Decision making in an uncertain environment: the scenario-based optimization approach

M.C. Campi (1) and G. Calafiore (2)

(1) Dipartimento di Elettronica per l'Automazione
Università di Brescia
Via Branze 38 – 25123 Brescia, Italy
E-Mail: campi@ing.unibs.it
(2) Dipartimento di Automatica e Informatica
Politecnico di Torino
Corso Duca degli Abruzzi 24 – 10129 Torino, Italy
E-Mail: giuseppe.calafiore@polito.it

Abstract. A central issue arising in financial, engineering and, more generally, in many applicative endeavors is to make a decision in spite of an uncertain environment. Along a *robust* approach, the decision should be guaranteed to work well in *all* possible realizations of the uncertainty. A less restrictive approach consists instead of requiring that the risk of failure associated to the decision should be *small* in some – possibly probabilistic – sense. From a mathematical viewpoint, the latter formulation leads to a *chance-constrained* optimization program, i.e. to an optimization program subject to constraints in probability. Unfortunately, however, both the robust approach as well as the chance-constrained approach are computationally intractable in general.

In this paper, we present a computationally efficient methodology for dealing with uncertainty in optimization based on sampling a finite number of instances (or *scenarios*) of the uncertainty. In particular, we consider uncertain programs with convexity structure, and show that the scenario-based solution is, with high confidence, a feasible solution for the chance-constrained problem. The proposed approach represents a viable way to address general convex decision making problems in a risk-adjusted sense.

Keywords: Convex optimization, scenario approach, confidence levels, decision making.

1 Introduction

In this paper we consider decision making problems that can be formalized as robust convex optimization programs, where the constraints are parameterized by an uncertain parameter δ which describes the situation or 'states of nature' that can possibly occur. Precisely, letting $\delta \in \mathbf{\Delta} \subseteq R^\ell$ be the vector of uncertain variables, the robust convex program writes:

$$\text{RCP} : \min_{x \in \mathcal{X} \subseteq R^n} c^T x \quad \text{subject to } f(x, \delta) \leq 0, \quad \forall \delta \in \Delta, \tag{1}$$

where $\mathcal{X}$ is a convex and closed set that represents the feasible set for the solutions and the function $f(x,\delta) : \mathcal{X} \times \mathbf{\Delta} \to R$ is convex in x for any fixed $\delta \in \mathbf{\Delta}$. In typical situations, (1) is a semi-infinite optimization problem since it contains an infinite number of constraints (i.e. $\mathbf{\Delta}$ has infinite cardinality), while x represents a finite number of optimization variables.

The function $f(x,\delta)$ is here assumed to be scalar-valued without loss of generality, since multiple scalar-valued convex constraints $f_i(x,\delta) \leq 0$, $i = 1, \ldots, n_f$, can always be converted into a single scalar-valued convex constraint by the position $f(x,\delta) = \max_{i=1,\ldots,n_f} f_i(x,\delta) \leq 0$. We also note that the cost function $c^T x$ is certain (i.e. it does not depend on δ) and is linear in x without loss of generality, since any convex uncertain program can be reformulated so that it exhibits a linear certain cost. To see this, simply note that a problem $\min_{\xi \in \Xi} g(\xi, \delta)$ subject to $h(\xi,\delta) \leq 0, \forall \delta \in \Delta$, can be rewritten as $\min_{\xi \in \Xi, \gamma} \gamma$ subject to $g(\xi,\delta) - \gamma \leq 0$ and $h(\xi,\delta) \leq 0, \forall \delta \in \Delta$. Finally, we remark that no assumption is made on the functional dependence of f on δ.

In general situations, finding a solution to (1) is a formidable task and no general solution methodology for (1) is to date available. In special cases, (1) can be solved by reconducting it to a problem with a finite number of constraints, possibly by conservative relaxation techniques. We refer the reader to (A. Nemirovski 1993) and (G. Calafiore and M.C. Campi 2003) and the references therein for a discussion on the complexity of solving (1) and on possible relaxation techniques within the specific context of robust control problems.

Program (1) is 'robust' since the optimal solution is guaranteed against all possible occurrences of the uncertain parameter δ, i.e. the constraint $f(x,\delta) \leq 0$ must be satisfied for all possible values of $\delta \in \Delta$. In certain problems it may be convenient to relax such a strong requirement by allowing that a small fraction of constraints can possibly be violated, so leaving a small chance that $f(x,\delta) \leq 0$ is not satisfied in correspondence of the found solution. The resulting optimization problem is named a chance-constrained convex optimization problem and is formalized as follows.

Let Prob denote a probability measure over Δ (i.e. Δ is endowed with a σ-algebra $\mathcal{D}$ and Prob is assigned for every set of $\mathcal{D}$). Depending on the situation at hand, Prob can have different interpretations. Sometimes, it is the actual probability with which the uncertainty parameter δ takes on value in a certain set, while other times Prob simply describes the relative importance we attribute to different uncertainty instances. The chance-constrained (or probabilistic) convex optimization problem then writes:

$$\text{PCP}(\epsilon): \min_{x \in \mathcal{X} \subseteq R^n} c^T x \quad \text{subject to Prob}\{f(x,\delta) > 0\} \leq \epsilon, \tag{2}$$

where the parameter $\epsilon \in (0,1)$ represents the admissible *probabilistic risk* of violating the constraint $f(x,\delta) \leq 0$. It is clear that allowing for a probability ϵ of violation results in an optimal solution that outperforms the optimal robust solution for the uncertainty instances that are indeed feasible at the optimum. Moreover, ϵ represents a 'tuning knob' that can be freely selected and the lower ϵ, the closer the chance-constrained optimization problem to the robust optimization problem.

Chance-constrained optimization has a long history, dating back to the work of Charnes and Cooper for linear programs in 1959 , (A. Charnes and W.W. Cooper 1959). Still to date, most of the available computational results on chance-constraints refer to the important but very specific case when $f(x,\delta)$ is linear in x, i.e. to probability-constrained *linear* programs. One reason for this is that chance-constrained problems are extremely difficult to solve exactly. This can be easily realized by noticing that the mere evaluation of the probability with which a constraint is satisfied amounts to solving a multi-dimensional integral, which is a formidable task in general. Moreover, even when $f(x,\delta)$ is convex (or even linear) in x for any fixed $\delta \in \mathbf{\Delta}$, the feasible set $\{x : \text{Prob}\{f(x,\delta) > 0\} \leq \epsilon\}$ may be nonconvex, and hence PCP is *not* a convex program in general. We direct the reader to the monograph by Prékopa (A. Prékopa 1995) and to (S. Vajda 1972) for an extensive presentation of many available results in this area.

Problems of the form (2) appear frequently in manufacturing and financial problems. For instance, in the Value-at-Risk (VaR) framework (see e.g. (T. J. Linsmeier and N. D. Pearson 1996, M. Pritsker 1997)) one considers the possible loss $-r(\xi,\delta)$ of a portfolio of risky assets, where ξ describes the allocations of assets in the portfolio, and δ represent the (uncertain) returns of the assets. The VaR is then defined as the minimal level γ such that the probability that the portfolio loss exceeds γ is below a fixed (small) level $\epsilon \in (0,1)$. Hence, the problem of minimizing the VaR over admissible portfolios takes the form

$$\min_{\xi,\gamma} \gamma \quad \text{subject to Prob}\{\gamma < -r(\xi,\delta)\} \leq \epsilon,$$

which is a PCP, with $x \doteq (\xi,\gamma)$ and $f(x,\delta) \doteq -r(\xi,\delta) - \gamma$.

The previous example indicates also that the class of PCP problems embodies probabilistic counterparts of classical min-max games. Indeed, a min-max game

$$\min_{\xi \in \Xi} \max_{\delta \in \mathbf{\Delta}} g(\xi,\delta)$$

may be expressed as the robust program

$$\min_{\xi \in \Xi, \gamma} \gamma \quad \text{subject to } g(\xi,\delta) \leq \gamma, \ \forall \delta \in \mathbf{\Delta}. \tag{3}$$

The probabilistic counterpart of this semi-infinite program is a PCP of the form

$$\min_{\xi \in \Xi, \gamma} \gamma \quad \text{subject to Prob}\{g(\xi,\delta) > \gamma\} \leq \epsilon, \quad \epsilon \in (0,1). \tag{4}$$

Notice that, for fixed ξ, the optimal γ in problem (3) is the max of $g(\xi,\delta)$ over δ, while in (4) it is the *probable approximate maximum* of $g(\xi,\delta)$ over δ, i.e. a value which is exceeded only over a set of δ's having small probability measure.

1.1 A computationally feasible paradigm: Scenario-based convex programs

Motivated by the computational complexity of RCP and PCP, in this paper we pursue a solution methodology which is based on randomization of the parameter δ.

To this end, we collect N samples $\delta^{(1)}, \ldots, \delta^{(N)}$ of the uncertain parameter randomly chosen in an independent fashion according to probability Prob (these instances shall be referred to as the *scenarios*), and construct the randomized convex program

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

This randomized program, called the *scenario program* in the sequel, has a distinctive advantage over RCP and PCP: it is a standard convex program with N constraints, and hence it is typically efficiently solvable. However, a fundamental question need be addressed: what can we say about the chance-constraint satisfaction for an optimal solution of RCP_N? More explicitly, the crucial question to which this paper is devoted is the following:

> How many samples (scenarios) need to be drawn in order to guarantee that the solution of the scenario problem violates at most a portion ϵ of the constraints (i.e. it is a feasible solution in a chance-constrained sense)?

Using statistical learning techniques, we provide an explicit bound on the measure (probability) of the set of original constraints that are possibly violated by the scenario solution. This measure rapidly decreases to zero as N is increased, and therefore the obtained randomized solution is feasible for the chance-constrained problem (2).

2 Scenario-Based Convex Programs

We first formalize the concept of violation probability.

Definition 1 (Violation probability). The *probability of violation* of $x \in \mathcal{X}$ is defined as

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

(here, it is assumed that $\{\delta \in \mathbf{\Delta} :\ f(x, \delta) > 0\}$ is an element of the σ-algebra $\mathcal{D}$). ⋆

For example, if a uniform (with respect to Lebesgue measure) probability density is assumed, then $V(x)$ measures the volume of the uncertainty instances δ such that the constraint $f(x, \delta) \leq 0$ is violated. We next have the following definition.

Definition 2 (ϵ-level solution). Let $\epsilon \in (0, 1)$. We say that $x \in \mathcal{X}$ is an ϵ-level robustly feasible solution if $V(x) \leq \epsilon$. ⋆

Notice that, by the above definition, the ϵ-level solutions are the feasible solutions for the chance-constrained optimization problem (2). Our goal is to devise an algorithm that returns a ϵ-level solution, where ϵ is any fixed small level. To this purpose, we now formally introduce the scenario convex program as follows.

Definition 3 (Scenario convex program). Let $\delta^{(1)}, \ldots, \delta^{(N)}$ be N independent identically distributed samples extracted according to probability Prob. The randomized convex program (or scenario convex program) derived from (2) is

$$\text{RCP}_N: \min_{x \in \mathcal{X} \subseteq R^n} c^T x \quad \text{subject to } f(x, \delta^{(i)}) \leq 0,\ i = 1, \ldots, N. \tag{5}$$

⋆

For the time being we assume that RCP_N admits a unique solution. The uniqueness assumption is temporarily made for clarity in the presentation and proof of the main result, and it is later removed in Appendix A.

Assumption 1 *For all possible extractions $\delta^{(1)}, \ldots, \delta^{(N)}$, the optimization problem (5) attains a unique optimal solution.* ⋆

Let then $\hat{x}_N$ be the unique solution of problem RCP$_N$. Since the constraints $f(x, \delta^{(i)}) \leq 0$ are randomly selected, the optimal solution $\hat{x}_N$ is a random variable that depends on the extraction of the multi-sample $\delta^{(1)}, \ldots, \delta^{(N)}$. The following key theorem pinpoints the properties of $\hat{x}_N$.

Theorem 1. *Fix two real numbers $\epsilon \in (0,1)$ (level parameter) and $\beta \in (0,1)$ (confidence parameter). If*

$$N \geq N(\epsilon, \beta) \doteq \frac{2}{\epsilon} \ln \frac{1}{\beta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon}, \tag{6}$$

then, with probability no smaller than $1 - \beta$, the optimal solution $\hat{x}_N$ of the scenario problem RCP$_N$ *is ϵ-level robustly feasible.* ⋆

In this theorem, probability $1-\beta$ refers to the probability Prob^N (= Prob $\times \cdots \times$ Prob, N times) of extracting a 'bad' multi-sample, i.e. a multi-sample $\delta^{(1)}, \ldots, \delta^{(N)}$ such that $\hat{x}_N$ does not meet the ϵ-level feasibility property. β cannot be sent to zero, since otherwise N goes to infinity. This corresponds to the natural fact that, no matter how large N is, a 'bad' multi-sample returning a $\hat{x}_N$ with poor violation properties can always be extracted. For any practical purpose, however, β plays a very marginal role. The reason is that β appears under the sign of logarithm in (6) so that it can be taken to be a really tiny number (10^{-10} or even 10^{-20}) without making N too large. If β is neglected, in simple words Theorem 1 states that if N (specified by (6)) random scenarios are drawn, the optimal solution of RCP$_N$ is a feasible solution of the chance-constrained problem (2). Thus, while problem (1) admits no solution methodologies in general, (5) represents a viable way to solve (1) in a risk-adjusted sense, namely the found solution is feasible for the corresponding chance-constrained problem.

The proof of Theorem 1 is postponed to Section 3 to avoid breaking the continuity of discourse.

2.1 Discussion on main result

We next comment more closely on the proposed randomized approach. Theorem 1 says that if we extract a *finite* number N of constraints, then the solution of the randomized problem satisfies most of the other unseen constraints. This is a *generalization* property: the explicit satisfaction of some scenarios generalizes automatically to the satisfaction of other scenarios. It is interesting to note that generalization calls for some kind of structure, and the only structure used here is convexity. So, convexity in the scenario approach is fundamental in two different respects: on the computational side, it allows for an efficient solution of the ensuing optimization problem, while on the theoretical side it allows for generalization.

Remark 1 (Sample complexity and VC-dimension). The 'sample complexity' of the scenario problem (i.e. the number N of random scenarios that need to be drawn in order to achieve the desired probabilistic level in the solution) exhibits a substantially linear scaling with $\frac{1}{\epsilon}$ and a logarithmic scaling with $\frac{1}{\beta}$. If e.g. $\epsilon = 0.01$, $\beta = 0.0001$ and $n = 10$, we have $N \geq 12459$. In general, for reasonable probabilistic levels, the required number of these constraints is manageable by current convex optimization numerical solvers. We also mention that the reader can find a tighter bound than (6) in Section 3.2: in Theorem 1 we have been well advised to provide bound (6) – derived from the bound in Section 3.2 – to improve readability.

Bound (6) depends on the problem structure through n, the number of optimization variables, only. It is not difficult to conceive situations where the class of sets $\{\delta \in \mathbf{\Delta} : f(x,\delta) > 0\} \subseteq \mathbf{\Delta}$, parameterized in x, has infinite VC-dimension, even for small n; see for instance (V. Vapnik 1996) for definition of VC dimension and an exposition of learning theory. In these situations, estimating Prob$\{\delta \in \mathbf{\Delta} : f(x,\delta) > 0\} = V(x)$ uniformly with respect to x is impossible and the VC-theory is of no use. Theorem 1 says that, if attention is restricted to $\hat{x}_N$, then estimating $V(\hat{x}_N)$ becomes possible at a low computational cost. ⋆

*Remark 2 (*Prob*-independent bound).* In some applications, probability Prob is not explicitly known, and the scenarios are directly made available as 'observations'. This could for example be the case when the instances of δ are actually related to various measurements or identification experiments made on a plant at different times and/or different operating conditions, see e.g. (G. Calafiore and M.C. Campi 2002, G. Calafiore and M.C. Campi 2003). In this connection, notice that the bound (6) is probability independent, i.e. it holds irrespective of the underlying probability Prob, and can therefore be applied even when Prob is unknown. ⋆

Remark 3 (A measurability issue). Theorem 1 states that Prob$^N\{V(\hat{x}_N) \leq \epsilon\} \geq 1-\beta$. However, without any further assumption, there is no guarantee that $V(\hat{x}_N)$ is measurable, so that Prob$^N\{V(\hat{x}_N) \leq \epsilon\}$ may not be well-defined. Similar subtle measurability issues are often encountered in learning theory, see e.g. (A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth 1989). Here and elsewhere, the measurability of $V(\hat{x}_N)$, as well as that of other variables defined over $\mathbf{\Delta}^N$, is taken as an assumption. ⋆

Remark 4 (Comparison between RCP_N *and* RCP*).* Since the solution of RCP_N is obtained by looking at N constraints only, it is certainly superoptimal for the robust convex program RCP. Thus, RCP_N alleviates the conservatism inherent in the robust approach to uncertain optimization. On the other hand, we also remark that in general the optimal solution of PCP outperforms the optimal solution of RCP_N when such a solution is feasible for PCP. We do not insist here on this point and refer the reader to (G. Calafiore and M.C. Campi 2003) for more discussion. ⋆

Remark 5 (Computational complexity). Through the scenario approach, the initial semi-infinite optimization problem is reduced to an optimization problem that can be solved efficiently. On the other hand, a side problem arising along the scenario approach is that one has to extract constraints out of the uncertainty set. This is not always a simple task to accomplish, see (G. Calafiore, F. Dabbene and R. Tempo 2000) for a discussion of this topic and polynomial-time algorithms for the sample generation. ⋆

2.2 A-priori and a-posteriori assessments

It is worth noticing that a distinction should be made between the a-priori and a-posteriori assessments that one can make regarding the probability of constraint violation. Indeed, *before* running the optimization, it is guaranteed by Theorem 1 that if $N \geq N(\epsilon, \beta)$ samples are drawn, the solution of the scenario program will be ϵ-level robustly feasible, with probability no smaller than $1 - \beta$. However, the a-priori parameters ϵ, β are generally chosen to be not too small, due to technological limitations on the number of constraints that one specific optimization software can deal with.

On the other hand, once a solution has been computed (and hence $x = \hat{x}_N$ has been fixed), one can make an a-posteriori assessment of the level of feasibility using standard Monte-Carlo techniques. In this case, a new batch of $\tilde{N}$ independent random samples of $\delta \in \boldsymbol{\Delta}$ is generated, and the *empirical probability* of constraint violation, say $\hat{V}_{\tilde{N}}(\hat{x}_N)$, is computed according to the formula $\hat{V}_{\tilde{N}}(\hat{x}_N) = \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} 1(f(\hat{x}_N, \delta^{(i)}) > 0)$, where $1(\cdot)$ is the indicator function. Then, the classical Hoeffding's inequality, (W. Hoeffding 1963), guarantees that

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

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

$$\tilde{N} \geq \frac{\ln 2/\tilde{\beta}}{2\tilde{\epsilon}^2} \tag{7}$$

test samples are drawn. This latter a-posteriori verification can be easily performed using a very large sample size $\tilde{N}$, because no numerical optimization is involved in such an evaluation.

2.3 Multi-participant decision making

A decision making has sometimes to be made in presence of collective opinions coming from different participants in the optimization problem. In general situations, moving from a single decision maker to a multiple decision maker situation introduces a great deal of complexity into the analysis. Here, we merely refer to a simple situation where the different participants share the same objective to be minimized, while having different opinions on the uncertain environment in which the decision has to be made.

Let us enumerate the participants with $j = 1, \ldots, M$. We assume that all participants aim at minimizing the same objective ($\min_{x \in \mathcal{X} \subseteq R^n} c^T x$) while having different perception of the uncertain environment. In mathematical terms, this is expressed by saying that the constraints are $f(x, \delta_1) \leq 0$, $\delta_1 \in \Delta_1$, for the first participant; $f(x, \delta_2) \leq 0$, $\delta_2 \in \Delta_2$, for the second participant; and so on for all other participants. In a probabilistic chance-constrained framework, letting Prob_j be the probability associated to Δ_j, the problem is as follows:

$$\begin{aligned} \min_{x \in \mathcal{X} \subseteq R^n} c^T x \quad \text{subject to} \quad & \text{Prob}_1\{f(x, \delta_1) > 0\} \leq \epsilon, \\ & \text{Prob}_2\{f(x, \delta_2) > 0\} \leq \epsilon, \\ & \vdots \\ & \text{Prob}_M\{f(x, \delta_M) > 0\} \leq \epsilon. \end{aligned} \tag{8}$$

As an example of this setting, consider the VaR framework of Section 1. Suppose that more financial partners (the participants) want to decide how to allocate the assets ξ so that the loss be minimized and each participant is willing to run a risk at most ϵ (according to his/her viewpoint on the uncertain reality) that the loss is bigger than the found minimum value γ. Then, the problem can be mathematically formulated as:

$$\begin{aligned} \min_{\xi,\gamma} \gamma \quad \text{subject to} \quad & \text{Prob}_1\{\gamma < -r(\xi,\delta_1)\} \leq \epsilon, \\ & \text{Prob}_2\{\gamma < -r(\xi,\delta_2)\} \leq \epsilon, \\ & \vdots \\ & \text{Prob}_M\{\gamma < -r(\xi,\delta_M)\} \leq \epsilon. \end{aligned}$$

A feasible solution for problem (8) can be found by treating the problem as a set of M separate problems and by applying Theorem 1 to each single problem. This corresponds to selecting a β and then drawing $N(\epsilon,\beta)$ constraints $\delta_1 \in \Delta_1$, an equal number of constraints $\delta_2 \in \Delta_2$ and so on for all participants. An application of Theorem 1 leads then to the conclusion that each participant has a confidence β that his constraints are satisfied with a violation probability at most ϵ. However, a different standpoint can be taken by asking a deeper question: to what extent the constraints associated to one participant are going to help the constraint satisfaction of other participants?

For the sake of simplicity, consider the case of two participants ($M = 2$). A very simple situation arises when the two participants share the same set of constraints ($\Delta_1 = \Delta_2 = \Delta$) with , say, Prob_2 absolutely continuous with respect to Prob_1 with a uniform coefficient α: $\text{Prob}_2(A) \leq \alpha\text{Prob}_1(A)$, $\forall A$ in the σ-algebra on Δ. A simple reasoning then reveals that if the first participant extracts $N(\epsilon_1,\beta_1)$ constraints according to Prob_1, then, with no extractions whatsoever, the second participants has, with confidence $\beta_2 = \alpha^N\beta_1$, a violation probability $\epsilon_2 = \alpha\epsilon_1$.

The above result is rather obvious since the extractions of the first participant correspond to extractions for the second participant, even though according to a different probability. A more intriguing situation occurs when Δ_1 and Δ_2 are different. Is then still true that the two participants are going to help each other? The following result holds: *fix ϵ and β and extract $N(\epsilon,\beta)$ constraints δ_1 according to* Prob_1 *and $N(\epsilon,\beta)$ constraints δ_2 according to* Prob_2. *Solve the optimization problem with all the constraints in place. Then, with probability no smaller than $1-\beta$, the optimal solution is robustly feasible for the first and the second participant with a violation probability ϵ_1 and ϵ_2 respectively that have to satisfy the relation:* $\epsilon_1+\epsilon_2-\epsilon_1\epsilon_2 \leq \epsilon$. Thus, if $\epsilon_1 = \epsilon$, then ϵ_2 has to be zero. This correspond to say that if the extractions of the second participant do not help the first participant, then the constraints of the second participant are 'dominated' by those of the first participant, and are therefore automatically satisfied. Intermediate situations can be studied as well. So, in all cases, the two participants are helping each other somehow, even though it may a-priory be unknown who is helping whom.

A sketch of proof of the above result is as follows. The $N(\epsilon,\beta)$ extractions of δ_1's and δ_2's can be seen as $N(\epsilon,\beta)$ extractions of couples (δ_1,δ_2) from $\Delta_1 \times \Delta_2$ with probability $\text{Prob}_1 \times \text{Prob}_2$. Theorem 1 can then be applied to this space, so concluding that, with confidence β, the violation probability in $\Delta_1 \times \Delta_2$ is at most ϵ. Consider a multi-sample such that the violation probability is $\leq \epsilon$. If a given $\bar{\delta}_1$ is violated, then $(\bar{\delta}_1,\delta_2)$ is violated in the product space, $\forall\delta_2$. So, if ϵ_1 is the probability of violation in Δ_1, this leads to a probability of violation ϵ_1 in the product space too. Similarly, if ϵ_2 is the probability of violation in Δ_2, this leads to a probability of violation ϵ_2 in the product space. This two sets overlap with a probability $\epsilon_1\epsilon_2$ and there union has a probability bounded by ϵ, so leading to the result above.

3 Technical preliminaries and proof of Theorem 1

This section is technical and contains the machinery needed for the proof of Theorem 1. The reader not interested in these mathematical aspects may skip to Section 4 without any loss of continuity.

3.1 Preliminaries

We first recall a classical result due to Helly, see (R.T. Rockafellar 1970).

Lemma 1 (Helly). *Let $\{\mathcal{X}_i\}_{i=1,\ldots,p}$ be a finite collection of convex sets in R^n. If every sub-collection consisting of $n+1$ sets has a non-empty intersection, then the entire collection has a non-empty intersection.*

⋆

Next, we prove a key instrumental result. Consider the convex optimization program

$$\mathcal{P} : \min_{x \in R^n} c^T x \quad \text{subject to } x \in \bigcap_{i \in \{1,\ldots,m\}} \mathcal{X}_i, \tag{9}$$

where $\mathcal{X}_i$, $i = 1, \ldots, m$, are closed convex sets. Define the convex programs $\mathcal{P}_k$, $k = 1, \ldots, m$, obtained from $\mathcal{P}$ by removing the k-th constraint:

$$\mathcal{P}_k : \min_{x \in R^n} c^T x \quad \text{subject to } x \in \bigcap_{i \in \{1,\ldots,m\} \setminus k} \mathcal{X}_i. \tag{10}$$

Let x^* be any optimal solution of $\mathcal{P}$ (assuming it exists), and let x_k^* be any optimal solution of $\mathcal{P}_k$ (again, assuming it exists). We have the following definition.

Definition 4 (support constraint). The k-th constraint $\mathcal{X}_k$ is a *support constraint* for $\mathcal{P}$ if problem $\mathcal{P}_k$ has an optimal solution x_k^* such that $c^T x_k^* < c^T x^*$. ⋆

The following theorem holds.

Theorem 2. *The number of support constraints for problem $\mathcal{P}$ is at most n.* ⋆

A proof of this result was originally given by the authors of the present paper in (G. Calafiore and M.C. Campi 2003). We here report an alternative and more compact proof whose outline was suggested us by A. Nemirovski in a personal communication.

Proof. Let problem $\mathcal{P}$ have q support constraints $\mathcal{X}_{s_1}, \ldots, \mathcal{X}_{s_q}$, where $\mathcal{S} \doteq \{s_1, \ldots, s_q\}$ is a subset of q indices from $\{1, \ldots, m\}$. We next prove (by contradiction) that $q \leq n$.

Let x^* be any optimal solution of $\mathcal{P}$, with corresponding optimal objective value $J^* = c^T x^*$, and let x_k^* be any optimal solution of $\mathcal{P}_k$, $k = 1, \ldots, m$, with corresponding optimal objective value $J_k^* = c^T x_k^*$. Consider the smallest objective improvement obtained by removing a support constraint

$$\eta_{\min} \doteq \min_{k \in \mathcal{S}} (J^* - J_k^*)$$

and, for some η with $0 < \eta < \eta_{\min}$, define the hyperplane

$$\mathcal{H} \doteq \{x : c^T x = J^* - \eta\}.$$

By construction, the q points x_k^*, $k \in \mathcal{S}$, lie in the half-space $\{x : c^T x < J^* - \eta\}$, while x^* lies in the half-space $\{x : c^T x > J^* - \eta\}$, and therefore $\mathcal{H}$ separates x_k^*, $k \in \mathcal{S}$, from x^*. Next, for all indices $k \in \mathcal{S}$, we denote with $\overline{x}_k^*$ the point of intersection between the line segment $\overline{x_k^* x^*}$ and $\mathcal{H}$.

Since $x_k^* \in \bigcap_{i \in \{1,\ldots,m\} \setminus k} \mathcal{X}_i$, $k \in \mathcal{S}$, and $x^* \in \bigcap_{i \in \{1,\ldots,m\}} \mathcal{X}_i$, then by convexity we have that $\overline{x}_k^* \in \bigcap_{i \in \{1,\ldots,m\} \setminus k} \mathcal{X}_i$, $k \in \mathcal{S}$, and therefore (since, by construction, $\overline{x}_k^* \in \mathcal{H}$)

$$\overline{x}_k^* \in \left(\bigcap_{i \in \{1,\ldots,m\} \setminus k} \mathcal{X}_i \right) \bigcap \mathcal{H}, \quad k \in \mathcal{S}.$$

For $i = 1, \ldots, m$, define the convex sets $\Omega_i \doteq \mathcal{X}_i \bigcap \mathcal{H}$, and consider any collection $\{\Omega_{i_1}, \ldots, \Omega_{i_n}\}$ of n of these sets.

Suppose now (for the purpose of contradiction) that $q > n$. Then, there must exist an index $j \notin \{i_1, \ldots, i_n\}$ such that $\mathcal{X}_j$ is a support constraint, and by the previous reasoning, this means that there exists a point $\overline{x}_j^*$ such that $\overline{x}_j^* \in \left(\bigcap_{i \in \{1,\ldots,m\} \setminus j} \mathcal{X}_i \right) \bigcap \mathcal{H}$. Thus, $\overline{x}_j^* \in \Omega_{i_1} \cap \cdots \cap \Omega_{i_n}$, that is the collection of convex sets $\{\Omega_{i_1}, \ldots, \Omega_{i_n}\}$ has at least a point in common. Now, since the sets Ω_i, $i = 1, \ldots, m$, belong to the hyperplane $\mathcal{H}$ — i.e. to R^{n-1}, modulo a fixed translation — and all collections composed of n of these sets have a point in common, by Helly's lemma (Lemma 1) there exists a point $\tilde{x}$ such that $\tilde{x} \in \bigcap_{i \in \{1,\ldots,m\}} \Omega_i$. Such a $\tilde{x}$ would therefore be feasible for problem $\mathcal{P}$; moreover, it would yield an objective value $\tilde{J} = c^T \tilde{x} < c^T x^* = J^*$ (since $\tilde{x} \in \mathcal{H}$). This is a contradiction, because x^* would no longer be an optimal solution for $\mathcal{P}$, and hence we conclude that $q \leq n$. ⋆

We are now ready to present a proof of Theorem 1.

3.2 Proof of Theorem 1

We prove that the conclusion in Theorem 1 holds with

$$N \geq \frac{1}{1-\gamma}\left(\frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\ln\frac{1}{\gamma\epsilon} + \frac{1}{\epsilon}\ln\left(\left(\frac{n}{e}\right)^n \frac{1}{n!}\right)\right), \tag{11}$$

where γ is a parameter that can be freely selected in $(0, 1)$. To prove that bound (6) follows from (11), proceed as follows. Since $n! \geq (n/e)^n$, the last term in (11) is not positive and can be dropped, leading to the bound

$$N \geq \frac{1}{1-\gamma}\left(\frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\ln\frac{1}{\gamma\epsilon}\right), \tag{12}$$

Bound (6) is then obtained from (12), by taking $\gamma = 1/2$. We also note that further optimizing (12) with respect to γ always leads to $\gamma \leq 1/2$ with a corresponding improvement by at most of a factor 2.

Proof of Theorem 1 with (11) in place of (6)

We shall prove that, with probability $1-\beta$, the solution of RCP$_N$ is ϵ-level robustly feasible. This part of the proof in inspired by a similar proof given in a different context in (S. Floyd and M. Warmuth 1995).

Given N scenarios $\delta^{(1)}, \dots, \delta^{(N)}$, select a subset $I = \{i_1, \dots, i_n\}$ of n indexes from $\{1, \dots, N\}$ and let $\hat{x}_I$ be the optimal solution of the program

$$\min_{x \in \mathcal{X} \subseteq R^n} c^T x \quad \text{subject to } f(x, \delta^{(i_j)}) \leq 0,\ j = 1, \dots, n. \tag{13}$$

Based on $\hat{x}_I$ we next introduce a subset $\mathbf{\Delta}_I^N$ of the set $\mathbf{\Delta}^N$ defined as

$$\mathbf{\Delta}_I^N \doteq \{(\delta^{(1)}, \dots, \delta^{(N)}) : \ \hat{x}_I = \hat{x}_N\} \tag{14}$$

($\hat{x}_N$ is the optimal solution with all N constraints $\delta^{(1)}, \dots, \delta^{(N)}$ in place).

Let now I range over the collection $\mathcal{I}$ of all possible choices of n indexes from $\{1, \dots, N\}$ (notice that $\mathcal{I}$ contains $\binom{N}{n}$ sets). We want to prove that

$$\mathbf{\Delta}^N = \bigcup_{I \in \mathcal{I}} \mathbf{\Delta}_I^N. \tag{15}$$

To show (15), take any $(\delta^{(1)}, \dots, \delta^{(N)}) \in \mathbf{\Delta}^N$. From the set of constraint $\delta^{(1)}, \dots, \delta^{(N)}$ eliminate a constraint which is not a support constraint (this is possible in view of Theorem 2 since $N > n$). The resulting optimization problem with $N-1$ constraints admits the same optimal solution $\hat{x}_N$ as the original problem with N constraints. Consider now the set of the remaining $N-1$ constraints and, among these, remove a constraint which is not a support constraint for the problem with $N-1$ constraints. Again, the optimal solution does not change. If we keep going this way until we are left with n constraints, in the end we still have $\hat{x}_N$ as optimal solution, showing that $(\delta^{(1)}, \dots, \delta^{(N)}) \in \mathbf{\Delta}_I^N$, where I is the set containing the n constraints remaining at the end of the process. Since this is true for any choice of $(\delta^{(1)}, \dots, \delta^{(N)}) \in \mathbf{\Delta}^N$, (15) is proven.

Next, let

$$B \doteq \{(\delta^{(1)}, \dots, \delta^{(N)}) : \ V(\hat{x}_N) > \epsilon\}$$

and

$$B_I \doteq \{(\delta^{(1)}, \dots, \delta^{(N)}) : \ V(\hat{x}_I) > \epsilon\}$$

We now have:

$$\begin{aligned} B &= B \cap \mathbf{\Delta}^N \\ &= B \cap (\cup_{I \in \mathcal{I}} \mathbf{\Delta}_I^N) \quad \text{(apply (15))} \\ &= \cup_{I \in \mathcal{I}} (B \cap \mathbf{\Delta}_I^N) \\ &= \cup_{I \in \mathcal{I}} (B_I \cap \mathbf{\Delta}_I^N). \quad \text{(because of (14))} \end{aligned} \tag{16}$$

A bound for Prob$^N(B)$ is now obtained by bounding Prob$(B_I \cap \mathbf{\Delta}_I^N)$ and then summing over $I \in \mathcal{I}$.

Fix any I, e.g. $I = \{1,\dots,n\}$ to be more explicit. The set $B_I = B_{\{1,\dots,n\}}$ is in fact a cylinder with base in the cartesian product of the first n constraint domains (this follows from the fact that condition $V(\hat{x}_{\{1,\dots,n\}}) > \epsilon$ only involves the first n constraints). Fix $(\bar{\delta}^{(1)},\dots,\bar{\delta}^{(n)}) \in$ base of the cylinder. For a point $(\bar{\delta}^{(1)},\dots,\bar{\delta}^{(n)},\delta^{(n+1)},\dots,\delta^{(N)})$ to be in $B_{\{1,\dots,n\}} \cap \mathbf{\Delta}^N_{\{1,\dots,n\}}$, constraints $\delta^{(n+1)},\dots,\delta^{(N)}$ must be satisfied by $\hat{x}_{\{1,\dots,n\}}$, for, otherwise, we would not have $\hat{x}_{\{1,\dots,n\}} = \hat{x}_N$, as it is required in $\mathbf{\Delta}^N_{\{1,\dots,n\}}$. But, $V(\hat{x}_{\{1,\dots,n\}}) > \epsilon$ in $B_{\{1,\dots,n\}}$. Thus, by the fact that the extractions are independent, we conclude that

$$\begin{aligned}\text{Prob}^{N-n}\{(\delta^{(n+1)},\dots,\delta^{(N)}):\ &(\bar{\delta}^{(1)},\dots,\bar{\delta}^{(n)},\delta^{(n+1)},\dots,\delta^{(N)}) \\ &\in B_{\{1,\dots,n\}} \cap \mathbf{\Delta}^N_{\{1,\dots,n\}}\} < (1-\epsilon)^{N-n}.\end{aligned}$$

The probability on the left hand side is nothing but the conditional probability that $(\delta^{(1)},\dots,\delta^{(N)}) \in B_{\{1,\dots,n\}} \cap \mathbf{\Delta}^N_{\{1,\dots,n\}}$ given $\delta^{(1)} = \bar{\delta}^{(1)},\dots,\delta^{(n)} = \bar{\delta}^{(n)}$. Integrating over the base of the cylinder $B_{\{1,\dots,n\}}$, we then obtain

$$\text{Prob}^N(B_{\{1,\dots,n\}} \cap \mathbf{\Delta}^N_{\{1,\dots,n\}}) < (1-\epsilon)^{N-n} \cdot \text{Prob}^n(\text{base of } B_{\{1,\dots,n\}}) \le (1-\epsilon)^{N-n}. \tag{17}$$

From (16), we finally arrive to the desired bound for $\text{Prob}^N(B)$:

$$\text{Prob}^N(B) \le \sum_{I\in\mathcal{I}} \text{Prob}^N(B_I \cap \mathbf{\Delta}_I) < \binom{N}{n}(1-\epsilon)^{N-n}. \tag{18}$$

The last part of the proof is nothing but algebraic manipulations on bound (18) to show that, if N is chosen according to (11), then

$$\binom{N}{n}(1-\epsilon)^{N-n} \le \beta, \tag{19}$$

so concluding the proof.

Any of the following inequality implies the next in a top-down fashion, where the first one is (11):

$$\begin{aligned}N &\ge \frac{1}{1-\gamma}\left(\frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\ln\frac{1}{\gamma\epsilon} + \frac{1}{\epsilon}\ln\left(\left(\frac{n}{e}\right)^n\frac{1}{n!}\right)\right) \\ (1-\gamma)N &\ge \frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\ln\frac{1}{\gamma\epsilon} + \frac{1}{\epsilon}\ln\left(\left(\frac{n}{e}\right)^n\frac{1}{n!}\right) \\ (1-\gamma)N &\ge \frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\left(\ln\frac{n}{\gamma\epsilon} - 1\right) - \frac{1}{\epsilon}\ln(n!) \\ N &\ge \frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\left(\ln\frac{n}{\gamma\epsilon} - 1 + \frac{\gamma N\epsilon}{n}\right) - \frac{1}{\epsilon}\ln(n!) \\ N &\ge \frac{1}{\epsilon}\ln\frac{1}{\beta} + n + \frac{n}{\epsilon}\ln N - \frac{1}{\epsilon}\ln(n!),\end{aligned} \tag{20}$$

where the last implication can be justified by observing that $\ln x \ge 1 - \frac{1}{x}$, for $x > 0$, and applying this inequality with $x = \frac{n}{\gamma N\epsilon}$. Proceeding from (20), the next inequalities in the chain are

$$\begin{aligned}\ln\beta &\ge -\epsilon N + \epsilon n + n\ln N - \ln(n!) \\ \beta &\ge \frac{N^n}{n!}e^{-\epsilon(N-n)} \\ \beta &\ge \frac{N(N-1)\cdot\dots\cdot(N-n+1)}{n!}(1-\epsilon)^{N-n},\end{aligned}$$

where, in the last implication, we have used the fact that $e^{-\epsilon(N-n)} \ge (1-\epsilon)^{N-n}$, as it follows by taking logarithm of the two sides and further noting that $-\epsilon \ge \ln(1-\epsilon)$. Proceeding,

$$\beta \ge \binom{N}{n}(1-\epsilon)^{N-n}, \tag{21}$$

which is (19).

4 A numerical example

As a simple numerical example of application of the scenario methodology, we consider the following linear program with uncertain constraint matrix

$$\min_{x \in R^5} c^T x \quad \text{subject to } (A + A_\delta)x \leq b, \tag{22}$$

where

$$A \begin{bmatrix} 13 & -3 & -24 & 7 & -4 \\ 19 & 2 & -11 & 7 & 14 \\ 7 & 6 & -4 & 6 & -6 \\ 8 & -6 & -21 & -1 & 2 \\ -2 & 2 & 15 & -12 & 7 \\ -1 & 3 & 2 & 21 & -10 \\ -9 & 5 & 6 & -14 & 6 \\ 4 & -7 & -12 & 4 & 17 \\ 12 & 13 & 1 & 3 & 0 \\ 12 & 9 & 16 & 20 & 25 \end{bmatrix}$$

$$c^T = \begin{bmatrix} 0 & -1 & -1 & 0 & 0 \end{bmatrix}$$

$$b^T = \begin{bmatrix} -23 & 39 & -5 & -18 & 51 & 61 & 23 & 17 & -22 & 1 \end{bmatrix},$$

and where the entries of the uncertainty matrix $A_\delta \in R^{10,5}$ are independent Normal random variables with zero mean and variance equal to 0.5. If the uncertain constraints are imposed up to a given level of probability ϵ, we obtain the chance-constrained problem

$$\min_{x \in R^5} c^T x \quad \text{subject to Prob}\{(A + A_\delta)x \not\leq b\} \leq \epsilon. \tag{23}$$

Notice that this problem is readily put in the form (2), by rewriting the element-wise constraints in scalar form, i.e. setting $f(x, \delta) = \max_j [A + A_\delta]_j x - [b]_j$, where $[\cdot]_j$ here denotes the j-th row of its argument, and vector δ contains the uncertain entries of matrix A_δ.

Now, fixing probability levels $\epsilon = 0.01$, $\beta = 0.001$, and using bound (6), we obtain

$$N(\epsilon, \beta) \simeq 6689.9.$$

Fixing then $N = 6690$, we solve the scenario problem

$$\min_{x \in R^5} c^T x \quad \text{subject to } (A + A_{\delta^{(i)}})x \leq b, \quad i = 1, \ldots, N = 6690, \tag{24}$$

where $A_{\delta^{(i)}}$ are the sampled uncertainty matrices, extracted in accordance with the assumed probability distribution (i.e. each entry is Normal with zero mean and variance equal to 0.5). One instance of the scenario problem then yielded an optimal objective $J_{RCP_N} = -3.2750$, which was attained for

$$\hat{x}_N = \begin{bmatrix} -2.0973 \\ -1.1436 \\ 4.4186 \\ 0.8752 \\ -2.4418 \end{bmatrix}.$$

Notice that the nominal problem (i.e. the problem obtained by fixing $A_\delta = 0$) yields optimal objective $J_{\text{nom}} = -5.3901$, which is attained for

$$x_{\text{nom}} \begin{bmatrix} -2.0958 \\ -0.0719 \\ 5.4620 \\ -0.4594 \\ -5.7843 \end{bmatrix}.$$

We next proceeded with an a-posteriori Monte-Carlo test on the obtained solutions. Fixing $\tilde{\epsilon} = 0.001$ and $\tilde{\beta} = 0.00001$, we have from (7) that the a-posteriori test should be run using at least 6.1030×10^6 samples. Setting $\tilde{N} = 6.1030 \times 10^6$ we obtained

$$\hat{V}_{\tilde{N}}(\hat{x}_N) = 0.0009,$$

leading to the estimate $V_{\tilde{N}}(\hat{x}_N) \leq \hat{V}_{\tilde{N}}(\hat{x}_N) + \tilde{\epsilon} = 0.0019$. Interestingly, running the same test on the nominal solution yielded

$$\hat{V}_{\tilde{N}}(x_{\rm nom}) = 0.969$$

showing, as expected, that while the nominal problem provides a better optimal objective with respect to the scenario problem, the nominal solution is actually infeasible for most of the scenarios that may happen in reality. Contrary, the scenario solution provides a remarkable level of robustness, being in practice feasible in all but $\sim 0.2\%$ of the possible situations.

A Relaxing the assumption that RCP_N has a unique solution

In Section 2, the theory has been developed under Assumption 1 requiring that RCP_N admits a unique optimal solution. Here, we drop Assumption 1 and consider the general case allowing for non-uniqueness of the solution or non-existence of the solution (i.e. the solution escapes to infinity), or even infeasibility of RCP_N.

A.1 Non-uniqueness of the solution

Suppose that when problem RCP_N admits more than one optimal solution we break the tie by a tie-break rule as follows:

Tie-break rule: *Let $t_i(x)$, $i = 1, \dots, p$, be given convex and continuous functions. Among the optimal solutions for RCP$_N$, select the one that minimizes $t_1(x)$. If indetermination still occurs, select among the x that minimize $t_1(x)$ the solution that minimizes $t_2(x)$, and so on with $t_3(x), t_4(x), \dots$. We assume that functions $t_i(x)$, $i = 1, \dots, p$, are selected so that the tie is broken within p steps at most. As a simple example of a tie-break rule, one can consider $t_1(x) = x_1, t_2(x) = x_2, \dots$.* $\star$

From now on, by ‘optimal solution’ we mean either the unique optimal solution, or the solution selected according to the Tie-break rule, if multiple optimal solutions occur.

Theorem 1 holds unchanged if we drop the uniqueness requirement in Assumption 1, provided that ‘optimal solution’ is intended in the indicated sense.

To see this, generalize Definition 4 of support constraints to: *The k-th constraint $\mathcal{X}_k$ is a support constraint for $\mathcal{P}$ if problem $\mathcal{P}_k$ has an optimal solution x_k^* such that $x_k^* \neq x^*$*. Indeed this definition generalizes Definition 4 since, in case of a single optimal solution, $x_k^* \neq x^*$ is equivalent to $c^T x_k^* < c^T x^*$. In (G. Calafiore and M.C. Campi 2003), Section 4.1, it is proven that Theorem 2 holds true with this extended definition of support constraint (i.e. the number of support constraints is at most n), and then an inspection of the proof of Theorem 1 reveals that this proof goes through unaltered in the present setting, so concluding that Theorem 1 still holds.

A.2 Infeasibility of RCP_N

It may happen that RCP_N is infeasible (i.e. the intersection of the domains where $f(x, \delta^{(i)}) \leq 0, i = 1, \dots, N$ is empty), in which case the initial RCP is clearly infeasible too. In this case, going through the proof of Theorem 1 reveals that this theorem remains valid with small amendments: the first part remains unchanged, while the final part reads: ”*..., with probability no smaller than $1 - \beta$, either the scenario problem* RCP_N *is infeasible; or it is feasible, and then its optimal solution $\hat{x}_N$ is ϵ-level robustly feasible.*”

A.3 Non-existence of the solution

Even when RCP_N is feasible, it may happen that no optimal solution exists since the set for x allowed by the extracted constraints is unbounded in such a way that the optimal solution escapes to infinity. In this section, we further generalize Theorem 1 so as to cope with this situation too and then provide a reformulation of Theorem 1 (Theorem 3 below) that covers all possible situations as indicated in Sections A.1, A.2 and the first part of this section.

Suppose that a random extraction of a multi-sample $\delta^{(1)}, \ldots, \delta^{(N)}$ is rejected when the problem is feasible but no optimal solution exists, and another extraction is performed in this case. Then, the result of Theorem 1 holds if attention is restricted to the accepted multi-samples. This idea is now formalized.

Let $D \subseteq \mathbf{\Delta}^N$ be the set where RCP_N is feasible but an optimal solution does not exist (it escapes to infinity) and assume that its complement $A = D^c$ has positive probability: $\mathrm{Prob}^N(A) > 0$. Moreover, let Prob_A^N be the probability Prob^N restricted to A: $\mathrm{Prob}_A^N(\cdot) \doteq \mathrm{Prob}(\cdot \cap A)/\mathrm{Prob}^N(A)$. Prob_A^N is therefore a conditional probability. In addition, assume that if a problem with, say, m constraints is feasible and admits optimal solution, then, after adding an extra $(m+1)$-th constraint, if the problem remains feasible, than an optimal solution continues to exist (this rules out the possibility of pathological situations where adding a constraint forces the solution to drift away to infinity).

The following theorem holds:

Theorem 3. *Fix two real numbers $\epsilon \in (0,1)$ (level parameter) and $\beta \in (0,1)$ (confidence parameter). If*

$$N \geq N(\epsilon, \beta) \doteq \frac{2}{\epsilon} \ln \frac{1}{\beta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon},$$

then, with probability Prob_A^N *no smaller than $1-\beta$, either the scenario problem* RCP_N *is infeasible; or it is feasible, and then its optimal solution $\hat{x}_N$ (unique after the Tie-break rule has been applied) is ϵ-level robustly feasible.* ⋆

The proof of this theorem is here omitted and can be found in (G. Calafiore and M.C. Campi 2003).

Acknowledgments

Research supported by MIUR under the project “New techniques for the identification and adaptive control of industrial systems” and under the FIRB project “Learning, randomization and guaranteed predictive inference for complex uncertain systems”.

References

A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth (1989), “Learnability and the Vapnik-Chervonenkis dimension,” *Journal of the ACM*, vol. 36, pp. 929–965.

G. Calafiore and M.C. Campi (2002), “Interval predictors for unknown dynamical systems: an assessment of reliability,” *Proc. IEEE Conf. on Decision and Control*, Las Vegas, USA.

G. Calafiore and M.C. Campi (2003), “A learning theory approach to the construction of predictor models,” *Discrete and Continuous Dynamical Systems*, pp. 156–166.

G. Calafiore and M.C. Campi (2003), “Robust convex programs: Randomized solutions and applications in control,” *Proc. IEEE Conf. on Decision and Control*, Maui, Hawaii, USA.

G. Calafiore and M.C. Campi (2003), “The scenario approach to robust control design,” *Technical Report*.

G. Calafiore and M.C. Campi (2003), “Uncertain convex programs: randomized solutions and confidence levels,” *Mathematical Programming*, in press.

G. Calafiore, F. Dabbene and R. Tempo (2000), “Randomized algorithms for probabilistic robustness with real and complex structured uncertainty,” *IEEE Trans. Aut. Control*, vol. 45, pp. 2218–2235.

A. Charnes and W.W. Cooper (1959), “Uncertain convex programs: randomized solutions and confidence levels,” *Management Sci.*, vol. 6, pp. 73–79.

S. Floyd and M. Warmuth (1995), “Sample compression, learnability, and the vapnik-chervonenkis dimension,” *Machine Learning*, pp. 1–36.

W. Hoeffding (1963), “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, vol. 58, pp. 13–30.

T. J. Linsmeier and N. D. Pearson (1996), “Risk measurement: An introduction to value-at-risk,” *Technical report 96-04, OFOR, University of Illinois, Urbana-Champaign*.

A. Nemirovski (1993), “Several NP-hard problems arising in robust stability analysis,” *SIAM J. on Matrix Analysis and Applications*, vol. 6, pp. 99–105.

A. Prékopa, G.M. (1995), *Stochastic Programming*, Kluwer.

M. Pritsker (1997), “Evaluating value-at-risk methodologies: Accuracy vs. computational time,” *J. Financial Services Res.*, vol. 12(2/3), pp. 201 242.

R.T. Rockafellar (1970), *Stochastic Programming*, Princeton University Press, Princeton, NJ.
S. Vajda (1972), *Probabilistic Programming*, Academic Press, New York.
V. Vapnik (1996), *Statistical Learning Theory*, Wiley, New York.