the data values (a_i, y_i) were known exactly, we obtain a linear equation in the unknown x, with Vandermonde structure

$$\begin{bmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_m & \cdots & a_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

which can be solved via standard least-squares.

Numerical Examples

Robust least-squares (2/5).

- Now assume that the interpolation points are not known exactly. For instance, assume interval uncertainty on the abscissae

$$a_i(\delta) = a_i + \delta_i, \; i = 1, \ldots, m,$$

where $\delta_i \in [-\rho, \rho]$, i.e.

$$\Delta = \{\delta = [\delta_1, \ldots, \delta_m]^T : \; \|\delta\|_\infty \leq \rho\}.$$

- We seek an interpolant that minimizes the worst-case interpolation error, i.e.

$$x^* = \arg \min_{x \in \mathbb{R}^n} \max_{\delta \in \Delta} \|A(\delta)x - y\|^2,$$

where

$$A(\delta) = \begin{bmatrix} 1 & a_1(\delta) & \cdots & a_1^{n-1}(\delta) \\ \vdots & \vdots & & \vdots \\ 1 & a_m(\delta) & \cdots & a_m^{n-1}(\delta) \end{bmatrix}.$$

Numerical Examples

Robust least-squares (3/5).

- The min-max problem can be cast as RCP

$$\min_{x\in\mathbb{R}^n,\gamma} \quad \gamma \quad \text{subject to}$$
$$\|A(\delta)x - y\|^2 \leq \gamma, \ \forall \delta \in \Delta.$$

- It is not known how to solve this exactly in polynomial time, but it is possible to efficiently minimize an upper bound on the optimal worst-case residual via semidefinite programming, as it is shown in (El Ghaoui & Lebret).

- Consider the numerical data

$$(a_1, y_1) = (1, 1), \ (a_2, y_2) = (2, -0.5), \ (a_3, y_3) = (4, 2),$$

with uncertainty level $\rho = 0.2$.

Numerical Examples

Robust least-squares (4/5).

- The semidefinite relaxation approach yields a sub-optimal solution with worst-case (guaranteed) residual error equal to 1.1573.

- To apply our randomized approach, we assumed uniform distribution for the uncertain parameters, and selected probabilistic levels $\epsilon = \beta = 0.1$, which requires $N = 401$ random samples of δ.

- The randomized counterpart of the problem can be expressed as the following semidefinite program

$$\min_{x \in \mathbb{R}^n, \gamma} \quad \gamma \quad \text{subject to}$$

$$\begin{bmatrix} \gamma & (A(\delta^{(i)})x - y)^T \\ (A(\delta^{(i)})x - y) & I \end{bmatrix} \succeq 0, \; i = 1, \ldots, N.$$

  This was easily solved on a PC using standard SDP software, and yielded the solution $\widehat{x}_N = [3.7539 \;\; -3.5736 \;\; 0.7821]^T$, with corresponding residual equal to 0.6993.

Numerical Examples

Robust least-squares (4/4).

- The obtained performance level (residual) makes a $\sim 40\%$ improvement upon the one resulting from the deterministic semidefinite relaxation approach

- Of course, this improvement comes at some cost: the computed residual is not guaranteed against *all possible* uncertainties, but only *for most* of them.

- Running an a-posteriori Monte-Carlo test with $\tilde{N} = 1 \times 10^6$, we obtained an estimated violation probability $\widehat{V}_{\tilde{N}}(\widehat{x}_N) = 0.0042$.

- By the Chernoff bound, we are 99.99% confident that the actual violation probability is close to the estimated one, up to $\tilde{\epsilon} = 0.002$.

- In conclusion, the randomized program yielded a solution which provides a 40% performance improvement in residual error, at the expense of a maximum 0.6% risk of constraint violation.

Conclusions

- A new(?) probabilistic concept in robust convex programming: the ϵ-level solution

- The approach is based on the assumption that the variable δ that parameterizes the constraint family is a random variable with assigned distribution

- A randomized version $\widetilde{RCP}_N$ of the robust problem returns an ϵ-level solution with high probability, provided that a sufficient number N of samples is drawn

- We provide an 'efficient' bound for N, which scales linearly with the problem dimension n, and it is inversely proportional to the product of the probability levels $\epsilon\beta$

- In contrast to the NP-hardness results for deterministic robust convex programs, once a small risk of failure is accepted, any robust convex program can be solved efficiently in the ϵ-level sense by a randomized algorithm, no matter the way in which the uncertainty enters the data.