SPXI Vienna August 27-31, 2007

ROBUST CONVEX OPTIMIZATION

Arkadi Nemirovski
nemirovs@isye.gatech.edu
School of Industrial and Systems Engineering
Georgia Institute of Technology

- **Robust Optimization – basic concepts**
- **Tractability of Robust Optimization models**
- **Robust Optimization and Chance Constraints**

- **Robust Optimization – basic concepts**
- **Tractability of Robust Optimization models**
- **Robust Optimization and Chance Constraints**

♣ **Data of real-world optimization programs typically are** *uncertain* **– not known exactly when the problem is solved.**

Small perturbations of uncertain data can make the "nominal" optimal solution heavily infeasible and as such – practically meaningless.

Example: Here is the constraint # 372 in LP problem `PILOT4` **from NETLIB library:**

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

♠ **"Ugly" coefficients like 1.362417 represent characteristics of real-life processes/devices and thus cannot be known to high accuracy.**
What happens with the constraint, as evaluated at the nominal optimal solution, when perturbing the "ugly coefficients" by 0.1% ?

- **The worst violation of the constraint is 450% of the rhs**
- **The mean violation of the constraint is 125% of the rhs**

♠ **This phenomenon is typical: in 19 (13) of 90 NETLIB problems, already 0.01%-violations of "clearly uncertain" data coefficients result in more than 5% (50%) violation of some of the constraints as evaluated at the nominal solution.**

$\Rightarrow$ *In applications of Optimization, there exists a real need of a methodology capable to "immunize" solutions against data uncertainty. In the case of non-stochastic uncertainty, such a a methodology is offered by Robust Optimization.*

♣ **Robust Optimization is a methodology for processing uncertain optimization problems**

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

- $x \in \mathbb{R}^n$ **is the** *decision vector*
- $\zeta \in \mathbb{R}^d$ **is the** *data* **(or** *data perturbation***)**
- $f(x,\zeta) : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ **and** $F(x,\zeta) : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^m$ **are given functions, and** $\mathbf{K} \subset \mathbb{R}^m$ **is a given set.**

$f(\cdot,\cdot), F(\cdot,\cdot), \mathbf{K}$ **form the** *structure* **of the uncertain problem.**

♣ **In contrast to Stochastic Programming, RO does not assume stochastic nature of data ζ and uses instead** *uncertain-but-bounded* **data model: ζ runs through a given** *uncertainty set* $\mathcal{Z} \subset \mathbb{R}^d$**. Thus, in RO an uncertain problem is a** *family of instances*

$$\left\{ \min_x \left\{ f(x,\zeta) : F(x,\zeta) \in \mathbf{K} \right\} : \zeta \in \mathcal{Z} \right\} \tag{Unc}$$

♠ **Assume that our "decision environment" is such that**

- **All decisions x_j should be made before ζ "reveals itself" and thus should be independent of ζ**
- **The constraints are "hard": their violations cannot be tolerated**

Under these assumptions, seemingly the only meaningful way to process (Unc) is to solve the Robust Counterpart

$$\min_{t,x} \left\{ t : \forall \zeta \in \mathcal{Z} : f(x,\zeta) \le t, F(x,\zeta) \in \mathbf{K} \right\} \tag{RC}$$

of (Unc) and to treat its solution as the optimal uncertainty-immunized solution of the uncertain problem.

♣ **Robust Optimization and Stochastic Programming:**

♠ **There are situations (e.g., the "ugly coefficients" example) when uncertainty in the data is not stochastic at all, or is stochastic, but with hardly identifiable distribution. Whenever this is the case, RO is a natural replacement of SP.**

♠ **As far as the crucial issue of** *computational tractability* **of resulting models is concerned, RO turns out to be "complementary" to SP:**

- **In Uncertain Linear Programming, RO, in contrast to SP, (nearly) always leads to computationally tractable models. Moreover, RO offers a way to process LP's with stochastic data uncertainty in a computationally efficient fashion.**

- **In more sophisticated situations, like Uncertain Conic Quadratic/Semidefinite Programming, stochastic models of uncertain data seem to be better suited for efficient numerical processing than the model of uncertain-but-bounded data perturbations.**

♣ **RO, people:**

- **1973: A.L. Soyster (LP)**
- **1997: P. Kouvelis & G. Yu (IP)**
- **1997 –: L. El Ghaoui & H. Lebret & F. Oustry, A. Ben-Tal & A. Nemirovski } (CP)**
- **2000 –: E. Adida, A. Atamturk, A. Beck, D. Bertsimas, C. Bhattacharyya, H.-G. Bock, S. Boyd, G. Calafiore, M. Diehl, Y. Eldar, E. Erdogan, L. Grate, E. Guslitzer, B. Golany, D. Goldfarb, C. Hol, G. Iyengar, M. Jordan, E. Kostina, O. Kostyukova, G. Lanckriet, A. Nilim, M. Sim, D. Pachamanova, C. Roos, C. Schrerer, A. Sood, A. Thiele, J.-Ph. Vial, M. Zhang,...**

♠ **RO, applications:**

• Structural/Circuit/Network Design • Control • Signal Processing • Machine Learning • Portfolio Optimization • Inventory...

$$\left\{\min_x \{f(x,\zeta) : F(x,\zeta) \in \mathbf{K}\} : \zeta \in \mathcal{Z}\right\} \qquad \text{(Unc)}$$

$$\Downarrow$$

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

♣ **How it works. The RC of an uncertain LP with** *"interval uncertainty"* **– coefficients run through given intervals – is an explicit LP of sizes polynomial in those of instances and thus is computationally tractable. When immunizing the NETLIB LP problems against uncertainty** $a_{ij} \to (1+\zeta_{ij})a_{ij}$, $|\zeta_{ij}| \le \epsilon$ **in "ugly coefficients"** a_{ij}**, the results are as follows:**

- **With** $\epsilon = 0.01\%$ **(**$\epsilon = 0.1\%$**), the RC's of all 90 NETLIB problems are feasible, and the** *price of robustness* **– the relative difference between the robust and the nominal optimal values – is** $\le 0.4\%$ **(**$\le 1\%$**);**

- **With** $\epsilon = 1\%$**, only 4 of 90 RC's become infeasible. For the remaining 86 problems, the price of robustness is**

  • always $\le 10\%$, • $\le 3\%$ in 84 cases, • $\le 1\%$ in 77 cases.

♣ **Extending the notion of RC: Adjustable/Affinely Adjustable Robust Counterparts.**

♠ **In many applications, the assumption** *"All decisions x_j should be independent of ζ"* **is too conservative:**

- **Some of x_j can be "analysis variables" which do not represent decisions at all and can therefore depend on the entire data.**

  Example: Converting the constraint $\sum_i |a_i^T x - b_i| \leq t$ with uncertain a_i, b_i into the LP form $-y_i \leq a_i^T x - b_i \leq y_i$, $\sum_i y_i \leq t$, it is natural to allow the analysis variables y_i to "adjust themselves" to the actual data.

- **In dynamical decision-making, some of the decisions x_j should be made when the actual data becomes partially known and thus can depend on the corresponding portions of the data.**

♣ **To account for adjustability, we can allow for every x_j to depend on a prescribed portion $P_j\zeta$ of ζ: $x_j = X_j(P_j\zeta)$, thus arriving at Adjustable Robust Counterpart**

$$\min_{t,\{X_j(\cdot)\}_{j=1}^n} \{t : \forall \zeta \in \mathcal{Z} : f(X(\zeta), \zeta) \leq t, F(X(\zeta), \zeta) \in \mathbf{K}\} \qquad \text{(ARC)}$$

$$[X(\zeta) = \{X_j(P_j\zeta)\}]$$

$$\min_{t,\{X_j(\cdot)\}_{j=1}^n} \left\{t : \forall \zeta \in \mathcal{Z} : f(X(\zeta),\zeta) \le t, F(X(\zeta),\zeta) \in \mathbf{K}\right\} \qquad \text{(ARC)}$$
$$\left[X(\zeta) = \{X_j(P_j\zeta)\}_{j=1}^n\right]$$

♣ **ARC is infinite-dimensional and thus is typically heavily computationally intractable. Seemingly the only applicable technique for ARC is Dynamic Programming ⇒"curse of dimensionality"**

♠ **To overcome, to some extent, intractability of ARC, we restrict the decision rules to be** ***affine*****:** $X_j(P_j\zeta) = \alpha_j + \beta_j^T P_j\zeta$**, thus arriving at the Affinely Adjustable Robust Counterpart**

$$\min_{t,\{\alpha_j,\beta_j\}_{j=1}^n} \left\{t : \forall \zeta \in \mathcal{Z} : f(X(\zeta),\zeta) \le t, F(X(\zeta),\zeta) \in \mathbf{K}\right\} \qquad \text{(AARC)}$$
$$\left[X(\zeta) = \{\alpha_j + \beta_j^T P_j\zeta\}_{j=1}^n\right]$$

♣ **Example: Optimizing management cost in multi-product inventory with backlogged demand and shared warehouse capacity**

$$\begin{array}{lll} \min\limits_{C,v_t,x_t,y_t,w_t} & C & \textbf{[inventory management cost]} \\ \textbf{s.t.} & \sum_{t=1}^{N}[c_{\rm h}^T y_t + c_{\rm b}^T w_t + c_{\rm o}^T v_t] \leq C & \textbf{[cost description]} \\ & x_{t+1} = x_t + v_t - \zeta_t & \textbf{[balance equations]} \\ & x_t \leq y_t,\ 0 \leq y_t & \textbf{[bounds on inventory levels]} \\ & -x_t \leq w_t,\ 0 \leq w_t & \textbf{[bounds on backlogged demand]} \\ & \underline{v}_t \leq v_t \leq \overline{v}_t & \textbf{[bounds on orders]} \\ & q^T y_t \leq Q & \textbf{[warehouse capacity bound]} \end{array}$$

Variables:

- $x_t \in \mathbb{R}^d$: inventory state at time t
- $v_t \in \mathbb{R}^d$: replenishment orders at time t
- $y_t \in \mathbb{R}^d$: inventory level at time t
- $w_t \in \mathbb{R}^d$: backlogged demand at time t

Certain data:

- $q \in \mathbb{R}^d_+$: storage coefficients
- $c_{\rm h} \in \mathbb{R}^d_+$: holding costs
- $c_{\rm b} \in \mathbb{R}^d_+$: backlog penalty
- $c_{\rm o} \in \mathbb{R}^d_+$: ordering costs

- **Uncertain data:** demands $\zeta = [\zeta_1; ...; \zeta_N] \in \mathcal{Z} \subset \mathbb{R}^{Nd}$

♠ **Assume that replenishment orders v_t can depend on the preceding demands $\zeta^{t-1} = [\zeta_1; ...; \zeta_{t-1}]$. To build the AARC of the Inventory problem, we**

• **introduce linear decision rules $v_t = v_t^0 + V_t\zeta^{t-1}$ for the orders;**

• **make the analysis variables x_t, y_t, w_t affine functions of ζ: $x_t = x_t^0 + X_t\zeta$, $y_t = y_t^0 + Y_t\zeta$, $w_t = w_t^0 + W_t\zeta$, thus ending up with the AARC**

$$\min_{C, v_t^0, V_t, \ldots, w_t^0, W_t} \quad C$$

$$\text{s.t.} \quad \left.\begin{array}{l} \sum_{t\leq N}[c_{\mathrm{h}}^T[y_t^0 + Y_t\zeta] + c_{\mathrm{b}}^T[w_t + W_t\zeta] + c_{\mathrm{p}}^T[v_t^0 + V_t\zeta^{t-1}]] \leq C \\ x_{t+1}^0 + X_{t+1}\zeta = x_t^0 + X_t\zeta + v_t^0 + V_t\zeta^{t-1} - \zeta_t \\ x_t^0 + X_t\zeta \leq y_t^0 + Y_t\zeta,\ 0 \leq y_t^0 + Y_t\zeta \\ -[x_t^0 + X_t\zeta] \leq w_t^0 + W_t\zeta,\ 0 \leq w_t^0 + W_t\zeta \\ \underline{v}_t \leq v_t^0 + V_t\zeta^{t-1} \leq \overline{v}_t,\ q^T[y_t^0 + Y_t\zeta] \leq Q \end{array}\right\} \forall \zeta \in \mathcal{Z}$$

♠ **The AARC is computationally tractable provided that $\mathcal{Z}$ is so. E.g., with polyhedral $\mathcal{Z}$, the complexity of solving AARC is polynomial in d, N and the size of polyhedral description of $\mathcal{Z}$.**

<u>Note:</u> *The complexity of solving ARC blows up exponentially as d grows.*

♠ **<u>How it works:</u> Consider** *single-product* **Inventory with** $N = 10$ **and a box uncertainty set:** $(1-\rho)\zeta^{\mathrm{n}} \leq \zeta \leq (1+\rho)\zeta^{\mathrm{n}}$. **Here the ARC is well within the grasp of Dynamic Programming.**

(?) *How large are gaps in the chain* $\mathrm{Opt}(\mathrm{ARC}) \leq \mathrm{Opt}(\mathrm{AARC}) \leq \mathrm{Opt}(\mathrm{RC})$ **?**

• **We generated a sample of 768 Inventory problems with uncertainty** ρ **in the range 10% – 50% by picking at random certain data and subsequent filtering out problems with infeasible ARC's.**

• **It turns out that** $\mathrm{Opt}(\mathrm{ARC}) = \mathrm{Opt}(\mathrm{AARC})$ **in** *every one* **of these 768 problems! At the same time,** $\mathrm{Opt}(\mathrm{RC})$ **was typically essentially worse than** $\mathrm{Opt}(\mathrm{ARC}) = \mathrm{Opt}(\mathrm{AARC})$**:**

Range of $\frac{\mathrm{Opt}(\mathrm{RC})}{\mathrm{Opt}(\mathrm{AARC})}$	1	$(1, 2]$	$(\mathbf{2}, \mathbf{10}]$	$(\mathbf{10}, \mathbf{1000}]$	∞
Frequency in the sample	38%	23%	**14%**	**11%**	**15%**

⇒*In the RO context, affine decision rules not necessarily are bad!*

• **<u>Note:</u> Affine decision rules do become bad when optimizing the expected inventory management cost rather than the worst-case one...**

- **Optimization programs with uncertain data and their Robust Counterparts**
- **Tractability of Robust Counterparts**
- **Robust Optimization and Chance Constraints**

♣ **Robust Counterparts of uncertain problems are** *semi-infinite* **programs and thus can be intractable even when all instances of the uncertain problem are easy to solve.**

$\Rightarrow$ *When Robust Counterparts are computationally tractable?*
What to do if it is not the case?

♠ **We focus on** *uncertain affinely perturbed LP/CQP/SDP problems*

$$\left\{\min_x \left\{{c_\zeta}^T x + d_\zeta : A^i_\zeta x + b^i_\zeta \in \mathbf{K}_i,\, 1 \leq i \leq m\right\} : \zeta \in \mathcal{Z}\right\}$$

with fixed recourse:

• $c_\zeta, d_\zeta, A^i_\zeta, b^i_\zeta$**: affine in** ζ

• $\mathbf{K}_i$**: nonnegative rays/Lorentz cones/semidefinite cones (uncertain LP, CQP, SDP, respectively)**

• *Fixed recourse* [**automatically valid for the RC**]**: all coefficients of the** *adjustable* **variables** x_j **(those with** $P_j \neq 0$**) are independent of** ζ

♠ **We always assume that** $\mathcal{Z}$ **is given by a strictly feasible semidefinite representation** $\mathcal{Z} = \{\zeta : \exists u : \mathcal{P}(\zeta, u) \succeq 0\}$ **(**$\mathcal{P}(\cdot)$**: affine in** (ζ, u)**).**

♣ **Investigating tractability of Robust Counterparts of uncertain affinely perturbed LP/CQP/SDP problems with fixed recourse reduces to investigating tractability of** *semi-infinite affinely perturbed conic inequalities*

$$\forall \zeta \in \mathcal{Z} : A_\zeta x + b_\zeta \in \mathbf{K} = \begin{cases} \text{nonnegative ray} & \textbf{[Uncertain LP]} \\ \text{Lorentz cone} & \textbf{[Uncertain CQP]} \\ \text{semidefinite cone} & \textbf{[Uncertain SDP]} \end{cases} \qquad \text{(C)}$$

$[A_\zeta, b_\zeta$: affine in $\zeta]$

♠ **Tractability of** (C) **depends on the tradeoff between the geometries of K and $\mathcal{Z}$ – the more complicated is $\mathcal{Z}$, the simpler should be K.**

♠ **<u>“Trivial case”:</u>** *Scenario-generated uncertainty set*

<u>Theorem</u> *The RC/AARC of an uncertain affinely perturbed LP/CQP/SDP problem with fixed recourse*

$$\left\{ \min_x \left\{ c_\zeta{}^T x + d_\zeta : A^i_\zeta x + b^i_\zeta \in \mathbf{K}_i,\ 1 \le i \le m \right\} : \zeta \in \mathcal{Z} \right\}$$

and with scenario-generated uncertainty set $\mathcal{Z} = \mathrm{Conv}\{\zeta^1, ..., \zeta^N\}$ *is computationally tractable.*

♠ **"Solvable case":** *Uncertain Linear Programming*

Theorem *The RC/AARC of uncertain affinely perturbed LP problem with fixed recourse*

$$\left\{\min_x \left\{{c_\zeta}^T x + d_\zeta : [a^i_\zeta]^T x + b^i_\zeta \geq 0,\, 1 \leq i \leq m\right\} : \zeta \in \mathcal{Z}\right\}$$

is computationally tractable. With the uncertainty set $\mathcal{Z}$ given by a strictly feasible LP/CQP/SDP representation, the RC/AARC is an LP/CQP/SDP program of sizes polynomial in those of instances and in the size of the representation of the uncertainty set.

♣ **Aside of a number of highly specific particular cases,** *semi-infinite conic quadratic/linear matrix inequalities are computationally intractable.* **Whenever it is the case, a natural course of actions in the RO context is to replace an intractable semi-infinite conic inequality with its** *safe tractable approximation***.**

Definition **Consider a semi-infinite conic inequality**

$$\forall \zeta \in \mathcal{Z} : A_\zeta x + b_\zeta \in \mathbf{K} \Leftrightarrow x \in \mathcal{X} \qquad (C)$$

and let $0 \in \mathcal{Z}$**. We embed** (C) **into the parametric family**

$$\forall (\zeta \in \rho\mathcal{Z}) : A_\zeta x + b_\zeta \in \mathbf{K} \Leftrightarrow x \in \mathcal{X}[\rho] \qquad (C[\rho])$$

of semi-infinite conic inequalities ($\rho \geq 0$**: uncertainty level).**

A system of constraints $(S[\rho])$ **in variables** x **is called a** *safe tractable approximation of* $(C[\rho])$ *with tightness factor* $\vartheta \geq 1$, **if**

- **[tractability]** *The constraints in* $(S[\rho])$ *are convex and efficiently computable*
- **[safety]** *Whenever* x *is feasible for* $(S[\rho])$, x *is feasible for* $(C[\rho])$
- **[tightness]** *Whenever* x *is not feasible for* $(S[\rho])$, x *is not feasible for* $(C[\vartheta\rho])$.

Equivalently: *The feasible set* $\mathcal{Y}[\rho]$ *of* $(S[\rho])$ *is computationally tractable and satisfies*

$$\mathcal{X}[\vartheta\rho] \subset \mathcal{Y}[\rho] \subset \mathcal{X}[\rho]$$

♣ Approximating semi-infinite Conic Quadratic Inequality

$$\forall \zeta = [\zeta_\ell; \zeta_r] \in \rho\mathcal{Z} : \|A_{\zeta_\ell} x + b_{\zeta_\ell}\|_2 \leq c_{\zeta_r}{}^T x + d_{\zeta_r}$$
$$\left[A_{\zeta_\ell}, b_{\zeta_\ell}, c_{\zeta_r}, d_{\zeta_r} : \text{ affine in } \zeta\right] \qquad (C[\rho])$$

Theorem **(i)** *when* $\mathcal{Z}$ *is an ellipsoid,* $(C[\rho])$ *is computationally tractable.*

(ii) *When uncertainty is side-wise:* $\mathcal{Z} = \mathcal{Z}^\ell \times \mathcal{Z}^r$ *and the lhs uncertainty set is intersection of* $M > 1$ *ellipsoids centered at the origin:*

$$\mathcal{Z}^\ell = \{\zeta^\ell : \zeta_\ell^T Q_j \zeta_\ell \leq 1, \, 1 \leq j \leq M\}, \qquad [Q_j \succeq 0, \sum_j Q_j \succ 0]$$

$(C[\rho])$ *admits tractable approximation tight within the factor* $\vartheta = O(1)\sqrt{\ln M}$.

♣ **Approximating semi-infinite Linear Matrix Inequality**

♠ **Aside of scenario-generated uncertainty set, the only known case when a semi-infinite LMI**

$$\forall(\zeta \in \rho\mathcal{Z}) : \mathcal{A}_\zeta(x) \succeq 0 \qquad [\mathcal{A}_\zeta(x) : \text{bi-affine in } x,\, \zeta] \qquad (C[\rho])$$

admits a tight tractable approximation is the case of *structured norm-bounded uncertainty***:**

$$\mathcal{Z} = \{\zeta = \{\zeta_j\}_{j=1}^{M} : \zeta_j \in \mathbb{R}^{p_j\times p_j}, \|\zeta_j\| \leq 1, \zeta_j = \xi_j I,\, j \in \mathcal{J}\}$$

$$\mathcal{A}_\zeta(x) = \mathcal{A}^{\rm n}(x) + \sum_j \left[\mathcal{L}_j^T \zeta_j \mathcal{R}_j(x) + \mathcal{R}_j^T(x)\zeta_j^T \mathcal{L}_j\right]$$

Theorem *The semi-infinite LMI* $(C[\rho])$ *with structured norm-bounded uncertainty admits a safe tractable approximation tight within factor* $\vartheta = \vartheta(\mu)$ *depending solely on the largest size* μ *of the scalar perturbation blocks:*

$$\mu = \begin{cases} 1, & \mathcal{J} = \emptyset \\ \max_{j\in\mathcal{J}} p_j, & \mathcal{J} \neq \emptyset \end{cases},$$

and such that $\vartheta(1) = \frac{\pi}{2}, \vartheta(2) = 2,\, \vartheta(k) \leq \pi\sqrt{k}$. *When* $M = 1$, *the approximation is exact:* $\vartheta = 1$**.**

Applications: *Robust Structural Design, Lyapunov Stability Analysis/Synthesis under interval uncertainty, etc.*

Corollary *Semi-infinite Least Squares inequality*

$$\forall \zeta \in \rho\mathcal{Z} : \|[A_\zeta; b_\zeta][x;1]\|_2 \leq \tau$$

with structured norm-bounded uncertainty:

$$\mathcal{Z} = \left\{\{\zeta_j \in \mathbb{R}^{p_j \times q_j}\}_{j=1}^M : \|\zeta_j\| \leq 1, 1 \leq j \leq M\right\}$$
$$[A_\zeta; b_\zeta][x;1] = A^{\rm n}(x) + \sum_{j=1}^M L_j(x)\zeta_j R_j(x)$$

($A^{\rm n}(x), L_j(x), R_j(x)$ are affine in x and for every j either $L_j(x)$, or $R_j(x)$ is independent of x) admits safe tractable SDP approximation tight within the factor $\vartheta = \frac{\pi}{2}$. When $M = 1$, this approximation is exact: $\vartheta = 1$.

- **Optimization programs with uncertain data and their Robust Counterparts**
- **Tractability of Robust Counterparts**
- **Robust Optimization and Chance Constraints**

♣ **RO does *not* assume stochastic nature of uncertain data and uses instead uncertain-but-bounded model of data perturbations.**

<u>However:</u> Stochastic nature of uncertainty, if any, can be utilized in the RO framework.

♣ **With stochastic data, the entity of primary interest is a** *randomly perturbed conic constraint*

$$w_0(x) + \sum_{\ell=1}^d \zeta_\ell w_\ell(x) \in \mathbf{K}$$

$$\left[\begin{array}{c} \bullet\ x\text{: decision vector } \bullet\ w_0(x),...,w_d(x)\text{: affine } \bullet\ \mathbf{K}\text{: convex cone} \\ \bullet\ \zeta_1,...,\zeta_d \in \mathbb{R}\text{: "primitive" random perturbations} \end{array}\right] \qquad (C)$$

♠ **A natural way to process** (C) **is to pass to a** *chance version* **of the constraint**

$$\mathrm{Prob}\left\{w_0(x) + \sum_{\ell=1}^d \zeta_\ell w_\ell(x) \notin \mathbf{K}\right\} \leq \epsilon \qquad (C_\epsilon)$$

(A. Charnes, W. Cooper, G. Symonds, 1958).

$$\pi(w) \equiv \mathrm{Prob}\{\zeta^w \equiv w_0 + \textstyle\sum_{\ell=1}^{d} \zeta_\ell w_\ell \notin \mathbf{K}\} \leq \epsilon \qquad (C_\epsilon)$$

♠ **There exists significant literature on chance constraints (T. Badics, D. Dentcheva, A. Dupačová, L. Miller, A. Prékopa, A. Ruszczynski, B. Vizvari, H. Wagner,...)**

<u>However:</u> In general, (C_ϵ) is difficult to process:

- **In many cases, the feasible set of (C_ϵ) is non-convex**

$\Rightarrow$*Optimization under the constraint is highly problematic...*

- **Even when convex, the feasible set of (C_ϵ) can be "computationally intractable":**

> When $\zeta \sim \mathrm{Uniform}([0,1]^d)$ and $\mathbf{K} = \mathbb{R}_-$, $\pi(w)$ is quasi-convex (C. Lagoa et al, 2005). However, *unless P=NP, it is impossible to compute* $\pi(w)$ *with accuracy* $\delta > 0$ *in time polynomial in d, total binary length of (rational)* w *and* $\ln(1/\delta)$ [L. Khachiyan, 1989].

$$\pi(w) \equiv \text{Prob}\{\zeta^w \equiv w_0 + \sum_{\ell=1}^d \zeta_\ell w_\ell \notin \mathbf{K}\} \leq \epsilon \qquad (C_\epsilon)$$

♣ **When (C_ϵ) "as it is" is difficult to process, one can look for a safe tractable approximation of (C_ϵ) – a computationally tractable convex set W_ϵ such that**

$$W_\epsilon \subset \{w : \pi(w) \leq \epsilon\} \qquad (*)$$

♣ **Approximating Scalar Chance Constraint**

In the scalar case $\mathbf{K} = \mathbb{R}_-$, a natural way to build a safe convex approximation of (C_ϵ) is as follows:

Take a convex "generating function" $\gamma : \mathbb{R} \to \mathbb{R}$ with $\gamma(-\infty) = 0$ and $\gamma(0) = 1$ and set $\Gamma(w) = \mathbf{E}\{\gamma(\zeta^w)\}$.

Observation: *For every $\alpha > 0$, $\gamma(\alpha^{-1}s)$ is an upper bound on $\chi_{[0,\infty]}(s)$*

$\Rightarrow \forall(\alpha > 0) : \pi(w) \equiv \text{Prob}\{\zeta^w > 0\} \leq \Gamma(\alpha^{-1}w)$

$\Rightarrow$ *The set* $W_\epsilon = \{w : G(w) \equiv \underbrace{\inf_{\alpha>0} \left[\alpha\Gamma(\alpha^{-1}w) - \alpha\epsilon\right]}_{\text{convex in } w} \leq 0\}$ *is a safe convex approximation of (C_ϵ). This approximation is tractable, provided that $\Gamma(\cdot)$ is efficiently computable.*

$$w\in W_\epsilon \equiv \{w: G(w)\le 0\} \Rightarrow \mathrm{Prob}\{\overbrace{w_0+\sum_{\ell=1}^d \zeta_\ell w_\ell}^{\zeta^w} > 0\} \le \epsilon$$

$$\left[\begin{array}{c}\bullet\ G(w)=\inf_{\alpha>0}\left[\alpha\Gamma(\alpha^{-1}w)-\alpha\epsilon\right]\quad \bullet\ \Gamma(w)\ge \mathbf{E}\left\{\gamma(\zeta^w)\right\}:\ \text{convex}\\ \bullet\ \gamma(\cdot):\mathbb{R}\to\mathbb{R}:\ \text{convex}, \gamma(-\infty)=0, \gamma(0)=1\end{array}\right]$$

Relation to RO: *The safe approximation* $w(x)\in W_\epsilon$ *of the chance constraint*

$$\mathrm{Prob}\{w_0(x)+\sum_{\ell=1}^d \zeta_\ell w_\ell(x)>0\}\le\epsilon$$

is the robust feasible set $\left\{x: w_0(x)+\sum_{\ell=1}^d \zeta_\ell w_\ell(x)\le 0\ \forall \zeta\in\mathcal{Z}_\epsilon\right\}$ *of the uncertain linear constraint*

$$w_0(x)+\sum_{\ell=1}^d \zeta_\ell w_\ell(x)\le 0 \qquad (*)$$

with properly chosen convex compact uncertainty set $\mathcal{Z}_\epsilon$. *This set is computationally tractable, provided* $\Gamma(w)$ *is so.*
Thus, immunizing a solution x *to* $(*)$ *against all perturbations from* $\mathcal{Z}_\epsilon$, *we ensure that* x *is feasible with probability* $\ge 1-\epsilon$ *under random perturbations.*

Note: *When* d *is large and* $\zeta_1,...,\zeta_d$ *are independent, the uncertainty set* $\mathcal{Z}_\epsilon$ *is incomparably smaller than "*$(1-\epsilon)$*-supports" of the distribution of* ζ. *It can even happen that* $\mathrm{Prob}\{\zeta\in\mathcal{Z}_\epsilon\}=0$*!*

$$\text{Prob}\{\zeta^w \equiv w_0 + \sum_{\ell=1}^d \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (C_\epsilon)$$

♣ **Choosing Generating Function** $\gamma(\cdot)$

♠ **The best, in terms of tightness of the approximation, choice is** $\gamma(s) = \max[1+s, 0]$**. This choice leads to the** *Conditional Value at Risk* **convex approximation of** (C_ϵ)**:**

$$\text{CVaR}_\epsilon(w) \equiv \min_{\beta \in \mathbb{R}} \left[\beta + \frac{1}{\epsilon}\mathbf{E}\{[\zeta^w - \beta]_+\}\right] \leq 0 \Rightarrow \text{Prob}\{\zeta^w > 0\} \leq \epsilon.$$

However: Typically, $\text{CVaR}_\epsilon(\cdot)$ **is difficult to compute...**

♠ **Choosing** $\gamma(s) = \exp\{s\}$**, we arrive at** *Bernstein approximation* **(Pinter, 1989; Nem.&Shapiro, 2005). This safe convex approximation of** (C_ϵ) **is tractable, provided that**

- **Primitive perturbations** ζ_ℓ **are independent with distributions belonging to given families** $\mathcal{P}_\ell$ **of probability distributions on the axis**
- **The functions** $F_\ell(t) = \sup_{P \in \mathcal{P}_\ell} \ln\left(\mathbf{E}_{\zeta_\ell \sim P}\{\exp\{t\zeta_\ell\}\}\right)$ **are efficiently computable.**

Example (one of many): *Range and Mean a priori information on* ζ_ℓ

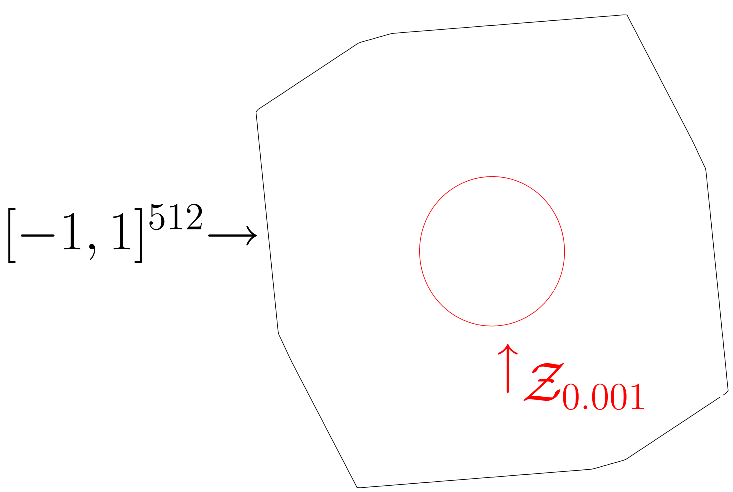
$$\mathrm{Prob}\{w_0(x) + \sum_{\ell=1}^d \zeta_\ell w_\ell(x) > 0\} \leq \epsilon \qquad (C_\epsilon)$$

$$\bullet\ \zeta_\ell \in [-1,1]\text{: independent } \bullet\ \mathbf{E}\{\zeta_\ell\} \in [\mu_\ell^-, \mu_\ell^+]$$

The Bernstein approximation of (C_ϵ) **is**

$$w_0(x) + \sum_{\ell=1}^d \zeta_\ell w_\ell(x) \leq 0\ \forall \zeta \in \mathcal{Z}_\epsilon = \{\zeta : \sum_{\ell=1}^d \phi_\ell(\zeta_\ell) \leq 2\ln(1/\epsilon)\} \qquad \text{(Br)}$$

$$\phi_\ell(s) = \begin{cases} (1+s)\ln\left(\frac{1+s}{1+\mu_\ell^-}\right) + (1-s)\ln\left(\frac{1-s}{1-\mu_\ell^-}\right) & , -1 \leq s \leq \mu_\ell^- \\ 0 & , \mu_\ell^- \leq s \leq \mu_\ell^+ \\ (1+s)\ln\left(\frac{1+s}{1+\mu_\ell^+}\right) + (1-s)\ln\left(\frac{1-s}{1-\mu_\ell^+}\right) & , \mu_\ell^+ \leq s \leq 1 \end{cases}$$



with $\zeta \sim$ **Uniform** $(\{-1;1\}^{512})$**,**
$\mathrm{Prob}\{\zeta \in \mathcal{Z}_{0.001}\} = 0$

Random 2D cross-section of $\mathcal{Z}_{0.001} \subset [-1,1]^{512}$ $(\mu_\ell^\pm = 0)$

♣ Bridging Bernstein and CVaR approximations

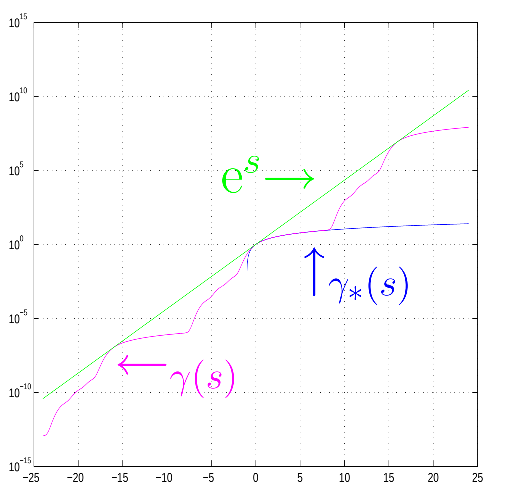
$$\mathrm{Prob}\{\zeta^w \equiv w_0 + \sum_{\ell=1}^{d} \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (C_\epsilon)$$

Assuming $\zeta_1, ..., \zeta_d$ **independent, we can handle efficiently not only the generating function** e^s**, but every generating function which is an exponential polynomial:**

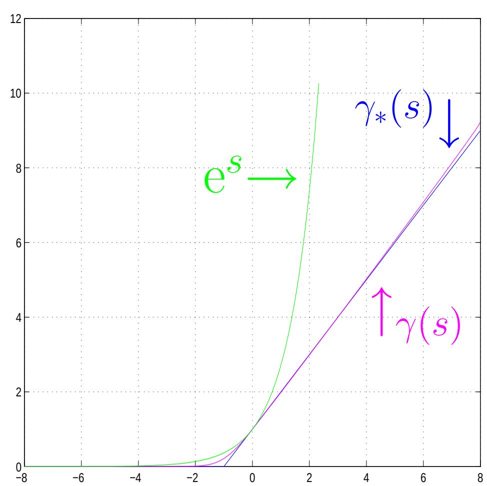
$$\gamma(s) = \sum_{i=1} a_i \exp\{\alpha_i s\} \qquad [a_i, \alpha_i \in \mathbb{C}]$$

Playing with a_i, α_i**, we can make** $\gamma(\cdot)$ **close to the optimal** CVaR**-generating function** $\gamma_*(s) = \max[1+s, 0]$**, thus reducing the conservatism of the approximation and keeping it tractable.**

$$\gamma(s) = e^s \cdot [\text{trigonometric polynomial of degree 11}]$$



Range $-24 \leq s \leq 24$
Logarithmic scale along y-axis

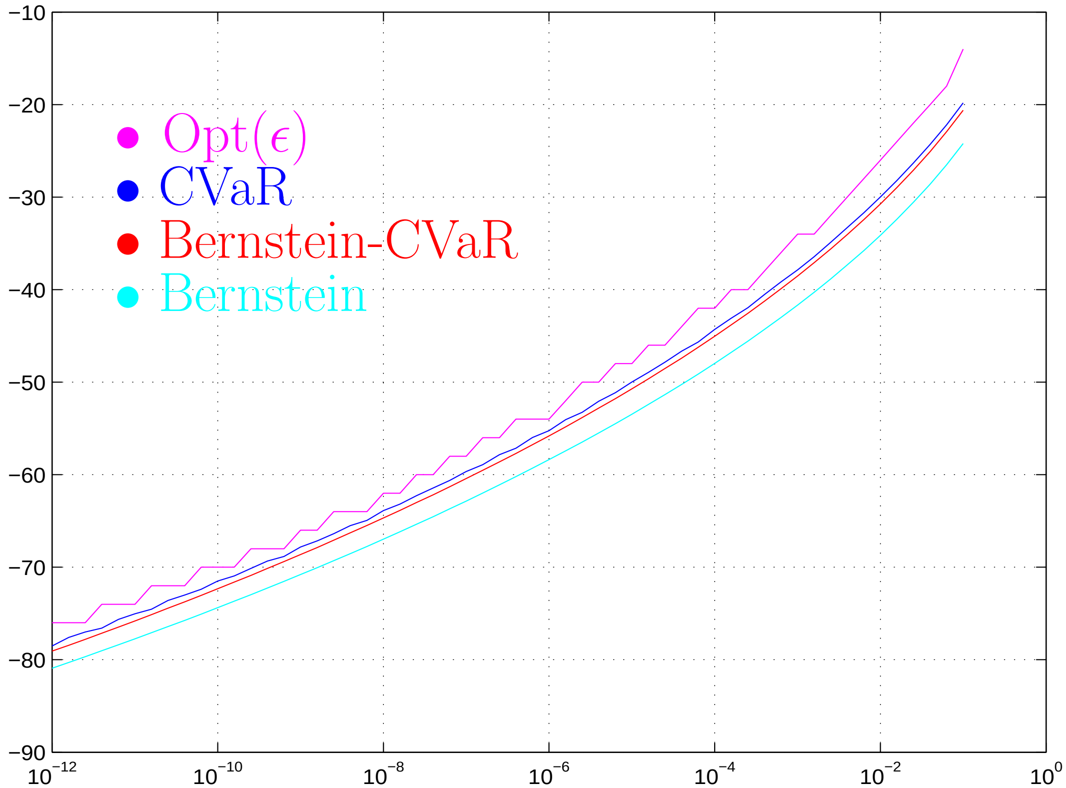


Range $-8 \leq s \leq 8$
Natural scale along y-axis

♠ <u>How it works:</u> Approximating

$$\mathrm{Opt}(\epsilon) = \max\left\{ w_0 : \mathrm{Prob}\{w_0 + \sum_{\ell=1}^{128} \zeta_\ell > 0\} \le \epsilon \right\} \qquad (*)$$

$[\zeta_1, ..., \zeta_{128}$: independent zero mean r.v.'s taking values in $[-1, 1]]$



$\mathrm{Opt}(\epsilon)$ and its approximations vs. ϵ,

$1.e\text{-}12 \le \epsilon \le 1.e\text{-}1$

♣ Approximating Chance Constrained Conic Quadratic/Linear Matrix Inequalities

♠ **CQI** $\|y\|_2 \le \tau$ **is equivalent to the LMI** $\left[\begin{array}{c|c} \tau & y^T \\ \hline y & \tau I \end{array}\right] \succeq 0$

$\Rightarrow$*Approximating Chance Constrained Conic Quadratic Inequality reduces to similar problem for Chance Constrained LMI.*

♠ **Consider Chance Constrained LMI with** $n \times n$ **matrices** $A_\ell[x]$**:**

$$\mathrm{Prob}\left\{A_0[x] + \zeta_1 A_1[x] + ... + \zeta_d A_d[x] \succeq 0\right\} \ge 1 - \epsilon \qquad (C_\epsilon[\rho])$$

and independent random ζ_ℓ **such that either** • $\zeta_\ell \sim \mathcal{N}(0, \rho^2)$**, or** • $|\zeta_\ell| \le \rho$ & $\mathbf{E}\{\zeta_\ell\} = \mathbf{E}\{\zeta_\ell^3\} = 0$**.**

Theorem *With* $\vartheta = O(n^{1/6}\sqrt{\ln(1/\epsilon)})$, *the LMI*

$$\left[\begin{array}{cccc} A_0[x] & \rho\vartheta A_1[x] & \cdots & \rho\vartheta A_d[x] \\ \rho\vartheta A_1[x] & A_0[x] & & \\ \vdots & & \ddots & \\ \rho\vartheta A_d[x] & & & A_0[x] \end{array}\right] \succeq 0$$

is a safe tractable approximation, tight within the factor ϑ, *of* $(C_\epsilon[\rho])$. *When* $(C_\epsilon[\rho])$ *comes from a Chance Constrained CQI,* ϑ *can be reduced to* $O(\sqrt{\ln(1/\epsilon)})$.