RANDOM CONVEX PROGRAMS[*]

GIUSEPPE CARLO CALAFIORE[†]

Abstract. Random convex programs (RCPs) are convex optimization problems subject to a finite number N of random constraints. The optimal objective value J^* of an RCP is thus a random variable. We study the probability with which J^* is no longer optimal if a further random constraint is added to the problem (violation probability, V^*). It turns out that this probability rapidly concentrates near zero as N increases. We first develop a theory for RCPs, leading to explicit bounds on the upper tail probability of V^*. Then we extend the setup to the case of RCPs with r a posteriori violated constraints (RCPVs): a paradigm that permits us to improve the optimal objective value while maintaining the violation probability under control. Explicit and nonasymptotic bounds are derived also in this case: the upper tail probability of V^* is upper bounded by a multiple of a beta distribution, irrespective of the distribution on the random constraints. All results are derived under no feasibility assumptions on the problem. Further, the relation between RCPVs and chance-constrained problems (CCP) is explored, showing that the optimal objective J^* of an RCPV with the generic constraint removal rule provides, with arbitrarily high probability, an upper bound on the optimal objective of a corresponding CCP. Moreover, whenever an optimal constraint removal rule is used in the RCPVs, then appropriate choices of N and r exist such that J^* approximates arbitrarily well the objective of the CCP.

Key words. scenario optimization, chance-constrained optimization, randomized methods, robust convex optimization

AMS subject classifications. 90C25, 90C15, 90C34, 68W20

DOI. 10.1137/090773490

1. Introduction. A random convex program (RCP) is a finite-dimensional optimization problem in which a linear objective is minimized under convex constraints of the form $f(x, \delta^{(i)}) \leq 0$, $i = 1, \ldots, N$, where $\delta^{(i)}$ are independently and identically distributed samples of a random vector of parameters δ; see section 3.1 for a formal definition. The optimal objective J^* of such a problem and a corresponding optimal solution x^* (when it exists) are random variables. A key feature of an RCP relates to the fact that its optimal objective J^* remains optimal with high probability also when a new constraint is added to the problem. Similarly, the optimal solution x^* (the so-called *scenario solution*), when it exists, remains optimal with high probability on a further "unseen" constraint. This is a *generalization* property in the learning-theoretic sense, since a solution based on a finite "training" batch of N sampled constraints has a high (depending on how large N is) probability of being feasible for a new generic random constraint. This fundamental property of scenario solutions has been seemingly pointed out for the first time in [9, 10], where bounds are derived on the number N of constraints that are needed for a scenario solution to achieve a desired a priori level of probabilistic feasibility.

Theoretical and practical interest toward RCPs stems from the fact that these problems are typically efficiently solvable while being closely related to important classes of "hard" optimization problems, such as robust convex programs, where the constraint $f(x, \delta) \leq 0$ is enforced for all admissible δ's, and to chance-constrained

programs (CCPs), where the constraint $f(x, \delta) \leq 0$ is enforced up to a given level of probability. RCPs also arise naturally in identification and prediction problems [7, 13], as well as in pattern analysis and data classification problems. RCPs have also been applied with success for approximating the solution to some hard problems arising in robust control; see [10].

The literature on RCPs is all quite recent: most of the relevant results can be found in the references [9, 10, 11, 15]. More specifically, in [10] an upper bound was first proved on the probability tail of $V_c(x^*)$, where $V_c(x)$ is the probability of constraint violation at x, that is, the probability of the event $\{\delta : f(x, \delta) > 0\}$. The main result of [10] states indeed that

$$\mathbb{P}^N\{V_c(x^*) > \epsilon\} \leq \binom{N}{d} (1 - \epsilon)^{N-d},$$

where d is the dimension of the decision variable x. Later, in [15], the authors refined this bound and actually proved that, for the restricted class of problems that *always* admits an optimal solution, the above probability is bounded by a binomial tail

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

with equality holding if the problem is fully supported (see section 2 for a definition of fully supported problems). The interest of these results lies in their potential generality: they hold for any type of convex program and irrespective of the probability distribution on δ. Further, this distribution need not be known by the user; all is needed for application of the scenario approach is the samples $\delta^{(i)}$ extracted from this underlying distribution. However, the previous bound was proved in [15] only under the assumption that every realization of the random program admits an optimal solution, which implies in particular that all realizations must be feasible (cf. Assumption 1 in [15]). This assumption is indeed quite restrictive, excluding, for instance, important classes of RCPs such as linear programs with Gaussian uncertainties. A hint of generalization, with no formal proof, is given in [15], but the generalized result claimed there appears to be incorrect; see section 3.5 for further discussion and a counterexample.

In this paper, we first develop a new theory for RCPs under no feasibility hypotheses. Then we extend this basic setup to the case when some of the N extracted constraints are purposely a posteriori violated, with the aim of improving the optimal objective value of the original RCP. Specifically, one selects an integer r and applies some optimal or suboptimal strategy in order to find a subset of the N sampled constraints of cardinality $N - r$ so that the objective of the optimization problem with this subset of constraints is significantly reduced. We call this modified class of problems *RCPs with violated constraints*, or RCPVs for short. RCPs with a posteriori violated constraints provide an effective tool for modulating the robustness of the solution (i.e., the violation probability) and the achievable objective level, and they have been studied seemingly for the first time in [12, 13], where they were introduced in the context of identification of predictor models from data; see also Remark 4.2. RCPVs are also related to CCPs. In particular, when a generic suboptimal strategy is applied for removing the constraints, the optimal RCPV objective gives, with arbitrarily high probability, an upper bound for the optimal objective of a corresponding CCP. When

an optimal strategy is applied for removing the constraints, the resulting RCPV precisely corresponds to a CCP in which the actual probability constraint is replaced by its empirical counterpart. In this case, the numerical solution of the RCPV may be numerically hard in general. However, we prove a theoretical result (Theorem 6.2) that provides an explicit assessment on "how close" the empirical problem is to the actual chance-constrained one in terms of the selected N and r.

Related results exist for specialized cases. For instance, in [19] the authors studied the case of constraint sampling in *linear* programs, whereas in [31] improved bounds on N are provided for the case when f is bi-affine in x and δ. In [30], Nemirovski and Shapiro also consider *analytical* approximations (as opposed to sampling approximations, which are used in the scenario approach) of a class of CCPs. An advantage of analytical approximation over scenario approximation is that the resulting approximated problem is a deterministic convex program. However, analytical approximations (such as the so-called Bernstein approximations proposed in [30]) currently appear to be possible only for problems with a special structure: where f is affine in δ and the components of δ are independent. Also, in [27] the authors propose a method of approximation of chance constraints based on replacing the true probability with its empirical counterpart, and then they provide results on the probability with which the empirical program yields upper and lower bounds on the original CCP. These bounds require introduction of some hypotheses such as finite cardinality of the search domain or uniform Lipschitz continuity of f with respect to x. An analogous idea, based on empirical means of binary functions, is used in [2], where probabilistic bounds on a so-called probability of one-sided constrained failure are derived using a statistical learning approach à la Vapnik [39]; the use of these bounds requires that the constraint function families have finite Vapnik–Chervonenkis (VC) dimension (a hypothesis which may not be fulfilled by generic RCPs), and in any case, a bound on this dimension needs be known in order to apply the results (see section 7 for a more in-depth discussion on the relation between the scenario approach and the VC theory).

Monte Carlo techniques for approximate solution of CCPs are also commonly studied in the stochastic programming literature; see, e.g., [37], although the bias in this context is toward the *asymptotic* behavior for $N \to \infty$, whereas the scenario theory is concerned with *finite* sample behavior.

The paper is organized as follows. In section 2 we introduce some basic notation and concepts (in particular, the notion of support constraints), which form the basis of the subsequent developments. RCPs are formally defined in section 3.1. Sections 3.2 and 3.3 contain two key results for RCPs (Theorem 3.3 and Corollary 3.4), derived under no feasibility assumptions. Section 4 then introduces the topic of RCPVs and provides a new result (in Theorem 4.1), giving an explicit upper bound on the violation probability for this class of problems. In section 5 both simplified and refined explicit lower bounds are derived on the minimum number N of samples required for an RCPV to attain the desired levels of probabilistic violation. Section 6 contains results highlighting the relations among RCPVs, CCPs, and empirical CCPs, and section 7 outlines a comparison between our results on RCPVs and similar ones that could be obtained by means of the VC theory. Finally, section 8 proposes an example of application of the RCPV theory to a classification context involving a classical geometric problem of circumscribing a set of points by a minimum radius circle. Some of the technical proofs are contained in Appendix A for better readability.

2. Definitions, assumptions, and preliminary facts. Consider a finite-dimensional convex optimization problem of the form

(2.1)
$$P[K] : \min_{x \in \Omega} \ c^\top x \text{ subject to}$$
$$f_j(x) \leq 0 \ \forall j \in K,$$

where x is the d-dimensional optimization variable, $\Omega \subset \mathbb{R}^d$ is a convex and compact domain, $c \neq 0$ is the objective direction, K is a *finite* set of indices, $|K| \geq d + 1$ (where $|\cdot|$ denotes the cardinality of a set), and $f_j(x) : \mathbb{R}^d \to \mathbb{R}$ are convex and lower semicontinuous (lsc) functions for each $j \in K$; each constraint thus defines a closed convex set $\{x : f_j(x) \leq 0\}$. With some abuse of notation, we shall use K to denote both the set of constraint indices as well as the actual set of constraints, depending on context. The shorthand notation $P[K, y_1, \ldots, y_n]$ is used to indicate a problem where constraints $y_1, \ldots, y_n$ are *added* to the constraint set K; i.e., $P[K, y_1, \ldots, y_n] = P[K \cup y_1 \cup \cdots \cup y_n]$. The notation $P[K \setminus y_1, \ldots, y_n]$ is used to indicate a problem where constraints $y_1, \ldots, y_n$ are *removed* from the constraint set K.

The feasible set of problem (2.1) is denoted by $\mathrm{Sat}\,(K)$: $\mathrm{Sat}\,(K) \doteq \{x \in \Omega : f_j(x) \leq 0 \ \forall j \in K\}$. We denote with $\mathrm{Obj}[K]$ the optimal objective value of problem $P[K]$ and with $\mathrm{Opx}[K]$ a corresponding optimal solution, when it exists. By convention, we set $\mathrm{Obj}[K] = +\infty$ if $P[K]$ is unfeasible. We postulate that $\mathrm{Opx}[K]$, when it exists, is unique. Uniqueness may be assumed essentially without loss of generality, since in case of multiple optimal solutions one may always introduce a suitable tie-breaking rule (for instance, select among optimal solutions the one having smallest entries in lexicographic order); see, e.g., Appendix A of [10]. Uniqueness is hence tacitly assumed to hold in the rest of this paper.

Let $K_1 \subseteq K$ and $y \in K$; then it is an intuitive fact that the optimal objective value is monotonically nondecreasing with the addiction of constraints; that is, $\mathrm{Obj}[K_1, y] \geq \mathrm{Obj}[K_1]$ (see Lemma 2.4 below for a formal statement). Also, from the uniqueness assumption, it follows that if $P[K_1]$ is feasible, then $\mathrm{Obj}[K_1, y] = \mathrm{Obj}[K_1]$ if and only if $P[K_1]$ and $P[K_1, y]$ have the same optimal solution.

DEFINITION 2.1 (support constraints). *A constraint $k \in K$ is a* support constraint *for problem $P[K]$ if $\mathrm{Obj}[K \setminus k] < \mathrm{Obj}[K]$. The set of support constraints for problem $P[K]$ is denoted by $\mathrm{Sc}\,(K) \subseteq K$.*

The next lemma is proved in [9].

LEMMA 2.2. *Any feasible problem $P[K]$ has at most d (the size of the decision variable x) support constraints; that is, $|\mathrm{Sc}\,(K)| \leq d$.*

However, it turns out that unfeasible problems may have more than d support constraints (it is, for instance, easy to devise an example of an unfeasible linear program in $\mathbb{R}^2$ having three support constraints). The next key lemma provides an extension of Lemma 2.2 to the case when $P[K]$ may be unfeasible; see section A.1 for its proof.

LEMMA 2.3. *Any problem $P[K]$ has at most $d + 1$ support constraints.*

Observe further that problem $P[K]$ satisfies the monotonicity and locality axioms defining the abstract class of so-called LP-type problems introduced in [38]. The following lemma holds; see section A.2 for its proof.

LEMMA 2.4 (structural properties of $P[K]$). *Let $K_1 \subseteq K_2 \subseteq K$. Then*
1. *(monotonicity) $\mathrm{Obj}[K_1] \leq \mathrm{Obj}[K_2]$;*
2. *(locality) if $\mathrm{Obj}[K_1] = \mathrm{Obj}[K_2]$ and $k \in K$, then $\mathrm{Obj}[K_2, k] > \mathrm{Obj}[K_2]$ if and only if $\mathrm{Obj}[K_1, k] > \mathrm{Obj}[K_1]$.*

Notice that a direct consequence of the locality property is that if $\mathrm{Obj}[K_1] = \mathrm{Obj}[K_2]$, then $K_1 \subseteq K_2$ implies $\mathrm{Sc}\,(K_1) \subseteq \mathrm{Sc}\,(K_2)$. The following definitions are useful.

DEFINITION 2.5 (fully supported problems). *Problem $P[K]$ is said to be* fully supported *if it possesses exactly d support constraints if it is feasible or $d+1$ support constraints if it is unfeasible.*

DEFINITION 2.6 (invariant constraint set). *A set of constraints $Y \subseteq K$ is an* invariant set *for problem $P[K]$ if $\mathrm{Obj}[Y] = \mathrm{Obj}[K]$.*

DEFINITION 2.7 (nondegenerate problems). *Problem $P[K]$ is said to be* nondegenerate *if $\mathrm{Obj}[K] = \mathrm{Obj}[\mathrm{Sc}\,(K)]$, that is, if the set of support constraints is invariant (note that $\mathrm{Sc}\,(K)$ may be empty).*

From the latter definition it follows that in nondegenerate problems the set of support constraints uniquely defines the optimal objective value. That is, solving the optimization problem with all constraints K in place yields the same optimal objective as if we solved the problem with only the support constraints in place. The following fact holds; see section A.3 for the proof of Lemma 2.8.

LEMMA 2.8. *If problem $P[K]$ is fully supported, then it is nondegenerate.*

Note that the converse of the previous proposition does not hold in general; that is, nondegenerate problems need not be fully supported.

DEFINITION 2.9 (essential constraint set). *An essential constraint set of problem $P[K]$ is a subset $\mathrm{Es}\,(K) \subseteq K$ such that*

$$\mathrm{Es}\,(K) = \arg\min\{|S|: \ \mathrm{Obj}[S] = \mathrm{Obj}[K], \quad S \subseteq K\}.$$

In other words, $\mathrm{Es}\,(K)$ is an invariant set of minimal cardinality (problem $P[K]$ may have more than one essential set).

The following results hold; see sections A.4, A.5, and A.6 for proofs of Lemmas 2.10, 2.11, and 2.12, respectively.

LEMMA 2.10. *Consider problem $P[K]$. Then*

$$(2.2) \qquad \mathrm{Sc}\,(K) \subseteq \mathrm{Es}\,(K),$$

$$(2.3) \qquad |\mathrm{Sc}\,(K)| \leq |\mathrm{Es}\,(K)| \leq d+1 \ (resp., \ d \ if \ P[K] \ is \ feasible).$$

LEMMA 2.11. *If $P[K]$ is nondegenerate, then it has a unique essential set: $\mathrm{Es}\,(K) = \mathrm{Sc}\,(K)$. Conversely, if $P[K]$ admits a unique essential set, then it is nondegenerate. Moreover, let $K_1 \subseteq K$: if $P[K]$ is nondegenerate, then $\mathrm{Obj}[K_1] = \mathrm{Obj}[K]$ if and only if $\mathrm{Es}\,(K) = \mathrm{Es}\,(K_1)$.*

LEMMA 2.12. *Let $Y \subseteq K$, let $J_Y^* = \mathrm{Obj}[Y]$, and let $h_1, \ldots, h_n \in K$ be additional constraints. Then*

$$\mathrm{Obj}[Y, h_j] = \mathrm{Obj}[Y], \ j = 1, \ldots, n \quad \Leftrightarrow \quad \mathrm{Obj}[Y, h_1, \ldots, h_n] = \mathrm{Obj}[Y].$$

2.1. Regularization. Let $P[K]$ be nondegenerate. This implies that the essential set of K is unique and coincides with the set of support constraints: $\mathrm{Es}\,(K) = \mathrm{Sc}\,(K)$. However, the cardinality of this set may well be smaller than $\zeta = d+1$ (for instance, whenever $P[K]$ is feasible, its essential set has cardinality no larger than d).

We next illustrate a procedure for constructing a problem $\check{P}[K]$ that is closely related to $P[K]$ and such that its essential set has cardinality exactly ζ. We call this procedure *regularization*, and $\tilde{P}[K]$ is a regularized version of $P[K]$.

Assume $|K| \geq \zeta$, and let $\bar{K} = K \setminus \mathrm{Es}\,(K)$, $\nu = \zeta - |\mathrm{Es}\,(K)|$ (ν is the essential cardinality drop). Define an arbitrary linear order on K (that is, put a unique numerical label on each constraint in K), and rank elements according to these labels.

Let $Z(K)$ be a list containing the ν largest elements in $\bar{K}$ (e.g., the ones with the ν largest labels). Introduce next an augmented objective,

$$\tilde{J}^*(K) = (\mathrm{Obj}[K], Z(K)),$$

and use this objective to rank the constraint sets according to a lexicographic criterion on the components of $\tilde{J}^*(\cdot)$. Let, for instance, $K_1 \subseteq K$. To check whether the clause $A = $ "$\tilde{J}^*(K) > \tilde{J}^*(K_1)$" is true, we first compare the first components (i.e., the $\mathrm{Obj}[\cdot]$ components): if $\mathrm{Obj}[K] > \mathrm{Obj}[K_1]$, then "A" is true. Otherwise (i.e., if the objectives are equal), the lexicographic order of the Z component in the augmented objective decides the comparison. That is, we set $\bar{K}_1 = K_1 \setminus \mathrm{Es}\,(K_1)$ and decide that "A" is true if the labels in $Z(K)$ are lexicographically larger than those in $Z(K_1)$. Regularization is thus nothing but a way for discriminating among sets of constraints that would otherwise yield the same objective value.

We define $\tilde{P}[K]$ as a version of $P[K]$ where comparisons on the objective values are resolved according to the $\tilde{J}^*$ criterion described above. All previous definitions (support constraints, essential sets, etc.) can be extended in an intuitive way to $\tilde{P}[K]$. It can be verified that the essential set of $\tilde{P}[K]$ is unique, and it is composed of the elements of $\mathrm{Es}\,(K)$ (the essential set of the original problem $P[K]$) plus the ν largest elements taken from $K \setminus \mathrm{Es}\,(K)$. A straightforward but important fact is that

$$(2.4) \qquad\qquad \mathrm{Obj}[K] > \mathrm{Obj}[K_1] \quad \Rightarrow \quad \tilde{J}^*(K) > \tilde{J}^*(K_1).$$

In the language of LP-type problems, $\tilde{P}[K]$ is a *refinement* of $P[K]$. The following result, originally stated in [29] for relating an LP-type problem to its refinements, can thus immediately be adapted to our context (recall $P[K]$ is LP-type); see also section 3 in [24].

LEMMA 2.13 (see [29, 24]). *If $P[K]$ is nondegenerate, then $\tilde{P}[K]$ is nondegenerate and its essential set has cardinality equal to ζ.*

It can be proved that if $P[K]$ is nondegenerate, then all results of section 2 also hold for the regularized version of $P[K]$, as formally stated next.

PROPOSITION 2.14 (extension to regularized objectives). *If $P[K]$ is nondegenerate, then Lemmas 2.3, 2.4, 2.11, and 2.12 hold also if the regularized objective $\tilde{J}(\cdot)$ is used in place of the standard objective $\mathrm{Obj}(\cdot)$.*

Full proofs are given in sections A.7 and A.8 for the extension to regularized objectives of Lemmas 2.4 and 2.12, respectively. The other proofs follow from similar arguments and are thus omitted.

3. Random convex problems and scenario solutions. Let $\delta \in \Delta$ denote a vector of random parameters, with $\Delta \subseteq \mathbb{R}^\ell$, and let $\mathbb{P}$ be a probability measure on Δ. Let $x \in \mathbb{R}^d$ be a design variable, and consider a family of functions $f(x, \delta) : (\mathbb{R}^d \times \Delta) \to \mathbb{R}$ that defines the design constraints. Specifically, for a given design vector x and realization δ of the uncertainty, the design constraints are satisfied if $f(x, \delta) \leq 0$. The following standing assumption is made on f for the rest of this paper.

Assumption 1 (convexity). $f(x, \delta) : (\mathbb{R}^d \times \Delta) \to \mathbb{R}$ is *convex* and lsc in x for any fixed $\delta \in \Delta$.

Define now

$$\omega \doteq (\delta^{(1)}, \ldots, \delta^{(N)}) \in \Delta^N,$$

where $\delta^{(i)} \in \Delta$, $i = 1, \ldots, N$, are independent random variables identically distributed according to $\mathbb{P}$ and where $\Delta^N = \Delta \times \Delta \cdots \Delta$ (N times). Let $\mathbb{P}^N$ denote the product

probability measure on Δ^N. To each $\delta^{(j)}$, we associate a constraint function

$$f_j(x) \doteq f(x, \delta^{(j)}), \quad j = 1, \ldots, N.$$

Therefore, to each randomly extracted ω there corresponds N random constraints $f_j(x)$, $j = 1, \ldots, N$.

3.1. The RCP. Given $\omega = (\delta^{(1)}, \ldots, \delta^{(N)}) \in \Delta^N$, we define the following convex optimization problem:

(3.1) $$P[\omega] : \min_{x \in \Omega} \; c^\top x \;\; \text{subject to}$$

(3.2) $$f_j(x) \leq 0, \; j = 1, \ldots, N,$$

where $f_j(x) = f(x, \delta^{(j)})$. It is clear that, for each random extraction of ω, problem (3.1) has the structure of the generic convex optimization problem (2.1) introduced in section 2. We denote with $J^* = J^*(\omega) = \mathrm{Obj}[\omega]$ the optimal objective value of $P[\omega]$, and with $x^* = x^*(\omega) = \mathrm{Opx}[\omega]$ the optimal solution of problem (3.1), when it exists. Problem (3.1) is here named an RCP, and the corresponding optimal solution x^* is named a *scenario solution*.

DEFINITION 3.1 (Helly's dimension). *Helly's dimension of $P[\omega]$ is the least integer ζ such that $\mathrm{ess\,sup}_{\omega \in \Delta^N} |\mathrm{Sc}\,(P[\omega])| \leq \zeta$ holds for any finite $N \geq 1$.*

In other words, ζ is the least integer such that $|\mathrm{Sc}\,(P[\omega])| \leq \zeta$ holds for almost all $\omega \in \Delta^N$ (i.e., possibly except for a set of $\mathbb{P}$-measure zero), for any $N \geq 1$. It follows from Lemmas 2.2 and 2.3 that if $P[\omega]$ is feasible with probability one, then $\zeta \leq d$, whereas $\zeta \leq d + 1$ in *all* cases. This is a fundamental fact which lies at the basis of all subsequent developments.

Remark 3.1 (on the generality of model (3.1)–(3.2)). Model (3.1)–(3.2) under consideration may seem specialized at first sight, but it actually encloses a quite general family of uncertain convex programs. First, we observe that problems with multiple uncertain (convex) constraints of the form

$$\min_{x \in \Omega} \; c^\top x \;\; \text{subject to}$$

$$f^{(1)}(x, \delta^{(j)}) \leq 0, \ldots, f^{(m)}(x, \delta^{(j)}) \leq 0, \qquad j = 1, \ldots, N,$$

can be readily cast in the form of (3.1)–(3.2) by condensing the multiple constraints into a single one, taking $f(x, \delta) \doteq \max_{i=1,\ldots,m} f^{(i)}(x, \delta)$. For instance, problems with multiple uncertain linear matrix inequality constraints of the form $F^{(i)}(x, \delta) \preceq 0$, where $F^{(i)}$ is a symmetric matrix affine in x and $\preceq$ means "negative semidefinite," fit in our framework by choosing $f^{(i)}(x, \delta) = \lambda_{\max}\{F^{(i)}(x, \delta)\}$, where $\lambda_{\max}$ denotes the maximum eigenvalue of its symmetric matrix argument.

Also, the case when the problem has an uncertain and nonlinear (but convex) objective function $g(x, \delta)$ can be fitted to the model under study at the price of adding one slack decision variable t and reformulating the problem with linear objective t and an additional constraint $g(x, \delta) - t \leq 0$.

Role of the Ω domain. The domain Ω for the decision variable can be used to model fixed deterministic (convex) constraints that do not depend on the uncertainty. If no such constraints are present, then Ω may just be, say, $\Omega = \{x : \|x\| \leq 10^{99}\}$. This compact domain is included in our model in order to avoid technical difficulties by guaranteeing that any feasible problem instance attains a solution. This can be done essentially with no loss of generality, since in any problem of practical interest there is a limit on how large a variable can grow.

The following assumptions are made on $P[\omega]$.

Assumption 2.
1. (Uniqueness) When problem $P[\omega]$ admits an optimal solution, this solution is unique.
2. (Nondegeneracy) Problem $P[\omega]$ is nondegenerate with probability one.

As we previously discussed in section 2, uniqueness of the solution can essentially be always obtained by imposing some suitable tie-breaking rule. Degeneracy is a singular situation arising when the essential set does not coincide with the set of support constraints. We are not ruling out this possibility, but we assume, for the sake of simplicity in the proofs, that the extractions of $\omega \in \Delta^N$ resulting in such pathological cases belong to a set of zero measure. This assumption can, however, be released in many cases; see the discussion in section 3.4.

The main focus of this paper is on the analysis of the a priori probability of objective violation of an RCP. As it will be made clear in the following, the optimal scenario objective has the property of remaining optimal, with high probability, if a further random constraint is added to the problem. Similarly, if the RCP has a solution, then this solution remains feasible (with high probability) on a further random constraint. That is, scenario solutions possess a *generalization property* in the learning-theoretic sense. More precisely, we study the probability with which the optimal objective $J^*(\omega)$ of an RCP with N constraints is no longer optimal when a new constraint is added to the problem, that is, the probability of the event $\{J^*(\omega, \delta) > J^*(\omega)\}$.

DEFINITION 3.2 (violation probability). *The violation probability of* $P[\omega]$ *is defined as*

$$V^*(\omega) \doteq \mathbb{P}\{\delta \in \Delta : \ J^*(\omega, \delta) > J^*(\omega)\}.$$

Note that for each random extraction of $\omega \in \Delta^N$, it corresponds to a value of V^*, which is therefore itself a random variable with values in $[0, 1]$. For a given $\epsilon \in (0, 1)$, let us define the "bad" event of having a violation larger than ϵ: $\mathcal{B} \doteq \{\omega \in \Delta^N : V^* > \epsilon\}$. We shall prove in the following section that it holds that $\mathbb{P}^N\{\mathcal{B}\} \leq \beta(\epsilon)$ for some explicitly given function $\beta(\epsilon)$ that goes to zero as N grows. In other words, if N is large enough, the scenario objective is a priori guaranteed with probability at least $1 - \beta(\epsilon)$ to have violation probability smaller than ϵ.

3.2. Probabilistic properties of RCPs. In this section we prove the following key result.

THEOREM 3.3. *Consider problem* (3.1) *with* $N \geq \zeta$ (ζ *is Helly's dimension). Let Assumptions* 1 *and* 2 *hold; then, for* $\epsilon \in (0, 1)$,

$$(3.3) \qquad \mathbb{P}^N\{\omega \in \Delta^N : V^*(\omega) > \epsilon\} \leq \Phi(\epsilon; \zeta - 1, N) \leq \Phi(\epsilon; d, N),$$

where

$$(3.4) \qquad \Phi(\epsilon; q, N) \doteq \sum_{j=0}^{q} \binom{N}{j} \epsilon^j (1 - \epsilon)^{N-j}$$

denotes the cumulative distribution of a binomial random variable: $\Phi(\epsilon; q, N) = \mathcal{I}(1 - \epsilon; N - q, q + 1) = 1 - \mathcal{I}(\epsilon; q + 1, N - q)$, *where* $\mathcal{I}(\epsilon; a, b)$ *is the* $\beta(a, b)$ *distribution (regularized incomplete beta function).*

3.2.1. Proof of Theorem 3.3. Every realization of problem $P[\omega]$ has the structure of the generic convex problem $P[K]$ introduced in section 2, where K represents the set of the N randomly extracted constraints. We let $\tilde{P}[\omega]$ denote a regularized

refinement of $P[\omega]$ such that $\tilde{P}[\omega]$ has, with probability one, an essential set with cardinality equal to Helly's dimension ζ, and let $\tilde{J}^*(\omega)$ denote the refined objective. This refinement can be obtained, as described in section 2.1, by associating to each δ a label obtained, for instance, by picking a real number uniformly at random from $[0, 1]$. When dealing with the refined problem, we thus understand that each extraction $\delta^{(i)}$ is actually augmented to represent both the uncertainty and its associated label, and the probabilities are to be intended accordingly. Let

$$\tilde{V}^*(\omega) \doteq \mathbb{P}\{\delta \in \Delta : \ \tilde{J}^*(\omega, \delta) > \tilde{J}^*(\omega)\}.$$

(Note that the ">" symbol in the previous expression is intended in the sense of the lexicographic order described in section 2.1.) We shall prove preliminarily that

$$(3.5) \qquad\qquad \mathbb{P}^N\{\omega \in \Delta^N : \tilde{V}^*(\omega) > \epsilon\} = \Phi(\epsilon; \zeta - 1, N).$$

To this end, start by noticing that, from the hypotheses, $P[\omega]$ is nondegenerate with probability one; therefore, by Lemma 2.13, with probability one its regularized version $\tilde{P}[\omega]$ is also nondegenerate and has a unique essential constraint set, which has cardinality exactly equal to ζ.

Let $I_\zeta^i(\omega)$, $i = 1, \ldots, C_{N,\zeta}$, with $C_{N,\zeta} = \binom{N}{\zeta}$, denote the subsets of ζ elements extracted from ω; say without loss of generality that $I_\zeta^1(\omega)$ is the set of the first ζ elements in ω: $I_\zeta^1(\omega) = \{\delta^{(1)}, \ldots, \delta^{(\zeta)}\}$. Consider the event where the essential set of $\tilde{P}[\omega]$ is $I_\zeta^i(\omega)$:

$$(3.6) \qquad S_i = \{\omega \in \Delta^N : \text{the essential set of } \tilde{P}[\omega] \text{ is } I_\zeta^i(\omega)\}, \quad i = 1, \ldots, C_{N,\zeta}.$$

Now, $\tilde{P}[\omega]$ has precisely one essential set (almost surely), and this set is of size ζ. Hence the events S_i are disjoint and exhaustive; i.e.,

$$(3.7) \qquad\qquad \Delta^N \doteq \bigcup_{i=1,\ldots,C_{N,\zeta}} S_i, \quad \mathbb{P}^N\{S_i \cap S_j\} = 0, \ i \neq j$$

(here $A \doteq B$ means that $A \setminus B$ has zero probability measure). Since the constraints are extracted independently at random and since the essential set does not depend on the order with which the constraints appear in $\tilde{P}[\omega]$, all S_i sets have the same probability (no I_ζ^i set is more likely than any other to be the essential set); hence, using (3.7),

$$1 = \mathbb{P}^N\{\Delta^N\} = \sum_{i=1}^{C_{N,\zeta}} \mathbb{P}^N\{S_i\} = C_{N,\zeta} \mathbb{P}^N\{S_1\},$$

whence

$$(3.8) \qquad\qquad \mathbb{P}^N\{S_i\} = C_{N,\zeta}^{-1} \quad \forall i = 1, \ldots, C_{N,\zeta}.$$

Define next, for $i = 1, \ldots, C_{N,\zeta}$, the following violation probabilities on the ith subset of constraints:

$$(3.9) \qquad\qquad \tilde{V}_i(\omega) \doteq \mathbb{P}\{\delta \in \Delta : \ \tilde{J}^*(I_\zeta^i(\omega), \delta) > \tilde{J}^*(I_\zeta^i(\omega))\}.$$

The $\tilde{V}_i$'s are random variables on $[0, 1]$—all with identical distribution (again, due to the fact that no constraint set is more likely than any other to be in I_ζ^i). Without loss

of generality we thus concentrate on $\tilde{V}_1$, that is, on the violation probability resulting from the extraction of the first ζ constraints $\omega_\zeta \doteq (\delta^{(1)}, \ldots, \delta^{(\zeta)})$, and we let the probability distribution of $\tilde{V}_1$ be

$$\tilde{F}_1(\alpha) \doteq \mathbb{P}^\zeta \{\omega_\zeta \in \Delta^\zeta : \tilde{V}_1 \leq \alpha\}, \quad \alpha \in [0,1].$$

Suppose now that the value of $\tilde{V}_1$ is given, say, $\tilde{V}_1 = v$. What would then be the probability of S_1? Recalling the definition (3.6), we have that

$$S_1 \Leftrightarrow \{\text{the essential set of } \tilde{P}[\omega] \text{ is } I_\zeta^1(\omega)\} \Leftrightarrow \{\tilde{J}^*(I_\zeta^1(\omega)) = \tilde{J}^*(I_\zeta^1(\omega), \delta^{(\zeta+1)}, \ldots, \delta^{(N)})\}.$$

Then, using Lemma 2.12 and Proposition 2.14, we would have that

$$(3.10) \qquad S_1 \Leftrightarrow \{\tilde{J}^*(I_\zeta^1(\omega)) = \tilde{J}^*(I_\zeta^1(\omega), \delta^{(\zeta+j)}), \; j = 1, \ldots, N - \zeta\}.$$

Since the δ extractions are independent, the probability of S_1 would then be the probability of realizing $\tilde{J}^*(I_\zeta^1(\omega)) = \tilde{J}^*(I_\zeta^1(\omega), \delta^{(i)})$ exactly $N - \zeta$ times in $N - \zeta$ independent extractions of $\delta^{(i)}$, where, from (3.9) and since $\tilde{V}_1 = v$, the probability of an individual success is $1 - v$. This probability would therefore be

$$\mathbb{P}^N \{S_1 | \tilde{V}_1 = v\} = (1 - v)^{N-\zeta}.$$

Now, deconditioning with respect to v, recalling that the distribution of $\tilde{V}_1$ is $\tilde{F}_1(\alpha)$ and using the continuous version of the total probability law, we write

$$\mathbb{P}^N \{S_1\} = \int_0^1 (1 - v)^{N-\zeta} \mathrm{d}\tilde{F}_1(v).$$

From (3.8) we thus obtain that

$$\int_0^1 (1 - v)^{N-\zeta} \mathrm{d}\tilde{F}_1(v) = \binom{N}{\zeta}^{-1}, \; N \geq \zeta.$$

It can be checked by solving an Euler integral that $\tilde{F}_1(\alpha) = \alpha^\zeta$ is a solution for this integral equation. Moreover, following a reasoning as in equation (9) in [15], we observe that this solution must be unique (a Hausdorff moment problem); therefore, we obtain that

$$(3.11) \qquad \tilde{F}_1(\alpha) = \alpha^\zeta.$$

Let us now consider the event in (3.5):

$$\tilde{B} \doteq \mathbb{P}^N \{\omega \in \Delta^N : \tilde{V}^*(\omega) > \epsilon\}.$$

Using (3.7), we have that

$$(3.12) \qquad \tilde{B} = \tilde{B} \cap \Delta^N = \bigcup_{i=1,\ldots,C_{N,\zeta}} \tilde{B} \cap S_i = \bigcup_{i=1,\ldots,C_{N,\zeta}} \tilde{B}_i \cap S_i,$$

where we define $\tilde{B}_i \doteq \mathbb{P}^N \{\omega \in \Delta^N : \tilde{V}_i(\omega) > \epsilon\}$. All probabilities $\mathbb{P}^N \{\tilde{B}_i \cap S_i\}$ are the same by the exchangeability of the measure; hence, we next evaluate $\mathbb{P}^N \{\tilde{B}_1 \cap S_1\}$. To this end, let us write formally

$$(3.13) \qquad \tilde{B}_1 = \bigcup_{\alpha \in (\epsilon, 1]} \{\omega \in \Delta^N : \tilde{V}_i(\omega) = \alpha\}.$$

Then

$$\mathbb{P}^N\{\tilde{B}_1 \cap S_1\} = \int_\epsilon^1 \mathbb{P}^N\{S_1 \cap \tilde{V}_1 = \alpha\}\mathrm{d}\alpha$$

$$= \int_\epsilon^1 \mathbb{P}^N\{S_1|\tilde{V}_1 = \alpha\}\mathrm{d}\tilde{F}_1(\alpha)$$

$$(\text{from } (3.11)) = \int_\epsilon^1 \mathbb{P}^N\{S_1|\tilde{V}_1 = \alpha\}\zeta\alpha^{\zeta-1}\mathrm{d}\alpha$$

(3.14) $$= \zeta \int_\epsilon^1 (1-\alpha)^{N-\zeta}\alpha^{\zeta-1}\mathrm{d}\alpha.$$

Let $B(x; a, b) \doteq \int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t$ be the incomplete beta function, let $B(a, b) \doteq B(1; a, b)$ be the beta function (Euler integral of the first kind), and recall that, for a, b integers,

$$B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}, \quad B(x; a, b) = B(a, b) \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} x^i(1-x)^{a+b-1-i}.$$

Then

(3.15) $$\int_\epsilon^1 \xi^{b-1}(1-\xi)^{a-1}\mathrm{d}\xi = B(1-\epsilon; a, b),$$

and (3.14) evaluates to

$$\mathbb{P}^N\{\tilde{B}_1 \cap S_1\} = \zeta B(1-\epsilon; N - \zeta + 1, \zeta)$$

$$= \frac{(N-\zeta)!\zeta!}{N!} \sum_{i=N-(\zeta-1)}^N \binom{N}{i} (1-\epsilon)^i \epsilon^{N-i}$$

$$= \binom{N}{\zeta}^{-1} \sum_{j=0}^{\zeta-1} \binom{N}{N-(\zeta-1)+j} (1-\epsilon)^{N-(\zeta-1)+j}\epsilon^{(\zeta-1)-j}$$

(3.16) $$= \binom{N}{\zeta}^{-1} \sum_{j=0}^{\zeta-1} \binom{N}{j} \epsilon^j(1-\epsilon)^{N-j} = \binom{N}{\zeta}^{-1} \Phi(\epsilon; \zeta - 1, N),$$

where $\Phi(\epsilon; \zeta - 1, N)$ is the cumulative distribution of a binomial random variable, that is, the probability of making no more than $\zeta - 1$ successes in N Bernoulli experiments having success probability ϵ. From (3.12) we thus obtain that

(3.17) $$\mathbb{P}^N\{\tilde{B}\} = \sum_{i=1}^{C_{N,\zeta}} \mathbb{P}^N\{\tilde{B}_i \cap S_i\} = C_{N,\zeta}\mathbb{P}^N\{\tilde{B}_1 \cap S_1\} = \Phi(\epsilon; \zeta - 1, N),$$

which proves (3.5). We are now ready to conclude our proof. Note from (2.4) that, for any δ, the following implication holds for the objectives of the original and of the refined problem:

$$J^*(\omega, \delta) > J^*(\omega) \quad \Rightarrow \quad \tilde{J}^*(\omega, \delta) > \tilde{J}^*(\omega).$$

Therefore, $V^* = \mathbb{P}\{J^*(\omega, \delta) > J^*(\omega)\} \leq \mathbb{P}\{\tilde{J}^*(\omega, \delta) > \tilde{J}^*(\omega)\} = \tilde{V}^*$, whence, for any ω, it holds that $V^* > \epsilon$ implies $\tilde{V}^* > \epsilon$; thus,

$$\mathbb{P}^N\{V^* > \epsilon\} \leq \mathbb{P}^N\{\tilde{V}^* > \epsilon\} = \Phi(\epsilon; \zeta - 1, N).$$

The proof is then concluded, observing that $\zeta \leq d + 1$ and that function Φ is nondecreasing in its second argument. $\quad\square$

3.3. The constraint violation probability. Whenever a solution $x^*(\omega)$ exists for $P[\omega]$, we can define the probability with which this solution is violated by a further randomly extracted constraint. More precisely, let

$$\Delta^{N*} \doteq \{\omega \in \Delta^N : \text{ the solution of } P[\omega] \text{ exists}\},$$

and define the constraint violation probability as

$$(3.18) \qquad V_c^*(\omega) = \begin{cases} \mathbb{P}\{\delta \in \Delta : f(x^*(\omega), \delta) > 0\} & \text{if } \omega \in \Delta^{N*}, \\ 1, & \text{otherwise.} \end{cases}$$

There is, of course, a close relationship between the constraint violation probability $V_c^*(\omega)$ and the objective violation probability $V^*(\omega)$. It can be readily inspected that

$$\omega \in \Delta^{N*} \quad \Rightarrow \quad \{f(x^*(\omega), \delta) > 0\} \Leftrightarrow \{J^*(\omega, \delta) > J^*(\omega)\};$$

therefore, $V^*(\omega)$ and $V_c^*(\omega)$ coincide on Δ^{N*}. Note further that $\omega \in \Delta^N \setminus \Delta^{N*}$ implies $V^*(\omega) = 0$; therefore, for $\epsilon \in (0, 1)$, $V^*(\omega) > \epsilon$ implies $\omega \in \Delta^{N*}$. Thus

$$\mathbb{P}^N\{\{V_c^*(\omega) > \epsilon\} \cap \Delta^{N*}\} = \mathbb{P}^N\{\{V^*(\omega) > \epsilon\} \cap \Delta^{N*}\} = \mathbb{P}^N\{V^*(\omega) > \epsilon\}$$
$$\leq \Phi(\epsilon; \zeta - 1, N),$$

where the last passage follows from Theorem 3.3. We thus proved the following corollary.

COROLLARY 3.4. *Consider problem* (3.1) *with* $N \geq \zeta$; *let Assumptions* 1 *and* 2 *hold, and let* $V_c^*(\omega)$ *be defined as in* (3.18). *Then, for* $\epsilon \in (0, 1)$,

$$(3.19) \quad \mathbb{P}^N\{\{V_c^*(\omega) > \epsilon\} \cap \Delta^{N*}\} = \mathbb{P}^N\{V^*(\omega) > \epsilon\} \leq \Phi(\epsilon; \zeta - 1, N) \leq \Phi(\epsilon; d, N).$$

3.4. Releasing the nondegeneracy assumption. Point 2 in Assumption 2 postulates that, with probability one, a realization of the RCP is nondegenerate. In practice, this assumption requires that the constraints are in "general position." It can be verified that this assumption, in particular, does not allow for discrete components (concentrated probability mass) in the distribution of δ. As a matter of fact, if one specific value of δ could be extracted with positive probability, then there would be a positive probability of extracting the very same constraint N times, which would result in a problem instance with an empty set of support constraints, thus in a degenerate instance. However, if the problem has (with probability one) a feasible region with nonempty interior, infinitesimal perturbation methods can be applied in order to obtain a nondegenerate refinement. For certain classes of geometric optimization problems, the technique of *simulation of simplicity* can also be used to remove nondegeneracies; see [22, 23]. For the important class of problems where the constraints are hyperplanes (linear programming), a specific perturbation technique for obtaining a nondegenerate refinement (also in the unfeasible cases) is described in detail by Matoušek; see Proposition 3.1 in [29]. A similar perturbation approach has been proposed in [15] for RCPs having feasible region with nonempty interior in any realization.

Summarizing, the nondegeneracy assumption has been here retained for the purpose of streamlining the proofs. Whenever a nondegenerate refinement can be constructed for $P[\omega]$ without increasing Helly's dimension (e.g., when the feasible set of $P[\omega]$ has nonempty interior in any instance or in all cases when the random constraints are hyperplanes), the nondegeneracy assumption can actually be released, and the results in Theorem 3.3 and Corollary 3.4 would hold unchanged.

3.5. Comments on related results. Campi and Garatti in [15] focus on the constraint violation probability V_c^* for a restricted class of RCPs where a hypothesis of existence holds for the solution x^* in any problem instance. Their main result (Theorem 1 in [15]) is to be compared with the result in Corollary 3.4 in the present paper. To put ourselves in the same setup, we shall thus assume that $\Delta^N = \Delta^{N*}$. An important fact that one needs to realize is that if the problem class is a priori restricted so that any problem instance is feasible, then Helly's dimension is bounded by d, whereas if one allows for unfeasible instances, the bound is $d+1$; see Lemmas 2.2 and 2.3. In the restricted case, the results in Theorem 3.3 and Corollary 3.4 hold, substituting $d-1$ in place of d in all the statements. Our result in (3.19) thus precisely coincides with Theorem 1 in [15] in the restricted class of feasible problems (note also that all inequalities in (3.19) actually become equalities if the $P[\omega]$ is further restricted to being fully supported, since no regularization is to be applied in such a case).

It appears, however, that in [15] the authors did not take into account the fact that Helly's dimension bound increases if unfeasible instances are allowed. Indeed, in section 2.1, point 5, page 1217 of [15], the authors claim that the existence assumption (Assumption 1 in [15]) can be released, and more precisely, they claim that

$$(3.20) \qquad \mathbb{P}^N\{\Delta^{N*} \cap V_c^* > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

but provide no proof of this statement in their paper. Therefore, according to [15], the probability $\mathbb{P}^N\{\Delta^{N*} \cap V_c^* > \epsilon\}$ is upper bounded by $\Phi(\epsilon; d-1, N)$, whereas according to our (3.19), the same quantity is upper bounded by $\Phi(\epsilon; d, N)$. However, the statement (3.20) reported in [15] is not correct in general, as is shown by the following simple counterexample.

Consider an RCP with $x \in \mathbb{R}$ (thus $d = 1$), $c = 1$, and

$$f(x, \delta) = \begin{cases} -x + 1 & \text{if } \delta = 1, \\ \max(x+1, -x-2) & \text{if } \delta = 0, \end{cases}$$

where δ is a Bernoulli variable taking value one with probability 0.5 and value zero with probability 0.5. One may assume $\Omega = \{x : |x| \leq 10^{99}\}$, but this is irrelevant for the purposes of the example. For any given value of δ, function $f(x, \delta)$ is convex in x. The random constraint $f(x, \delta) \leq 0$ simply describes the set $\Xi_1 = \{x \geq 1\}$ when $\delta = 1$ and the set $\Xi_0 = \{x \in [-2, -1]\}$ when $\delta = 0$.

Consider $N = 2$, that is, two independent extractions of the constraints. With probability 0.5 both constraints are extracted, and the problem instance is unfeasible. If Ξ_1 is extracted twice (which happens with probability $1/4$), then the optimal solution is $x^* = x_1^* = 1$. If Ξ_0 is extracted twice (which happens with probability $1/4$), then the optimal solution is $x^* = x_0^* = -2$. We thus have that $\mathbb{P}^N\{\Delta^{N*}\} = 0.5$. Whenever a solution x^* exists, the probability of violating the constraint on a further extraction is clearly 0.5; that is, $\mathbb{P}\{\delta : f(x^*, \delta) > 0\} = 0.5$, which means that the constraint violation is $V_c^* = 0.5$. Thus if $\epsilon < 0.5$, the event $\{V_c^* > \epsilon | \Delta^{N*}\}$ is the certain event. Then

$$\mathbb{P}^N\{\Delta^{N*} \cap V_c^* > \epsilon\} = \mathbb{P}^N\{V_c^* > \epsilon|\Delta^{N*}\}\mathbb{P}\{\Delta^{N*}\} = 0.5 \cdot \mathbb{P}^N\{V_c^* > \epsilon|\Delta^{N*}\}$$

$$= \begin{cases} 0.5 & \text{if } \epsilon < 0.5, \\ 0 & \text{if } \epsilon \geq 0.5. \end{cases}$$

Now (3.20) would prescribe that the previous probability is upper bounded by $(1-\epsilon)^2$, which is clearly not the case; see Figure 3.1.
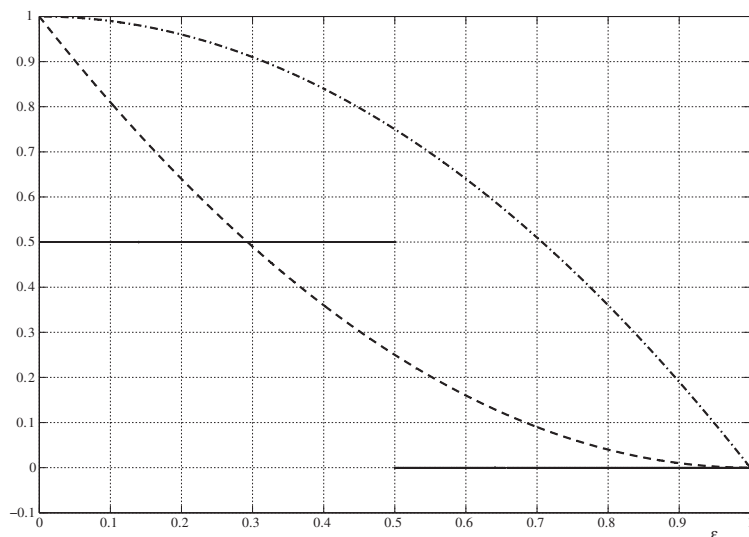
FIG. 3.1. *Solid line is $\mathbb{P}^N\{\Delta^{N*} \cap V_c^* > \epsilon\}$; dashed line is (3.20); dash-dotted line is the upper bound in (3.19).*

4. Random convex programs with violated constraints.

In this section we consider an important generalization of the basic problem introduced in section 3. Namely, we consider a situation where, given a sampled batch of N constraints, an "algorithm" takes this batch as input and returns an objective value J^*, which is optimal on a selection of $N - r$ of these constraints and which is violated on the remaining r constraints. We call this extended class of problems RCPVs. The reason for considering such a generalization resides in the fact that violating a small number r of the sampled constraints may drastically improve the optimal objective value while maintaining an acceptable level of violation probability. We start by assuming that a generic procedure or "rule" is assigned that behaves as follows.

PROCEDURE 4.1. *Given* $\omega = (\delta^{(1)}, \dots, \delta^{(N)}) \in \Delta^N$ *and* $r \le N - \zeta$ *(ζ is Helly's dimension of $P[\omega]$), the procedure returns a partition* $\{\bar{R}_r, R_r\}$ *of* $\{1, \dots, N\}$,

$$R_r = \{i_1, \dots, i_r\} \subseteq \{1, \dots, N\}, \quad \bar{R}_r = \{1, \dots, N\} \setminus R_r,$$

where $J^* \doteq \mathrm{Obj}[\bar{R}_r(\omega)]$ *is such that, with probability one,* $J^* < \mathrm{Obj}[\bar{R}_r(\omega), k] \; \forall k \in R_r(\omega)$. *The procedure has the only restriction of being permutation invariant, and it is otherwise completely* arbitrary. *By permutation invariant we mean that*

$$R_r(\pi(\omega)) = \pi(R_r(\omega)) \quad \text{for any permutation } \pi \text{ of } \{1, \dots, N\}.$$

Examples of constraint removal procedures are described in section 5.1. The following key theorem holds.

THEOREM 4.1. *Let* ζ *be Helly's dimension of* $P[\omega]$ *(recall* $\zeta \le d$ *if almost all realizations of* $P[\omega]$ *are known a priori to be feasible or* $\zeta \le d+1$ *in the general case). Consider Procedure 4.1 with* $N - r \ge \zeta$, *let Assumption 1 hold, and let Assumption 2 hold for problem* $P[\bar{R}_r(\omega)]$. *Define*

$$(4.1) \qquad V^*(\omega) \doteq \mathbb{P}\{\delta \in \Delta : \mathrm{Obj}[\bar{R}_r(\omega), \delta] > \mathrm{Obj}[\bar{R}_r(\omega)]\}.$$

Then, for $\epsilon \in (0,1)$,

$$(4.2) \qquad \mathbb{P}^N\{\omega \in \Delta^N : V^*(\omega) > \epsilon\} \leq \beta_{r,\zeta}(\epsilon) \leq \beta_{r,d+1}(\epsilon),$$

where

$$(4.3) \qquad \beta_{r,\zeta}(\epsilon) \doteq \binom{r+\zeta-1}{r} \Phi(\epsilon; r+\zeta-1, N)$$

with Φ being the cumulative binomial distribution (beta distribution) defined in (3.4).

4.1. Preliminaries and proof of Theorem 4.1. We start in a way similar to section 3.2.1: every instance of problem $P[\bar{R}_r(\omega)]$ has the structure of the generic convex problem $P[K]$ introduced in section 2, where K here represents the set of the $N - r$ constraints selected by the procedure among the N constraints extracted at random. We let $\tilde{P}$ denote a regularized refinement of P that brings the essential sets to have cardinality ζ with probability one, obtained as described in section 2.1 and in the beginning of section 3.2.1, and let $\tilde{J}^*(\bar{R}_r(\omega))$ denote the refined objective. Start by noticing that, from the hypotheses, $P[\bar{R}_r(\omega)]$ is nondegenerate with probability one; therefore, by Lemma 2.13, with probability one, its regularized version $\tilde{P}$ is also nondegenerate and has a unique essential constraint set, which has size exactly ζ. Then let $I_\zeta^i(\omega)$, $i = 1, \ldots, C_{N,\zeta}$, be the subsets of ω containing ζ elements, as defined in section 3.2.1, and consider the following event where the essential set of $\tilde{P}[\bar{R}_r(\omega)]$ is $I_\zeta^i(\omega)$:

$$S_i \doteq \{\omega \in \Delta^N : \text{the essential set of } \tilde{P}[\bar{R}_r(\omega)] \text{ is } I_\zeta^i(\omega)\}, \quad i = 1, \ldots, C_{N,\zeta}.$$

The same reasoning previously done below (3.6) leads to

$$(4.4) \qquad \Delta^N = \bigcup_{i=1,\ldots,C_{N,\zeta}} S_i, \quad \mathbb{P}^N\{S_i \cap S_j\} = 0,\, i \neq j.$$

Further, let $I_{N-r}^i(\omega)$, $i = 1, \ldots, C_{N,r}$, be the subsets of ω containing $N - r$ elements, and observe that, since the constraints are extracted independently at random and since the rule for selection of the subset $\bar{R}_r(\omega)$ is permutation invariant, this rule selects any of the I_{N-r}^i subsets with equal probability. Moreover, one of these subsets is selected with probability one; therefore, defining the events

$$\mathcal{R}_j \doteq \{\omega \in \Delta^N : \bar{R}_r(\omega) = I_{N-r}^j(\omega)\}, \quad i = 1, \ldots, C_{N,r},$$

we have that

$$(4.5) \qquad \Delta^N = \bigcup_{j=1,\ldots,C_{N,r}} \mathcal{R}_j, \quad \mathbb{P}^N\{\mathcal{R}_j \cap \mathcal{R}_k\} = 0,\, k \neq j,$$

and

$$\mathbb{P}^N\{\mathcal{R}_j\} = C_{N,r}^{-1}, \quad j = 1, \ldots, C_{N,r}.$$

Note that the overall mechanism leading to the essential set being I_ζ^i is permutation invariant: First, a subset $\bar{R}_r(\omega)$ is selected by the procedure, and this selection is permutation invariant by hypothesis. Then one subset of ζ of the selected constraints will happen to be the essential set of $P[\bar{R}_r(\omega)]$, and this event also does not depend

on the constraints order; therefore, the overall choice of I_ζ^i is permutation invariant. Therefore,

$$\mathbb{P}^N\{S_i\} = C_{N,\zeta}^{-1}, \quad j = 1, \ldots, C_{N,\zeta}.$$

Consider then the $C_{N,r}$ sets I_{N-r}^j (i.e., the possible outcomes of $\bar{R}_r$). We denote with Υ_ζ^i the subset of indices such that $j \in \Upsilon_\zeta^i$ implies $I_\zeta^i \subseteq I_{N-r}^j$. That is, Υ_ζ^i contains the indices of the I_{N-r}^j sets that admit I_ζ^i as a subset; in particular, Υ_ζ^1 are the indices of the sets containing $I_\zeta^1 = \{1, \ldots, \zeta\}$. Notice that, by a simple counting argument, the cardinality of Υ_ζ^i is

(4.6) $$|\Upsilon_\zeta^i| = C_{N-\zeta,r} \doteq \binom{N-\zeta}{r}.$$

4.1.1. Proof of Theorem 4.1. Let

$$\tilde{V}^*(\omega) \doteq \mathbb{P}\{\delta \in \Delta : \tilde{J}^*(\bar{R}_r(\omega), \delta) > \tilde{J}^*(\bar{R}_r(\omega))\},$$

and define, for $i = 1, \ldots, C_{N,\zeta}$, the violation probabilities on the ith subset of constraints as follows:

$$\tilde{V}_i(\omega) \doteq \mathbb{P}\{\delta \in \Delta : \tilde{J}^*(I_\zeta^i(\omega), \delta) > \tilde{J}^*(I_\zeta^i(\omega))\}.$$

The $\tilde{V}_i$'s are random variables on $[0, 1]$, all with identical distribution (again, due to the fact that the problem is invariant to permutations of the order of the constraints). Without loss of generality we thus concentrate on $\tilde{V}_1$ (that is, the violation probability resulting from the first ζ constraints $\omega_\zeta \doteq (\delta^{(1)}, \ldots, \delta^{(\zeta)})$), and we let the probability distribution of $\tilde{V}_1$ be

$$\tilde{F}_1(\alpha) \doteq \mathbb{P}^\zeta\{\omega_\zeta \in \Delta^\zeta : \tilde{V}_1 \leq \alpha\}, \quad \alpha \in [0, 1].$$

From (3.11) we already know that $\tilde{F}_1(\alpha) = \alpha^\zeta$. Let us now consider the event

$$\tilde{B} \doteq \mathbb{P}^N\{\omega \in \Delta^N : \tilde{V}^*(\omega) > \epsilon\}.$$

Using (4.4), we have that

(4.7) $$\tilde{B} = \tilde{B} \cap \Delta^N = \bigcup_{i=1,\ldots,C_{N,\zeta}} \tilde{B} \cap S_i = \bigcup_{i=1,\ldots,C_{N,\zeta}} \tilde{B}_i \cap S_i,$$

where we defined

$$\tilde{B}_i \doteq \mathbb{P}^N\{\omega \in \Delta^N : \tilde{V}_i(\omega) > \epsilon\}.$$

All probabilities $\mathbb{P}^N\{\tilde{B}_i \cap S_i\}$ are the same, by the exchangeability of the measure; hence, we next evaluate $\mathbb{P}^N\{\tilde{B}_1 \cap S_1\}$. Using (4.5), we have that

$$\tilde{B}_1 \cap S_1 = \tilde{B}_1 \cap S_1 \cap \Delta^N = \bigcup_{j=1,\ldots,C_{N,r}} S_1 \cap \mathcal{R}_j \cap \tilde{B}_1,$$

and then

$$\mathbb{P}_m^N\{\tilde{B}_1 \cap S_1\} = \sum_{j=1}^{C_{N,r}} \mathbb{P}^N\{\mathcal{S}_1 \cap \mathcal{R}_j \cap \tilde{B}_1\}$$

$$(\text{using (3.13)}) = \sum_{j=1}^{C_{N,r}} \int_\epsilon^1 \mathbb{P}^N\{S_1 \cap \mathcal{R}_j \cap \tilde{V}_1 = \alpha\} d\alpha$$

$$= \sum_{j=1}^{C_{N,r}} \int_\epsilon^1 \mathbb{P}^N\{S_1 \cap \mathcal{R}_j | \tilde{V}_1 = \alpha\} d\tilde{F}_1(\alpha)$$

$$= \sum_{j=1}^{C_{N,r}} \int_\epsilon^1 \mathbb{P}^N\{\mathcal{S}_1 | \mathcal{R}_j, \tilde{V}_1 = \alpha\} \mathbb{P}^N\{\mathcal{R}_j | \tilde{V}_1 = \alpha\} d\tilde{F}_1(\alpha)$$

$$\leq \sum_{j=1}^{C_{N,r}} \int_\epsilon^1 \mathbb{P}^N\{S_1 | \mathcal{R}_j, \tilde{V}_1 = \alpha\} d\tilde{F}_1(\alpha)$$

$$(\text{from (3.11)}) = \sum_{j=1}^{C_{N,r}} \int_\epsilon^1 \mathbb{P}^N\{S_1 | \mathcal{R}_j, \tilde{V}_1 = \alpha\} \zeta \alpha^{\zeta-1} d\alpha$$

$$(4.8) \qquad = \sum_{j \in \Upsilon_\zeta^1} \int_\epsilon^1 \mathbb{P}^N\{S_1 | \mathcal{R}_j, \tilde{V}_1 = \alpha\} \zeta \alpha^{\zeta-1} d\alpha.$$

The last passage follows from the fact that the probability appearing in the integral represents the probability with which $J^*(I_\zeta^1(\omega))$ (that is, the optimal objective constructed on I_ζ^1) is optimal on the $N - r$ constraints in $\bar{R}_r = I_{N-r}^j$ and violates the r constraints in R_r, given that the violation probability of $J^*(I_\zeta^1(\omega))$ is α. Notice that if I_ζ^1 is not included in I_{N-r}^j, then this probability is zero, since at least one constraint in I_ζ^1 (which is by definition not violated by $J^*(I_\zeta^1(\omega))$) will belong to R_r. Thus, the elements in the sum are nonzero only for $j \in \Upsilon_\zeta^1$. Also, for $j \in \Upsilon_\zeta^1$, the objective $J^*(I_\zeta^1(\omega))$ certainly does not violate ζ constraints in $\bar{R}_r = I_{N-r}^j$; therefore, since the constraint extractions are independent, the probability inside the integral in (4.8) is equal to $(1 - \alpha)^{N-r-\zeta} \alpha^r$, whence

$$\mathbb{P}^N\{\tilde{B}_1 \cap S_1\} \leq \sum_{j \in \Upsilon_\zeta^1} \int_\epsilon^1 (1 - \alpha)^{N-r-\zeta} \alpha^r \zeta \alpha^{\zeta-1} d\alpha$$

$$(\text{due to (4.6) and (3.15)}) = C_{N-\zeta,r} \frac{\zeta(N - r - \zeta)!(r + \zeta - 1)!}{N!} \Phi(\epsilon; r + \zeta - 1, N),$$

and then, from (4.7),

$$\mathbb{P}^N\{\tilde{B}\} \leq C_{N,\zeta} C_{N-\zeta,r} \zeta \frac{(N - r - \zeta)!(r + \zeta - 1)!}{N!} \Phi(\epsilon; r + \zeta - 1, N),$$

$$(4.9) \qquad = \binom{r + \zeta - 1}{r} \Phi(\epsilon; r + \zeta - 1, N).$$

The proof is then concluded by observing that the following relation holds between the violation of the original problem and the one of its refined version: $\mathbb{P}^N\{V^* > \epsilon\} \leq$

$\mathbb{P}^N\{\tilde{V}^* > \epsilon\}$; see the passages below (3.17). Finally, observe that $\zeta \leq d + 1$ and the right-hand side of the last equation is nondecreasing in ζ. □

The next corollary can then be easily obtained by following the same lines as in the proof of Corollary 3.4.

COROLLARY 4.2 (on the constraint violation probability). *Under the setup and assumptions of Theorem* 4.1*, let*

$$(4.10) \qquad \Delta^{(N-r)*} \doteq \{\omega \in \Delta^N : \text{ the solution of } P[\bar{R}_r(\omega)] \text{ exists}\},$$

and define the constraint violation probability

$$(4.11) \qquad V_c^*(\omega) \doteq \begin{cases} \mathbb{P}\{\delta \in \Delta : f(\text{Opx}[\bar{R}_r(\omega)], \delta) > 0\} & \text{if } \omega \in \Delta^{(N-r)*}, \\ 1, & \text{otherwise.} \end{cases}$$

Then

$$(4.12) \qquad \mathbb{P}^N\{\omega \in \Delta^N : \{V_c^*(\omega) > \epsilon\} \cap \Delta^{(N-r)*}\} \leq \beta_{r,\zeta}(\epsilon) \leq \beta_{r,d+1}(\epsilon),$$

where $\beta_{r,\zeta}(\epsilon)$ *is given in* (4.3).

Remark 4.1 (generalizations). The nondegeneracy assumption can be released in all cases mentioned in section 3.4. The details of this extension are left to the reader; they can be obtained following the ideas outlined in section 3.4.

Remark 4.2 (related results). RCPs with a posteriori discarded constraints have been studied seemingly for the first time in [12, 13], where they were introduced in the context of a problem of identification of predictor models from data; see also [7] for new results on linear interval predictor models developed along the lines described in the present paper. The result in Theorem 3 in [13] establishes precisely the following upper bound on the probability of $\{V_c^* > \epsilon\}$ for an RCP with r discarded constraints (under the assumption of existence of solution in all realizations; thus, $\zeta \leq d$): $\mathbb{P}^N\{V_c^* > \epsilon\} \leq \tilde{\beta}_{r,d}(\epsilon)$ with

$$(4.13) \qquad \tilde{\beta}_{r,d}(\epsilon) \doteq \min\left(\binom{N}{d}\sum_{i=0}^{r}\binom{N-d}{i}\epsilon^i(1-\epsilon)^{N-d-i}, 1\right).$$

It is worth comparing the above bound with the new result given in Theorem 4.1 in the present paper. Figure 4.1 shows that $\beta_{r,d}(\epsilon)$ in (4.12) is typically orders of magnitude better (smaller) than the older bound $\tilde{\beta}_{r,d}(\epsilon)$.

After the first version of this manuscript was submitted, I was advised that results similar to the ones reported in section 4 of this paper were independently derived by Campi and Garatti and appear in an unpublished document currently available in an open repository [14]. I was unaware of the existence of this reference, but I gratefully acknowledge it. Corollary 4.2 in the present paper is similar to Theorem 1 in [14], although it is obtained via a different line of proof and under a different set of assumptions (compare Assumption 1 in [14] and Assumption 2 in the present paper), and it yields a different bound in the general case of possibly unfeasible problem realizations. Also, the explicit sample complexity bounds reported in section 5 of the present paper are derived using the Chernoff bounding ideas previously used in [8] and improve upon the result reported in section 3.4 of [14].

It is worth noting that Assumption 1 in [14] is so demanding that results based on this assumption are inapplicable in all but some quite specialized problem classes. In particular, all results in [14] hinge on the hypothesis not only that every realization
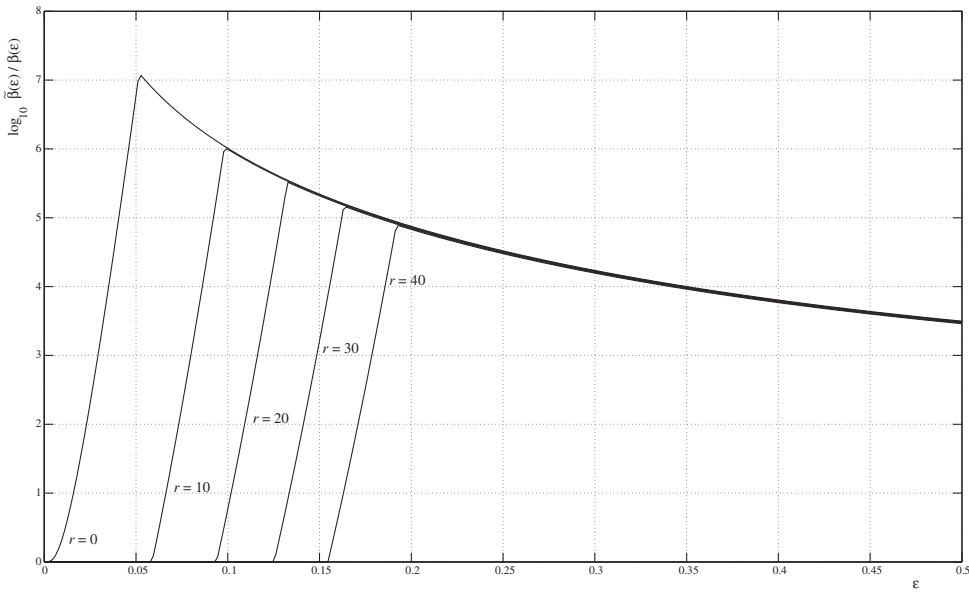
FIG. 4.1. *Logarithmic plot of the ratio between* $\tilde{\beta}_{r,d}(\epsilon)$ *in (4.13) and* $\beta_{r,d}(\epsilon)$ *in (4.3), for* $N = 500$, $\zeta \leq d = 5$, *and* $r = \{0, 10, 20, 30, 40\}$.

of the RCP with fixed N is feasible and attains an optimal solution but also that all possible realizations, with any $N \geq d$, admit a solution. To better appreciate this matter, consider a simple random linear programming problem in dimension $d = 2$ with, say, $N = 5$. Suppose that a realization of the random constraints is the one depicted in Figure 4.2. Clearly, such an instance admits a solution. However, the results in [14] are not applicable, since not all subproblems retaining a subset of two constraints admit a solution (these problems are all feasible, but for some of them the solution drifts to infinity). A similar situation arises with respect to feasibility, even if the problem is compactified so that every feasible instance admits a solution. Again, Assumption 1 in [14] requires that all possible realizations of N constraints lead to a nonempty feasible set, a condition that is usually not fulfilled when the support
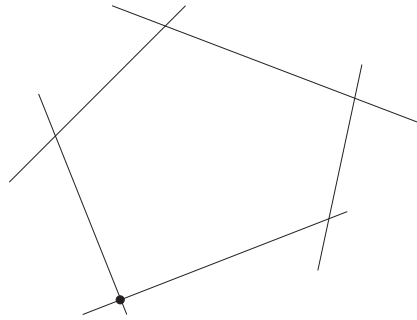

FIG. 4.2. *An LP-type example; objective direction is downward.*

of the distribution of δ is unbounded (e.g., in classical uncertain linear programming problems with Gaussian distribution on δ).

5. Explicit bounds on the sample complexity. In the phase of designing the optimization experiment, it is often useful to fix a priori the removal cardinality r, the violation probability level ϵ, and a level $\beta \in (0,1)$ and to determine a value of N which guarantees that $\mathbb{P}^N\{V^* > \epsilon\} \le \beta$. The resulting value of N is a lower bound on the number of constraints that need be considered in the random optimization problem in order for the resulting scenario objective to possess the desired properties in terms of violation probability. This number is usually referred to as the "sample complexity" of the RCPV. Clearly, a value of N that guarantees $\mathbb{P}^N\{V^* > \epsilon\} \le \beta$ can always be obtained by numerical "inversion" of the upper bound in (4.2). The same numerical approach can actually be used also to solve for other quantities (the remaining ones being fixed to given values): for instance, one may fix ϵ, β, and N, and search for the maximum r such that $\mathbb{P}^N\{V^* > \epsilon\} \le \beta$ holds. It is, however, interesting to have an idea of how N grows as a function of the problem parameters, that is, to have an analytical assessment of the sample complexity of the scenario approach. To this end, we next provide a simple lower bound on N.

COROLLARY 5.1. *Given* $\epsilon \in (0,1)$, $\beta \in (0,1)$, *and integer* $r \le N - \zeta$, *let the hypotheses of Theorem 4.1 be satisfied, and let* V^* *be defined as in (4.1). If* N *is an integer such that*

$$(5.1) \qquad N \ge \frac{2}{\epsilon} \ln \beta^{-1} + \frac{4}{\epsilon}(r + \zeta - 1), \quad r > 0,$$

then it holds that $\mathbb{P}^N\{V^* > \epsilon\} \le \beta$.

For $r = 0$, *the lower bound simplifies to*

$$(5.2) \qquad N \ge \frac{2}{\epsilon}\left(\ln \beta^{-1} + \zeta - 1\right) \quad for \ r = 0.$$

Proof. Recall from Theorem 4.1 that probability of V^* exceeding ϵ is upper bounded by

$$(5.3) \qquad \beta_{r,\zeta}(\epsilon) = \binom{r + \zeta - 1}{r} \Phi(\epsilon; r + \zeta - 1, N).$$

We use a technique similar to that introduced in [8]: the classical Chernoff's inequality [17] for the binomial tail yields the bound

$$\Phi(\epsilon; s, N) \le \exp\left(-\frac{(N\epsilon - s)^2}{2N\epsilon}\right) \quad \text{for } N\epsilon \ge s,$$

whereas the following well-known bound holds for the binomial coefficient

$$(5.4) \qquad \binom{r + \zeta - 1}{r} \le \left(\frac{\mathrm{e}(r + \zeta - 1)}{r}\right)^r, \quad r > 0.$$

We therefore have that, for $r > 0$,

$$\mathbb{P}^N\{V^* > \epsilon\} \le \binom{r + \zeta - 1}{r} \Phi(\epsilon; r + \zeta - 1, N)$$

$$\le \frac{\mathrm{e}^r(r + \zeta - 1)^r}{r^r} \exp\left(-\frac{(N\epsilon - \zeta - r + 1)^2}{2N\epsilon}\right) \quad \text{for } N\epsilon \ge r + \zeta - 1.$$

Now let $\beta < 1$ be given as follows: the latter expression is upper bounded by β if and only if

(5.5)
$$\frac{(N\epsilon - \zeta - r + 1)^2}{2N\epsilon} \geq \ln \beta^{-1} + r \ln \frac{e(\zeta + r - 1)}{r}.$$

This latter expression is a second-order condition on N, which can be easily solved analytically in order to provide an explicit lower bound on N; see Remark 5.1. We prefer, however, to further simplify the expression, with the purpose of obtaining a very simple, albeit more relaxed, form for this lower bound. First, observe that $r \ln \frac{e(\zeta + r - 1)}{r} = r + r \ln(1 + (\zeta - 1)/r) \leq r + (\zeta - 1)$. Then we have that (5.5) $\Leftarrow$ $1/2N\epsilon - (\zeta + r - 1) \geq \ln \beta^{-1} + r \ln \frac{e(\zeta + r - 1)}{r} \Leftarrow 1/2N\epsilon - (\zeta + r - 1) \geq \ln \beta^{-1} + \zeta + r - 1$, which is equivalent to (5.1).

The simplified bound (5.2) valid for $r = 0$ is finally obtained through the very same reasoning as above, considering that the binomial coefficient appearing in (5.3) is equal to one for $r = 0$.   ☐

Remark 5.1 (refined lower bounds on N). Two improved versions of bound (5.1) can readily be obtained as follows: notice preliminarily that $\binom{r+\zeta-1}{r} = \binom{r+\zeta-1}{\zeta-1}$; hence, two different bounds could be used to limit this binomial coefficient from above. The first is the one already used in (5.4), and the second is

(5.6)
$$\binom{r + \zeta - 1}{\zeta - 1} \leq \left(\frac{e(r + \zeta - 1)}{\zeta - 1} \right)^{\zeta - 1}.$$

It can be verified that $(\frac{e(r+\zeta-1)}{\zeta-1})^{\zeta-1} \geq (\frac{e(r+\zeta-1)}{r})^r$, whenever $\zeta - 1 \geq r$; therefore, bound (5.4) is tighter than bound (5.6) when $r \leq \zeta - 1$ and vice versa for $r > \zeta - 1$.

The first refined lower bound on N is obtained by expanding the square in (5.5), obtaining

$$N^2 \epsilon^2 + z^2 - 2zN\epsilon \geq 2N\epsilon \ln \beta^{-1} + 2N\epsilon \, r \ln(ez/r),$$

where we set $z \doteq \zeta + r - 1$. Solving this inequality for N, under the condition $N\epsilon \geq z$, we obtain

(5.7)
$$N \geq N_{\text{lower}}^{(1)}(\epsilon, \beta, r, \zeta) \doteq \frac{1}{\epsilon} \left(z + \ln \Upsilon_1 + \sqrt{\ln^2 \Upsilon_1 + 2z \ln \Upsilon_1} \right), \quad \Upsilon_1 \doteq \left(\frac{ez}{r} \right)^r \beta^{-1}.$$

The second refined lower bound on N is obtained by following the same route as before, but by using the binomial bound (5.6) instead of (5.4). It is straightforward to verify that the bound in this case is

(5.8)
$$N \geq N_{\text{lower}}^{(2)}(\epsilon, \beta, r, \zeta) \doteq \frac{1}{\epsilon} \left(z + \ln \Upsilon_2 + \sqrt{\ln^2 \Upsilon_2 + 2z \ln \Upsilon_2} \right), \quad \Upsilon_2 \doteq \left(\frac{ez}{\zeta - 1} \right)^{\zeta - 1} \beta^{-1}.$$

Clearly, bound (5.7) is preferable when $r \leq \zeta - 1$, whereas bound (5.8) is preferable when $r > \zeta - 1$.

Remark 5.2. Equation (5.1) shows that guaranteeing a small β is "cheap" in terms of the sample complexity N, since β appears under a logarithm. In practice, the level β can be fixed to a very small value, say $\beta = 10^{-9}$, and still the required number of sampled constraints remains reasonable. In Table 5.1 we use exact numerical inversion

of $\beta_{r,\zeta}(\epsilon) \leq \beta$ to show, for various values of N and ϵ, the maximum allowed number r of constraints that can be a posteriori removed while still guaranteeing that $\mathbb{P}^N\{V^* > \epsilon\} \leq \beta$ with $\beta = 10^{-9}$. Similarly, in Table 5.2 we show, for various values of r and ϵ, the minimum required number N of random constraints needed for guaranteeing that $\mathbb{P}^N\{V^* > \epsilon\} \leq \beta$ with $\beta = 10^{-9}$.

TABLE 5.1
Maximum number r of a posteriori removable constraints guaranteeing $\mathbb{P}^N\{V^* > \epsilon\} \leq 10^{-9}$ for problems of Helly's dimension $\zeta = 5$.

	$\epsilon = 0.5$	$\epsilon = 0.4$	$\epsilon = 0.3$	$\epsilon = 0.2$	$\epsilon = 0.1$	$\epsilon = 0.01$	$\epsilon = 0.001$
$N = 60$	2	-	-	-	-	-	-
$N = 100$	11	5	0	-	-	-	-
$N = 200$	42	26	13	3	-	-	-
$N = 500$	153	109	69	34	6	-	-
$N = 1000$	358	265	179	100	32	-	-
$N = 2000$	793	602	421	250	96	-	-
$N = 5000$	2160	1673	1201	748	322	5	-
$N = 10000$	4506	3523	2563	1629	735	30	-
$N = 40000$	18959	14992	11071	7206	3426	237	2

TABLE 5.2
Minimum number N of random constraints guaranteeing $\mathbb{P}^N\{V^* > \epsilon\} \leq 10^{-9}$ for problems of Helly's dimension $\zeta = 5$.

	$\epsilon = 0.5$	$\epsilon = 0.4$	$\epsilon = 0.3$	$\epsilon = 0.2$	$\epsilon = 0.1$	$\epsilon = 0.01$	$\epsilon = 0.001$
$r = 0$	48	64	91	144	301	3134	31459
$r = 5$	74	99	139	219	458	4757	47737
$r = 10$	95	125	176	277	577	5980	60005
$r = 20$	130	172	240	376	782	8075	81001
$r = 50$	223	292	405	629	1301	13370	134058
$r = 100$	362	469	647	1000	2057	21060	211083
$r = 200$	617	795	1090	1676	3430	34974	350394

5.1. Ways for discarding constraints. The a posteriori removal of r constraints (that is, removal of constraints *after* they have been extracted randomly) in Procedure 4.1 is useful to improve the optimal objective value of the problem while keeping $\mathbb{P}^N\{V^* > \epsilon\}$ under control. Clearly, the larger the number r of discarded constraints, the lower the attained optimal objective value of the problem. On the other hand, removing too many constraints, while beneficial from the point of view of the optimal objective, is detrimental from the point of view of the achievable confidence level β such that $\beta_{r,\zeta}(\epsilon) \leq \beta$ is guaranteed. However, level $\mathbb{P}^N\{V^* > \epsilon\}$ is under our control through formula (4.2); hence, we can keep removing constraints (and improving the objective) as long as $\beta_{r,\zeta}(\epsilon)$ remains below the a priori desired value β.

It is important to remark that the actual way in which the constraints are selected for removal is irrelevant—as long as this removal rule is invariant with respect to the order in which constraints appear. Here we mention two suboptimal heuristics and a globally optimal strategy for reducing the objective value by removing constraints. The advantage of suboptimal heuristics resides, of course, in avoiding the combinatorial complexity which may be encountered in the globally optimal approach.

Greedy constraint removal. This procedure finds the partition $\bar{R}_r$, R_r as follows: First, solve the problem with all the N extracted constraints, and then remove the

constraint that produces the greatest decrease in objective value (at most N optimization problems need be solved for finding such a constraint—and usually much less—since the only constraints that may reduce the objective are the ones that are active at the optimum of the original problem, which are usually in number of d). Then, consider the remaining $N-1$ constraints and, among them, remove the one that produces the greatest decrease in objective value (at most $N-1$—and typically d—optimization problems need be solved for finding such a constraint). Proceed in this way until all the required r constraints have been removed. A total of at most $(2N-r)(r+1)/2$ (but typically only $rd+1$) convex programs need be solved to apply this procedure and find the optimal solution of problem $P[\bar{R}_r(\omega)]$. This approach is clearly suboptimal in the sense that it does not guarantee that the *best possible* (in the sense of objective reduction) set of r constraints is removed.

Constraint removal based on marginal costs. Another heuristic for removing the constraints with the aim of reducing the optimal objective value is to select for removal those constraints that, at the optimum, are associated to large values of the corresponding Lagrange multipliers (dual variables). Suppose we first solve the problem with all the N extracted constraints, and let λ_j^* be the Lagrange multipliers associated to the constraints f_j at optimum. It is well known (complementary slackness) that if $\lambda_j^* > 0$, then f_j is active at x^* and that the value of λ_j^* provides the sensitivity of the optimal objective to variations of the jth constraint (λ_j^* is usually referred to as the marginal cost or shadow price). We then remove the constraint with largest value of λ_j^*, solve the resulting problem with $N-1$ constraints, and repeat the same reasoning until all the r constraints are removed. A total of $r+1$ convex programs need to be solved to apply this removal rule and to find the optimal solution of problem $P[\bar{R}_r(\omega)]$.

Optimal constraint removal. There is, of course, a globally optimal way for selecting $N-r$ constraints among the extracted N such that the largest possible decrease in objective is achieved. A direct brute-force approach would require solving $N!/(N-r)!r!$ convex optimization problems: one for each possible subset of $N-r$ constraints out of the N extracted ones. For nondegenerate problems, a smarter approach can be used that reduces the number of problems to be solved to a number smaller than $N(d^r-1)/(d-1)$; see section 3.2.2 in [13] for details on such an algorithm. Such direct approaches are mainly of theoretical interest, since they result, in general, in an unaffordable combinatorial number of problems. However, since the problem class we are dealing with belongs to the special class of LP-type optimization problems (see [38]), specific algorithms can be devised with complexity $O(Nr^d)$, which are hence efficient for fixed dimension d; see, e.g., [4, 29]. Further efficiency can be gained for certain specialized classes of geometric optimization problems. For instance, Matoušek in [28] reports an algorithm with $O(N \log N + N(N-r))$ complexity for the classical problem of computing the minimum radius circle containing all N but r points in the plane.

6. Connections with robust and chance-constrained optimization.

Robust optimization. A robust convex optimization problem [5, 6, 25] may take the following form:

$$P_\infty : \min_{x \in \Omega} \ c^\top x \text{ subject to}$$
$$f(x, \delta) \leq 0 \ \forall \delta \in \Delta.$$

Except for some specific problem classes (such as linear or conic programs) in which, moreover, the dependence of f on δ is of a special type, such as affine or rational,

the above robust (or semi-infinite) convex optimization problem entails a possibly continuum infinity of constraints, and it is very hard to solve numerically. The randomized approach in problem (3.1) may hence be useful for determining a solution which solves approximately a "probabilistic relaxation" of the robust problem: instead of seeking a solution that is *feasible for all possible* instances of $\delta \in \Delta$, in the scenario approach we fix a small ϵ and seek a solution which is *feasible for most*, albeit not all, the instances. Specifically, Theorem 4.1 guarantees that the scenario objective is, with high probability $1 - \beta_{r,\zeta}(\epsilon)$, still optimal *for all but the rarest* outcomes of $\delta \in \Delta$ (those having probability measure smaller than ϵ). In turn, $\beta_{r,\zeta}(\epsilon)$ may be rendered arbitrarily small by choosing a sufficiently high number N of random constraints. Any scenario solution thus provides a lower bound on the optimal objective of the robust problem.

When RCPs are used as a tool for obtaining approximate solutions to robust convex programs, therefore, one usually sets ϵ to quite a small number, say $\epsilon = 10^{-3}$, and then uses (3.3), (3.4), or the explicit bound in Corollary 5.1 to obtain the necessary number N of constraints to be used in (3.1). In this case, since robustness is the main issue, one usually does not wish to a posteriori remove constraints; hence, r is usually set to zero. RCPs have been successfully used in accordance to the above philosophy, for instance, in the context of robust control analysis and design; see [10].

Chance-constrained optimization. A CCP [16, 32, 34] usually takes the standard form

$$\text{(6.1)} \qquad \text{CCP}(\nu) : \min_{x \in \Omega} \; c^\top x \text{ subject to}$$

$$\text{(6.2)} \qquad V_c(x) \leq \nu,$$

where $V_c(x) = \mathbb{P}\{\delta \in \Delta : f(x, \delta) > 0\}$ is the constraint violation probability of x and $\nu \in (0, 1)$ is the admissible level of constraint violation. Besides very specific cases where the distribution has special symmetries and concavity properties (see, e.g., [20, 26, 33, 36]), CCPs are known to be very hard to solve and may be nonconvex even if the function f is convex in x for each given δ. Notice that in chance-constrained optimization it is not always the case that ν be very close to zero, and it may be completely reasonable, depending on the application at hand, to ask that a constraint be violated with probability smaller than, say, 0.4. In the following paragraphs we discuss the relations between RCPVs and CCPs.

To this end, we next compare the solutions of the CCP (6.1) with that of an RCPV $P[\bar{R}_r(\omega)]$. Given $\epsilon \in (0, 1)$, consider an RCPV where N and r are chosen so that $\beta_{r,\zeta}(\epsilon) < 1$. Then we notice from (4.12) that it must hold that $\mathbb{P}^N\{\omega \in \Delta^N : \{V_c^*(\omega) \leq \epsilon\} \cup \bar{\Delta}^{(N-r)*}\} \geq 1 - \beta_{r,\zeta}(\epsilon)$, where $\bar{\Delta}^{(N-r)*}$ is the set of extractions for which the solution does not exist, that is, the complement of $\Delta^{(N-r)*}$ in Δ^N. Notice also from (4.11) that, for $\epsilon < 1$, we have the inclusion $\{V_c^*(\omega) \leq \epsilon\} \subseteq \Delta^{(N-r)*}$; thus, $\{V_c^*(\omega) \leq \epsilon\} = \{V_c^*(\omega) \leq \epsilon \cap \Delta^{(N-r)*}\}$, and hence,

$$\mathbb{P}^N\{\omega \in \Delta^N : \{V_c^*(\omega) \leq \epsilon \cap \Delta^{(N-r)*}\} \cup \bar{\Delta}^{(N-r)*}\} \geq 1 - \beta_{r,\zeta}(\epsilon).$$

In other words, with probability at least $1 - \beta_{r,\zeta}(\epsilon)$ such an RCPV either (a) does not admit a solution (unfeasible, $J^* = +\infty$) or (b) admits a solution and $V_c^* \leq \epsilon$ holds for this solution (i.e., this solution is feasible—albeit possibly suboptimal—for CCP(ϵ); thus, $J^* \geq J_{\text{ccp}}^*(\epsilon)$). We thus have that $J^* \geq J_{\text{ccp}}^*(\epsilon)$ holds in both the (a) and (b) cases; therefore, $\mathbb{P}^N\{J^* \geq J_{\text{ccp}}^*(\epsilon)\} \geq 1 - \beta_{r,\zeta}(\epsilon)$. For suitably small ϵ_1, a converse inequality $J^* \leq J_{\text{ccp}}^*(\epsilon_1)$ also holds with high probability, as shown in Theorem 6.1

below. Before stating this result, we formalize the setup and conditions under which we compare the optimal objectives of RCPV and CCP.

STATEMENT 1 (setup for comparison of RCPV and CCP). *Given $\epsilon \in (0,1)$, consider an RCPV problem $P[\bar{R}_r(\omega)]$ with N and r chosen so that $\beta_{r,\zeta}(\epsilon) < 1$. Let J^* be the optimal objective of $P[\bar{R}_r(\omega)]$. Let the hypotheses of Theorem 4.1 be satisfied, and, for any $\nu \in (0,1)$, let $J_{\mathrm{ccp}}^*(\nu)$ be the optimal objective of (6.1) and $x_{\mathrm{ccp}}^*(\nu)$ be the corresponding optimal solution, when it exists.*

THEOREM 6.1. *Consider the setup in Statement 1, and let $\epsilon_1 = 1 - (1 - \beta_{r,\zeta}(\epsilon))^{1/(N-r)}$. Then*

$$\mathbb{P}^N\{\omega \in \Delta^N : J^* \geq J_{\mathrm{ccp}}^*(\epsilon)\} \geq 1 - \beta_{r,\zeta}(\epsilon), \tag{6.3}$$

$$\mathbb{P}^N\{\omega \in \Delta^N : J^* \leq J_{\mathrm{ccp}}^*(\epsilon_1)\} \geq 1 - \beta_{r,\zeta}(\epsilon). \tag{6.4}$$

Proof. Equation (6.3) has already been proved in the preceding discussion; thus, we only need to prove (6.4). Suppose first that $\mathrm{CCP}(\epsilon_1)$ is unfeasible: then $J_{\mathrm{ccp}}^*(\epsilon_1) = +\infty$, $J^* \leq J_{\mathrm{ccp}}^*(\epsilon_1)$ is the certain event, and (6.4) holds. Suppose then that $\mathrm{CCP}(\epsilon_1)$ is feasible: we show that $x_{\mathrm{ccp}}^*(\epsilon_1)$ is feasible for RCPV with probability at least $1 - \beta_{r,\zeta}(\epsilon)$; hence, (6.4) holds. Consider thus the event $\{x_{\mathrm{ccp}}^*(\epsilon_1) \in \mathrm{Sat}\,(\bar{R}_r(\omega))\}$, which clearly implies the event $\{J^* \leq J_{\mathrm{ccp}}^*(\epsilon_1)\}$, and let

$$\mathcal{R}_j \doteq \{\omega \in \Delta^N : \bar{R}_r(\omega) = I_{N-r}^j(\omega)\}, \quad j = 1, \ldots, C_{N,r}.$$

Then

$$
\begin{aligned}
\mathbb{P}^N\{J^* \leq J_{\mathrm{ccp}}^*(\epsilon_1)\} &\geq \mathbb{P}^N\{x_{\mathrm{ccp}}^*(\epsilon_1) \in \mathrm{Sat}\,(\bar{R}_r(\omega))\} \\
&= \sum_{j=1}^{C_{N,r}} \mathbb{P}^N\{x_{\mathrm{ccp}}^*(\epsilon_1) \in \mathrm{Sat}\,(\bar{R}_r(\omega)) \mid \mathcal{R}_j\}\mathbb{P}^N\{\mathcal{R}_j\} \\
&= C_{N,r}^{-1} \sum_{j=1}^{C_{N,r}} \mathbb{P}^N\{x_{\mathrm{ccp}}^*(\epsilon_1) \in \mathrm{Sat}\,(\bar{R}_r(\omega)) \mid \mathcal{R}_j\} \\
&= C_{N,r}^{-1} \sum_{j=1}^{C_{N,r}} \mathbb{P}^N\left\{x_{\mathrm{ccp}}^*(\epsilon_1) \in \mathrm{Sat}\left(I_{N-r}^j(\omega)\right)\right\}.
\end{aligned}
$$

The probability in the latter sum is the probability that, over N extractions, $x_{\mathrm{ccp}}^*(\epsilon_1)$ is feasible at least for the $N - r$ constraints in $I_{N-r}^j(\omega)$. Since the probability of being feasible for a single constraint is at least $1 - \epsilon_1$ (this follows from the fact that $x_{\mathrm{ccp}}^*(\epsilon_1)$ is optimal, hence feasible, for the CCP (6.1)) and since the constraint extractions are independent, the probability we are seeking is no smaller than $(1 - \epsilon_1)^{N-r}$; therefore,

$$\mathbb{P}^N\{J^* \leq J_{\mathrm{ccp}}^*(\epsilon_1)\} \geq (1 - \epsilon_1)^{N-r}.$$

The proof is concluded by taking ϵ_1 such that $1 - (1 - \epsilon_1)^{N-r} = \beta_{r,\zeta}(\epsilon)$. □

The result of Theorem 6.1 can be strengthened in the special case when a globally optimal strategy is applied in Procedure 4.1, that is, when the set of r constraints that provide the *best possible improvement* in the objective is removed. The following theorem holds.

THEOREM 6.2. *Consider the setup in Statement 1, and let an optimal constraint removal procedure be applied to the RCPV. Then, for any $\epsilon_1 < \epsilon$,*

$$\mathbb{P}^N\{\omega \in \Delta^N : J^* \geq J_{\mathrm{ccp}}^*(\epsilon)\} \geq 1 - \beta_{r,\zeta}(\epsilon), \tag{6.5}$$

$$\mathbb{P}^N\{\omega \in \Delta^N : J^* \leq J_{\mathrm{ccp}}^*(\epsilon_1)\} \geq 1 - \bar{\beta}(\epsilon_1), \tag{6.6}$$

where $\beta_{r,\zeta}(\epsilon)$ is defined in (4.3) and

$$(6.7) \qquad \bar{\beta}(\epsilon_1) \doteq \sum_{j=r+1}^{N} \binom{N}{j} \epsilon_1^j (1-\epsilon_1)^{N-j}.$$

Moreover,

$$(6.8) \qquad \mathbb{P}^N \{\omega \in \Delta^N : J_{\text{ccp}}^*(\epsilon) \leq J^* \leq J_{\text{ccp}}^*(\epsilon_1)\} \geq 1 - (\beta_{r,\zeta}(\epsilon) + \bar{\beta}(\epsilon_1)).$$

Proof. The statement in (6.5) is proved exactly as in Theorem 6.1. To prove (6.6), we proceed as follows. Suppose first that $\text{CCP}(\epsilon_1)$ is unfeasible: then $J_{\text{ccp}}^*(\epsilon_1) = +\infty$, $J^* \leq J_{\text{ccp}}^*(\epsilon_1)$ is the certain event, and (6.6) holds. Suppose then that $\text{CCP}(\epsilon_1)$ is feasible, and consider the event $\{x_{\text{ccp}}^*(\epsilon_1) \in \text{Sat}\,(\bar{R}_r(\omega))\}$, which clearly implies the event $\{J^* \leq J_{\text{ccp}}^*(\epsilon_1)\}$. Define the following event:

$$\mathcal{F} = \bigcup_{i=0}^{r} \mathcal{F}_i,$$

where

$$\mathcal{F}_i \doteq \{\omega \in \Delta^N : x_{\text{ccp}}^*(\epsilon_1) \text{ is feasible on a subset of } N - r + i \text{ constraints}\}.$$

In other words, $\mathcal{F}$ is the event where $x_{\text{ccp}}^*(\epsilon_1)$ is feasible on some subset of the constraints having cardinality $N-r$ or larger. Since J^* is the minimal achievable objective over all possible subsets of the constraints of cardinality $N - r$ (hence, it is also the minimal achievable objective over subsets of larger cardinality), it is immediate to conclude that $\omega \in \mathcal{F}$ implies that $\omega \in \{J^* \leq J_{\text{ccp}}^*(\epsilon_1)\}$; therefore,

$$\mathbb{P}^N \{J^* \leq J_{\text{ccp}}^*(\epsilon_1)\} \geq \mathbb{P}^N \{\mathcal{F}\} = \sum_{i=0}^{r} \mathbb{P}^N \{\mathcal{F}_i\}.$$

Now denote with ν_1 the probability of $x_{\text{ccp}}^*(\epsilon_1)$ being feasible for the constraint $f(\cdot, \delta) \leq 0$, and observe from (6.2) that $\nu_1 \geq 1 - \epsilon_1$. Since the constraint extractions are independent, we have that

$$\mathbb{P}^N \{\mathcal{F}_i\} = \binom{N}{r-i} \nu_1^{N-r+i} (1-\nu_1)^{r-i},$$

and therefore

$$\mathbb{P}^N \{J^* \leq J_{\text{ccp}}^*(\epsilon_1)\} \geq \sum_{i=0}^{r} \binom{N}{r-i} \nu_1^{N-r+i} (1-\nu_1)^{r-i} = \sum_{j=0}^{r} \binom{N}{j} \nu_1^{N-j} (1-\nu_1)^j.$$

This latter probability is the probability of having $N - r$ or more successes in N Bernoulli trials, each having success probability ν_1: this quantity is clearly increasing in ν_1. Hence $\nu_1 \geq 1 - \epsilon_1$ implies that

$$\sum_{j=0}^{r} \binom{N}{j} \nu_1^{N-j} (1-\nu_1)^j \geq \sum_{j=0}^{r} \binom{N}{j} \epsilon_1^j (1-\epsilon_1)^{N-j}$$

$$= 1 - \sum_{j=r+1}^{N} \binom{N}{j} \epsilon_1^j (1-\epsilon_1)^{N-j} = 1 - \bar{\beta}(\epsilon_1),$$

which concludes the first part of the proof. The statement in (6.8) then follows from Bonferroni's inequality. □

6.1. Finding the tradeoff between N and r. We next show that, for any given $\epsilon_1 < \epsilon$, it is always possible to find suitable values of N and r such that the probability in (6.8) is larger than $1 - \beta$ for any assigned $\beta \in (0, 1)$. This implies that the optimal value of an RCPV may, in principle, approximate to arbitrary level the optimal value of a CCP.

Observe that $\beta_{r,\zeta}(\epsilon)$ increases as the removal cardinality r increases, whereas $\bar{\beta}(\epsilon_1)$ decreases as r increases.[1] Well-known bounds exist for the upper binomial tail $\bar{\beta}(\epsilon_1)$ and for the lower binomial tail $\beta_{r,\zeta}(\epsilon)$. In particular, the Hoeffding–Chernoff bound (see, for instance, Chapter 4 in [21]) yields, for $\epsilon_1 N \leq r + 1$,

$$(6.9) \qquad \ln \bar{\beta}(\epsilon_1) \leq -\frac{2(r + 1 - N\epsilon_1)^2}{N},$$

whereas the Chernoff bound (see, for instance, [17]) for the lower tail, together with the standard bound on the binomial coefficient $\ln C_{r+\zeta-1,r} = \ln C_{r+\zeta-1,\zeta-1} \leq (\zeta - 1)\ln(ez/(\zeta - 1))$ (where we set for convenience $z \doteq r + \zeta - 1$) yields, for $N\epsilon \geq z$ and $\zeta > 1$,

$$(6.10) \qquad \ln \beta_{r,\zeta}(\epsilon) \leq (\zeta - 1)\ln\left(\frac{ez}{(\zeta - 1)}\right) - \frac{(N\epsilon - z)^2}{2N\epsilon}.$$

Using (6.9), we easily obtain that

$$(6.11) \qquad r \geq N\epsilon_1 - 1 + \sqrt{\frac{N}{2}\ln\frac{2}{\beta}} \; \Rightarrow \; \bar{\beta}(\epsilon_1) \leq \frac{\beta}{2}.$$

Similarly, from (6.10) we obtain that

$$(6.12) \qquad z^2 - 2N\epsilon(z + (\zeta - 1)\ln(ez/(\zeta - 1))) + N^2\epsilon^2 - 2N\epsilon\ln 2/\beta \geq 0$$

implies that $\beta_{r,\zeta}(\epsilon) \leq \beta/2$. Therefore, if both (6.11) and (6.12) are satisfied, it holds that $\bar{\beta}(\epsilon_1) + \beta_{r,\zeta}(\epsilon) \leq \beta$, as desired. In practice, we may set $r = N\epsilon_1 + \sqrt{\frac{N}{2}\ln\frac{2}{\beta}}$, substitute in (6.12), and search numerically this latter inequality for an admissible N. Once an admissible pair (N, r) is found, if desired, a numerical optimization can be performed iteratively on the N and r coordinates, with constraint $\beta_{r,\zeta}(\epsilon) + \bar{\beta}(\epsilon_1) \leq \beta$, for finding an admissible pair with minimal N; some numerical values are reported in Table 6.1. The following reasoning shows that the search for an initial admissible (N, r) pair will always be successful.

To see this fact, note that for "large" N, (6.11) requires that r (hence z) is also large. For sufficiently large z, the logarithmic term in the first parenthesis in (6.12) is negligible compared to z; hence, condition (6.12) tends, for large N and z to the condition

$$(6.13) \qquad z^2 - 2N\epsilon z + N^2\epsilon^2 - 2N\epsilon\ln 2/\beta \geq 0,$$

which, for $N\epsilon \geq z$, is satisfied if

$$(6.14) \qquad z \leq N\epsilon - \sqrt{2N\epsilon\ln 2/\beta}.$$

---

[1] In Matlab, $\beta_{r,\zeta}(\epsilon)$ and $\bar{\beta}(\epsilon)$ can be computed as follows: let `epsil=` ϵ, `hdim=` ζ, `q=r+hdim-1`, `rp=1:r`, `C=prod((hdim-1-rp)./rp)` (or `C=1` if $r = 0$ or $\zeta = 1$). Then $\beta_{r,\zeta}(\epsilon) = $ `C*betacdf(1-epsil,N-q,q+1)`, $\bar{\beta}(\epsilon) = $ `betacdf(epsil,r+1,N-r)`.

TABLE 6.1
Examples of (N, r) pairs guaranteeing $\bar{\beta}(\epsilon_1) + \beta_{r,\zeta}(\epsilon) \leq \beta$ for $\beta = 10^{-6}$, $\epsilon_1 = 0.6\epsilon$.

	$\epsilon = 0.6$	$\epsilon = 0.5$	$\epsilon = 0.4$	$\epsilon = 0.3$	$\epsilon = 0.2$	$\epsilon = 0.1$
$\zeta = 1$	[403, 193]	[562, 223]	[798, 252]	[1190, 281]	[1974, 310]	[4326, 339]
$\zeta = 2$	[496, 231]	[694, 268]	[991, 305]	[1483, 341]	[2467, 377]	[5421, 413]
$\zeta = 3$	[578, 265]	[811, 308]	[1160, 351]	[1740, 393]	[2902, 436]	[6385, 478]
$\zeta = 5$	[726, 325]	[1022, 379]	[1466, 433]	[2207, 487]	[3688, 541]	[8133, 595]
$\zeta = 10$	[1053, 455]	[1489, 533]	[2145, 612]	[3240, 691]	[5431, 770]	[12010, 849]

Putting together (6.11) and (6.14), we thus have that, for large N, the condition

$$(6.15) \qquad N\epsilon_1 - 1 + \sqrt{\frac{N}{2}\ln\frac{2}{\beta}} \leq r \leq 1 - \zeta + N\epsilon - \sqrt{2N\epsilon\ln\frac{2}{\beta}}$$

implies $\bar{\beta}(\epsilon_1) + \beta_{r,\zeta}(\epsilon) \leq \beta$. For a suitable r to exist, the gap between the upper and the lower bound needs be positive; that is, $N(\epsilon - \epsilon_1) + 2 - \zeta - \sqrt{2N\epsilon\ln 2/\beta} - \sqrt{N/\ln 2/\beta} \geq 0$, which, for $\epsilon_1 < \epsilon$, is always feasible for sufficiently large N. This fact is important in principle, since it guarantees that, no matter how close ϵ_1 is to ϵ, a suitable pair (N, r) can always be found such that the resulting optimal RCPV objective (computed according to a globally optimal constraint removal strategy) is between $J_{\text{ccp}}^*(\epsilon)$ and $J_{\text{ccp}}^*(\epsilon_1)$ with arbitrarily large probability.

Remark 6.1 (empirical CCPs). It is not difficult to observe that an optimal objective J^* of an RCPV resulting from Procedure 4.1 (for any constraint removal rule) is suboptimal for the following (nonconvex) optimization problem:

$$(6.16) \qquad P_{\text{emp}}[\omega] : \qquad \min_{x \in \Omega} c^\top x \qquad \text{subject to}$$

$$(6.17) \qquad \frac{1}{N}\sum_{j=1}^{N}\mathbb{I}(f(x,\delta^{(i)}) > 0) \leq \varphi,$$

where $\varphi \doteq r/N$ and $\mathbb{I}$ is the indicator function, which is one when the clause is true, and zero otherwise. The left-hand side of (6.17) represents the empirical probability of constraint violation; therefore, problem $P_{\text{emp}}[\omega]$ is the empirical counterpart of the CCP (6.1)–(6.2). Also, the optimal objective J^* resulting from Procedure 4.1, *when a globally optimal constraint removal rule is employed*, exactly coincides with the optimal objective of $P_{\text{emp}}[\omega]$. RCPVs with an optimal removal rule are hence equivalent to empirical CCPs, and the results of Theorem 6.2 thus directly apply to $P_{\text{emp}}[\omega]$.

7. RCPVs and the VC theory. Statistical learning techniques based on the VC theory [39] could in principle also be used to derive sample-size bounds for certain classes of RCPVs (specifically, those having a finite VC dimension). These bounds, however, are generally considered to be so conservative as to be hardly useful in practice. Indeed, standard application of VC theory would lead to sample-size bounds that scale essentially with $1/\epsilon^2$, which is clearly much worse than the $1/\epsilon$ dependence that one obtains from scenario theory; see (5.1). Nevertheless, a new approach has been recently proposed in [2, 3] along the learning-theoretic framework that leads to less conservative results. In this section, we show how to adapt the results in [2, 3] to RCPVs and provide a direct comparison of the corresponding bounds.

As we observed in Remark 6.1, an RCPV with optimal constraint removal is equivalent to the empirical version (6.16) of a CCP. Given $\epsilon \in (0,1)$ and $\varphi = r/N$, the *probability of one-sided constrained failure* is defined in [2] as follows:

(7.1)
$$P_{\mathrm{oscf}}(\epsilon, \varphi) \doteq \mathbb{P}^N\{\omega \in \Delta^N : \exists x \text{ such that } \hat{V}_c(x,\omega) \leq \varphi \text{ and } V_c(x) > \hat{V}_c(x,\omega) + \epsilon\},$$

where $V_c(x) = \mathbb{P}\{\delta : f(x,\delta) > 0\}$ is the constraint violation probability and $\hat{V}_c(x,\omega) = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}(f(x,\delta^{(i)}) > 0)$ is its empirical counterpart. Then, under the standing assumption that the function $g(x,\delta) = \mathbb{I}(f(x,\delta) > 0)$ has *VC dimension* bounded by a known constant q, from the proof of Theorem 7 in [2] it results that

(7.2)
$$P_{\mathrm{oscf}}(\epsilon, \varphi) \leq \beta_{\mathrm{oscf}} \doteq 4\left(\frac{2eN}{q}\right)^q \exp\left(-\frac{N\epsilon^2}{4(\epsilon + \varphi)}\right).$$

In order to adapt this result to the RCPV setting, we establish the following lemma.

LEMMA 7.1 (violation probability bound from the VC approach). *For $\omega \in \Delta^N$, let x^* be the optimal solution of (6.16), assuming that such a solution always exists. Then*

(7.3)
$$\mathbb{P}^N\{\omega \in \Delta^N : V_c(x^*) > \varphi + \epsilon\} \leq 4\left(\frac{2eN}{q}\right)^q \exp\left(-\frac{N\epsilon^2}{4(\epsilon + \varphi)}\right).$$

Proof. Considering the complementary event in (7.1), we have that

$$1 - P_{\mathrm{oscf}}(\epsilon, \varphi) = \mathbb{P}^N\{\omega \in \Delta^N : \hat{V}_c(x,\omega) > \varphi \text{ or } V_c(x) \leq \hat{V}_c(x,\omega) + \epsilon \; \forall x\}$$
$$\text{(from (7.2))} \geq 1 - \beta_{\mathrm{oscf}}.$$

Since this inequality holds uniformly in x, in particular, it should hold for $x = x^*$; therefore,

$$1 - \beta_{\mathrm{oscf}} \leq \mathbb{P}^N\{\omega \in \Delta^N : \hat{V}_c(x^*,\omega) > \varphi \text{ or } V_c(x^*) \leq \hat{V}_c(x^*,\omega) + \epsilon\}$$
$$\text{(by Boole's inequality)} \leq \mathbb{P}^N\{\omega \in \Delta^N : \hat{V}_c(x^*,\omega) > \varphi\}$$
$$+ \mathbb{P}^N\{\omega \in \Delta^N : V_c(x^*) \leq \hat{V}_c(x^*,\omega) + \epsilon\}$$
$$\leq \mathbb{P}^N\{\omega \in \Delta^N : V_c(x^*) \leq \varphi + \epsilon\},$$

where the last inequality follows from (6.17). □

Note that the probability in (7.3) is the probability with which the constraint violation V_c at the optimal scenario solution exceeds by ϵ the empirical probability of constraint violation, which is $\varphi = r/N$. We now apply the scenario theory to bound the probability in (7.3), under the assumption that $\Delta^N = \Delta^{(N-r)*}$; hence, $\zeta = d$. To this end, inspecting the proof of Corollary 5.1 and using the binomial bound (5.6), we obtain that

(7.4)
$$\mathbb{P}^N\{\omega \in \Delta^N : V_c(x^*) > \varphi + \epsilon\} \leq \left(\frac{e(d + \varphi N - 1)}{d - 1}\right)^{d-1} \exp\left(-\frac{(N\epsilon - d + 1)^2}{2N(\varphi + \epsilon)}\right).$$

The scenario bound (7.4) and the VC bound (7.3) can now be compared directly. A first observation is that in order to apply (7.3), the VC dimension of $g(x,\delta)$ must be finite, and moreover, one needs to know an upper bound q on the VC dimension, which is a very complex task in general. In the very special case when the constraint family

$f(x, \delta) \leq 0$ describes half-spaces in $\mathbb{R}^d$ (this is the case, for instance, in uncertain linear programs), then the VC dimension of the binary function family $g(x, \delta)$ is exactly $q = d + 1$. Considering this case, for simplicity, it can be verified that β_{oscf} is exponentially worse than β_{rcpv} as N increases. In particular, for $N > (d-1)/(1-\varphi)$, it holds that

$$\frac{\beta_{\text{oscf}}}{\beta_{\text{rcpv}}} \geq \alpha_d \cdot N \cdot \exp\left(\frac{N\epsilon^2}{4(\varphi + \epsilon)}\right),$$

where α_d is a constant that depends on the space dimension d.

8. Example: The smallest enclosing circle problem. We next illustrate an application of the RCPV theory on a classic problem: given N points in the plane, find the minimum radius circle that encloses all but r of these points. This is an archetypal problem in geometric optimization for which there exist several efficient solution algorithms; see, for instance, [1, 28] and the references therein. In particular, [28] reports a solution algorithm with $O(N \log N + N(N - r))$ expected running time. However, we are not interested here in the specific algorithm used to solve the problem; the scenario approach is not intended to provide an "algorithm," rather it is a theoretical framework within which one may analyze the probabilistic properties of the output of an optimization problem when the input data is random. To fix ideas, consider the enclosing circle problem where the N points to be encircled are drawn independently on the plane according to some (unknown) probability distribution. Let $\delta^{(i)} \in \mathbb{R}^2$, $i = 1, \ldots, N$, be the random points, suppose that any suitable algorithm is applied on a realization of these points, and call $\mathcal{C}_N$ the resulting minimum radius circle that encloses all but r of these points. Clearly, $\mathcal{C}_N = \mathcal{C}_N(\omega)$ is a random object that depends on the realization of the multiextraction $\omega = (\delta^{(1)}, \ldots, \delta^{(N)})$. The fundamental question we ask is, *What is the probability that $\mathcal{C}_N$ will contain a new unseen point dropped on the plane according to the same probability distribution?*

To answer this question, we show first that the optimal circle, independently of the actual way in which it is computed, can be viewed as the output of an RCPV. To see this fact, call $c \in \mathbb{R}^2$ the center and ρ the radius of the circle. Then the center and radius of $\mathcal{C}_N$ are the optimal solution of the RCPV (a convex second-order cone program with $d = 3$ variables),

$$\min_{c,\rho} \rho \text{ subject to } \|c - \delta^{(i)}\| \leq \rho, \quad \delta^{(i)} \in \bar{R}_r(\omega),$$

where $\bar{R}_r(\omega)$ is the subset of $N - r$ enclosed points as selected by the specific constraint selection rule (for instance, the one subset yielding the lowest radius if an optimal rule is applied). It is an intuitive fact that this problem is feasible, and it admits a solution in any realization of the random points. Hence $\Delta^N = \Delta^{(N-r)*}$ in this case, and Helly's dimension of the problem is $\zeta \leq d = 3$. Moreover, if the distribution generating the random points is continuous, the problem is nondegenerate almost surely (this problem can be degenerate only if the optimal circle has four cocircular points or if it is determined by only two diametral points and it passes through a third point, which is a situation that arises with zero probability—see section 3.3 of [29]). The probability with which a newly extracted point is in $\mathcal{C}_N$ is thus precisely given by the constraint violation probability V_c^*, and from Corollary 4.2, we have that $\mathbb{P}^N\{V_c^* > \epsilon\} \leq \beta_{r,\zeta}(\epsilon)$.

For instance, let $N = 500$, $r = 42$, and $\epsilon = 0.2$. Then $\beta_{r,\zeta}(\epsilon) = 5.26 \times 10^{-9}$, which means that $V_c^* \leq \epsilon$ holds with practical certainty (i.e., with probability larger than

$1 - 5.26 \times 10^{-9}$). In other words, given any $N = 500$ random points, the minimum radius circle enclosing $N-r = 458$ of these points will contain a further unseen random point with probability at least $1-\epsilon = 0.8$. The reader may verify by simple numerical calculations that if instead we want this probability to be higher (say, $1-\epsilon = 0.9$), up to a comparable level of confidence ($\beta_{r,\varsigma}(\epsilon) \leq 10^{-8}$), then we need to decrease r to $r = 11$ or increase N to $N = 1024$. It is a remarkable fact that such exact, nonasymptotic results hold independently of the distribution generating the random points (which can be completely general and totally unknown to the user) and depend on the specific problem structure only through Helly's dimension parameter.

As a numerical example, we considered $N = 500$ points distributed on the plane according to a random normal mixture. Precisely, let w_1 be a random variable distributed according to a standard (zero mean, unit covariance) two-dimensional normal distribution, and let w_2 be uniform in $[-4, 4]$. Then we set $\delta = w_1$ with probability $p = 0.7$ and $\delta = w_2[1\,0]^\top + 2w_1$ with probability $1-p = 0.3$. Computing the minimum enclosing circle that rejects $r = 42$ points on an instance of this problem, we obtained a circle with center $c = [0.2904, 0.1401]$ and radius $\rho = 3.5428$; see Figure 8.1.

The condition $V_c^* \leq 0.2$ is a priori guaranteed to hold with practical certainty for a circle resulting from this procedure. Note again that the same results hold independently of the algorithm and for any selection of a suboptimal set of r points to be removed; that is, $\mathbb{P}^N\{V_c^* > \epsilon\} \leq \beta_{r,\varsigma}(\epsilon)$ holds for any minimum radius circle that encloses $N - r$ points and rejects the remaining r points, even if the subset of enclosed points is not the one leading to the smallest possible radius. This is due
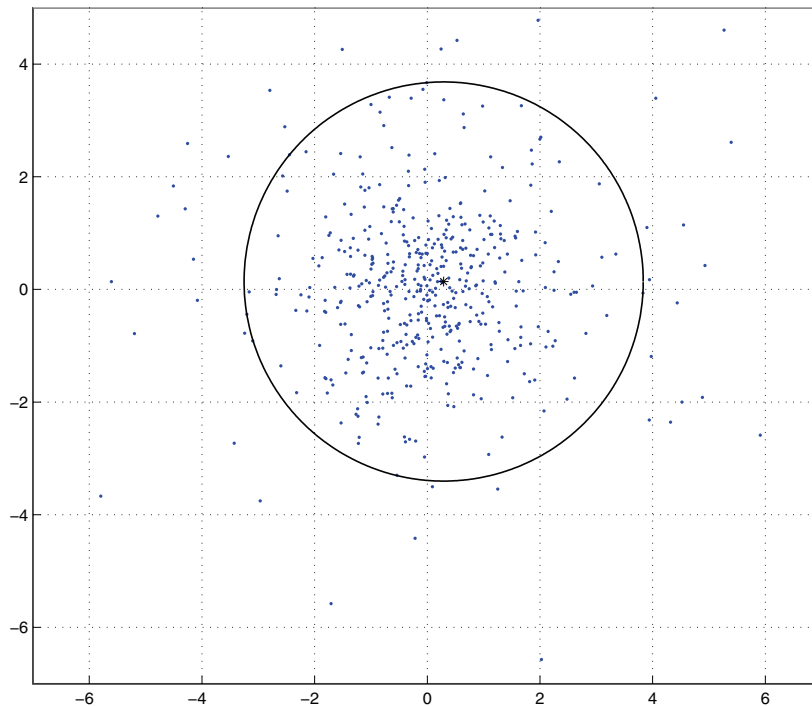

FIG. 8.1. *A minimum radius circle with $N = 500$, $r = 42$.*

to the fact that Theorem 4.1 and Corollary 4.2 hold independently of the constraint removal procedure.

This example highlights the fact that the RCPV framework can be employed, among many other contexts in which it can be useful, for assessing the generalization properties of pattern classification schemes that can be reformulated in the form of a convex program with violated constraints. These include, for instance, classification problems with linear separation surfaces that are commonly encountered in kernel methods and support vector machine classifiers; see, e.g., [18, 39].

Appendix A. Proofs.

A.1. Proof of Lemma 2.3. If problem $P[K]$ is feasible, then it has at most $d < d+1$ support constraints, by virtue of Lemma 2.2. We next consider the situation of unfeasible $P[K]$. Let problem $P[K]$ be unfeasible and have q support constraints. For simplicity, and without loss of generality, assume that the support constraints are the first q: $\mathrm{Sc}\,(K) = \{1, \ldots, q\} \subseteq K$. We next prove that $q \leq d+1$. Suppose, for the purpose of contradiction, that $q > d+1$. By definition of support constraints, each problem $P[K \setminus k]$, $k \in \mathrm{Sc}\,(K)$, must have optimal objective value $J_k^* < \infty$; hence, it must be feasible. Let $\mathcal{X}_i \doteq \mathrm{Sat}\,(i)$ and $\mathcal{S}_i \doteq \mathrm{Sat}\,(K \setminus i)$, $i \in K$, and let $\Psi \doteq \cap_{i=d+3}^{|K|} \mathcal{X}_i$. Clearly,

$$(\mathrm{A.1}) \qquad \left(\cap_{i=1}^{d+2} \mathcal{X}_i\right) \cap \Psi \equiv \mathrm{Sat}\,(K) = \emptyset,$$

since problem $P[K]$ is unfeasible by assumption. Given the following collection of $d+3$ convex sets in $\mathbb{R}^d$

$$C = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{d+1}, \mathcal{X}_{d+2}, \Psi\},$$

consider all subcollections composed of $d+1$ of these sets. Each of these subcollections will lack at least one of the $\mathcal{X}_i$ sets, $i = 1, \ldots, d+2$. Suppose for instance that a subcollection lacks $\mathcal{X}_i$; then the intersection of all sets in this subcollection must contain the set $\mathcal{S}_i$, since this set is the intersection of all elements of C, except for $\mathcal{X}_i$. But $\mathcal{S}_i$ is nonempty by assumption, since i is a support constraint; hence, it follows that the intersection of the elements in any subcollection of $d+1$ elements from C has a nonempty intersection. From Helly's theorem (see, e.g., Theorem 21.6 in [35]), it then follows that the whole collection C must have a nonempty intersection. But this would contradict (A.1); hence, we conclude that $q \leq d+1$. □

A.2. Proof of Lemma 2.4. Part 1 is immediate, since $J_1^* = \mathrm{Obj}[K_1]$ is a global minimum; hence, $J_1^* \leq c^\top x \ \forall \ x \in \mathrm{Sat}\,(K_1)$, and $\mathrm{Sat}\,(K_2) \subseteq \mathrm{Sat}\,(K_1)$.

For part 2, first let $\mathrm{Obj}[K_1, k] > \mathrm{Obj}[K_1]$, and suppose by contradiction that $\mathrm{Obj}[K_2, k] = \mathrm{Obj}[K_2]$. Then

$$\mathrm{Obj}[K_1, k] > \mathrm{Obj}[K_1] = \mathrm{Obj}[K_2] = \mathrm{Obj}[K_2, k] \geq \mathrm{Obj}[K_1, k] \ (\text{since } K_1 \subseteq K_2),$$

which is impossible; hence, it must be that $\mathrm{Obj}[K_2, k] > \mathrm{Obj}[K_2]$. Conversely, let $\mathrm{Obj}[K_2, k] > \mathrm{Obj}[K_2]$, and suppose by contradiction that $\mathrm{Obj}[K_1, k] = \mathrm{Obj}[K_1]$. Note that in the case of K_1 unfeasible, the second point of the lemma is uninformative. Thus we let K_1 be feasible; hence, $x^* = \mathrm{Opx}[K_1]$ is well defined and unique. Since $\mathrm{Obj}[K_2] = \mathrm{Obj}[K_1] = \mathrm{Obj}[K_1, k]$, then x^* is also the unique optimal solution on K_2 and on $[K_1, k]$. Thus x^* is optimal on K_2 and feasible on k, which implies it is also optimal on $[K_2, k]$, which would imply, in turn, that $\mathrm{Obj}[K_2] = \mathrm{Obj}[K_2, k]$, which contradicts the hypothesis. We then conclude that it must be $\mathrm{Obj}[K_1, k] > \mathrm{Obj}[K_1]$. □

A.3. Proof of Lemma 2.8. Suppose first that $P[K]$ is fully supported and feasible, and let $x^* = \mathrm{Opx}[K]$. By definition, x^* is feasible for all constraints in K. Suppose next we remove one constraint of index j which is not a support constraint (that is, $j \in K \setminus \mathrm{Sc}\,(K)$) and solve the resulting optimization problem. Since we removed a constraint, the optimal objective cannot increase with respect to $J^* = c^\top x^*$. However, it cannot decrease either, for otherwise the removed constraint would be a support one. Therefore, the optimal objective remains the same; hence, x^* is still optimal for the problem with the removed constraint. Consider now the problem obtained by removing a second constraint $k \neq j$, which was not of support for the original problem. We claim that x^* is still an optimal solution for this problem. Indeed, for the optimal objective to improve upon J^*, it would be necessary that $k \in \mathrm{Sc}\,(K \setminus j)$. However, for any $i \in \mathrm{Sc}\,(K)$, we have that $\mathrm{Obj}[K \setminus i]$ improves with respect to J^*—hence also $\mathrm{Obj}[K \setminus i, j]$ would improve with respect to J^*—hence all support constraints of $P[K]$ are also support constraints of $P[k \setminus j]$. Since the original problem is fully supported, this means that there are d constraints (different from the kth) that are of support for problem $P[K \setminus j]$. But from Lemma 2.2 no feasible problem can have more than d support constraints; hence, k cannot be a support constraint for $P[K \setminus j]$, and therefore, x^* is still an optimal solution for $P[K \setminus j, k]$. We can go on by removing further constraints that are not support constraints for the original problem and, by the same reasoning, conclude that x^* remains an optimal solution. In the end, we remove all the constraints that were not support ones, and find x^* as an optimal solution: this means that the optimal solution with all constraints in place coincides with the optimal solution obtained with only the support constraints in place, which permits us to conclude that the problem was not degenerate. The situation of unfeasible $P[K]$ can be treated in an analogous way, considering that in this case a fully supported problem has $d + 1$ support constraints: all nonsupport constraints can be removed without rendering the problem feasible; thus, the problem is still unfeasible on the set of the remaining $d + 1$ support constraints, which means that $\mathrm{Obj}[K] = \mathrm{Obj}[\mathrm{Sc}\,(K)]$ (i.e., that $P[K]$ is nondegenerate). □

A.4. Proof of Lemma 2.10. If $\mathrm{Sc}\,(K)$ is empty, the results are obvious, so we consider next $\mathrm{Sc}\,(K)$ nonempty. We first prove the inclusion $\mathrm{Sc}\,(K) \subseteq \mathrm{Es}\,(K)$ in (2.2). Suppose for the purpose of contradiction that $k \in \mathrm{Sc}\,(K)$ and $k \notin \mathrm{Es}\,(K)$. Then, since k is a support constraint, it holds that

$$(\mathrm{A.2}) \qquad \mathrm{Obj}[K \setminus k] < \mathrm{Obj}[K] = \mathrm{Obj}[\mathrm{Es}\,(K)].$$

Removing from $K \setminus k$ all further constraints that are not in $\mathrm{Es}\,(K)$, we are left with $\mathrm{Es}\,(K)$ itself, and since removing constraints cannot increase the optimal value of the problem, we would have $\mathrm{Obj}[\mathrm{Es}\,(K)] \leq \mathrm{Obj}[K \setminus k]$, which is in contradiction with (A.2), thus proving that it must be $k \in \mathrm{Es}\,(K)$.

The left part of (2.3) is obvious from the inclusion (2.2). We finally prove the right part of (2.3) by constructing an invariant set S such that $|S| \leq d + 1$ (resp., $|S| \leq d+1$ if $P[K]$ is feasible): since $\mathrm{Es}\,(K)$ is the invariant set of minimal cardinality, (2.3) would remain proved.

Given problem $P[K]$, consider the following procedure that takes as input argument K and returns as output S:

   0. Let $K_0 \doteq K$, $S_0 \doteq \mathrm{Sc}\,(K)$, and $\bar{K}_0 = K_0 \setminus S_0$. Let $i = 0$.
   1. If $|S_i| = d + 1$ (resp., $|S| \leq d + 1$ if $P[K]$ is feasible) or $|\bar{K}_i| = 0$, then stop and return $S = S_i$.

2. Select a $k \in \bar{K}_i$, and let

$$K_{i+1} = K_i \setminus k,$$
$$S_{i+1} = \mathrm{Sc}\,(K_{i+1}),$$
$$\bar{K}_{i+1} = K_{i+1} \setminus S_{i+1}.$$

3. Let $i = i + 1$; go to 1.

This procedure constructs a sequence of problems $P_0 = P[K_0]$, $P_1 = P[K_1]$, ..., in which each problem P_{i+1} has as constraints the constraints of the preceding problem, P_i, minus one (note that if P_0 is feasible, then all P_i remain feasible). The removed constraint is chosen among the constraints that were not of support for problem P_i. From the definition of support constraints, this implies that J^* remains the optimal objective value for all $P_0, P_1, \ldots$, and thus $K_0, K_1, \ldots$ are all invariant.

Note that $S_i \subseteq S_{i+1}$ (support constraints are "inherited"). To see this, suppose by contradiction that $j \in S_i$ and $j \notin S_{i+1}$. But $j \in S_i$ implies that $\mathrm{Obj}[K_i \setminus j] < J^*$, and certainly $\mathrm{Obj}[K_i \setminus j, k] \leq \mathrm{Obj}[K_i \setminus j]$. Thus $\mathrm{Obj}[K_i \setminus j, k] < J^*$, and this would imply that j is a support constraint for P_{i+1}, which is a contradiction. At each iteration the cardinality of S_i may grow, but, by virtue of Lemma 2.3, it must always remain bounded by $d + 1$ (resp., d if $P[K]$ is feasible).

The procedure may terminate for two reasons: either (a) $|S_i| = d + 1$ (resp., d if $P[K]$ is feasible) at some iteration i or (b) $|\bar{K}_i| = 0$. In case (a) problem P_i is fully supported, hence nondegenerate (Lemma 2.8), thus S_i is an invariant set of cardinality no larger than $d + 1$ (resp., d). In case (b) the set $\bar{K}_i$ is empty, which means that all constraints of problem P_i are support ones: $S_i \equiv K_i$. Thus again, S_i is an invariant set of cardinality no larger than $d + 1$ (resp., d).   □

A.5. Proof of Lemma 2.11. From the inclusion (2.2) we have that an essential set is a minimal-cardinality invariant set of the form $\mathrm{Es}\,(K) = \mathrm{Sc}\,(K) \cup Y$ for some $Y \subseteq \bar{K}$, $\bar{K} = K \setminus \mathrm{Sc}\,(K)$. Since $\mathrm{Sc}\,(K)$ is invariant by hypothesis and it is unique, the minimum cardinality invariant set is obtained by taking $Y = \emptyset$.

Conversely, we next show that if the essential set is unique, then $P[K]$ is nondegenerate. We prove this claim by showing that the negated implication holds; that is, that $P[K]$ being degenerate implies that the essential set is not unique. Suppose then that $P[K]$ is degenerate. This means that $\mathrm{Sc}\,(K)$ is not invariant; hence, for any essential set we must have $\mathrm{Es}\,(K) = \mathrm{Sc}\,(K) \cup Y$ for some nonempty Y. Now take $y \in Y$, and consider problem $P[K \setminus y]$. Let $\mathrm{Es}\,(K \setminus y) = \mathrm{Sc}\,(K) \cup \tilde{Y}$ be an essential set for this problem for some $\tilde{Y} \neq Y$. Since y is not a support constraint, problem $P[K \setminus y]$ has the same optimal objective as $P[K]$. Moreover, $\tilde{Y}$ is nonempty; otherwise, the optimal objective obtained on $\mathrm{Sc}\,(K)$ would be the overall optimal one, which is not possible since the problem is degenerate. Therefore, we must conclude that $\mathrm{Sc}\,(K) \cup \tilde{Y}$ is also an essential set for problem $P[K]$ and that this problem has at least two essential sets ($\mathrm{Sc}\,(K) \cup Y$ and $\mathrm{Sc}\,(K) \cup \tilde{Y}$), and the first claim remains proved.

For the second claim, let $P[K]$ be nondegenerate (hence $\mathrm{Es}\,(K) = \mathrm{Sc}\,(K)$), and let $K_1 \subseteq K$. Suppose first that $\mathrm{Es}\,(K) \subseteq \mathrm{Es}\,(K_1)$. Then, by monotonicity,

$$\mathrm{Obj}[\mathrm{Es}\,(K)] \leq \mathrm{Obj}[\mathrm{Es}\,(K_1)] = \mathrm{Obj}[K_1] \leq \mathrm{Obj}[K].$$

Then, since by definition $\mathrm{Obj}[\mathrm{Es}\,(K)] = \mathrm{Obj}[K]$, it must be that $\mathrm{Obj}[K_1] = \mathrm{Obj}[K]$. Conversely, suppose that $\mathrm{Obj}[K_1] = \mathrm{Obj}[K]$, and assume for the purpose of contradiction that $\mathrm{Es}\,(K) \nsubseteq \mathrm{Es}\,(K_1)$. Then there exists $h \in \mathrm{Es}\,(K) = \mathrm{Sc}\,(K)$ such that

$h \notin \mathrm{Es}\,(K_1)$. Therefore

$$\mathrm{Obj}[K_1] = \mathrm{Obj}[K_1 \setminus h] \leq \text{ (by monotonicity) } \mathrm{Obj}[K \setminus h] < \mathrm{Obj}[K] \text{ (since } h \in \mathrm{Sc}\,(K)),$$

which is a contradiction. We thus showed that $\mathrm{Obj}[K_1] = \mathrm{Obj}[K]$ if and only if $\mathrm{Es}\,(K) \subseteq \mathrm{Es}\,(K_1)$. We next show that one cannot have $\mathrm{Es}\,(K) \subset \mathrm{Es}\,(K_1)$; therefore, it must be that $\mathrm{Es}\,(K) = \mathrm{Es}\,(K_1)$. To show this, suppose for the purpose of contradiction that $\mathrm{Es}\,(K) \subset \mathrm{Es}\,(K_1)$, and let $Y = \mathrm{Es}\,(K_1) \setminus \mathrm{Es}\,(K)$. Then, by definition, $\mathrm{Obj}[K] = \mathrm{Obj}[\mathrm{Es}\,(K)]$. If we add Y to the set of constraints $\mathrm{Es}\,(K)$, the objective cannot increase; otherwise, for adding further constraints up to completing K, by monotonicity, we would get an objective larger than $\mathrm{Obj}[\mathrm{Es}\,(K)]$. Therefore

$$\mathrm{Obj}[K] = \mathrm{Obj}[\mathrm{Es}\,(K)] = \mathrm{Obj}[\mathrm{Es}\,(K), Y].$$

But, by definition of Y, $[\mathrm{Es}\,(K), Y] = \mathrm{Es}\,(K_1)$; hence,

$$\mathrm{Obj}[K_1] = \mathrm{Obj}[\mathrm{Es}\,(K_1)] = \mathrm{Obj}[\mathrm{Es}\,(K), Y] > \mathrm{Obj}[\mathrm{Es}\,(K)];$$

otherwise, K_1 would have an invariant set of cardinality smaller than $\mathrm{Es}\,(K_1)$, which is not possible since the essential set is the minimal cardinality invariant set. Since by hypothesis $\mathrm{Obj}[K] = \mathrm{Obj}[K_1]$, the last two displayed statements are in contradiction; thus, we conclude that it must be $Y = \emptyset$. □

A.6. Proof of Lemma 2.12. Suppose first that $J_Y^* = \mathrm{Obj}[Y]$ is the optimal value also for all problems $P[Y, h_j]$, $j = 1, \ldots, n$. Let $K_0 = Y$, $K_1 = K_0 \cup h_1$, $K_2 = K_1 \cup h_2$, etc. Then, by hypothesis, $\mathrm{Obj}[K_0] = \mathrm{Obj}[K_1]$, and also $\mathrm{Obj}[K_0] = \mathrm{Obj}[K_0, h_2]$. This implies, by point 2 in Lemma 2.4 (locality), that it must be $\mathrm{Obj}[K_0] = \mathrm{Obj}[K_0, h_2] = \mathrm{Obj}[K_1, h_2] = \mathrm{Obj}[K_2]$; thus, $\mathrm{Obj}[K_0] = \mathrm{Obj}[K_1] = \mathrm{Obj}[K_2]$. This reasoning holds, of course, independently of the choice of h_1 and h_2, and $\mathrm{Obj}[Y, h_i, h_j] = \mathrm{Obj}[Y]$ for all pairs $(h_i,\ h_j)$. We can iterate the same reasoning and conclude that $\mathrm{Obj}[Y, h_i, h_j, h_z] = \mathrm{Obj}[Y]$ for all triples, etc., thus proving that $\mathrm{Obj}[Y, h_1, \ldots, h_n] = \mathrm{Obj}[Y]$. The converse statement is immediate: suppose $\mathrm{Obj}[Y, h_1, \ldots, h_n] = \mathrm{Obj}[Y]$ and remove $h_2, \ldots, h_n$. The objective $\mathrm{Obj}[Y, h_1]$ cannot increase with respect to $\mathrm{Obj}[Y, h_1, \ldots, h_n]$, since we removed some constraints. However, it cannot decrease either; otherwise, also removing h_1 would lead to $\mathrm{Obj}[Y] < \mathrm{Obj}[Y]$, which contradicts our hypothesis. Therefore, it must be $\mathrm{Obj}[Y, h_1] = \mathrm{Obj}[Y]$. This reasoning holds irrespective of the choice of h_1; hence, the result remains proved. □

A.7. Proof of Lemma 2.4 for regularized objectives. Let $\tilde{P}$ denote the regularized version of P, and denote by J the optimal objective of the standard problem and by $\tilde{J}$ the regularized objective. Let $K_1 \subseteq K_2 \subseteq K$, assign a unique numerical label to each element of K, and let $P[K_1]$ and $P[K_2]$ be nondegenerate. We first prove that $\tilde{P}$ satisfies the monotonicity property. Observe that $J(K_1) < J(K_2)$ implies $\tilde{J}(K_1) < \tilde{J}(K_2)$ (see (2.4)). Thus, we only have to consider the case where $J(K_1) = J(K_2)$. In this case, from Lemma 2.11 we have that $\mathrm{Sc}\,(K_1) = \mathrm{Sc}\,(K_2)$. Then let $\bar{K}_1 = K_1 \setminus \mathrm{Sc}\,(K_1)$, $\bar{K}_2 = K_2 \setminus \mathrm{Sc}\,(K_2)$, and let $\nu_1 = \min(d+1, |K_1|) - |\mathrm{Sc}\,(K_1)|$ and $\nu_2 = \min(d+1, |K_2|) - |\mathrm{Sc}\,(K_2)|$. Note that, by definition of the regularized objective, since $J(K_1) = J(K_2)$, then $\tilde{J}(K_1) \leq \tilde{J}(K_2)$ holds if and only if the ν_1 largest labels in $\bar{K}_1$ are lexicographically less than or equal to the ν_2 largest labels in $\bar{K}_2$, which we indicate with the notation $Z(K_1) \leq Z(K_2)$. Since $\mathrm{Sc}\,(K_1) = \mathrm{Sc}\,(K_2)$, then $\bar{K}_1 \subseteq \bar{K}_2$; hence, the largest labels in $\bar{K}_2$ are no smaller than the largest labels in $\bar{K}_1$, and we

indeed conclude that $\tilde{J}(K_1) \leq \tilde{J}(K_2)$, thus proving monotonicity for the regularized objective.

We next prove locality of $\tilde{P}$: let $k \in K$ and

$$\text{(A.3)} \qquad\qquad\qquad \tilde{J}(K_1) = \tilde{J}(K_2).$$

We need to prove that $\tilde{J}(K_2, k) > \tilde{J}(K_2)$ if and only if $\tilde{J}(K_1, k) > \tilde{J}(K_1)$. Notice that (A.3) implies that

$$\text{(A.4)} \qquad\qquad J(K_1) = J(K_2) \quad \text{and} \quad Z(K_1) = Z(K_2)$$

(recall $Z(K_i)$ is the vector of ν_i largest labels in $\bar{K}_i$, and the comparison is in lexicographic sense).

First, let $\tilde{J}(K_2, k) > \tilde{J}(K_2)$. Then, either $J(K_2, k) > J(K_2)$, which would immediately imply that $J(K_1, k) > J(K_1)$ by the locality of P (hence also $\tilde{J}(K_1, k) > \tilde{J}(K_1)$) or $J(K_2, k) = J(K_2)$ and $Z(K_2, k) > Z(K_2)$. In this latter case $k \notin K_2$, since otherwise, $Z(K_2, k) = Z(K_2)$, and $k \notin \mathrm{Sc}\,(K_2, k)$, since otherwise, $J(K_2, k) > J(K_2)$. Thus, $k \in K \setminus K_2$ and $\mathrm{Sc}\,(K_2, k) = \mathrm{Sc}\,(K_2)$; $Z(K_2, k) > Z(K_2)$ implies that the label of k is larger than all labels in $\bar{K}_2$. Since $\bar{K}_1 \subseteq \bar{K}_2$, the label of k is also larger than all labels in $\bar{K}_1$. Now we previously proved monotonicity of $\tilde{P}$; therefore, $\tilde{J}(K_1, k) \geq \tilde{J}(K_1)$. But it cannot be that $\tilde{J}(K_1, k) = \tilde{J}(K_1)$ for otherwise $J(K_1, k) = J(K_1)$ and $Z(K_1, k) = Z(K_1)$, which is not possible due to the fact that k is larger than all labels in $\bar{K}_1$. We then conclude that it must be that $\tilde{J}(K_1, k) > \tilde{J}(K_1)$, which concludes the first part of the proof.

Conversely, let $\tilde{J}(K_1, k) > \tilde{J}(K_1)$. Again, if $J(K_1, k) > J(K_1)$, then we immediately conclude that $J(K_2, k) > J(K_2)$ by the locality of P; hence, also $\tilde{J}(K_2, k) > \tilde{J}(K_2)$. Consider then the case where $J(K_1, k) = J(K_1)$ and $Z(K_1, k) > Z(K_1)$. By the same reasoning as before we obtain that the label of k must be larger than all labels in $\bar{K}_1$. But (A.4) implies that $\nu_1 = \nu_2 \doteq \nu$ and that the ν largest labels in $\bar{K}_1$ and in $\bar{K}_2$ are the same. Therefore, k must also be larger than all labels in $\bar{K}_2$. Now, by monotonicity of $\tilde{P}$, $\tilde{J}(K_2, k) \geq \tilde{J}(K_2)$. But it cannot be that $\tilde{J}(K_2, k) = \tilde{J}(K_2)$, since otherwise, $J(K_2, k) = J(K_2)$ and $Z(K_2, k) = Z(K_2)$, which is not possible due to the fact that k is larger than all labels in $\bar{K}_2$. We then conclude that it must be that $\tilde{J}(K_2, k) > \tilde{J}(K_2)$, which concludes the proof. ☐

A.8. Proof of Lemma 2.12 for regularized objectives. Denote by J the optimal objective of the standard problem and by $\tilde{J}$ the regularized objective. Let $Y \subseteq K$, assign a unique numerical label to each element of K, let $h_1, \ldots, h_n \in K$, and let $P[Y, h_i]$ be nondegenerate. We next prove that $\tilde{J}(Y, h_i) = \tilde{J}(Y)$, for $i = 1, \ldots, n$, if and only if $\tilde{J}(Y, h_1, \ldots, h_n) = \tilde{J}(Y)$.

Suppose $\tilde{J}(Y, h_i) = \tilde{J}(Y)$ for $i = 1, \ldots, n$. Then by definition of regularized objective, this implies that $J(Y, h_i) = J(Y)$ and $Z(Y, h_i) = Z(Y)$ for $i = 1, \ldots, n$. Lemma 2.11 guarantees that $\mathrm{Es}\,(Y, h_i) = \mathrm{Es}\,(Y)$ $\forall i$; thus, $Z(Y, h_i) = Z(Y)$ implies that the label of h_i must be no larger than any of the labels in $\bar{Y} = Y \setminus \mathrm{Es}\,(Y)$. Therefore, it holds that $Z(Y, h_1, \ldots, h_n) = Z(Y)$. Moreover, by Lemma 2.12, which has already been proved for the standard objective, it must be $J(Y, h_1, \ldots, h_n) = J(Y)$. Hence we proved that $\tilde{J}(Y, h_1, \ldots, h_n) = \tilde{J}(Y)$. The converse implication follows from an analogous line. ☐

REFERENCES

[1] A. Aggarwal, H. Imai, N. Katoh, and S. Suri, *Finding k points with minimum diameter and related problems*, J. Algorithms, 12 (1991), pp. 38–56.

[2] T. Alamo, R. Tempo, and E. F. Camacho, *Revisiting statistical learning theory for uncertain feasibility and optimization problems*, in Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, 2007.

[3] T. Alamo, R. Tempo, and E. F. Camacho, *A randomized strategy for probabilistic solutions of uncertain feasibility and optimization problems*, IEEE Trans. Automat. Control, 54 (2009), pp. 2545–2559.

[4] E. Bai, H. Cho, and R. Tempo, *Optimization with few violated constraints for linear bounded error parameter estimation*, IEEE Trans. Automat. Control, 47 (2002), pp. 1067–1077.

[5] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769–805.

[6] A. Ben-Tal and A. Nemirovski, *Robust optimization: Methodology and applications*, Math. Program., 92 (2002), pp. 453–480.

[7] G. C. Calafiore, *Learning noisy functions via interval models*, Systems Control Lett., 59 (2010), pp. 404–413.

[8] G. C. Calafiore, *On the expected probability of constraint violation in sampled convex programs*, J. Optim. Theory Appl., 143 (2009), pp. 405–412.

[9] G. C. Calafiore and M. C. Campi, *Uncertain convex programs: Randomized solutions and confidence levels*, Math. Program., 102 (2005), pp. 25–46.

[10] G. C. Calafiore and M. C. Campi, *The scenario approach to robust control design*, IEEE Trans. Automat. Control, 51 (2006), pp. 742–753.

[11] M. C. Campi and G. C. Calafiore, *Notes on the scenario design approach*, IEEE Trans. Automat. Control, 54 (2009), pp. 382–385.

[12] M. C. Campi, G. C. Calafiore, and S. Garatti, *New results on the identification of interval predictor models*, in Proceedings of the IFAC World Congress, Prague, 2005.

[13] M. C. Campi, G. C. Calafiore, and S. Garatti, *Interval predictor models: Identification and reliability*, Automatica J. IFAC, 45 (2009), pp. 382–392.

[14] M. C. Campi and S. Garatti, *Chance-Constrained Optimization via Randomization: Feasibility and Optimality*, preprint, http://www.optimization-online.org, 2008.

[15] M. C. Campi and S. Garatti, *The exact feasibility of randomized solutions of uncertain convex programs*, SIAM J. Optim., 19 (2008), pp. 1211–1230.

[16] A. Charnes and W. W. Cooper, *Deterministic equivalents for optimizing and satisfying under chance constraints*, Oper. Res., 11 (1963), pp. 18–39.

[17] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Stat., 23 (1952), pp. 493–507.

[18] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, 2000.

[19] D. P. de Farias and B. Van Roy, *On constraint sampling in the linear programming approach to approximate dynamic programming*, Math. Oper. Res., 29 (2004), pp. 462–478.

[20] D. Dentcheva, A. Prékopa, and A. Ruszczyński, *Concavity and efficient points of discrete distributions in probabilistic programming*, Math. Program., 89 (2000), pp. 55–77.

[21] A. Desolneux, L. Moisan, and J.-M. Morel, From Gestalt Theory to Image Analysis, Interdiscip. Appl. Math. Springer, New York, 2008.

[22] H. Edelsbrunner and E. P. Mucke, *Simulation of simplicity: A technique to cope with degenerate cases in geometric algorithms*, ACM Trans. Graph., 9 (1990), pp. 66–104.

[23] I. Emiris and J. Canny, *An efficient approach to removing geometric degeneracies*, in Proceedings of the 8th Annual ACM Symposium on Computational Geometry, 1992, pp. 74–82.

[24] B. Gärtner and E. Welzl, *A simple sampling lemma: Analysis and applications in geometric optimization*, Discrete Comput. Geom., 25 (2001), pp. 569–590.

[25] L. El Ghaoui, F. Oustry, and H. Lebret, *Robust solutions to uncertain semidefinite programs*, SIAM J. Optim., 9 (1998), pp. 33–52.

[26] C. M. Lagoa, X. Li, and M. Sznaier, *Probabilistically constrained linear programs and risk-adjusted controller design*, SIAM J. Optim., 15 (2005), pp. 938–951.

[27] J. Luedtke and S. Ahmed, *A sample approximation approach for optimization with probabilistic constraints*, SIAM J. Optim., 19 (2008), pp. 674–699.

[28] J. Matoušek, *On enclosing k points by a circle*, Inform. Proces. Lett., 53 (1995), pp. 217–221.

[29] J. Matoušek, *On geometric optimization with few violated constraints*, Discrete Comput. Geom., 14 (1994), pp. 365–384.

[30] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM J. Optim., 17 (2006), pp. 969–996.

[31] A. Nemirovski and A. Shapiro, *Scenario approximations of chance constraints*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. C. Calafiore and F. Dabbene, eds., Springer-Verlag, London, 2006, pp. 3–47.

[32] A. Prékopa, *On probabilistic constrained programming*, in Proceedings of the Princeton Symposium on Mathematical Programming, Princeton University Press, Princeton, NJ, 1970, pp. 113–138.

[33] A. Prékopa, *Stochastic Programming*, Kluwer Academic Publishers, Dordrecht, 1995.

[34] A. Prékopa, *Probabilistic programming*, in Stochastic Programming, Handbooks Oper. Res. Management Sci. 10, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, 2003.

[35] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[36] A. Ruszczyński and A. Shapiro, *Optimization of risk measures*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. C. Calafiore and F. Dabbene, eds., Springer-Verlag, London, 2006, pp. 117–158.

[37] A. Ruszczyński and A. Shapiro, eds., *Stochastic Programming*, Handbooks Oper. Res. Management Sci. 10, Elsevier, Amsterdam, 2003.

[38] M. Sharir and E. Welzl, *A combinatorial bound for linear programming and related problems*, in Proceedings of STACS 92, 9th Annual Symposium on Theoretical Aspects of Computer Science, Lecture Notes in Comput. Sci. 577, Springer-Verlag, Berlin, 1992, pp. 569–579.

[39] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.