On Tractable Approximations of Randomly Perturbed Convex Constraints

Arkadi Nemirovski
Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology
nemirovs@ie.technion.ac.il

Abstract

We consider a chance constraint $\mathrm{Prob}\{\xi : A(x,\xi) \in \mathbf{K}\} \geq 1-\epsilon$ (x is the decision vector, ξ is a random perturbation, $\mathbf{K}$ is a closed convex cone, and $A(\cdot,\cdot)$ is bilinear). While important for many applications in Optimization and Control, chance constraints typically are "computationally intractable", which makes it necessary to look for their tractable approximations. We present these approximations for the cases when the underlying conic constraint $A(x,\xi) \in \mathbf{K}$ is (a) scalar inequality, or (b) conic quadratic inequality, or (c) linear matrix inequality, and discuss the level of conservativeness of the approximations.

1 The problem

Consider a randomly perturbed convex constraint in the conic form:

$$A_{\xi,\sigma}(x) = A_0(x) + \sigma \sum_{i=1}^{k} \xi_i A_i(x) \in \mathbf{K}, \qquad (1)$$

where

• $A_i(\cdot)$ are affine mappings from $\mathbf{R}^n$ to finite-dimensional real vector space E, and $x \in \mathbf{R}^n$ is the decision vector;

• ξ_i are scalar random perturbations satisfying the relations

$$\begin{array}{ll} (a): & \xi_i \text{ are mutually independent;} \\ (b): & \mathbf{E}\{\xi_i\} = 0; \\ (c): & \mathbf{E}\left\{\exp\{\xi_i^2/4\}\right\} \leq \sqrt{2} \end{array} \qquad (2)$$

Cases of primary interest:
— $\xi_i \sim \mathcal{N}(0,1)$ ("Gaussian noise"; the absolute constants in (2.c) come exactly from the desire to make the relation valid for the standard Gaussian perturbations);
— $\mathbf{E}\{\xi_i\} = 0$, $|\xi_i| \leq 1$ ("bounded random noise").

• $\sigma \geq 0$ is the level of perturbations,

• $\mathbf{K}$ is a closed pointed convex cone in E.
Cases of primary interest:
— $E = \mathbf{R}$, $\mathbf{K} = \mathbf{R}_+$; here (1) is a scalar linear inequality;
— $E = \mathbf{R}^{m+1}$, $\mathbf{K} = \mathbf{L}^m = \{x \in \mathbf{R}^{m+1} : x_{m+1} \geq \sqrt{x_1^2 + ... + x_m^2}\}$; here (1) is a randomly perturbed *Conic Quadratic Inequality* (CQI)

$$\|A[\sigma\xi]x + b[\sigma\xi]\|_2 \leq c^T[\sigma\xi]x + d[\sigma\xi], \qquad (3)$$

where the "data" $A[\cdot]$, $b[\cdot]$, $c[\cdot]$, $d[\cdot]$ are affine in the perturbations. (Systems of) CQI's arise in many important applications, see, e.g., [6, 2];
— $E = \mathbf{S}^m$ is the space of $m \times m$ symmetric matrices, $\mathbf{K} = \mathbf{S}^m_+$ is the cone of positive semidefinite matrices from $\mathbf{S}^m$; here (1) is a randomly perturbed Linear Matrix Inequality (LMI).

We are interested to describe x's which satisfy (1) with a given high probability, that is, are such that

$$\mathrm{Prob}\{\xi : A_{\xi,\sigma}(x) \notin \mathbf{K}\} \leq \epsilon, \qquad (4)$$

for a given $\epsilon << 1$, with ultimate goal to optimize over the resulting set, perhaps under additional constraints on x (the latter problem arises in numerous situations coming from Optimization and Control). The *chance constraint* (4) usually is "computationally intractable". E.g., among the examples above, the only one where (4) can be rewritten equivalently as an explicit convex constraint on x is the simplest case of scalar linear inequality affected by Gaussian noise; even here, replacing the Gaussian perturbations ξ_i with uniformly distributed ones (4) becomes intractable. In the case when the chance constraint (4) is computationally intractable, it makes sense to look for its "computationally tractable approximations", specifically, systems $\mathcal{S} = \mathcal{S}_{\sigma,\epsilon}(x,u)$ of explicit convex constraints on x and additional "analysis" variables u such that

$$\begin{array}{ll} P_{\sigma,\epsilon}(x): & \exists u : (x,u) \text{ satisfies } \mathcal{S}_{\sigma,\epsilon}(x,u) \\ & \Downarrow \\ Q_{\sigma,\epsilon}(x): & x \text{ satisfies (4), whatever is the} \\ & \text{distribution of } \xi \text{ satisfying (2)} \end{array} \qquad (5)$$

so that the possibility to extend a given x to a solution to $\mathcal{S}$ implies that x satisfies (4). Besides this, we require the possibility to build $\mathcal{S}$, given σ, ϵ and the data $A_i(\cdot)$, $i = 0, 1, ..., k$, in time polynomial in the size of these data and $\ln(1/\epsilon)$. The latter requirement, modulo minor additional technical assumptions, implies that one can check efficiently whether or not a given x satisfies $P_{\sigma,\epsilon}(x)$, so that $\mathcal{S}$ gives an efficiently verifiable *sufficient condition* for the validity of (4). Moreover, the set of x's satisfying this sufficient condition is convex, so that one can minimize efficiently convex objectives over x's satisfying this sufficient condition and, perhaps, additional "explicit" convex constraints.

The validity of $P_{\sigma,\epsilon}(x)$ is just a sufficient condition for the validity of (4), and as such it can be very conservative. A natural way to quantify the associated "level of conservativeness" is offered by the following

Definition 1 *We say that $P_{\sigma,\epsilon}(\cdot)$ is a tight, within factor $\kappa \geq 1$, approximation of $Q_{\sigma,\epsilon}(\cdot)$ if, first, implication* (5) *holds true and, second, if an x does not satisfy $P_{\sigma,\epsilon}(\cdot)$, then, for properly chosen distribution of ξ satisfying* (2)*, x does not satisfy $Q_{\kappa\sigma,\epsilon/\kappa}(\cdot)$ as well. In other words, the set of x's satisfying $P_{\sigma,\epsilon}(\cdot)$ is in-between the feasible set of the original chance constraint* (4) *(whatever be a distribution of ξ satisfying* (2)*) and the feasible set of the chance constraint obtained from* (4) *by increasing by factor κ the level of perturbations and reducing by the same factor the "unreliability" ϵ.*

In reality the levels of uncertainty and/or unreliability are usually given "by order of magnitude", so that tight, within moderate factors, approximations of chance constraints can be treated as reasonable substitutions of these constraints.

2 The results

2.1 Scalar Linear Inequality

This case is trivial: a tight, within factor $O(1)$, tractable approximation of (1) is the conic quadratic inequality

$$A_0(x) - \Theta\sigma\sqrt{\sum_{i=1}^{k} A_i^2(x)} \geq 0, \quad \Theta = O(1)\sqrt{\ln(1/\epsilon)}.$$

2.2 Conic Quadratic Inequality

In the case when (1) is a CQI (3), we know how to build a tight tractable approximation only in the particular case when the random perturbations in the left hand side and in the right hand side of (3) are independent of each other. Thus, assume that ξ is partitioned into subvectors $\xi' = (\xi_1, ..., \xi_p)$ and $\xi'' = (\xi_{p+1}, ..., \xi_k)$, with $A[\cdot]$, $b[\cdot]$ depending affinely on $\sigma\xi'$, and $c[\cdot]$, $d[\cdot]$ depending affinely on $\sigma\xi''$. Under this assumption, (3) can be rewritten equivalently as

$$\|r(x) + \sigma R(x)\xi'\|_2 \leq s(x) + \sigma q^T(x)\xi'', \qquad (6)$$

with $r(\cdot)$, $R(\cdot)$, $s(\cdot)$, $q(\cdot)$ affinely depending on x. An $O(\sqrt{\ln(1/\epsilon)})$-tight approximation of the associated chance constraint (4) turns out to be given by the following system of constraints in variables x, τ, μ:

$$\begin{array}{ll}
(a) & \tau \leq s(x) - \Theta\sigma\|q(x)\|_2 \\
(b) & \tau \geq \sqrt{r^T(x)r(x) + \sigma^2\Theta^2\mathrm{Tr}(R^T(x)R(x))} \\
(c) & 0 \preceq \left[\begin{array}{c|c|c}
\tau - \Theta^2\sigma^2\mu & & r^T(x) \\ \hline
 & & R_1^T(x) \\
 & \mu I & \vdots \\
 & & R_p^T(x) \\ \hline
r(x) & R_1(x)\cdots R_p(x) & \tau I
\end{array}\right]
\end{array} \qquad (7)$$

where $R_1(x), ..., R_p(x)$ are columns of $R(x)$ and $\Theta = O(1)\sqrt{\ln(1/\epsilon)}$.

The informal outline of the derivation is as follows. First, independence of the perturbations in the left and in the right hand sides implies that in order for (4) to be valid with $\epsilon << 1$, there should exist $\tau \geq 0$ (depending on x only) such that the right hand side in (3) is, with probability $\geq 1 - O(\epsilon)$, $\geq \tau$, while the left hand side is, with the same probability, $\leq \tau$. The former of these requirements is expressed by (7.a); the essence of the matter is how to express the fact that $\|r(x) + \sigma R(x)\zeta\|_2^2 \leq \tau^2$, $\zeta = \xi'$,z with probability close to 1. Assuming the noise Gaussian, the inequality in question becomes

$$r^T(x)r(x) + 2\sigma\zeta^T R^T(x)r(x) + \sigma^2\zeta^T R^T(x)R(x)\zeta \leq \tau^2. \qquad (8)$$

In order for this inequality to be satisfied with probability close to 1, one clearly should have

$$r^T(x)r(x) + 2\sigma\|R^T(x)r(x)\|_2 + \sigma^2\mathrm{Tr}(R^T(x)R(x)) \leq \tau^2, \qquad (9)$$

whence

$$\begin{array}{ll}
(a) & r^T(x)r(x) + 2\sigma\|R^T(x)r(x)\|_2 \\
 & \qquad +\sigma^2\mathrm{Tr}(R^T(x)R(x)) \leq \tau^2, \\
(b) & r^T(x)r(x) + \sigma^2\mathrm{Tr}(R^T(x)R(x)) \leq \tau^2.
\end{array} \qquad (10)$$

Note that from (10) it follows that

$$\begin{array}{ll}
(a) & \max\limits_{\eta:\|\eta\|_2 \leq 1} \left\{r^T(x)r(x) + 2\sigma\eta^T R^T(x)r(x) \right. \\
 & \qquad \left. +\sigma^2\eta^T R^T(x)R(x)\eta\right\} \leq \tau^2, \\
(b) & r^T(x)r(x) + \sigma^2\mathrm{Tr}(R^T(x)R(x)) \leq \tau^2.
\end{array} \qquad (11)$$

Vice versa, applying results on large deviations for vector sums (see Section 3), it is easily seen that whenever x, τ satisfy the "reliable" version of (11), specifically,

$$\begin{array}{ll}
(a) & \max\limits_{\eta:\|\eta\|_2 \leq \Theta} \left\{r^T(x)r(x) + 2\sigma\eta^T R^T(x)r(x) \right. \\
 & \qquad \left. +\sigma^2\eta^T R^T(x)R(x)\eta\right\} \leq \tau^2, \\
(b) & r^T(x)r(x) + \sigma^2\Theta^2\mathrm{Tr}(R^T(x)R(x)) \leq \tau^2.
\end{array} \qquad (12)$$

where $\Theta = O(1)\sqrt{\ln(1/\epsilon)}$, x, τ satisfy (9) with probability at least $1 - \epsilon$. It remains to note that (12.b)

is exactly (7.b), while (12.a) states that the uncertain CQI

$$\|r(x) + R(x)\eta\|_2 \le \tau$$

is satisfied for *all* perturbations η from the ball $\|\eta\|_2 \le \sigma\Theta$. It is known (see [5, 1]) that the latter requirement is equivalent to the existence of μ satisfying (7.c).

2.3 Linear Matrix Inequality

What we know in this most difficult case can be summarized as follows.

A. In the case of nondegenerate $A_0(x)$ (which can be assumed without loss of generality) a sufficient condition for the validity of (4) is that $A_0(x) \succ 0$ and

$$\begin{array}{c}\Theta\sigma\sqrt{\sum\limits_{i=1}^{k}\lambda_{\max}^2(A_0^{-1/2}(x)A_i(x)A_0^{-1/2}(x))} \le 1,\\ \Theta = O(1)\sqrt{\ln(m)\ln(1/\epsilon)},\end{array} \quad (13)$$

where $\lambda_{\max}(A)$ is the maximum eigenvalue of a symmetric matrix A and m is the size of the matrices $A_i(x)$. A bad news about this sufficient condition is that in general it is a nonconvex restriction on x and thus is computationally intractable. Sometimes this difficulty does not occur, e.g., when $A_1(x), ..., A_k(x)$ are independent of x ("LMI with randomly perturbed constant term"); here the above condition becomes

$$\exists(\mu_1, ..., \mu_k > 0): \left\{\begin{array}{l} -A_0(x) \preceq \mu_i A_i \preceq A_0(x),\ i = 1, ..., k \\ O(1)\sigma\sqrt{\ln(m)\ln(1/\epsilon)}\sqrt{\sum\limits_{i=1}^{k}\mu_i^{-2}} \le 1 \end{array}\right.$$

In more complicated cases, where (13) is intractable, one can replace this condition by its tractable approximation, e.g., by the system of constraints

$$\begin{array}{rcl} A_0(x) & \succeq & \tau D \\ \left[\begin{array}{cc}\lambda_i D & A_i(x)\\ A_i(x) & \lambda_i D\end{array}\right] & \succeq & 0,\ i = 1, ..., k \\ O(1)\sigma\sqrt{\ln(m)\ln(1/\epsilon)}\sqrt{\sum_i \lambda_i^2} & \le & \tau \end{array} \quad (14)$$

in variables x, τ, λ ($D \succ 0$ is the "free parameter" of the construction), or the system

$$\begin{array}{rcl} A_0(x) & \succeq & 0 \\ \left[\begin{array}{cc}\alpha_i A_0(x) & A_i(x)\\ A_i(x) & \alpha_i A_0(x)\end{array}\right] & \succeq & 0,\ i = 1, ..., k \end{array} \quad (15)$$

in variables x, where positive parameters α_i satisfy the restriction

$$O(1)\sigma\sqrt{\ln(m)\ln(1/\epsilon)}\sqrt{\sum_i \alpha_i^2} \le 1.$$

Note that the tightness factor of (13) does not exceed $O(\sqrt{m})$, while tightness factors of (14) and (15) seemingly cannot bounded solely in terms of data size.

B. A simple way to build "safe versions" of randomly perturbed constraints is offered by the *scenario approach*: we generate a sample ξ^1,..., ξ^S of random perturbations, choose a "safety parameter" $\Theta > 1$ and replace (4) with the system of constraints

$$A_{\xi^s,\Theta\sigma}(x) \equiv A_0[x] + \Theta\sigma\sum_{i=1}^{k}\xi_i^s A_i(x) \in \mathbf{K},\ s = 1, ..., S. \quad (16)$$

The advantage of this approach is that it guarantees computational tractability of "approximation" (16) of chance constraint (4), provided that $\mathbf{K}$ itself is computationally tractable (which indeed is the case in all our examples) and that the number S of "scenarios" is polynomial in the size of the original data and $\ln(1/\epsilon)$. Note also that system (16) is random, so that we cannot require from it to possess a particular property (e.g., to be an approximation of (4)) for *all* scenario samples); all we may hope for is that this system possesses a desired property with probability at least $1 - \delta$, where $\delta << 1$ is a given "(un)reliability level".

Recently, Calafiore and Campi [4] have obtained an extremely elegant and deep result on the "power" of the scenario approach as applied to a general-type randomly perturbed convex program. We are about to present an incomparably more specialized result, which has, however, the advantage that it deals with the samples of scenarios of polynomial in $\ln(\epsilon^{-1}\delta^{-1})$ size, while the result of [4] requires the sample to be of cardinality inverse proportional to $\delta\epsilon$. For the sake of definiteness, consider the case of randomly perturbed LMI and Gaussian perturbations (the latter restriction seems to be crucial). Our result is as follows:

Proposition 1 *Consider randomly perturbed LMI with Gaussian perturbations:*

$$A_0(x) + \sigma\sum_{i=1}^{k}\xi_i A_i(x) \succeq 0 \quad [A_i(\cdot) \in \mathbf{S}^m, \xi \sim \mathcal{N}(0, I_k)]$$

along with its "scenario approximation" given by (16) *(where* $\mathbf{K} = \mathbf{S}^m_+$*). Then*

(i) *With safety parameter* $\Theta = O(1)\sqrt{\ln(/\epsilon)}$ *and sample size*

$$S \ge O(1)[m^2 k\ln(k) + \ln(1/\delta)],$$

the feasible set $\mathcal{Y}_{\Theta\sigma}[\xi^1, ..., \xi^S]$ *of* (16) *is contained in the set* $\mathcal{X}[\sigma, \epsilon]$ *of solutions to the chance constraint*

$$\text{Prob}\left\{\xi : A_0(x) + \sigma\sum_{i=1}^{k}\xi_i A_i(x) \succeq 0\right\} \ge 1 - \epsilon$$

with probability at least $1 - \delta$.

(ii) *For all* δ, ϵ, $0 < \delta, \epsilon < 1/4$, *every* $\Theta > 1$ *and every sample size* S, *with* $\kappa \ge O(1)\sqrt{\ln(S/\delta)}$, *every point* x

of the set $\mathcal{X}[\kappa\Theta\sigma,\epsilon]$ with probability at least $1-\delta$ belongs to the set $\mathcal{Y}_{\Theta\sigma}[\xi^1,...,\xi^S]$.

In particular, when approximating an optimization problem

$$c_*(\sigma,\epsilon) = \min_x \left\{e^T x : x \in \mathcal{X} \cap \mathcal{X}[\sigma,\epsilon]\right\} \qquad (17)$$

with its "scenario approximation"

$$c^* = \min_x \left\{e^T x : x \in \mathcal{X},\ A_{\xi^s,\Theta\sigma}(x) \succeq 0,\ s=1,...,S\right\}$$
$$\left[\begin{array}{c} \Theta = O(1)\sqrt{\ln(1/\epsilon)} \\ S = O(1)[m^2 k \ln(k) + \ln(1/\delta)] \end{array}\right]$$

the solution to the approximation, with probability at least $1-\delta$, is feasible for the problem of interest (17)*, and the optimal value c^* of the approximation, with probability at least $1-\delta$, satisfies the inequalities*

$$c_*(\sigma,\epsilon) \le c^* \le c_*(\sigma\sqrt{\ln(1/\epsilon)\ln(mk/\delta)},\epsilon).$$

When considering the scenario approximations, one should mention the following phenomenon. By Proposition 1, in the case of Gaussian noise, for "moderate" Θ and S, a "scenario-feasible" domain $\mathcal{Y}_{\Theta\sigma}$ is, with probability close to 1, inside the feasible set $\mathcal{X}[\sigma,\epsilon]$ of the chance constraint, and with probability close to 1 contains every *fixed* feasible solution of the "strengthened" chance constraint (with the level of perturbations increased by a moderate factor). However, the set of *all* solutions to the strengthened chance constraint can be much larger than a "typical" scenario-feasible domain. Indeed, consider a simple randomly perturbed linear constraint in $\mathbf{R}^n$:

$$\sigma x^T \xi \le 1 \qquad [\xi \in \mathcal{N}(0, I_n)]$$

The feasible set of the corresponding chance constraint is the centered at the origin ball B of radius $O(\sigma^{-1}/\sqrt{\ln(1/\epsilon)})$. In contrast to this, a typical "scenario" ξ^s is of Euclidean norm $O(\sqrt{n})$, so that a typical set of the form $\mathcal{Y}_{\Theta\sigma}$ is contained in the intersection of $O(S)$ "strips" of width $O(n^{-1/2}\Theta^{-1}\sigma^{-1})$ each, and every one of these strips, for large n, is pretty "thin" as compared to B.

3 The techniques

The techniques underlying the outlined results might be of interest by their own right. Our main "building blocks" are as follows:

3.1 Exponential bounds on probabilities of large deviations in normed spaces

Let $(E,\|\cdot\|)$ be a separable Banach space with the following "smoothness" property: there exists a norm $p(x)$ which is compatible, up to factor 2, with the norm $\|\cdot\|$ (i.e., $\|x\| \le p(x) \le 2\|x\|$ for all x) which is smooth outside of the origin, specifically, the function $P(x) = \frac{1}{2}p^2(x)$ is continuously differentiable and satisfies the relation

$$P(x+y) \le P(x) + \langle P'(x), y\rangle + \frac{\kappa^2}{2}P(y).$$

Now let ξ^s, $s=1,...,S$, be independent random vectors in E with zero mean, such that $\mathbf{E}\left\{\exp\{\|\xi^s\|^2/\sigma_s^2\}\right\} \le \exp\{1\}$. Then for properly chosen positive *absolute* constant c and for all $\rho > 0$ one has

$$\mathrm{Prob}\left\{\|\sum_{s=1}^S \xi^s\| > \rho\kappa\sqrt{\sum_{s=1}^S \sigma_s^2}\right\} \le c^{-1}\exp\{-c\rho^2\}.$$

The outlined result seems to be very natural; however, we were unable to find this result in the literature, even in the simplest case of $(E,\|\cdot\|) = (\mathbf{R}^n, \|\cdot\|_2)$ (where $\kappa = 1$). Note that this latter case underlies the analysis of the tractable approximation (7) of a randomly perturbed CQI. Similarly, the sufficiency of condition (13) for the validity of (4) in the case of a randomly perturbed LMI is a direct consequence of the above "large deviation" result as applied to the space $E = \mathbf{M}^{m,n}$ of $m\times n$ matrices equipped with the standard matrix norm $\|\cdot\|$ (maximum singular value); the (not completely trivial) fact underlying this application is that this particular normed space is smooth with $\kappa = O(1)\sqrt{\ln(\min[m,n]+1)}$.

3.2 Convex sets and Gaussian distributions

The results stated in Proposition 1 are straightforward consequences of the following statement (which seems to be new; it is an easy consequence of the "isoperimetric inequality" for Gaussian distribution, see [3]):

Let ξ be a Gaussian vector with zero mean in $\mathbf{R}^n$, and B be a closed convex set in $\mathbf{R}^n$ such that $\mathrm{Prob}\{\xi \in B\} \ge \theta > 1/2$. *Then*

$$0 < \alpha < 1 \Rightarrow \mathrm{Prob}\{\alpha\xi \in B\} \ge 1 - \exp\left\{-\frac{\phi^2(\theta)}{2\alpha^2}\right\},$$
$$\phi(r) : \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\phi(r)} \exp\{-s^2/2\}ds = r.$$

References

[1] Ben-Tal, A., Nemirovski, A., "Robust Convex Optimization" – *Mathematics of Operations Research* **23** (1998).

[2] Ben-Tal, A., Nemirovski, A., *Lectures on Modern Convex Optimization*, AMS-SIAM Series on Optimization, SIAM, Philadelphia, 2002.

[3] Borell, C., "The Brunn-Minkowski inequality in Gauss space" – *Inventiones Mathematicae* **30** (2) (1975), 207-216.

[4] Calafiore, G., Campi, M.C., "Uncertain convex programs: randomized solutions and confidence levels" – to appear in *Mathematical Programming*.

[5] El-Ghaoui, L., Lebret, H., "Robust solutions to least-square problems with uncertain data matrices" – *SIAM J. of Matrix Anal. and Appl.* **18** (1997), 1035-1064.

[6] Lobo, M.S., Vanderbeghe, L., Boyd, S., Lebret, H., "Second-Order Cone Programming",*Linear Algebra and Applications* **284** (1998), 193 – 228.