Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Average and Worst-Case Techniques in Convex Optimization with Stochastic Uncertainty

Giuseppe Calafiore and Fabrizio Dabbene

Abstract— **We consider two standard philosophies for finding minimizing solutions of convex objective functions affected by uncertainty. In a first approach, the solution should minimize the expected value of the objective w.r.t. uncertainty (*average* approach), while in a second one it should minimize the worst-case objective (*worst-case*, or min-max approach). Both approaches are however numerically hard to solve exactly, for general dependence of the cost function on the uncertain data. Here, we discuss two techniques based on uncertainty randomization that permit to solve efficiently some suitable probabilistic relaxation of the indicated problems, with full generality with respect to the way in which the uncertainty enters the problem data. A specific application to uncertain Least-Squares problems is also examined in the paper.**

I. INTRODUCTION

A convex optimization program is usually formulated as the problem of minimizing a convex cost function $f(x)$ over a convex set $\mathcal{X}$, i.e.

$$\min_{x\in\mathcal{X}} f(x).$$

In this paper, we consider the situation where the objective function may be affected by uncertainty. In other words, the function f depends not only on the decision vector x but also on a vector of uncertain parameters δ, that are assumed to be random with known distribution over a compact set $\Delta \subset \mathbb{R}^\ell$. In this setting, the formulation of the optimization problem and the meaning of solution need to be clarified. In fact, it is possible to devise different paradigms, that consider the effect of the uncertainty from different viewpoints. One may consider the problem from a min-max viewpoint, and look for a solution that minimizes the "worst-case" value (w.r.t. the uncertainty) of the objective function. Alternatively, one may be interested in considering the effect of the uncertainty "on average," and this corresponds to minimizing an uncertainty-averaged version of the cost function. In this paper, we present these two standard philosophies and we discuss recent techniques based on uncertainty randomization that allow for efficient solution of some suitable probabilistic relaxation of the indicated problems.

This work has been supported by FIRB funds from the Italian Ministry of University and Research.

G. Calafiore is with the faculty of Dipartimento di Automatica e Informatica, Politecnico di Torino – Italy giuseppe.calafiore@polito.it

F. Dabbene is with IEIIT-CNR, Politecnico di Torino – Italy fabrizio.dabbene@polito.it

II. AVERAGE AND WORST-CASE DESIGN

We formally state the two classes of optimization problems that are the object of our study.

Consider an objective function $f(x,\delta) : \mathcal{X} \times \Delta \to \mathbb{R}$, where $\mathcal{X} \subseteq \mathbb{R}^n$ is a convex set and $\Delta \subset \mathbb{R}^\ell$ is a compact set. Denote by $\mathbb{P}$ the probability measure on δ.

We make the standing assumption that the objective is convex in the decision variables for any fixed value of the uncertainty:

Assumption 1: The function $f(x,\delta)$ is convex in x for every $\delta \in \Delta$.

A. Worst-case approach

In the worst-case optimization approach one seeks for a solution guaranteed for all possible values taken by the uncertainty $\delta \in \Delta$. This leads to the following *min-max* optimization problem

$$\min_{x\in\mathcal{X}} \max_{\delta\in\Delta} f(x,\delta). \tag{1}$$

Problem (1) is usually restated in an epigraphic form as a robust optimization problem:

$$\min_{t,x\in\mathcal{X}} \quad t \quad \text{subject to} \quad f(x,\delta) \le t, \quad \forall \delta \in \Delta \tag{2}$$

These problems are in general hard to solve numerically (see e.g. [3], [4]) and computable solutions are available only when the uncertainty enters the function f in a simple form (such as affine), see for instance [10]. This motivates our approximate approach based on sampling relaxations.

B. Average approach

In the average approach one aims at solving the stochastic program

$$\min_{x\in\mathcal{X}} E_\delta[f(x,\delta)] \tag{3}$$

where $E_\delta[f(x,\delta)]$ is the expectation of $f(x,\delta)$ with respect to the probability measure $\mathbb{P}$ defined on $\delta \in \Delta$. For notation ease, we define the function

$$\phi_{\text{AV}}(x) \doteq E_\delta[f(x,\delta)]. \tag{4}$$

We denote by x^*_{AV} a minimizer of $\phi_{\text{AV}}(x)$, and define the achievable minimum as $\phi^*_{\text{AV}} \doteq \min_{x\in\mathcal{X}} \phi_{\text{AV}}(x) = \phi_{\text{AV}}(x^*_{\text{AV}})$. Problem (3) is a classical stochastic optimization problem, and it has been extensively treated in the literature,

see e.g. [13]. Again, determining an exact solution to a stochastic program is in general computationally prohibitive (indeed, just evaluating the expectation for a fixed x amounts to computing a multi-dimensional integral, which is numerically hard).

In the next section we follow a probabilistic approach, and propose approximate solutions for both worst-case and average problems, based on random sampling of the uncertainty.

III. SAMPLING-BASED APPROXIMATE SOLUTIONS

The main idea in randomized relaxations is a simple one: we collect a finite number N of random samples of the uncertainty

$$\delta^{(1)}, \delta^{(1)}, \ldots, \delta^{(N)} \tag{5}$$

extracted independently according to $\mathbb{P}$, and we construct a suitable "sampled" approximation of the problems previously considered.

We next show how to construct these approximations, and we discuss the theoretical properties of the solutions obtained by the approximated problems.

A. Worst-case design

A sampled approximation of problem (1) can be naturally stated as

$$\min_{x \in \mathcal{X}} \max_{i=1,\ldots,N} f(x, \delta^{(i)}). \tag{6}$$

This problem may be also rewritten in epigraph form as

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & t \quad \text{subject to} \\ & f(x, \delta^{(i)}) \leq t, \text{ for } i = 1, \ldots, N. \end{aligned} \tag{7}$$

Notice that in practice, in this latter formulation, the possibly infinite number of constraints of problem (2) is substituted by a finite number N of *sampled scenarios* of the uncertainty. This scenario approximation of the worst-case problem has been first introduced in [6], where it has been shown that the solution obtained from the scenario problem is actually approximately feasible for the original worst-case problem, in a sense explained next.

Define the *probability of violation* of x as

$$P_V(x) \doteq \mathbb{P}\{\delta \in \Delta : \ f(x, \delta) - t > 0\}.$$

For example, if a uniform (with respect to Lebesgue measure) probability distribution is assumed, then $P_V(x)$ measures the volume of 'bad' parameters δ such that the constraint $f(x, \delta) \leq t$ is violated. Clearly, a solution x with small associated $P_V(x)$ is feasible for most of the problem instances, i.e. it is *approximately feasible* for the robust problem.

Definition 1 (ϵ-level solution): Let $\epsilon \in (0, 1)$. We say that $x \in \mathcal{X}$ is an ϵ-level robustly feasible (or, more simply, an ϵ-level) solution, if $P_V(x) \leq \epsilon$.

The following theorem establishes the probabilistic properties of the scenario solution.

Theorem 1 (Corollary 1 of [5]): Assume that, for any extraction of $\delta^{(1)}, \ldots, \delta^{(N)}$, the scenario problem (7) attains a unique optimal solution $\hat{x}_{\text{WC}}^{(N)}$. Fix two real numbers $\epsilon \in (0, 1)$ (level parameter) and $\beta \in (0, 1)$ (confidence parameter) and let[1]

$$N \geq N_{\text{WC}}(\epsilon, \beta) \doteq \left\lceil \frac{2}{\epsilon} \ln \frac{1}{\beta} + 2(n{+}1) + \frac{2(n{+}1)}{\epsilon} \ln \frac{2}{\epsilon} \right\rceil \tag{8}$$

then, with probability no smaller than $(1 - \beta)$, $\hat{x}_{\text{WC}}^{(N)}$ is ϵ-level robustly feasible.

B. Average design

For the developments in this subsection, we need to state a further assumption on the function f, namely that the total-variation of the function is bounded.

Assumption 2: Let $f^*(\delta) \doteq \min_{x \in \mathcal{X}} f(x, \delta)$, and assume that the total variation of f is bounded by a constant $V > 0$, i.e.

$$f(x, \delta) - f^*(\delta) \leq V, \ \forall x \in \mathcal{X}, \forall \delta \in \Delta.$$

This implies that the total variation of the expected value is also bounded by V, i.e.

$$\phi_{\text{AV}}(x) - \phi_{\text{AV}}^* \leq V, \ \forall x \in \mathcal{X}.$$

Notice that we only assume that there exist a constant V such that the above holds, but do not need to actually know its numerical value.

An approximate solution to (3) may be obtained constructing an empirical estimate of ϕ_{AV} based on the random samples

$$\hat{\phi}_{\text{AV}}^{(N)}(x) \doteq \frac{1}{N} \sum_{i=1}^{N} f(x, \delta^{(i)}).$$

The function $\hat{\phi}_{\text{AV}}^{(N)}$ is convex in x, being the sum of convex functions.

Now, we study the convergence of the empirical minimum $\hat{\phi}_{\text{AV}}^{(N)}(\hat{x}_k)$ to the actual unknown minimum $\phi_{\text{AV}}(x^*)$. To this end, notice first that as x varies over $\mathcal{X}$, $f(x, \cdot)$ spans a family $\mathcal{F}$ of measurable functions of δ, namely

$$\mathcal{F} \doteq \{f(x, \delta) : x \in \mathcal{X}\}. \tag{9}$$

A key step for assessing convergence is to bound (in probability) the relative deviation between the actual mean $\phi_{\text{AV}}(x) = E_\delta[f(x, \delta)]$ and the empirical mean $\hat{\phi}_{\text{AV}}^{(N)}(x)$ *for all* $f(\cdot, \delta)$ belonging to the family $\mathcal{F}$. In other words, for given relative scale error $\epsilon \in (0, 1)$, we require that

$$\text{PR}\left\{ \sup_{x \in \mathcal{X}} \frac{|\phi_{\text{AV}}(x) - \hat{\phi}_{\text{AV}}^{(N)}(x)|}{V} > \epsilon \right\} \leq \alpha(N), \tag{10}$$

[1]The notation $\lceil \ \rceil$ denotes the smallest integer greater than or equal to the argument.

with $\alpha(N) \to 0$ as $N \to \infty$. Notice that the uniformity of the bound (10) with respect to x is crucial, since x is *not* fixed and known in advance: the uniform "closeness" of $\hat{\phi}_{\text{AV}}^{(N)}(x)$ to $\phi_{\text{AV}}(x)$ is the feature that allows us to perform the minimization on $\hat{\phi}_{\text{AV}}^{(N)}(x)$ instead of on $\phi_{\text{AV}}(x)$. Property (10) is usually referred to as the Uniform Convergence of the Empirical Mean (UCEM) property. A fundamental result of Learning Theory states that the UCEM property holds for a function class $\mathcal{F}$ whenever a particular measure of the complexity of the class, called the P-dimension of $\mathcal{F}$ (P-DIM $(\mathcal{F})$), is finite. The interested reader can refer to the monographs [16], [17] for formal definitions and further details.

The key result for average optimization is given in the next theorem.

Theorem 2: Let $\alpha, \epsilon \in (0,1)$, let d be an upper bound on the P-dim of $\mathcal{F}$ and let

$$N \geq N_{\text{AV}} \doteq \frac{128}{\epsilon^2}\left[\ln\frac{8}{\alpha} + d\left(\ln\frac{32\text{e}}{\epsilon} + \ln\ln\frac{32\text{e}}{\epsilon}\right)\right]. \tag{11}$$

Let x_{AV}^* be a minimizer of $\phi_{\text{AV}}(x)$, and let $\hat{x}_{\text{AV}}^{(N)}$ be a minimizer of the empirical mean $\hat{\phi}_{\text{AV}}^{(N)}(x)$. Then, it holds with probability at least $(1-\alpha)$ that

$$\frac{\phi_{\text{AV}}(\hat{x}_{\text{AV}}^{(N)}) - \phi_{\text{AV}}(x_{\text{AV}}^*)}{V} \leq \epsilon. \tag{12}$$

That is, $\hat{x}_{\text{AV}}^{(N)}$ is an ϵ-suboptimal solution (in the relative scale), with high probability $(1-\alpha)$. A solution $\hat{x}_{\text{AV}}^{(N)}$ such that the above holds is called an $(1-\alpha)$-probable ϵ-near minimizer of $\phi_{\text{AV}}(x)$, in the relative scale V.

Proof. See Appendix.

IV. APPLICATION: UNCERTAIN LEAST-SQUARES PROBLEMS

As an important application of the robust optimization techniques discussed in this paper, we here examine the case of uncertain Least-Squares problems.

In the standard Least-Squares (LS) problem, the objective is to determine a solution vector x^* such that the residual norm $\|Ax - y\|^2$ of a (usually over-determined) system of linear equations is minimized. However, in many practical applications the data matrices A, y are not exactly known. This uncertainty in the data can be modeled assuming A, y to be generic, possibly nonlinear functions of a vector of uncertain real parameters

$$A(\delta) \in \mathbb{R}^{m,n}, \quad y(\delta) \in \mathbb{R}^m, \quad \delta = [\delta_1\ \delta_2\ \cdots\ \delta_\ell]^T,$$

where the uncertain parameter δ is assumed to belong to a given compact set $\Delta \subset \mathbb{R}^\ell$. Thus, the uncertain LS problem falls in the problem family discussed in the previous sections, by taking

$$f(x,\delta) \doteq \|A(\delta)x - y(\delta)\|^2. \tag{13}$$

Notice that $f(x,\delta)$ is convex quadratic in x, for any fixed $\delta \in \Delta$.

The worst-case approach to uncertain LS is discussed for instance in the papers [2], [9], [14], and is closely related to Tikhonov-type regularization [15]. With the exception of special cases where the uncertainty has simple structure and enters f in a specific fashion (such as linear, see [11]), an exact worst-case solution for the uncertain LS problem is numerically hard to determine. Similarly, the average approach gives rise to a numerically hard problem, since the mere evaluation of the objective function, for fixed x, can be numerically prohibitive.

The randomized methods that we propose next provide efficient numerical techniques to solve uncertain LS problems, both in the average and worst-case setting, in all cases that cannot be tackled by means of the cited exact methods.

A. Approximate average solution of uncertain LS

To apply the general results of Section III-B, we simply need to determine a bound on the P-dimension of the function class

$$\mathcal{F} \doteq \{f(x,\delta) : x \in \mathcal{X}\}, \quad f(x,\delta) = \|A(\delta)x - y(\delta)\|^2. \tag{14}$$

This is established in the following lemma.

Lemma 1 (P-dimension of $\mathcal{F}$): Consider the function family $\mathcal{F}$ defined in (14). Then,

$$\text{P-DIM}\,(\mathcal{F}) \leq 9n.$$

Proof. Let $M = \sup_{x\in\mathcal{X},\delta\in\Delta} f(x,\delta)$, and define the family of binary valued functions $\bar{\mathcal{F}}$, whose elements are the functions

$$\bar{f}(x,\delta,c) \doteq \begin{cases} 1, & \text{if } f(x,\delta) \geq c \\ 0, & \text{otherwise,} \end{cases}$$

for $c \in [0, M]$. Then, from Lemma 10.1 in [17], we have that $\text{P-DIM}\,(\mathcal{F}) = \text{VC}\left(\bar{\mathcal{F}}\right)$, where $\text{VC}\left(\bar{\mathcal{F}}\right)$ denotes the Vapnik-Chervonenkis dimension of the class $\bar{\mathcal{F}}$. Notice that the functions in $\bar{\mathcal{F}}$ are quadratic in the parameter vector $x \in \mathbb{R}^n$, therefore a bound on the VC-dimension can be derived from a result of Karpinski and Macintyre, [12]

$$\text{VC}\left(\bar{\mathcal{F}}\right) \leq 2n\log_2(8e) < 9n.$$

□

The sample size bound for solving in the ϵ-approximate sense the average uncertain LS problem then simply obtains by plugging the P-dimension bound of Lemma 1 into Theorem 2. A result along these lines has been previously derived in [7].

1) Numerical computation of $\hat{x}_N$: While Theorem 2 provides the theoretical properties of $\hat{x}_N$, in this section we briefly discuss a simple numerical technique to compute it.

Notice that, in the specific LS context, the objective function $\hat{\phi}_{\text{AV}}^{(N)}(x)$ has a sum-of-squares structure

$$\hat{\phi}_{\text{AV}}^{(N)}(x) = \frac{1}{N}\sum_{i=1}^{N} \|A(\delta^{(i)})x - y(\delta^{(i)})\|^2 = \frac{1}{N}\|\mathcal{A}x - \mathcal{Y}\|^2$$

where

$$\mathcal{A} \doteq \begin{bmatrix} A(\delta^{(1)}) \\ A(\delta^{(2)}) \\ \vdots \\ A(\delta^{(N)}) \end{bmatrix}, \quad \mathcal{Y} \doteq \begin{bmatrix} y(\delta^{(1)}) \\ y(\delta^{(2)}) \\ \vdots \\ y(\delta^{(N)}) \end{bmatrix}.$$

Therefore, an exact minimizer of $\hat{\phi}_{\text{AV}}^{(N)}(x)$ can be readily computed as $\hat{x}_N = \mathcal{A}^{\dagger}\mathcal{Y}$, where $\mathcal{A}^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\mathcal{A}$. Remark that, since $\mathcal{A}, \mathcal{Y}$ are functions of $\delta^{(i)}$, $i = 1, \ldots, N$, the resulting solution $\hat{x}_N$ is a random quantity, whose probability distribution is defined over the product space $\Delta \times \Delta \times \cdots \times \Delta$ (N times). The solution $\hat{x}_N$ can be alternatively defined as the result given at the N-th iteration by the following standard recursive form of the LS algorithm.

Algorithm 1: Assuming that $A(\delta^{(1)})$ is full-rank, an exact minimizer $\hat{x}^{(N)}$ of the empirical mean $\hat{\phi}_{\text{AV}}^{(N)}(x)$ can be recursively computed as

$$\hat{x}^{(k)} = \hat{x}^{(k-1)} + K_k^{-1} A^T(\delta^{(k)}) \left(y(\delta^{(k)}) - A(\delta^{(k)})\hat{x}^{(k)} \right),$$

where

$$K_k = K_{k-1} + A^T(\delta^{(k)})A(\delta^{(k)}),$$

and the recursion for $k = 1, \ldots, N$ is started with $K_0 = 0$, $\hat{x}^{(0)} = 0$.

B. Approximate worst-case solution of uncertain LS

In the worst-case LS problem, we seek a solution x_{WC}^* that solves the min-max problem

$$\min_x \max_{\delta \in \Delta} \|A(\delta)x - y(\delta)\|. \tag{15}$$

Assuming a probability distribution on Δ, the scenario counterpart of this problem is

$$\min_{x,t} \quad t \quad \text{subject to:} \tag{16}$$

$$\|A(\delta^{(i)})x - y(\delta^{(i)})\| \le t, \quad i = 1, \ldots, N \tag{17}$$

where $\delta^{(i)}$, $i = 1, \ldots, N$ are iid random samples of $\delta \in \Delta$. Call $\hat{x}_N, \hat{t}_N$ the resulting optimal solution of (16).

The results of Theorem 1 can be applied directly to this problem, concluding that if the scenario solution is computed using a number of scenarios compatible with bound (8), then $\hat{x}_N$ guarantees (with high probability $1-\beta$) that all residuals, except possibly those in an exceptional subset with small probability measure ($\le \epsilon$), are smaller than $\hat{t}_N$. From a numerical point of view, solving (16) amounts to solving a convex second-order cone program (SOCP), which can be done in a computationally efficient way, see for instance [1].

V. Example

In this section, we consider a numerical example of a least-squares problem with affine uncertainty on the matrix A. Since in this case both the worst-case solution x_{WC}^* and x_{AV}^* can be computed exactly as shown in Subsection V-A, we can directly test the quality of the randomized solutions proposed in Section III against the exact solution. Let

$$A(\delta) = A_0 + \sum_{i=1}^{3} \delta_i A_i, \quad y^T = \begin{bmatrix} 0 & 2 & 1 & 3 \end{bmatrix},$$

with

$$A_0 = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 1 & 1 \\ -2 & 5 & 3 \\ 1 & 4 & 5.2 \end{bmatrix}, A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & .1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0 & 0 & .2 \\ 0 & .2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ .4 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Consider the uncertainties vector $\delta = [\delta_1 \quad \delta_2 \quad \delta_3]^T$ to be uniformly distributed over the set $[-1, 1]^3$.

A. Exact solutions

In the special case of affine uncertainty, it is possible to compute the exact min–max solution. In fact, it has been shown in [9], [11] that x_{WC}^* is the optimal solution of the following semi-definite programming problem

$$\min \ \lambda \qquad \text{subject to}$$

$$\begin{bmatrix} \lambda - \tau & 0 & (A_0 x - y)^T \\ * & \tau I & M(x)^T \\ * & * & I \end{bmatrix} \succeq 0$$

where $M(x) = [A_1 x - y \quad A_2 x - y \quad A_3 x - y]$. The optimal min-max solution hence resulted to be

$$x_{\text{WC}}^* = \begin{bmatrix} -0.0700 \\ 0.1909 \\ 0.2576 \end{bmatrix}$$

and the corresponding worst-case residual is given by

$$r_{WC} = \lambda^{1/2} = 2.2781.$$

In can be easily shown that also for the average problem, the minimizing solutions can be computed exactly. In particular, in [11], [8] it is shown that x_{AV}^* can be computed as the solution of the modified normal equations

$$(A_0^T A_0 + \sum_{i=1}^{\ell} \sigma_i^2 A_i^T A_i)x = A_0^T y, \tag{18}$$

being $\sigma_i^2 \doteq E_{\delta_i}[\delta_i^2]$, $i = 1, \ldots, \ell$ the covariances of δ_i. In our case, since the covariance of a uniform variable in $[-1, 1]$ is $\sigma^2 = 1/3$, we immediately obtained

$$x^*_{\text{AV}} = \begin{bmatrix} -1.9247 \\ -1.6490 \\ 2.0626 \end{bmatrix}.$$

B. Approximate solutions

We set $\epsilon = 0.1$, $\beta = 0.001$ and computed the scenario bound of Theorem 1 as

$$N_{WC} = 386.$$

This number of samples guarantees that[2] the approximate solution $\hat{x}^{(N)}_{\text{WC}}$ will satisfy the constraints with probability larger than 0.9. Solving the sampled scenarios approximate problem (7) yielded the solution

$$\hat{x}^{(N)}_{\text{WC}} = \begin{bmatrix} -0.0678 \\ 0.1857 \\ 0.2622 \end{bmatrix}$$

with (approximate) worst case residual

$$\hat{r}_{WC} = 2.2827.$$

The probability of violation of the residual $\hat{r}_{WC}$ was then estimated *a posteriori* via Monte-Carlo analysis using a very large number of samples, obtaining

$$\hat{P}_V = \frac{N_V}{N} = 0.0051,$$

showing that the the scenario solution had indeed a very low associated violation probability.

As for the average paradigm, setting $\epsilon = 0.1$ and $\alpha = 0.001$, the bound of Theorem 2 would require

$$N = 3,115,043$$

uncertainty samples. We remark however that the Learning Theory bounds are usually rather conservative, and we expect practical convergence of the algorithm for much smaller sample sizes. Indeed, in the example at hand, we observed practical convergence of the recursive LS algorithm (Algorithm 1) in less than 1×10^5 iterations to the solution

$$\hat{x}^{(N)}_{\text{AV}} = \begin{bmatrix} -1.9200 \\ -1.6450 \\ 2.0583 \end{bmatrix}.$$

VI. CONCLUSIONS

This paper presented a brief overview of available techniques based on uncertainty sampling for solving in a relaxed sense convex design problems in presence of uncertainty. The key point is that both average and worst-case design can be solved efficiently in the discussed probabilistic setup. Moreover, the sampling approximations are simple, and rigorous sample complexity bounds are available. In our opinion, these two features make these techniques very appealing for application to practical engineering design problems.

[2]Unless the scenarios $\delta^{(1)}, \ldots, \delta^{(N)}$ are chosen in an "unfortunate" way, an event that is guaranteed to occur only with very low probability β.

REFERENCES

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[2] S. Chandrasekaran ans G.H. Golub, M. Gu, and A.H. Sayed. Parameter estimation in the presence of bounded data uncertainties. *SIAM J. on Matrix Analysis and Applications*, 19:235–252, 1998.

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

[4] A. Ben-Tal and A. Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12(3):811–833, 2002.

[5] G. Calafiore and M.C. Campi. A new bound on the generalization rate of sampled convex programs. In *Proc. IEEE Conf. on Decision and Control*, Paradise Island, Bahamas, 2004.

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

[7] G. Calafiore and F. Dabbene. Near optimal stochastic solution to uncertain least squares problems. In *Proc. American Control Conf.*, Denver, 2003.

[8] G. Calafiore and F. Dabbene. Least-squares problems with stochastic uncertainty. In *Probabilistic and Randomized Methods for Design under Uncertainty (G. Calafiore and F. Dabbene, Editors)*. Springer, London, to appear: 2005.

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

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

[11] H.A. Hindi and S.P. Boyd. Robust solutions to l_1, l_2, and l_∞ uncertain linear approximation problems using convex optimization. In *Proc. American Control Conf.*, volume 6, pages 3487–3491, 1998.

[12] M. Karpinski and A.J. Macintyre. Polynomial bounds for VC dimension of sigmoidal neural networks. In *Proc. 27th ACM Symp. Thy. of Computing*, pages 200–208, 1995.

[13] A. Ruszczyński and A. Shapiro (editors). *Stochastic Programming*. Handbooks in OR and MS, Vol. 10. North-Holland Publishing Company, Amsterdam, 2003.

[14] A.H. Sayed, V.H. Nascimento, and S. Chandrasekaran. Estimation and control with bounded data uncertainties. *Linear Algebra and Its Applications*, 248:259–306, 1999.

[15] A. Tikhonov and V. Arsenin. *Solution to Ill-posed Problems*. Wiley, New York, 1977.

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

[17] M. Vidyasagar. *A Theory of Learning and Generalization*. Springer-Verlag, London, 1997.

[18] M. Vidyasagar. Randomized algorithms for robust controller synthesis using statistical learning theory. *Automatica*, 37(10):1515–1528, October 2001.

APPENDIX

PROOF OF THEOREM 2. Consider the function family $\mathcal{G}$ generated by the functions

$$g(x, \delta) \doteq \frac{f(x, \delta) - f^*(\delta)}{V},$$

as x varies over $\mathcal{X}$. The family $\mathcal{G}$ is a simple rescaling of $\mathcal{F}$ and maps Δ into the interval $[0, 1]$, therefore the P-dimension of $\mathcal{G}$ is the same as that of $\mathcal{F}$. Define

$$\phi_g(x) \doteq E_\delta[g(x, \delta)] = \frac{\phi_{\text{AV}}(x) - K}{V}, \qquad (19)$$

and

$$\hat{\phi}_g^{(N)}(x) \doteq \frac{1}{N}\sum_{i=1}^{N} g(x,\delta^{(i)}) = \frac{\hat{\phi}_{\text{AV}}^{(N)}(x) - \hat{K}}{V}, \qquad (20)$$

where

$$K \doteq E_\delta[f^*(\delta)], \quad \hat{K} \doteq [f^*(\delta)] = \frac{1}{N}\sum_{i=1}^{N} f^*(\delta^{(i)}).$$

Notice that a minimizer $\hat{x}_{\text{AV}}$ of $\hat{\phi}_{\text{AV}}^{(N)}(x)$ is also a minimizer of $\hat{\phi}_g^{(N)}(x)$. Then, Theorem 2 in [18] guarantees that, for $\alpha, \nu \in (0,1)$,

$$\text{PR}\left\{\sup_{g\in\mathcal{G}}\left|E_\delta[g(\delta)] - \frac{1}{N}\sum_{i=1}^{N} g(\delta^{(i)})\right| > \nu\right\} \leq \alpha,$$

provided that

$$N \geq \frac{32}{\nu^2}\left[\ln\frac{8}{\alpha} + \text{P-DIM}\,(\mathcal{G})\left(\ln\frac{16\text{e}}{\nu} + \ln\ln\frac{16\text{e}}{\nu}\right)\right]. \qquad (21)$$

Applying this theorem with $\nu = \epsilon/2$, and using the bound $\text{P-DIM}\,(\mathcal{G}) = \text{P-DIM}\,(\mathcal{F}) \leq d$, we have that, for all $x \in \mathcal{X}$, it holds with probability at least $(1-\alpha)$ that

$$|\phi_g(x) - \hat{\phi}_g^{(N)}(x)| \leq \frac{\epsilon}{2}. \qquad (22)$$

From (22), evaluated in $x = x_{\text{AV}}^*$ it follows that

$$\phi_g(x_{\text{AV}}^*) \geq \hat{\phi}_g^{(N)}(x_{\text{AV}}^*) - \frac{\epsilon}{2} \geq \hat{\phi}_g^{(N)}(\hat{x}_{\text{AV}}^{(N)}) - \frac{\epsilon}{2}, \qquad (23)$$

where the last inequality follows since $\hat{x}_{\text{AV}}^{(N)}$ is a minimizer of $\hat{\phi}_g$. From (22), evaluated in $x = \hat{x}_{\text{AV}}^{(N)}$ it follows that

$$\hat{\phi}_g^{(N)}(\hat{x}_{\text{AV}}^{(N)}) \geq \phi_g(\hat{x}_{\text{AV}}^{(N)}) - \frac{\epsilon}{2},$$

which substituted in (23), gives

$$\phi_g(\hat{x}_{\text{AV}}^{(N)}) \geq \phi_g(\hat{x}_{\text{AV}}^{(N)}) - \epsilon.$$

From the last inequality and (19) it follows that

$$\phi(\hat{x}_{\text{AV}}^{(N)}) - \phi(x_{\text{AV}}^*) \leq \epsilon V,$$

which concludes the proof. □