The Exact Feasibility of Randomized Solutions of Robust Convex Programs

M.C. Campi[*] S. Garatti[†]

Abstract

Robust optimization programs are hard to solve even when the constraints are convex. In previous contributions, it has been shown that approximately robust solutions (i.e. solutions feasible for all constraints but a small fraction of them) to convex programs can be obtained at low computational cost through constraints randomization.

In this paper, we establish new feasibility results for randomized algorithms. Specifically, the *exact* feasibility for the class of the so-called *fully-supported* problems is obtained. It turns out that all fully-supported problems shares the same feasibility properties, revealing a deep kinship among problems of this class. It is further proven that the feasibility of the randomized solutions for all other convex programs can be bounded based on the feasibility for the prototype class of fully-supported problems.

The feasibility result of this paper outperforms previous bounds, and is not improvable because it is exact for fully-supported problems.

[*]Dipartimento di Elettronica per l'Automazione - Università di Brescia, via Branze 38, 25123 Brescia, Italia. E-mail: marco.campi@ing.unibs.it, web-site: http://bsing.ing.unibs.it/∼campi/

[†]Dipartimento di Elettronica e Informatica - Politecnico di Milano, P.zza Leonardo da Vinci 32, 20133 Milano, Italia. E-mail: sgaratti@elet.polimi.it, web-site: http://www.elet.polimi.it/upload/sgaratti/

1 Introduction

Uncertain convex optimization, [2, 13, 3, 4], deals with convex optimization in which the constraints are imprecisely known:

$$\begin{aligned} &\min_{x \in \mathcal{X} \subseteq \mathbb{R}^d} c^T x \\ &\text{subject to: } x \in \mathcal{X}_\delta, \end{aligned} \tag{1}$$

where $\mathcal{X}$ and $\mathcal{X}_\delta$ are convex and closed sets, and $\delta \in \Delta$ is an uncertain parameter. Often times, Δ is a set of infinite cardinality. The fact that the optimization objective is linear and does not carry any dependence on δ, that is it is certain, is without loss of generality.

A *robust* solution for (1) is a solution that satisfies all constraints simultaneously, namely a solution of the robust program

$$\begin{aligned} \text{RP} : \quad &\min_{x \in \mathcal{X} \subseteq \mathbb{R}^d} c^T x \\ &\text{subject to: } x \in \bigcap_{\delta \in \Delta} \mathcal{X}_\delta. \end{aligned} \tag{2}$$

RP encompasses as special cases robust LP (linear programs), robust QP (quadratic programs), robust SOCP (second-order cone programs), and robust SDP (semi-definite programs) and plays a central role in many design endeavors such as [1, 12, 14, 11, 8, 9, 7].

When Δ has infinite cardinality, RP is called a *semi-infinite* optimization program since it is a program with a finite number d of optimization variables and an infinite number of constraints. Semi-infinite programs are reportedly extremely hard to solve in general, [16, 2, 5]. To circumvent this computational issue, recently methodologies relying on the randomization over the set of constraints have been introduced, [9, 6, 17, 7, 10]. Specifically, in [6, 7], it is proposed to approximate RP by the following randomized robust program RRP_N where N constraints $\delta^{(1)}, \ldots, \delta^{(N)}$ are randomly extracted from Δ, in an independent fashion, according to a given probability $\mathbb{P}$:

$$\begin{aligned} \text{RRP}_N : \quad &\min_{x \in \mathcal{X} \subseteq \mathbb{R}^d} c^T x \\ &\text{subject to: } x \in \bigcap_{i \in \{1, \ldots, N\}} \mathcal{X}_{\delta^{(i)}}. \end{aligned}$$

RRP_N is also known as "scenario program". Depending on the optimization problem at hand, $\mathbb{P}$ can have different interpretations. Sometimes, it is the actual probability with which the uncertainty parameter δ takes on value in Δ. Other times, it simply describes the relative importance attributed to different instances of δ.

The distinctive feature of RRP_N is that it is a program with a *finite* number of constraints and, as such, it can be solved at low computational cost provided that N is not too large; it is indeed a fact that RRP_N has opened up new resolution avenues in robust optimization. On the other hand, the obvious question to ask is how robust the solution of RRP_N is, that is how large the fraction of original constraints in Δ that are possibly violated by the solution x_N^* of RRP_N is. Papers [6, 7] have pioneered a feasibility theory showing that x_N^* is feasible for the vast majority of the other unseen constraints – those that have not been used when optimizing according to RRP_N – and this result holds in full generality, independently of the structure of the set of constraints Δ and the probability $\mathbb{P}$. So to say, the vast majority of constraints takes care of itself, without explicitly accounting for them.

To allow for a sharper comparison with the results presented in this paper, we feel advisable to first recall in some detail the results in [6, 7]. The following notion of violation probability from [6] is central.

Definition 1 (violation probability) *The* violation probability *of a given $x \in \mathcal{X}$ is defined as $V(x) = \mathbb{P}\{\delta \in \Delta : x \notin \mathcal{X}_\delta\}$.* $\square$

The problem addressed in [6, 7] is to evaluate if and when the violation probability of x_N^*, namely $V(x_N^*)$, is below a satisfying level ϵ. To state the result precisely, note that $V(x_N^*)$ is a random variable since the solution x_N^* of RRP_N is, due to that it depends on the random extractions $\delta^{(1)}, \ldots, \delta^{(N)}$. Thus, $V(x_N^*) \leq \epsilon$ may hold for certain extractions $\delta^{(1)}, \ldots, \delta^{(N)}$, while $V(x_N^*) > \epsilon$ may be true for others. The following quantification of the "bad" extractions where $V(x_N^*) > \epsilon$ is the key result of [7]:

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \binom{N}{d}(1-\epsilon)^{N-d}. \tag{3}$$

Moving a fundamental step forward with respect to [7], we in this paper establish the validity of relation

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \tag{4}$$

(note that (4) holds with "=", that is it is not a bound) for the prototype class of *fully-supported* problems according to Definition 3 in Section 2. This result sheds new light on a deep kinship among all fully-supported problems, proving that their randomized solutions share the same violation properties, and writes a final word on the violation assessment for this type of problems.

It is further proven in this paper that the right-hand-side of (4) is an upper bound for all convex problems, that is

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \tag{5}$$

holds for all convex problems. This result (5) **(i)** cannot be improved (being tight for the prototype class of fully-supported problems), and **(ii)** outperforms the previous bound from [7], at times by a huge extent (note that when $\epsilon \to 0$, the previous bound (3) tends to $\binom{N}{d}$ while the new bound (5) goes to 1!).

2 Main result

The technical result of this paper is precisely stated in this section, followed by a discussion on the significance of the result.

For a fixed integer m and fixed given constraints $\delta^{(1)}, \ldots, \delta^{(m)}$, program

$$\begin{aligned} \mathrm{RP}_m : \quad & \min_{x \in \mathcal{X} \subseteq \mathbb{R}^d} c^T x \\ & \text{subject to: } x \in \bigcap_{i \in \{1,\ldots,m\}} \mathcal{X}_{\delta^{(i)}} \end{aligned} \tag{6}$$

is called a *finite instance with m constraints* of the optimization program RP in (2). We assume that every RP_m is feasible and that its feasibility domain has nonempty interior. Moreover, existence and uniqueness of the solution x_m^* of RP_m is also assumed.[1]

[1]Existence and uniqueness are here assumed to streamline the presentation. The reader is referred to point (5) in the discussion section 2.1 for relaxations of this assumption.

We recall the following fundamental definition and proposition from [6].

Definition 2 (support constraint) *Constraint $\delta^{(r)}$, $r \in \{1,\ldots,m\}$, is a* support constraint *for RP_m if its removal changes the solution of RP_m.*

Proposition 1 *The number of support constraints for RP_m is at most d, the size of x.*

Suppose now that Δ is endowed with a σ-algebra $\mathcal{D}$ and that a probability $\mathbb{P}$ over $\mathcal{D}$ is assigned. Further assume that m constraints $\delta^{(1)},\ldots,\delta^{(m)}$ are randomly extracted from Δ according to $\mathbb{P}$ in an independent fashion. Differently stated, the multi-extraction $(\delta^{(1)},\ldots,\delta^{(m)})$ is a random element from the probability space Δ^m equipped with the product probability $\mathbb{P}^m$. Each multi-extraction $(\delta^{(1)},\ldots,\delta^{(m)})$ generates a program RP_m and the map from Δ^m to RP_m programs is a *randomized robust program* RRP_m, see Figure 1. Note that this is the same as

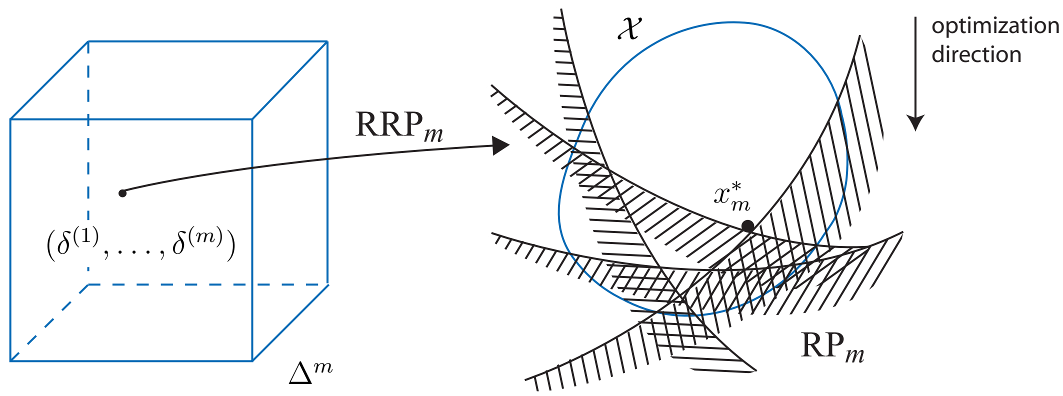

Figure 1: RRP_m, a map from constraint multi-extractions to finite instances RP_m of the optimization problem.

RRP_N in Section 1 with the only difference that we have used here m to indicate the number of constraints, a choice justified by the fact that in this section m plays the role of a generic running argument taking on any integer value, while N represents in Section 1 the fixed number of constraints picked by the user for the implementation of the randomized scheme.

We are now ready to introduce the notion of *fully-supported* problem.

Definition 3 (fully-supported problem) *A finite instance RP_m, with $m \geq d$, is fully-supported if the number of support constraints of RP_m is d.*
Problem RP is fully-supported if, for any $m \geq d$, RRP_m is fully-supported with probability 1.

The main result of this paper is now stated in the following theorem.

Theorem 1 *It holds that*

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}; \tag{7}$$

moreover, the bound is tight for all fully-supported *robust optimization problems, that is*

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}. \tag{8}$$

The proof is given in Section 3. The measurability of $\{V(x_N^*) > \epsilon\}$, as well as the measurability of other sets, is assumed for granted in this paper.

2.1 Discussion

(1) Equation (8) delivers the exact feasibility for all fully supported problems independently of their nature and characteristics and establishes a fundamental kinship among problems of this prototype class.
Bound (7) further asserts that all possible sources of non-fully-supportedness can only improve the feasibility properties of the problem.

(2) The quantity $\beta := \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ in the right-hand-side of equations (7) and (8) is the tail of a Binomial distribution and goes rapidly (exponentially) to zero as N increases. Letting $\beta_{\text{old}} := \binom{N}{d}(1-\epsilon)^{N-d}$ (bound in (3) from [7]), Table 1 provides a comparison between

N	150	300	450	600	750	900	1050	1200	1350	1500
β	0.78	0.06	$8.8\ 10^{-4}$	$4.8\ 10^{-6}$	$1.5\ 10^{-8}$	$3.5\ 10^{-11}$	$6.2\ 10^{-14}$	$9.2\ 10^{-17}$	$1.2\ 10^{-19}$	$1.4\ 10^{-22}$
β_{old}	$8.8\ 10^{11}$	$4.8\ 10^{11}$	$1.3\ 10^{10}$	$1.1\ 10^{8}$	$4.8\ 10^{5}$	$1.3\ 10^{3}$	2.9	$5.1\ 10^{-3}$	$7.5\ 10^{-6}$	$9.9\ 10^{-9}$

Table 1: β vs. β_{old} for different values of N ($\epsilon = 0.05$, $d = 10$).

β and β_{old}.

(3) A typical use of Theorem 1 consists in selecting ϵ (violation parameter) and β (confidence parameter) in $(0,1)$, and then computing the smallest number N of constraints to be extracted in order to guarantee that $\mathbb{P}^N\{V(x_N^*) > \epsilon\} \le \beta$ by solving equation $\beta = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ for N. In Table 2, the values of N and of N_{old} obtained by using the bound in (3) are displayed

ϵ	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
N	285	581	1171	2942	5895	11749	29513
N_{old}	579	1344	3035	8675	18943	41008	112686

Table 2: N vs. N_{old} for different values of ϵ ($\beta = 10^{-5}$, $d = 10$).

for different values of ϵ, $\beta = 10^{-5}$ and $d = 10$.

(4) A simple example illustrates Theorem 1.

$N = 1650$ points are independently extracted in $\mathbb{R}^2$ according to an unknown probability density $\mathbb{P}$, and the strip of smaller vertical width that contains all the points is constructed, see

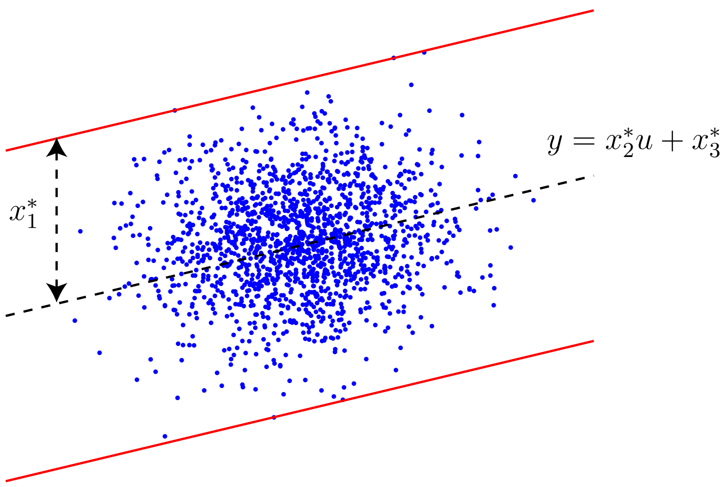

Figure 2: strip of smaller vertical width.

Figure 2.

In mathematical terms – letting the points be $(u^{(i)}, y^{(i)})$, $i = 1, \ldots, N$, where u is horizontal coordinate and y vertical coordinate – this amounts to solve the following program:

$$\begin{aligned} \mathrm{RP}_N: \quad & \min_{x_1,x_2,x_3 \in \mathbb{R}^3} x_1 \\ & \text{subject to: } \left| y^{(i)} - [x_2 u^{(i)} + x_3] \right| \leq x_1, \quad i = 1, \dots, N, \end{aligned}$$

where $[x_2 u^{(i)} + x_3]$ is the median line of the strip and x_1 is the semi-width of the strip.

What guarantee do we have that the strip contains at least 99% of the probability mass of $\mathbb{P}$? One can easily recognize that this question is the same as asking for a guarantee, or a probability, that the violation is less than $\epsilon = 0.01$, and the answer can be found in Theorem 1: this probability is no less than $1 - \sum_{i=0}^{2} \binom{1650}{i} 0.01^i (1-0.01)^{1650-i} \approx 1 - 10^{-5}$. As a matter of fact, this probability is exact since, as it can be verified, this problem is fully supported.

We can further ask for a different geometrical construction and look for the disk of smaller radius that contains all points, see Figure 3. Again, we are facing a finite convex program:

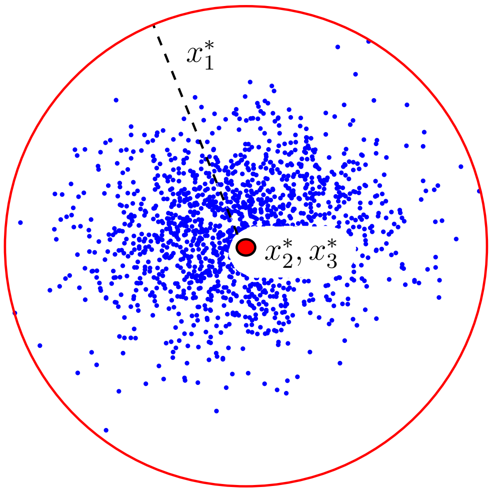

Figure 3: disk of smaller radius.

$$\begin{aligned} \mathrm{RP}_N: \quad & \min_{x_1,x_2,x_3 \in \mathbb{R}^3} x_1 \\ & \text{subject to: } \sqrt{(u^{(i)} - x_2)^2 + (y^{(i)} - x_3)^2} \leq x_1, \quad i = 1, \dots, N, \end{aligned}$$

where (x_2, x_3) is the center of the disk and x_1 is its radius, and again we can claim with confidence $1 - 10^{-5}$ that the constructed disk will contain at least 99% of the probability mass. In this disk case, figure $1-10^{-5}$ is a lower bound since the problem is not fully-supported, as it can be easily recognized by noting that a configuration with two points away from each other and all the other points concentrated near the mid-position of the first two points generates a disk where the segment joining the first two points is a diameter and only these two points are of support.

Finally, let us compare probability $1-10^{-5}$ with the probability that would have been obtained by applying the previous bound (3) from [7]. Applying the latter, we find $1-48.4$, a figure clearly devoid of any meaning and that does not allow to draw any conclusion as far as the confidence is concerned.

(5) We here discuss the assumption of existence and uniqueness of the solution of RP_m.
Suppose first that the solution exists but it may be non-unique. Then, the tie can be broken by selecting among the optimal solutions the one with minimum Euclidian norm, and one can prove that Theorem 1 holds unchanged.
If we further relax the assumption that the solution exists (note that the solution may not exist even if RP_m is feasible since the solution can drift away to infinity), extending Theorem 1 one can show that

$$\mathbb{P}^N\{x_N^* \text{ exists, and } V(x_N^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i},$$

where x_N^* is unique after applying the tie-break rule as above. In words, this result says that, when a solution is found, its violation exceeds ϵ with small probability only. In normal problems non-existence of the solution is a rare event whose probability exponentially vanishes with N.

3 Proof of Theorem 1

We first prove that $\mathbb{P}^N\{V(x_N^*) > \epsilon\} = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ for *fully-supported* problems and then that $\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ for every problem.

<u>PART 1</u>: $\mathbb{P}^N\{V(x_N^*) > \epsilon\} = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ FOR FULLY-SUPPORTED PROBLEMS

Consider the solution x_d^* of RP_d (recall that d is the size of x) and let

$$F(\alpha) := \mathbb{P}^d\{V(x_d^*) \leq \alpha\} \tag{9}$$

be the probability distribution of the violation of x_d^*. It is a remarkable fact that this distribution is

$$F(\alpha) = \alpha^d, \tag{10}$$

independently of the problem type.

To prove (10), we have to consider multi-extractions of m elements, where m is a generic integer bigger than or equal to d. To each multi-extraction $(\delta^{(1)}, \ldots, \delta^{(m)}) \in \Delta^m$, associate the indexes of the corresponding d support constraints (this is always possible except for a probability 0 set because the problem is *fully-supported* [2]). Further, group all multi-extractions having the same indexes. In this way, $\binom{m}{d}$ sets S_i are constructed forming a partition (up to a probability 0 set) of Δ^m. We claim that the probability of each of these sets is

$$\mathbb{P}^m\{S_i\} = \int_0^1 (1-\alpha)^{m-d} F(\mathrm{d}\alpha), \tag{11}$$

where $F(\alpha)$ is defined in (9); using (11), later on in the proof, we shall show that $F(\alpha)$ must have the expression in (10).

To establish (11), consider e.g. the set where the support constraints indexes are $1, \ldots, d$ and name it S_1. Also let $\widetilde{S}_1$ be the set where $\delta^{(d+1)}, \ldots, \delta^{(m)}$ are not violated by the solution generated by $\delta^{(1)}, \ldots, \delta^{(d)}$. It is an intuitive fact that S_1 is the same as $\widetilde{S}_1$ up to a probability 0 set. To streamline the presentation, we accept here this fact for granted; however, the interested reader can find full details at the end of this PART 1 of the proof.

[2] The fact that a fully-supported problem is one where the RRP_m are fully supported with probability 1, as opposed to *always* fully-supported, is a source of a bit of complication in the proof. On the other hand, requiring always fully-supportedness is too limitative since e.g. extracting the same constraint m times results in a non fully-supported RP_m.

We next compute $\mathbb{P}^m\{\widetilde{S}_1\}$, which is the same as $\mathbb{P}^m\{S_1\}$.

Select fixed values for $\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(d)}$ and let α be the violation of the solution with these d constraints only. Then, the probability that $\delta^{(d+1)}, \ldots, \delta^{(m)}$ fall in the non-violated set, that is $(\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(d)}, \delta^{(d+1)}, \ldots, \delta^{(m)}) \in \widetilde{S}_1$, is $(1-\alpha)^{m-d}$. Integrating over all possible values for α, we have

$$\mathbb{P}^m\{\widetilde{S}_1\} = \int_0^1 (1-\alpha)^{m-d} F(\mathrm{d}\alpha).$$

Since $\mathbb{P}^m\{S_1\} = \mathbb{P}^m\{\widetilde{S}_1\}$ and this probability is the same for any other set S_i, (11) remains proven.

Turn now back to (10). Recalling that the sets S_i form a partition of Δ^m up to a probability 0 set and that $\mathbb{P}^m\{\Delta^m\} = 1$, (11) yields

$$\binom{m}{d} \int_0^1 (1-\alpha)^{m-d} F(\mathrm{d}\alpha) = 1, \quad \forall m \geq d. \tag{12}$$

Expression $F(\alpha) = \alpha^d$ in (10) is indeed a solution of (12) (integration by parts); on the other hand, no other solutions exist since determining an F satisfying (12) is a moment problem whose solution is unique, see e.g. Corollary 1, §12.9, Chapter II of [19]. Thus, it remains proven that $F(\alpha)$ must have the expression (10).

To conclude the proof of PART 1, consider now the problem with N constraints and partition set $\{(\delta^{(1)}, \ldots, \delta^{(N)}) : \quad V(x_N^*) > \epsilon\}$ by intersecting it with the $\binom{N}{d}$ sets S_i grouping multi-extractions such that the d support constraints have the same indexes. We then have

$$
\begin{aligned}
&\mathbb{P}^N\{V(x_N^*) > \epsilon\} \\
&= \binom{N}{d} \int_\epsilon^1 (1-\alpha)^{N-d}\ F(\mathrm{d}\alpha) \\
&= [\text{since } F(\mathrm{d}\alpha) = d\alpha^{d-1}\ \mathrm{d}\alpha] \\
&= \binom{N}{d} \int_\epsilon^1 \left[(1-\alpha)^{N-d} d\alpha^{d-1}\right]\ \mathrm{d}\alpha \\
&= [\text{integrating by parts}] \\
&= \binom{N}{d} \left[-\frac{(1-\alpha)^{N-d+1}}{N-d+1} d\alpha^{d-1}\Big|_\epsilon^1 + \int_\epsilon^1 \frac{(1-\alpha)^{N-d+1}}{N-d+1} d(d-1)\alpha^{d-2}\ \mathrm{d}\alpha \right] \\
&= \binom{N}{d-1} \epsilon^{d-1}(1-\epsilon)^{N-d+1} + \binom{N}{d-1} \int_\epsilon^1 (1-\alpha)^{N-d+1}(d-1)\alpha^{d-2}\ \mathrm{d}\alpha \\
&= \ldots \\
&= \binom{N}{d-1} \epsilon^{d-1}(1-\epsilon)^{N-d+1} + \ldots + \binom{N}{1} \epsilon(1-\epsilon)^{N-1} + \binom{N}{1} \int_\epsilon^1 (1-\alpha)^{N-1}\ \mathrm{d}\alpha \\
&= \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}.
\end{aligned}
$$

Proof of the fact that $\boldsymbol{S_1} = \widetilde{\boldsymbol{S}}_{\boldsymbol{1}}$ up to a probability zero set

$\boldsymbol{S_1} \subseteq \widetilde{\boldsymbol{S}}_{\boldsymbol{1}}$: take a $(\delta^{(1)}, \ldots, \delta^{(m)}) \in S_1$ and eliminate a constraint among $\delta^{(d+1)}, \ldots, \delta^{(m)}$. Since this constraint is not of support, the solution remains unchanged; moreover, it is easy to see that the first d constraints are still the support constraints for the problem with $m-1$ constraints. If we now remove another constraint among those which are not of support, the conclusion is similarly drawn that the solution remains unchanged and that the first d constraints are still the support constraints for the problem with $m-2$ constraints. Proceeding this way until all constraints but the first d are removed, we obtain that the solution with the sole d support constraints $\delta^{(1)}, \ldots, \delta^{(d)}$ in place is the same as the solution with all m constraints. Since no constraint among $\delta^{(d+1)}, \ldots, \delta^{(m)}$ can be violated by the solution with all m constraints and such solution is the same as the one with only the first d constraints, it follows that $(\delta^{(1)}, \ldots, \delta^{(m)}) \in \widetilde{S}_1$.

$\widetilde{\boldsymbol{S}}_{\boldsymbol{1}} \subseteq \boldsymbol{S_1}$ **up to a probability 0 set**: suppose now that $\delta^{(d+1)}, \ldots, \delta^{(m)}$ are not violated by the solution generated by $\delta^{(1)}, \ldots, \delta^{(d)}$, i.e. $(\delta^{(1)}, \ldots, \delta^{(m)}) \in \widetilde{S}_1$. A simple reasoning reveals that $(\delta^{(1)}, \ldots, \delta^{(m)})$ does not belong to anyone of sets $S_2, S_3, \ldots$. In fact, adding non-violated constraints to $\delta^{(1)}, \ldots, \delta^{(d)}$ does not change the solution and each of the added constraints can be removed back without altering the solution. Therefore, none of the constraints $\delta^{(d+1)}, \ldots, \delta^{(m)}$ can be of support and hence the multi-extraction is not in S_2 or in S_3, etc. It follows that $\widetilde{S}_1$ is a subset of the complement of $S_2 \cup S_3 \cup \cdots$, which is S_1 up to a probability 0 set.

PART 2: $\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ FOR EVERY PROBLEM

A non-fully-supported problem admits with non-zero probability randomized instances where the number of support constraints is less than d. A support constraint has to be an active constraint, and the typical reason for a lack of support constraints is that at the optimum the active constraints are less than d, see Figure 4. To carry on a proof along lines akin to

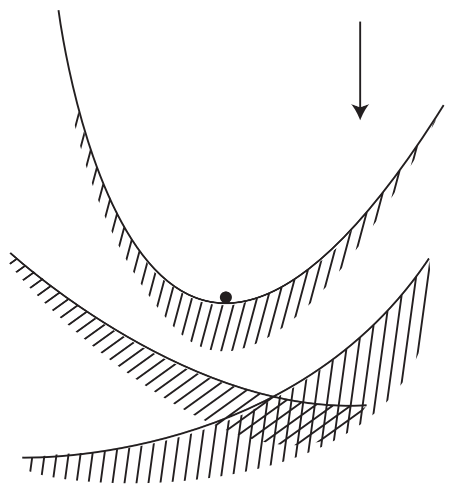

Figure 4: A 2-dimensional problem with only 1 active constraint which is of support.

those for the fully-supported case, we are well-advised to generalize the notion of solution to that of ball-solution; a ball-solution has always at least d active constraints. For simplicity, we henceforth assume that constraints are not trivial, i.e. $\mathcal{X}_\delta \neq \mathbb{R}^d$, $\forall \delta \in \Delta$.

Definition 4 (ball-solution) *Consider a finite instance RP_m of RP with $m \geq d$, and let x_m^* be its solution. The* ball-solution $\mathcal{B}(x_m^*, r_m^*)$ *of RP_m is the largest closed ball centered in x_m^**

fully contained in the feasibility domain of all constraints with the exception of at most $d-1$ of them, i.e. $\mathcal{X}_{\delta^{(i)}} \cap \mathcal{B}(x_m^*, r_m^*) = \mathcal{B}(x_m^*, r_m^*)$ *for all i's, except at most $d-1$ of them.*

See Figure 5 for an example of ball-solution. Note also that, when active constraints are d or

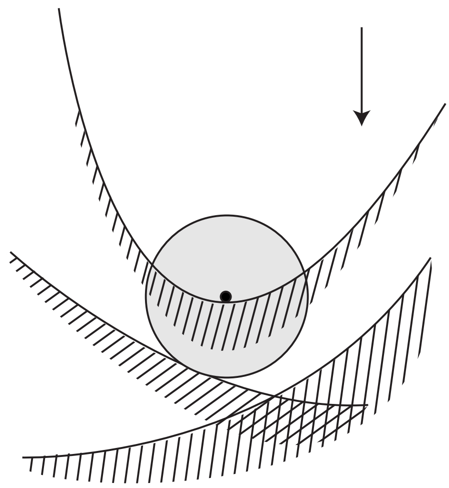

Figure 5: Ball-solution.

more, $r_m^* = 0$ and $\mathcal{B}(x_m^*, r_m^*)$ reduces to the standard solution x_m^*. Moreover, a ball-solution $\mathcal{B}(x_m^*, r_m^*)$ need not be contained in $\mathcal{X}$, although its center x_m^* does.

The notion of active constraint can be generalized to balls as follows.

Definition 5 (active constraint for a ball) *A constraint δ is* active *for a ball $\mathcal{B}(x,r)$ if $\mathcal{X}_\delta \cap \mathcal{B}(x,r) \neq \emptyset$ and $\mathcal{X}_\delta \cap \mathcal{B}(x, r+h) \neq \mathcal{B}(x, r+h)$, $\forall h > 0$. If in addition $\mathcal{X}_\delta \cap \mathcal{B}(x,r) = \mathcal{B}(x,r)$, $\mathcal{X}_\delta$ is said to be* strictly active.

This definition of active constraint for a ball is illustrated in Figure 6. If the ball is a single point, active and strictly active is the same and reduces to the standard notion of active.

By construction, a ball-solution has at least d active constraints. To go back to the track of the proof in PART 1, however, we need d support constraints, not just active constraints. The following definition naturally extends the notion of support constraint to the case of ball-solutions.

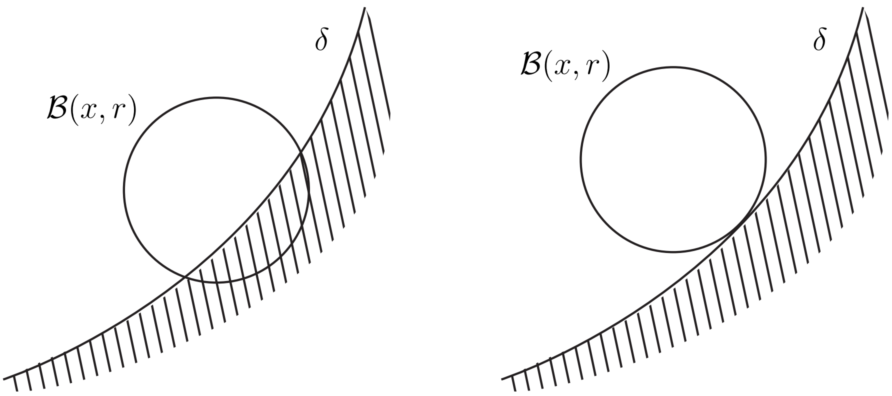

Figure 6: Active and strictly-active constraint for a ball.

Definition 6 (ball-support constraint) *Constraint* $\delta^{(r)}$*,* $r \in \{1, \dots, m\}$*, is a* ball-support constraint *for* RP_m *if its removal changes the ball-solution of* RP_m*.*

An active constraint need not be of ball-support, nor an RP_m has always d ball-support constraints (see Figure 7 where $\delta^{(2)}$ and $\delta^{(3)}$ are not of support). It is clear that the number

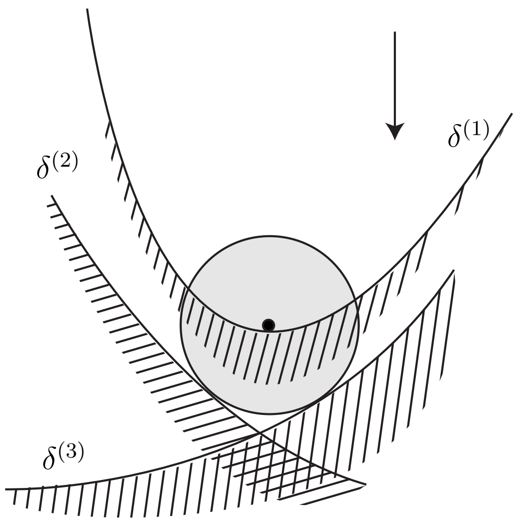

Figure 7: Only $\delta^{(1)}$ is a ball-support constraint.

of ball-support constraints is less than or equal to d. The case with less than d ball-support constraints is regarded as *degenerate* and needs to be treated separately. We thus split the remaining part of the proof in two sections: PART 2.a "Fully-ball-supported problems" and

PART 2.b "Degenerate problems". Before proceeding, we are well-advised to give a formal definition of fully-ball-supported problems.

Definition 7 (fully-ball-supported problem) *A finite instance RP_m, with $m \geq d$, is fully-ball-supported if the number of ball-support constraints of RP_m is d.*
Problem RP is fully-ball-supported if, for any $m \geq d$, RRP_m is fully-ball-supported with probability 1.

PART 2.a: FULLY-BALL-SUPPORTED PROBLEMS

We start by introducing the notion of constraint violated by a ball: a constraint δ is violated by $\mathcal{B}(x,r)$ if $\mathcal{X}_\delta \cap \mathcal{B}(x,r) \neq \mathcal{B}(x,r)$. The definition of probability of violation then generalizes naturally to the ball case:

Definition 8 (violation probability of a ball) *The* violation probability *of a ball $\mathcal{B}(x,r)$, $x \in \mathcal{X}$, is defined as $V_{\mathcal{B}}(x,r) = \mathbb{P}\{\delta \in \Delta : \mathcal{X}_\delta \cap \mathcal{B}(x,r) \neq \mathcal{B}(x,r)\}$.*

Clearly, for any x, $V_{\mathcal{B}}(x,r) \geq V(x)$. Hence, if $\mathcal{B}(x_N^*, r_N^*)$ is the ball-solution of RRP_N we have

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \mathbb{P}^N\{V_{\mathcal{B}}(x_N^*, r_N^*) > \epsilon\}. \tag{13}$$

Below, we show that a result similar to (8) holds for fully-ball-supported problems, namely

$$\mathbb{P}^N\{V_{\mathcal{B}}(x_N^*, r_N^*) > \epsilon\} = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}, \tag{14}$$

and this result together with (13) leads to the thesis

$$\mathbb{P}^N\{V(x_N^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}.$$

The proof of (14) is verbatim the same as the proof of PART 1 provided that one substitutes

- *solution* with *ball-solution*
- *support constraint* with *ball-support constraint*
- *violation probability V* with *violation probability of a ball $V_{\mathcal{B}}$*,

with only one exception: the part where we proved that $\boldsymbol{S_1} \subseteq \widetilde{\boldsymbol{S}}_{\boldsymbol{1}}$ has to be modified in a way that we spell out in the following.

The first rationale to conclude that "the solution with only the d support constraints

$\delta^{(1)}, \ldots, \delta^{(d)}$ in place is the same as the solution with all m constraints" is still valid and leads in our present context to the fact that the *ball*-solution with only the d *ball*-support constraints $\delta^{(1)}, \ldots, \delta^{(d)}$ in place is the same as the *ball*-solution with all m constraints. Instead, the last argument with which we concluded that $\boldsymbol{S_1} \subseteq \widetilde{\boldsymbol{S}}_{\boldsymbol{1}}$ is no longer valid since ball-solutions can violate constraints.

To amend it, suppose for the purpose of contradiction that a constraint among $\delta^{(d+1)}, \ldots, \delta^{(m)}$, say $\delta^{(d+1)}$, is violated by the ball-solution with d constraints. Two cases can occur: (i) the ball-solution has only 1 strictly active constraint among $\delta^{(1)}, \ldots, \delta^{(d)}$; or (ii) it has more than one. In case (i), $d-1$ constraints among $\delta^{(1)}, \ldots, \delta^{(d)}$ are violated by the ball solution, so that, with the extra $\delta^{(d+1)}$ violated constraint, the number of violated constraints of the ball-solution with m constraints would add up to at least d and this contradicts the definition of ball-solution. If instead (ii) is true, a simple thought reveals that, with one more constraint $\delta^{(d+1)}$ violated by the ball-solution, the strictly active constraints (which, in this case, are more than 1) cannot be of ball-support for the problem with m constraints and this contradicts the fact that $(\delta^{(1)}, \ldots, \delta^{(m)}) \in S_1$.

<u>PART 2.b</u>: DEGENERATE PROBLEMS

For not being fully-ball-supported, a finite problem RP_m needs to have more than one strictly active constraint, a circumstance which requires that constraints are not "generically" distributed. This observation is at the basis of the rather technical proof of this PART 2.b, which proceeds along the following steps:

- STEP 1 a constraints "heating" is introduced; heating scatters constraints around and the resulting heated problem is shown to be fully-ball-supported; by resorting to the result in PART 2.a, conclusions are derived about the violation properties of the heated problem;
- STEP 2 it is shown that the solution of the original problem is recovered by cooling the heated problem down;
- STEP 3 the violation properties of the original (non-heated) problem are determined from the violation properties of the heated problem by a limiting process.

STEP 1 [Heating]

Let $\Delta' := \Delta \times \mathcal{B}_\rho$, where $\rho > 0$ is the *heating parameter* and $\mathcal{B}_\rho \subset \mathbb{R}^d$ is the closed ball centered in the origin with radius ρ, and let $\mathbb{P}' := \mathbb{P} \times \mathbb{U}$ be the probability in Δ' obtained as the product probability between $\mathbb{P}$ and the uniform probability $\mathbb{U}$ in $\mathcal{B}_\rho$. Each $z \in \mathcal{B}_\rho$ represents a constraint translation and the heated robust program is defined as

$$\begin{aligned} \mathrm{HRP}: \quad & \min_{x \in \mathcal{X} \subseteq \mathbb{R}^d} c^T x \\ & \text{subject to: } x \in \bigcap_{(\delta,z) \in \Delta'} [\mathcal{X}_\delta + z], \end{aligned}$$

where $[\mathcal{X}_\delta + z]$ is set $\mathcal{X}_\delta$ translated by z. We show that HRP is fully-ball-supported.

To start with, consider a given deterministic ball $\mathcal{B}(x,r)$. We first prove that the strictly active constraints $\delta' \in \Delta'$ for $\mathcal{B}(x,r)$ form a set of zero-probability $\mathbb{P}'$, and later on from this we shall conclude that HRP is fully-ball-supported.

Let $\delta' = (\delta, z)$ and let $\mathbb{I}_A$ indicate the indicator function of set A, and write

$$\begin{aligned} & \mathbb{P}'\{\delta' \text{ is strictly active for } \mathcal{B}(x,r)\} \\ = \; & \int_{\Delta'} \mathbb{I}_{\{\delta' \text{ is strictly active for } \mathcal{B}(x,r)\}} \mathbb{P}'(\mathrm{d}\delta') \\ = \; & \text{[by Fubini's theorem [18]]} \\ = \; & \int_\Delta \left[\int_{\mathcal{B}_\rho} \mathbb{I}_{\{(\delta,z) \text{ is strictly active for } \mathcal{B}(x,r)\}} \frac{\mathrm{d}z}{\mathrm{Vol}(\mathcal{B}_\rho)} \right] \mathbb{P}(\mathrm{d}\delta). \end{aligned} \tag{15}$$

The result that

$$\mathbb{P}'\{\delta' \text{ is strictly active for } \mathcal{B}(x,r)\} = 0 \tag{16}$$

is established by showing that the term within square brackets in formula (15) is null for all δ's.

Fix a δ and let $C = \{z \in \mathcal{B}_\rho : \mathcal{B}(x,r) \subseteq [\mathcal{X}_\delta + z]\}$ be the set of translations not violating $\mathcal{B}(x,r)$. We show that C is convex and that the set $\{z \in \mathcal{B}_\rho : (\delta,z) \text{ is strictly active for } \mathcal{B}(x,r)\}$ belongs to ∂C, the boundary of C. Since the boundary of a convex set has zero Lebesgue measure[3],

[3] This simple fact follows from the observation that a convex set C in $\mathbb{R}^d$ either belongs to a flat of dimension $d-1$ – and therefore C has zero $\mathbb{R}^d$ Lebesgue measure – or it admits an interior point $\bar{z}$ and every half-line from $\bar{z}$ crosses the boundary of C in only one point (see e.g. Propositions 1.1.13 and 1.1.14 in [15]).

the desired result that the term within square brackets in formula (15) is null follows, viz.

$$\int_{\mathcal{B}_\rho} \mathbb{I}_{\{(\delta,z) \text{ is strictly active for } \mathcal{B}(x,r)\}} \frac{\mathrm{d}z}{\mathrm{Vol}(\mathcal{B}_\rho)} = 0. \tag{17}$$

The convexity of C is immediate: let $z_1, z_2 \in C$, that is $\mathcal{B}(x,r) \subseteq [\mathcal{X}_\delta + z_1]$ and $\mathcal{B}(x,r) \subseteq [\mathcal{X}_\delta + z_2]$ or, equivalently, $\mathcal{B}(x,r) - z_1 \subseteq \mathcal{X}_\delta$ and $\mathcal{B}(x,r) - z_2 \subseteq \mathcal{X}_\delta$. From convexity of $\mathcal{X}_\delta$, it follows that $\mathcal{B}(x,r) - \alpha z_1 - (1-\alpha) z_2 \subseteq \mathcal{X}_\delta$, $\forall \alpha \in [0,1]$, that is $\alpha z_1 + (1-\alpha) z_2 \in C$ and C is convex.

Consider now an interior point z of C (if any), i.e. it exists a ball centered in z all contained in C. This means that $[\mathcal{X}_\delta + z]$ can be moved around in all directions by a small quantity and $\mathcal{B}(x,r)$ remains contained in it. It easily follows that (δ, z) cannot be strictly active and, thus, $\{z \in \mathcal{B}_\rho : (\delta, z) \text{ is strictly active for } \mathcal{B}(x,r)\}$ has to belong to ∂C.

Wrapping up, (17) is established and, substituting in (15), equation (16) is obtained.

We next prove that (16) entails that HRP is fully-ball-supported.

Consider a finite instance HRP_m of HRP with $m \geq d$. One by one, eliminate $m-d$ constraints choosing any time a constraint among those non-violated by the ball-solution in such a way that the ball-solution does not change. This is certainly possible because the ball-support constraints are at most d. In the end, we are left with d constraints, say the first d $\delta'^{(1)}, \ldots, \delta'^{(d)}$. A simple thought reveals that these d constraints are actually of ball-support for HRP_m provided that none of the other $m-d$ constraints that have been removed was strictly active.

Repeat the same above procedure for every m-ple of constraints (that is for every HRP_m generated by HRP), and group together all the m-ples for which the procedure returns in the end the first d constraints $\delta'^{(1)}, \ldots, \delta'^{(d)}$. Call this group of m-ples G. We shall show that the probability of the m-ples in G such that HRP_m is not fully-ball-supported is zero, and from this – by the observation that only a finite number $\binom{m}{d}$ of groups of m-ples can be similarly constructed – the final conclusion that HRP is fully-ball-supported will be secured.

Select fixed values $\bar{\delta}'^{(1)}, \ldots, \bar{\delta}'^{(d)}$ for the first d constraints and consider the ball-solution $\mathcal{B}$ these constraints generate. Let the other $m-d$ constraints vary in such a way that the m-

ple $\bar{\delta}'^{(1)}, \ldots, \bar{\delta}'^{(d)}, \delta'^{(d+1)}, \ldots, \delta'^{(m)}$ belongs to G. For one such m-ple to correspond to a *non* fully-ball-supported HRP$_m$ at least one among the $m-d$ constraints $\delta'^{(d+1)}, \ldots, \delta'^{(m)}$ must be strictly active for $\mathcal{B}$, but we have proven in (16) that this happens with probability zero. Integrating over all possible values $\bar{\delta}'^{(1)}, \ldots, \bar{\delta}'^{(d)}$ for the first d constraints, the conclusion is drawn that the *non* fully-ball-supported HRP$_m$ in G have zero probability.
Hence, by the above observation that there are only a finite number $\binom{m}{d}$ of groups and by the fact that $\binom{m}{d}$ times zero is zero, we obtain that HRP is fully-ball-supported.

To conclude STEP 1, note that if we suppose to extract N constraints $\delta'^{(1)}, \ldots, \delta'^{(N)}$ from Δ' according to probability $\mathbb{P}'$ and in an independent fashion, and we denote by x'^*_N the corresponding solution, the result of PART 2.a can be invoked to establish that

$$(\mathbb{P}')^N\{V'(x'^*_N) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}, \tag{18}$$

where $V'(x)$ is the probability of violation for the heated problem (i.e. $V'(x) = \mathbb{P}'\{(\delta, z) \in \Delta' : x \notin [\mathcal{X}_\delta + z]\}$). (18) is the final result to which we wanted to arrive in this heating STEP 1.

<u>STEP 2</u> [Cooling]

Fix a multi-extraction $(\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}) \in \Delta^N$, and consider x^*_N, the solution of the original optimization problem RP$_N$ with such constraints. We remark that in all this STEP 2 the multi-extraction $(\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)})$ is kept fixed and never changed throughout. Consider a closed ball $\mathcal{B}(x_f, r_f)$, $r_f > 0$, in the feasibility domain of RP$_N$, which exists because the feasibility domain of RP$_N$ has non-empty interior. Further, let $\rho_k \downarrow 0$ be a sequence of heating parameters monotonically decreasing to zero (*cooling of the heating parameter*) and such that $\rho_1 < \frac{r_f}{2}$. For all ρ_k, consider the heated versions of $(\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)})$, namely $\big((\bar{\delta}^{(1)}, z_k^{(1)}), \ldots, (\bar{\delta}^{(N)}, z_k^{(N)})\big)$ where $z_k^{(1)}, \ldots, z_k^{(N)} \in \mathcal{B}_{\rho_k}$, and let $x'^*_N(z_k^{(1)}, \ldots, z_k^{(N)})$ be the solution of the heated optimization problem HRP$_N$ with heated constraints $(\bar{\delta}^{(1)}, z_k^{(1)}), \ldots, (\bar{\delta}^{(N)}, z_k^{(N)})$. The goal of this STEP 2 is to prove that

$$\sup_{z_k^{(1)}, \ldots, z_k^{(N)} \in \mathcal{B}_{\rho_k}} \left\| x'^*_N(z_k^{(1)}, \ldots, z_k^{(N)}) - x^*_N \right\| \longrightarrow 0, \quad \text{as } k \to \infty, \tag{19}$$

that is, the solution of the original problem is recovered by cooling the heated problem down[4].

For brevity, from now on we omit the arguments $z_k^{(1)}, \ldots, z_k^{(N)}$ and write x'^*_N for $x'^*_N(z_k^{(1)}, \ldots, z_k^{(N)})$.

We first show that

$$\limsup_{k\to\infty} \sup_{z_k^{(1)},\ldots,z_k^{(N)} \in \mathcal{B}_{\rho_k}} c^T x'^*_N \le c^T x^*_N. \tag{20}$$

Following Figure 8, consider the convex hull $\mathrm{co}[\mathcal{B}(x_f, r_f) \cup x^*_N]$ generated by the feasibility ball $\mathcal{B}(x_f, r_f)$ and the solution x^*_N of the original problem with constraints $\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}$. By

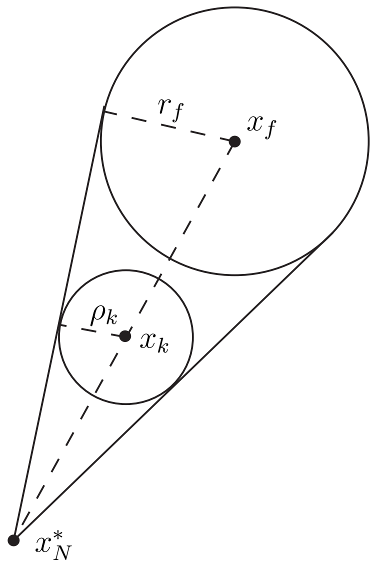

Figure 8: Convex hull of $\mathcal{B}(x_f, r_f)$ and x^*_N, and construction of $\mathcal{B}(x_k, \rho_k)$.

convexity, $\mathrm{co}[\mathcal{B}(x_f, r_f) \cup x^*_N]$ is feasible for the original problem RP_N. Construct the closed ball $\mathcal{B}(x_k, \rho_k) \subset \mathrm{co}[\mathcal{B}(x_f, r_f) \cup x^*_N]$ with radius ρ_k, whose center x_k is as close as possible to x^*_N and lies on the line segment connecting x_f with x^*_N (this ball exists since $\rho_1 < r_f$; the assumed stricter condition that $\rho_1 < \frac{r_f}{2}$ is required in a next construction). Clearly, $x_k \to x^*_N$ as $k \to \infty$. Since x_k is in the feasibility domain of RP_N at a distance at least ρ_k from where $\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}$ are violated, x_k is also in the feasibility domain of every heated problem HRP_N

[4]Although result (19) has an intuitive appeal, its proof is rather technical. The reader not interested in these technical details can jump to Step 3 from here without loss of continuity.

with heating parameter ρ_k. Thus,

$$\limsup_{k\to\infty} \sup_{z_k^{(1)},\dots,z_k^{(N)}\in\mathcal{B}_{\rho_k}} c^T x_N'^* \le \limsup_{k\to\infty} c^T x_k = c^T x_N^*,$$

that is (20) holds.

Next, we construct a new convex hull which will allow us to reformulate goal (19) in a different, handier, way. Based on this reformulation, (19) will then be established in the light of (20).

The new convex hull is $\mathrm{co}[\mathcal{B}(x_f, r_f - \rho_k) \cup x_N'^*]$, see Figure 9. Note that, for a given k,

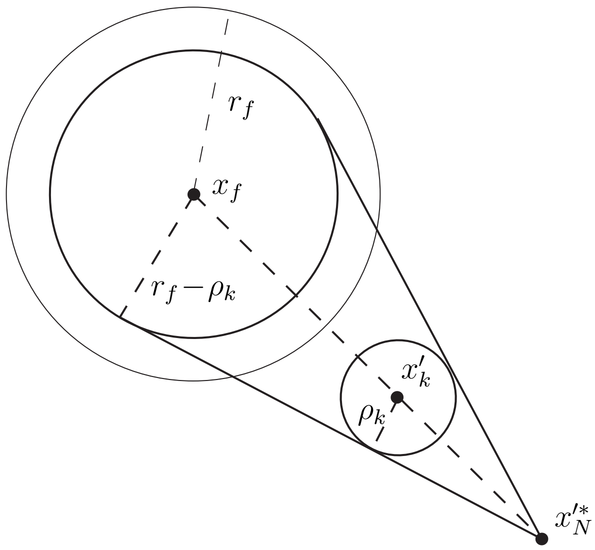

Figure 9: Convex hull of $\mathcal{B}(x_f, r_f - \rho_k)$ and $x_N'^*$, and construction of $\mathcal{B}(x_k', \rho_k)$.

$\mathcal{B}(x_f, r_f - \rho_k)$ is a fixed ball, instead $x_N'^*$ depends on the specific choice of $z_k^{(1)}, \dots, z_k^{(N)} \in \mathcal{B}_{\rho_k}$; this means that there are actually as many convex hulls as choices of $z_k^{(1)}, \dots, z_k^{(N)}$. Moreover, $\mathrm{co}[\mathcal{B}(x_f, r_f - \rho_k) \cup x_N'^*]$ is feasible for problem HRP_N with constraints translated by $z_k^{(1)}, \dots, z_k^{(N)}$ since $\mathcal{B}(x_f, r_f - \rho_k)$ and $x_N'^*$ are. Construct then the closed ball $\mathcal{B}(x_k', \rho_k) \subseteq \mathrm{co}[\mathcal{B}(x_f, r_f - \rho_k) \cup x_N'^*]$ with radius ρ_k, whose center x_k' is as close as possible to $x_N'^*$ and lies on the line segment connecting x_f with $x_N'^*$ (this ball exists since $\rho_1 < \frac{r_f}{2}$). Note that x_k' depends on $z_k^{(1)}, \dots, z_k^{(N)}$ too.

Since x_k' is in the feasibility domain of HRP_N with constraints translated by $z_k^{(1)}, \dots, z_k^{(N)}$ at a distance at least ρ_k from where these translated constraints are violated, x'_k is also in the feasibility domain of RP_N.

What is different from the previous convex hull construction is that we cannot here easily conclude that $x'_k \to x'^*_N$ as $k \to \infty$ since x'^*_N is not a fixed point (it depends on $z_k^{(1)}, \ldots, z_k^{(N)} \in \mathcal{B}_{\rho_k}$, a ball that changes with k). We can still, however, secure a result that goes along a similar line, namely that

$$x'_k = \alpha_k x_f + (1-\alpha_k) x'^*_N, \qquad \text{where } \alpha_k = \frac{\rho_k}{r_f - \rho_k} \longrightarrow 0 \text{ as } k \to \infty, \tag{21}$$

as it results from Figure 9 by a simple proportion argument[5]. Reorganizing terms in this equation, we obtain $x'^*_N - x^*_N = -\frac{\alpha_k}{1-\alpha_k}(x_f - x^*_N) + \frac{1}{1-\alpha_k}(x'_k - x^*_N)$, from which

$$\left\| x'^*_N - x^*_N \right\| \leq \frac{\alpha_k}{1-\alpha_k} \left\| x_f - x^*_N \right\| + \frac{1}{1-\alpha_k} \left\| x'_k - x^*_N \right\|.$$

We are now ready to reformulate goal (19) in a different way.

Note that the norm in (19) is the same as the left-hand-side of the latter equation. In the right-hand-side, $\|x_f - x^*_N\|$ is a fixed quantity multiplied by scalar $\frac{\alpha_k}{1-\alpha_k}$ which goes to zero. So, this first term vanishes. In the second term, scalar $\frac{1}{1-\alpha_k} \to 1$, and hence (19) is equivalent to:

$$\sup_{z_k^{(1)}, \ldots, z_k^{(N)} \in \mathcal{B}_{\rho_k}} \left\| x'_k - x^*_N \right\| \longrightarrow 0, \quad \text{as } k \to \infty. \tag{22}$$

The goal of establishing (19) is finally achieved by proving equation (22) by contradiction.

Suppose that (22) is false; then, for a given $\mu > 0$, we can choose translations $\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)} \in \mathcal{B}_{\rho_k}$, $k = 1, 2, \ldots$, such that

$$\left\| x'_k(\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)}) - x^*_N \right\| > \mu, \quad \forall k,$$

[5]Note that (21) does not imply that $x'_k \to x'^*_N$ since x'^*_N could in principle escape to infinity.

where we have here preferred to explicitly indicate dependence of x'_k on $\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)}$. Note that, $x'_k(\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)})$ is asymptotically super-optimal for problem RP_N:

$$\begin{aligned}
&\limsup_{k\to\infty} c^T x'_k(\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)}) \\
&\quad \leq \ [\text{using (21) and since } \alpha_k \to 0] \\
&\quad \leq \ \limsup_{k\to\infty} \sup_{z_k^{(1)},\ldots,z_k^{(N)}} c^T x'^*_N \\
&\quad \leq \ [\text{using (20)}] \\
&\quad \leq \ c^T x^*_N.
\end{aligned} \tag{23}$$

The line segment connecting $x'_k(\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)})$ with x^*_N intersects the surface of the ball with center x^*_N and radius μ in a point that we name x'^S_k. x'^S_k is still feasible for RP_N being a convex combination of x^*_N and $x'_k(\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)})$, both feasible points for RP_N. In addition, since $x'_k(\bar{z}_k^{(1)}, \ldots, \bar{z}_k^{(N)})$ is asymptotically super-optimal for RP_N (see (23)) and x^*_N is the solution of RP_N, x'^S_k is asymptotically super-optimal for RP_N too, i.e. $\limsup_{k\to\infty} c^T x'^S_k \leq c^T x^*_N$. Finally, since x'^S_k belongs to a compact, it admits a convergent subsequence to, say, x'^S_∞, a point which is still feasible for RP_N due to that the feasibility domain of RP_N is closed. x'^S_∞ would thus be feasible and super-optimal for RP_N, so contradicting the uniqueness of the solution of RP_N.

This concludes STEP 2.

<u>STEP 3</u> [Drawing the conclusions]

The theorem statement that $\mathbb{P}^N\{V(x^*_N) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$ is established in this STEP 3 along the following line: by the convergence result (19) in STEP 2, a bad multi-extraction $(\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)})$ (i.e. one such that $V(x^*_N) > \epsilon$) is shown to generate bad heated multi-extractions $\big((\bar{\delta}^{(1)}, z_k^{(1)}), \ldots, (\bar{\delta}^{(N)}, z_k^{(N)})\big)$ for k large enough; we thus have that the probability of bad multi-extractions can be bounded by the probability of bad heated multi-extractions; by then using the bound for the probability of bad heated multi-extractions derived in STEP 1, the thesis follows.

Fix a *bad* multi-extraction $(\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}) \in \Delta^N$, and consider x^*_N, the solution of the optimization problem RP_N with constraints $\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}$. For an additional constraint $\delta \in \Delta$ to be violated by x_N^*, x_N^* must belong to the complement of $\mathcal{X}_\delta$, i.e. $\mathcal{X}_\delta^c$. Since $\mathcal{X}_\delta^c$ is open, we then have that there exists a small enough ball centered in x_N^* fully contained in $\mathcal{X}_\delta^c$. Thus,

$$\{\delta \in \Delta : \ x_N^* \notin \mathcal{X}_\delta\} = \bigcup_{n=1,2,\ldots} \{\delta \in \Delta : \ \mathcal{B}(x_N^*, 1/n) \subseteq \mathcal{X}_\delta^c\}, \tag{24}$$

and

$$\begin{aligned}
\epsilon &< \text{[since } (\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}) \text{ is bad]} \\
&< V(x_N^*) \\
&= \mathbb{P}\{\delta \in \Delta : \ x_N^* \notin \mathcal{X}_\delta\} \\
&= \text{[using (24)]} \\
&= \mathbb{P}\Big\{ \bigcup_{n=1,2,\ldots} \{\delta \in \Delta : \ \mathcal{B}(x_N^*, 1/n) \subseteq \mathcal{X}_\delta^c\} \Big\} \\
&= \lim_{n \to \infty} \mathbb{P}\{\delta \in \Delta : \ \mathcal{B}(x_N^*, 1/n) \subseteq \mathcal{X}_\delta^c\},
\end{aligned}$$

from which there exists a $\bar{n}$ such that

$$\mathbb{P}\{\delta \in \Delta : \ \mathcal{B}(x_N^*, 1/\bar{n}) \subseteq \mathcal{X}_\delta^c\} > \epsilon. \tag{25}$$

Let us now heat the constraints $\bar{\delta}^{(1)}, \ldots, \bar{\delta}^{(N)}$ up by translation parameters $z_k^{(1)}, \ldots, z_k^{(N)} \in \mathcal{B}_{\rho_k}$, and ask the following question: is it true that the heated multi-extraction $\big((\bar{\delta}^{(1)}, z_k^{(1)}), \ldots, (\bar{\delta}^{(N)}, z_k^{(N)})\big)$ is bad for HRP with heating parameter ρ_k? It turns out that the answer is positive for k large enough, a fact that is proven next.

Recall that $x_N'^*$ is the solution with constraints $(\bar{\delta}^{(1)}, z_k^{(1)}), \ldots, (\bar{\delta}^{(N)}, z_k^{(N)})$ and define $d_k := \sup_{z_k^{(1)}, \ldots, z_k^{(N)} \in \mathcal{B}_{\rho_k}} \|x_N'^* - x_N^*\|$ which, by (19), goes to 0 as $k \to \infty$. Pick a $\bar{k}$ such that

$$d_k + \rho_k < 1/\bar{n}, \quad \forall k \geq \bar{k}.$$

All heated solutions $x_N'^*$ are apart from x_N^* by at most d_k and all heated constraints $(\delta, z) \in \Delta \times \mathcal{B}_{\rho_k}$ are apart from the corresponding unheated constraint δ by at most ρ_k. Thus, if $k \geq \bar{k}$, all heated versions of a constraint δ in the set $\{\delta \in \Delta : \ \mathcal{B}(x_N^*, 1/\bar{n}) \subseteq \mathcal{X}_\delta^c\}$ in the left-hand-side of (25) are violated by $x_N'^*$. That is,

$$\{\delta \in \Delta : \ \mathcal{B}(x_N^*, 1/\bar{n}) \subseteq \mathcal{X}_\delta^c\} \times \mathcal{B}_{\rho_k} \subseteq \{(\delta, z) \in \Delta \times \mathcal{B}_{\rho_k} : \ x_N'^* \notin [\mathcal{X}_\delta + z]\}, \quad \forall k \geq \bar{k}. \tag{26}$$

Then, for any $z_k^{(1)}, \dots, z_k^{(N)} \in \mathcal{B}_{\rho_k}$ and for any $k \geq \bar{k}$, we have that

$$\begin{aligned}
V'(x_N'^*) &= \mathbb{P}'\{(\delta, z) \in \Delta \times \mathcal{B}_{\rho_k} :\ x_N'^* \notin [\mathcal{X}_\delta + z]\} \\
&\geq \text{[using (26)]} \\
&\geq \mathbb{P}'\Big\{\{\delta \in \Delta :\ \mathcal{B}(x_N^*, 1/\bar{n}) \subseteq \mathcal{X}_\delta^c\} \times \mathcal{B}_{\rho_k}\Big\} \\
&= \text{[recalling that } \mathbb{P}' = \mathbb{P} \times \mathbb{U}] \\
&= \mathbb{P}\{\delta \in \Delta :\ \mathcal{B}(x_N^*, 1/\bar{n}) \subseteq \mathcal{X}_\delta^c\} \cdot \mathbb{U}\{\mathcal{B}_{\rho_k}\} \\
&> \text{[since } \mathbb{U}\{\mathcal{B}_{\rho_k}\} = 1 \text{ and using (25)]} \\
&> \epsilon,
\end{aligned}$$

i.e. $\big((\bar{\delta}^{(1)}, z_k^{(1)}), \dots, (\bar{\delta}^{(N)}, z_k^{(N)})\big)$ is bad for HRP with heating parameter ρ_k for any $z_k^{(1)}, \dots, z_k^{(N)} \in \mathcal{B}_{\rho_k}$ when $k \geq \bar{k}$. In turn, this entails that

$$\int_{\mathcal{B}_{\rho_k}^N} \mathbb{I}_{\{V'(x_N'^*) > \epsilon\}} \frac{\mathrm{d}z^N}{\mathrm{Vol}(\mathcal{B}_{\rho_k}^N)} = 1, \quad \forall k \geq \bar{k}. \tag{27}$$

Finally,

$$\begin{aligned}
&\sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \\
&\quad \geq \quad \text{[using (18)]} \\
&\quad \geq \quad (\mathbb{P}')^N \{V'(x_N'^*) > \epsilon\} \\
&\quad = \quad \int_{\Delta^N} \left[\int_{\mathcal{B}_{\rho_k}^N} \mathbb{I}_{\{V'(x_N'^*) > \epsilon\}} \frac{\mathrm{d}z^N}{\mathrm{Vol}(\mathcal{B}_{\rho_k}^N)} \right] \mathbb{P}^N(\mathrm{d}\delta^N) \\
&\quad \geq \quad \int_{\{V(x_N^*) > \epsilon\}} \left[\int_{\mathcal{B}_{\rho_k}^N} \mathbb{I}_{\{V'(x_N'^*) > \epsilon\}} \frac{\mathrm{d}z^N}{\mathrm{Vol}(\mathcal{B}_{\rho_k}^N)} \right] \mathbb{P}^N(\mathrm{d}\delta^N) \\
&\quad \xrightarrow[k \to \infty]{} \quad \text{[recalling (27) and by the dominated convergence theorem, [19]]} \\
&\quad \xrightarrow[k \to \infty]{} \quad \int_{\{V(x_N^*) > \epsilon\}} \mathbb{P}^N(\mathrm{d}\delta^N) \\
&\quad = \quad \mathbb{P}^N\{V(x_N^*) > \epsilon\}.
\end{aligned}$$

This concludes the proof.

References

[1] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semidefinite programming. *SIAM J. on Optimization*, 7(4):991–1016, 1997.

[2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[4] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions to uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.

[5] V.D. Blondel and J.N. Tsitsiklis. A survey of computational complexity results in systems and control. *automatica*, 36:1249–1274, 2000.

[6] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[7] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Trans. Aut. Control*, 51:742–753, 2006.

[8] G. Calafiore and B. Polyak. Stochastic algorithms for exact and approximate feasibility of robust LMIs. *IEEE Trans. Aut. Control*, 46(11):1755–1759, November 2001.

[9] D.P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.

[10] E. Erdogan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming - ser. B*, 107(1-2):37–61, 2006.

[11] L. El Ghaoui and G. Calafiore. Robust filtering for discrete-time systems with bounded noise and parametric uncertainty. *IEEE Trans. Aut. Control*, 46(7):1084–1089, July 2001.

[12] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[13] L. El Ghaoui and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.*, 9(1):33–52, 1998.

[14] L. El Ghaoui and S.-I. Niculescu. Robust decision problems in engineering: A Linear Matrix Inequality approach. In L. El Ghaoui and S.-I. Niculescu, editors, *Advances in Linear Matrix Inequality Methods in Control*, chapter 1. SIAM, 2000.

[15] R. Lucchetti. *Convexity and well-posed problems.* CMS Books in Mathematics. Springer, New York, NY, USA, 2006.

[16] A. Nemirovski. Several np-hard problems arising in robust stability analysis. *SIAM J. on Matrix Analysis and Applications*, 6:99–195, 1993.

[17] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 3–48, London, 2005. Springer.

[18] W. Rudin. *Real and complex analysis.* McGraw-Hill, New York, NY, USA, 3rd edition, 1987.

[19] A.N. Shiryaev. *Probability.* Springer, New York, NY, USA, 1996.