HYPOTHESES TESTING ON THE OPTIMAL VALUES OF SEVERAL RISK-NEUTRAL OR RISK-AVERSE CONVEX STOCHASTIC PROGRAMS AND APPLICATION TO HYPOTHESES TESTING ON SEVERAL RISK MEASURE VALUES

VINCENT GUIGUES

Abstract. Given an arbitrary number of risk-averse or risk-neutral convex stochastic programs, we study hypotheses testing problems aiming at comparing the optimal values of these stochastic programs on the basis of samples of the underlying random vectors. We propose non-asymptotic tests based on confidence intervals on the optimal values of the stochastic programs obtained using the Robust Stochastic Approximation and the Stochastic Mirror Descent algorithms. When the objective functions are uniformly convex, we also propose a multi-step version of the Stochastic Mirror Descent algorithm and obtain confidence intervals on both the optimal values and optimal solutions. The results are applied to compare, using tests of hypotheses, the (extended polyhedral) risk measure values of several distributions.

AMS subject classifications: 90C15, 90C90.

1. Introduction

Let $\xi_1, \ldots, \xi_m \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ be m random vectors. For $i = 1, \ldots, m$, let $\rho_i : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ where $\rho_i(\xi)$ is the optimal value of a risk-averse convex stochastic optimization problem where the underlying random vector $\xi \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ appears in the objective function:

$$\rho_i(\xi) = \left\{ \begin{array}{l} \min \ f(x) := \mathcal{R}_i \left[g(x, \xi) \right], \\ x \in X \end{array} \right. \tag{1.1}$$

with

- $g(x, \xi) : E \times \Omega \to \mathbb{R}$ a Borel function which is convex in x for every ξ and $\mathbb{P}$-summable in ξ for every x;
- X a closed and bounded convex set in a Euclidean space E;
- $\mathcal{R}_i$ an extended polyhedral risk measure [3]; and
- $f : X \to \mathbb{R}$ a Lipschitz continuous function which is convex in Sections 2.1 and 2.2, and uniformly convex in Section 2.3.

Given samples from the distributions of $\xi_1, \ldots, \xi_m$, our goal is to compare the optimal values $\rho_i(\xi_i)$, $i = 1, \ldots, m$, studying the following statistical tests:

$$\begin{array}{lll} (a)\ H_0 : & \rho_1(\xi_1) = \rho_2(\xi_2) = \ldots = \rho_m(\xi_m) & \text{against } \overline{H_0}, \\ (b)\ H_0^i : & \rho_i(\xi_i) \leq \rho_j(\xi_j), \ \ 1 \leq j \neq i \leq m & \text{against } \overline{H_0^i}, \\ (c)\ H_0 : & \rho_1(\xi_1) \leq \rho_2(\xi_2) \leq \ldots \leq \rho_m(\xi_m) & \text{against } \overline{H_0}. \end{array} \tag{1.2}$$

Key words and phrases. Stochastic Optimization, Risk measures, Hypothesis Testing, Multi-step Stochastic Mirror Descent, Robust Stochastic Approximation.

In Section 2, we start providing a confidence interval for the optimal value of (1.1) in the case when $\mathcal{R}_i = \mathbb{E}$ is the expectation. Various algorithms are considered. In each case, on the basis of a sample $(\xi_1, \ldots, \xi_N)$ of ξ, the algorithm produces an approximate optimal value g^N for (1.1) and a confidence interval for that optimal value. In Sections 2.1 and 2.2, the Robust Stochastic Approximation (RSA) and the Stochastic Mirror Descent (SMD) algorithms are considered. In the particular case when the objective function f is uniformly convex, we additionally provide confidence intervals for the optimal solution of (1.1). Applying the techniques discussed in [5] to the SMD algorithm, multi-step versions of the Stochastic Mirror Descent algorithm are studied in Section 2.3 in the case when f is uniformly convex. Confidence intervals for the optimal value of (1.1) obtained using these multi-step algorithms are also given. Section 2.4 specializes the results to two-stage risk-neutral convex stochastic optimization problems. The results of Section 2 are then extended in Section 3 taking for $\mathcal{R}_i$ an extended polyhedral risk measure (EPRM), introduced in [3]. In Section 4, we apply the results of Sections 2 and 3 to study tests (1.2)-(a), (b), (c).

Finally, we discuss an application of this work to hypotheses testing of several risk measure values. Risk measures [1] have become a standard tool to assess the risk of one dimensional loss or income distributions. In particular, one can be interested in comparing the risk measures of two or more unknown distributions, on the basis of samples from these distributions, knowing that the lower the risk measure value, the better. These distributions can correspond to the distributions of returns for different investments, among which we wish to choose one. They can also correspond to the cost distributions of different policies for a multistage stochastic optimization problem; the objective being here to choose the best policy-the one with smallest risk measure value-among a set of candidate policies. One approach to compare different risk measure values is to perform a test of hypothesis, as in [4], which, to our knowledge has been the only work dealing with this problem so far. In [4], the special case of spectral risk measures was considered.

More precisely, in our case, given an EPRM $\mathcal{R}$ and an arbitrary number m of random variables $\xi_1, \xi_2, \ldots, \xi_m$, we are interested for $i = 1, \ldots, m$, in the tests

$$\begin{array}{lll} (a)\ H_0: \ \mathcal{R}(\xi_1) = \mathcal{R}(\xi_2) = \ldots = \mathcal{R}(\xi_m) & \text{against } \overline{H_0}, \\ (b)\ H_0^i: \ \mathcal{R}(\xi_i) \leq \mathcal{R}(\xi_j), \ \ 1 \leq j \neq i \leq m & \text{against } \overline{H_0^i}, \\ (c)\ H_0: \ \mathcal{R}(\xi_1) \leq \mathcal{R}(\xi_2) \leq \ldots \leq \mathcal{R}(\xi_m) & \text{against } \overline{H_0}, \end{array} \tag{1.3}$$

aiming at comparing the risks of $\xi_1, \ldots, \xi_m$, and at deciding whether random variable ξ_i has the lowest risk or not. This test can be seen as an extension of the classical ANOVA which aims at comparing the means of several distributions (corresponding to $\mathcal{R} = \mathbb{E}$ in (1.3)-(a)). Examples of the EPRM we consider are the CVaR, some spectral risk measures, the optimzed certainty equivalent and the expected utility with piecewise affine utility function, making the study applicable to various popular risk measures. The last section 5 specializes the previous results to tests (1.3) and discusses possible extensions.

Throughout the paper, we use the following notation. We will denote by $f'(x)$ a subgradient of f at $x \in X$. For a norm $\|\cdot\|$ of a Euclidean space E associated to a scalar product $\langle \cdot, \cdot \rangle$, the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$ is given by

$$\|y\|_* = \max_{x : \|x\| \leq 1} \langle x, y \rangle.$$

We denote the ℓ_p norm of a vector x in $\mathbb{R}^n$ by $\|x\|_p$. The ball of center x_0 and radius R is denoted by $B(x_0, R)$. By Π_Y, we denote the metric projection operator onto the set Y, i.e., $\Pi_Y(x) = \arg\min_{y\in Y} \|y-x\|_2$. For a nonempty set $X \subseteq \mathbb{R}^n$, the polar cone X^* is defined by $X^* = \{x^* : \langle x, x^*\rangle \le 0 \ \forall x \in X\}$, where $\langle \cdot, \cdot \rangle$ is the standard scalar product on $\mathbb{R}^n$. By $\xi^t = (\xi_1, \ldots, \xi_t)$, we denote the history of the process (ξ_t) up to time t and by $\mathcal{F}_t$ the sigma-algebra generated by ξ^t. Finally, unless stated otherwise, all relations between random variables are supposed to hold almost surely.

2. Quality of approximate solutions in the risk-neutral case

In this section we take the expectation for $\mathcal{R}_i$ in problem (1.1) and make the following assumption:

Assumption 1. All subgradients of the objective function are bounded:

$$\|f'(x)\|_* \le L \text{ for every } x \in X.$$

We also assume the existence of a *stochastic oracle*: at t-th call to the oracle, $x \in X$ being the query point, the oracle returns $g(x, \xi_t) \in \mathbb{R}$ and $G(x, \xi_t) \in \partial_x g(x, \xi_t)$, where $\xi_1, \xi_2, \ldots$ are i.i.d. "oracle noises." We treat $g(x, \xi)$ as an estimate of $f(x)$ and $G(x, \xi)$ as an estimate of a subgradient of f at x.

We assume that our estimates are *unbiased*:

$$\forall x \in X : f(x) = \mathbb{E}_\xi\left[g(x, \xi)\right] \quad \text{and} \quad f'(x) := \mathbb{E}_\xi\left[G(x, \xi)\right] \in \partial f(x).$$

From now on, we set

$$\delta(x, \xi) = g(x, \xi) - f(x), \ \Delta(x, \xi) = G(x, \xi) - f'(x), \tag{2.4}$$

so that

$$\mathbb{E}_\xi\left[\delta(x, \xi)\right] = 0, \ \mathbb{E}_\xi\left[\Delta(x, \xi)\right] = 0.$$

Assumptions on the Stochastic Oracle. In the sequel, we assume that the observation errors of our oracle satisfy some assumptions (introduced in [6]) in addition to having zero means. Specifically, our *minimal* assumption is the following:

Assumption 2. For some $M_1, M_2 \in (0, \infty)$ and for all $x \in X$

$$\begin{array}{lrcl} (a) & \mathbb{E}\Big[\delta^2(x, \xi)\Big] & \le & M_1^2, \\ (b) & \mathbb{E}\Big[\|\Delta(x, \xi)\|_*^2\Big] & \le & M_2^2. \end{array} \tag{2.5}$$

Under our minimal assumption, we will obtain an upper bound on the average error on the optimal value of (1.1). To obtain a confidence interval on this optimal value, we will need a stronger assumption:

Assumption 3. For some $M_1, M_2 \in (0, \infty)$ and for all $x \in X$ it holds that

$$\begin{array}{lrcl} (a) & \mathbb{E}\Big[\exp\{\delta^2(x, \xi)/M_1^2\}\Big] & \le & \exp\{1\}, \\ (b) & \mathbb{E}\Big[\exp\{\|\Delta(x, \xi)\|_*^2/M_2^2\}\Big] & \le & \exp\{1\}. \end{array} \tag{2.6}$$

Note that condition (2.6) is indeed stronger than condition (2.5): if a random variable Y satisfies $\mathbb{E}\Big[\exp\{Y\}\Big] \leq \exp\{1\}$ then by Jensen inequality, using the concavity of the logarithmic function, $\mathbb{E}\Big[Y\Big] = \mathbb{E}\Big[\ln\Big(\exp\{Y\}\Big)\Big] \leq \ln\Big(\mathbb{E}\Big[\exp\{Y\}\Big]\Big) \leq 1$.

For a given confidence level, a smaller confidence interval can be obtained under an even stronger assumption:

Assumption 4. For some $M_1, M_2 \in (0, \infty)$ and for all $x \in X$ it holds that

$$\begin{array}{lrcl} (a) & \mathbb{E}\Big[\exp\{\delta^2(x,\xi)/M_1^2\}\Big] & \leq & \exp\{1\}, \\ (b) & \|\Delta(x,\xi)\|_* & \leq & M_2 \text{ almost surely.} \end{array} \tag{2.7}$$

Observe that the validity of (2.7) for all $x \in X$ and some M_1, M_2 implies the validity of (2.6) for all $x \in X$ with the same M_1, M_2.

2.1. **Robust Stochastic Approximation Algorithm.** In this section, we use the scalar product $\langle x, y\rangle = x^\top y$ with corresponding norm $\|x\| = \|x\|_2 = \sqrt{x^\top x}$ and dual norm $\|x\|_* = \|x\|_2$, meaning that (2.5), (2.6), and (2.7) hold with $\|\cdot\|_* = \|\cdot\|_2$. The Robust Stochastic Approximation algorithm solves (1.1) as follows:

Algorithm 1: Robust Stochastic Approximation.

Initialization. Take x_1 in X. Fix the number of iterations $N-1$ and positive deterministic stepsizes $\gamma_1, \ldots, \gamma_N$.

Loop. For $t = 1, \ldots, N-1$, compute

$$x_{t+1} = \Pi_X(x_t - \gamma_t G(x_t, \xi_t)). \tag{2.8}$$

Outputs:

$$x^N = \frac{1}{\Gamma_N}\sum_{\tau=1}^N \gamma_\tau x_\tau \text{ and } g^N = \frac{1}{\Gamma_N}\left[\sum_{\tau=1}^N \gamma_\tau g(x_\tau, \xi_\tau)\right] \text{ with}$$
$$\Gamma_N = \sum_{\tau=1}^N \gamma_\tau.$$

Note that by convexity of X, we have $x^N \in X$ and after $N-1$ iterations, x^N is an approximate solution of (1.1). The value $f(x^N)$ is an approximation of the optimal value of (1.1) but it is not computable since f is not known. Denoting by x_* an optimal solution of (1.1), we introduce after $N-1$ iterations the computable approximation[1]

$$g^N = \frac{1}{\Gamma_N}\left[\sum_{\tau=1}^N \gamma_\tau g(x_\tau, \xi_\tau)\right] \tag{2.9}$$

of the optimal value $f(x_*)$ of (1.1) obtained using the points generated by the algorithm and information from the stochastic oracle. Our goal is to obtain exponential bounds on large deviations of this estimate g^N of $f(x_*)$ from $f(x_*)$ itself, i.e., a confidence interval on the optimal value of (1.1) using the information provided by

[1]Note that the approximation depends on $(x_1, \ldots, x_N, \xi_1, \ldots, \xi_N, \gamma_1, \ldots, \gamma_N)$ so we could write $g^N(x_1, \ldots, x_N, \xi_1, \ldots, \xi_N, \gamma_1, \ldots, \gamma_N)$ but we choose, for the moment, to suppress this dependence to alleviate notation

the RSA algorithm along iterations. We need two technical lemmas. The first one gives an $O(1/\sqrt{N})$ upper bound on the first absolute moment of the estimation error (the average distance of g^N to $f(x_*)$):

Lemma 2.1. *Let Assumptions 1 and 2 hold and assume that the number of iterations $N-1$ of the RSA algorithm is fixed in advance with stepsizes given by*

$$\gamma_\tau = \frac{D_X}{\sqrt{2(M_2^2+L^2)}\sqrt{N}},\ \tau=1,\ldots,N, \tag{2.10}$$

where

$$D_X = \max_{x\in X}\ \|x-x_1\|. \tag{2.11}$$

Let g^N be the approximation of $f(x_*)$ given by (2.9). *Then*

$$\mathbb{E}\left[\left|g^N - f(x_*)\right|\right] \le \frac{M_1 + D_X\sqrt{2(M_2^2+L^2)}}{\sqrt{N}}. \tag{2.12}$$

Proof. Letting

$$f^N = \frac{1}{\Gamma_N}\sum_{\tau=1}^{N}\gamma_\tau f(x_\tau), \tag{2.13}$$

it is known (see [6], sec. 2.2) that under our assumptions

$$\mathbb{E}\left[f(x^N)-f(x_*)\right] \le \mathbb{E}\left[f^N - f(x_*)\right] \le \frac{D_X\sqrt{2(M_2^2+L^2)}}{\sqrt{N}}. \tag{2.14}$$

Since the main steps of the proof of (2.14) will be useful for our further developments, we rewrite them here. Setting $A_\tau = \frac{1}{2}\|x_\tau - x_*\|_2^2$, we can show (see [6] for instance) that

$$\sum_{\tau=1}^{N}\ \gamma_\tau\langle G(x_\tau,\xi_\tau), x_\tau - x_*\rangle \le A_1 + \frac{1}{2}\sum_{\tau=1}^{N}\ \gamma_\tau^2\|G(x_\tau,\xi_\tau)\|_*^2. \tag{2.15}$$

To save notation, let us set

$$\delta_\tau = g(x_\tau,\xi_\tau) - f(x_\tau),\ \Delta_\tau = \Delta(x_\tau,\xi_\tau),\ G_\tau = G(x_\tau,\xi_\tau) = f'(x_\tau)+\Delta_\tau. \tag{2.16}$$

Inequality (2.15) can be rewritten

$$\sum_{\tau=1}^{N}\ \gamma_\tau\langle f'(x_\tau), x_\tau - x_*\rangle \le \frac{D_X^2}{2} + \frac{1}{2}\sum_{\tau=1}^{N}\ \gamma_\tau^2\|G_\tau\|_*^2 + \sum_{\tau=1}^{N}\ \gamma_\tau\langle\Delta_\tau, x_* - x_\tau\rangle. \tag{2.17}$$

Taking into account that by convexity of f we have $f(x_\tau)-f(x_*) \le \langle f'(x_\tau), x_\tau - x_*\rangle$, we get
(2.18)

$$\begin{array}{rcl} f(x^N)-f(x_*) & \le & f^N - f(x_*) = \dfrac{1}{\Gamma_N}\displaystyle\sum_{\tau=1}^{N}\gamma_\tau\Big(f(x_\tau)-f(x_*)\Big) \\ & \le & \dfrac{1}{\Gamma_N}\left[\dfrac{D_X^2}{2} + \dfrac{1}{2}\displaystyle\sum_{\tau=1}^{N}\ \gamma_\tau^2\|G_\tau\|_*^2 + \sum_{\tau=1}^{N}\ \gamma_\tau\langle\Delta_\tau, x_* - x_\tau\rangle\right] \end{array}$$

where the first inequality is due to the origin of x^N and to the convexity of f.

Next, note that under Assumptions 1 and 2,

$$\mathbb{E}\Big[\|G_\tau\|_*^2\Big] = \mathbb{E}\Big[\|f'(x_\tau)+\Delta_\tau\|_*^2\Big] \le 2\mathbb{E}\Big[\|f'(x_\tau)\|_*^2 + \|\Delta_\tau\|_*^2\Big] \le 2\Big[M_2^2+L^2\Big]. \tag{2.19}$$

Passing to expectations in (2.18), and taking into account that the conditional, $\xi^{\tau-1} := (\xi_1, ..., \xi_{\tau-1})$ being fixed, expectation of Δ_τ is zero, while x_τ by construction is a deterministic function of $\xi^{\tau-1}$, we get

$$\begin{array}{rcl} \mathbb{E}\Big[f(x^N) - f(x_*)\Big] & \leq & \mathbb{E}\Big[f^N - f(x_*)\Big] \leq \dfrac{D_X^2 + \sum_{\tau=1}^{N} \gamma_\tau^2 \mathbb{E}\left[\|G_\tau\|_*^2\right]}{2\Gamma_N} \\ & \leq & \dfrac{1}{\Gamma_N}\Big[\dfrac{D_X^2}{2} + (M_2^2 + L^2)\sum_{\tau=1}^{N}\gamma_\tau^2\Big]. \end{array} \tag{2.20}$$

Using stepsizes (2.10), we have $\Gamma_N = \frac{D_X\sqrt{N}}{\sqrt{2(M_2^2+L^2)}}$. Plugging this value of Γ_N into (2.20), we obtain the announced inequality (2.14).

We now show that

$$\mathbb{E}\Big[\Big|g^N - f^N\Big|\Big] \leq \frac{M_1}{\sqrt{N}}. \tag{2.21}$$

First, note that

$$g^N - f^N = \frac{1}{\Gamma_N}\sum_{\tau=1}^{N}\gamma_\tau\delta_\tau = \frac{1}{N}\sum_{\tau=1}^{N}\delta_\tau. \tag{2.22}$$

By the same argument as above, the conditional, $\xi^{\tau-1}$ being fixed, expectation of δ_τ is 0, whence

$$\mathbb{E}\Big[\Big(\sum_{\tau=1}^{N}\delta_\tau\Big)^2\Big] = \sum_{\tau=1}^{N}\mathbb{E}\Big[\delta_\tau^2\Big] \leq NM_1^2,$$

where the concluding inequality is due to $(2.5.a)$. We conclude that

$$\mathbb{E}\Big[\Big|g^N - f^N\Big|\Big] \leq \frac{1}{N}\sqrt{\mathbb{E}\left[\left(\sum_{\tau=1}^{N}\delta_\tau\right)^2\right]} \leq \frac{1}{N}\sqrt{NM_1^2} = \frac{M_1}{\sqrt{N}},$$

which is the announced inequality (2.21). Next, observe that by convexity of f, $f^N \geq f(x^N)$ and since $x^N \in X$, we have $f(x^N) \geq f(x_*)$, i.e., $f^N - f(x_*) \geq f(x^N) - f(x_*) \geq 0$, so that (2.14) and (2.21) imply

$$\begin{array}{rcl} \mathbb{E}\Big[|g^N - f(x_*)|\Big] & \leq & \mathbb{E}\Big[|g^N - f^N| + |f^N - f(x_*)|\Big] = \mathbb{E}\Big[|g^N - f^N|\Big] + \mathbb{E}\Big[f^N - f(x_*)\Big] \\ & \leq & \Big[M_1 + D_X\sqrt{2(M_2^2 + L^2)}\Big]\dfrac{1}{\sqrt{N}}, \end{array}$$

which achieves the proof of (2.12). □

Lemma 2.2. *Let $\eta_\tau, \tau = 1, \ldots, N$, be a sequence of real-valued random variables with η_τ $\mathcal{F}_\tau$-measurable. Let $\mathbb{E}_{|\tau-1}[\cdot]$ be the conditional expectation $\mathbb{E}\left[\cdot|\xi^{\tau-1}\right]$. Assume that*

$$\mathbb{E}_{|\tau-1}\Big[\eta_\tau\Big] = 0, \;\; \mathbb{E}_{|\tau-1}\Big[\exp\{\eta_\tau^2\}\Big] \leq \exp\{1\}. \tag{2.23}$$

Then, for any $\Theta > 0$,

$$\mathbb{P}\left(\sum_{\tau=1}^{N}\eta_\tau > \Theta\sqrt{N}\right) \leq \exp\{-\Theta^2/4\}. \tag{2.24}$$

Proof. We first show that for any $\gamma > 0$ and $\tau = 1, \ldots, N$, we have

$$\mathbb{E}_{|\tau-1}\Big[\exp\{\gamma\eta_\tau\}\Big] \le \exp\{\gamma^2\}. \tag{2.25}$$

Let us fix $0 < \gamma \le 1$. Observing that

$$e^x \le x + e^{x^2} \text{ for every } x \in \mathbb{R}, \tag{2.26}$$

we obtain

$$\begin{array}{lll}\mathbb{E}_{|\tau-1}\Big[\exp\{\gamma\eta_\tau\}\Big] & \le & \mathbb{E}_{|\tau-1}\Big[\gamma\eta_\tau\Big] + \mathbb{E}_{|\tau-1}\Big[\exp\{\gamma^2\eta_\tau^2\}\Big] \\ & \le & \mathbb{E}_{|\tau-1}\Big[\exp\{\gamma^2\eta_\tau^2\}\Big] \text{ using (2.23)} \\ & \le & \mathbb{E}_{|\tau-1}\Big[(\exp\{\eta_\tau^2\})^{\gamma^2}\Big] \le \Big(\mathbb{E}_{|\tau-1}\Big[\exp\{\eta_\tau^2\}\Big]\Big)^{\gamma^2},\end{array}$$

where the last inequality is Jensen inequality applied to the concave function x^{γ^2}. Plugging (2.23) into the above inequality shows that (2.25) holds for $0 < \gamma \le 1$.

For $\gamma > 1$,

$$\begin{array}{lll}\mathbb{E}_{|\tau-1}\Big[\exp\{\gamma\eta_\tau\}\Big] & \le & \mathbb{E}_{|\tau-1}\Big[\exp\{\frac{1}{2}\gamma^2 + \frac{1}{2}\eta_\tau^2\}\Big] \\ & \le & \exp\{\frac{\gamma^2}{2}\}\sqrt{\mathbb{E}_{|\tau-1}\Big[\exp\{\eta_\tau^2\}\Big]} \le \exp\{\frac{\gamma^2+1}{2}\} \le \exp\{\gamma^2\},\end{array}$$

where we have used (2.23) for the third inequality and the fact that $\gamma > 1$ for the last one. We have thus shown that (2.25) holds for every $\gamma > 0$. As a result, for $\gamma > 0$, setting $S_\tau = \sum_{s=1}^{\tau}\eta_s$, we have

$$\begin{array}{lll}\mathbb{E}\Big[\exp\{\gamma S_\tau\}\Big] & = & \mathbb{E}\Big[\exp\{\gamma S_{\tau-1}\}\mathbb{E}_{|\tau-1}\Big[\exp\{\gamma\eta_\tau\}\Big]\Big] \\ & \le & \exp\{\gamma^2\}\mathbb{E}\Big[\exp\{\gamma S_{\tau-1}\}\Big] \text{ using (2.25).}\end{array}$$

It follows that for $\gamma > 0$

$$\mathbb{E}\Big[\exp\{\gamma S_\tau\}\Big] \le \exp\{\gamma^2(\tau-1)\}\mathbb{E}\Big[\exp\{\gamma\eta_1\}\Big] \le \exp\{\gamma^2\tau\} \text{ using (2.25).} \tag{2.27}$$

Next, for $\gamma > 0$,

$$\begin{array}{lll}\mathbb{P}\Big(S_N > \Theta\sqrt{N}\Big) & = & \mathbb{P}\Big(\exp\{\gamma S_N\} > \exp\{\Theta\sqrt{N}\gamma\}\Big) \\ & \le & \min\limits_{\gamma>0} \exp\{-\Theta\sqrt{N}\gamma\}\mathbb{E}\Big[\exp\{\gamma S_N\}\Big] \text{ using Chernoff bound,} \\ & \le & \exp\{\min\limits_{\gamma>0}\Big[\gamma^2 N - \Theta\sqrt{N}\gamma\Big]\} = \exp\{-\Theta^2/4\} \text{ using (2.27).}\end{array}$$

This achieves the proof of inequality (2.24). □

We are now in a position to provide a confidence interval for the optimal value of (1.1) using the RSA algorithm:

Proposition 2.3. *Assume that the number of iterations $N-1$ of the RSA algorithm is fixed in advance with stepsizes given by (2.10). Let g^N be the approximation of $f(x_*)$ given by (2.9). Then*

(i) *if Assumptions 1 and 3 hold, for any $\Theta > 0$, we have*

$$\mathbb{P}\left(\left|g^N - f(x_*)\right| > \frac{K_1(X) + \Theta K_2(X)}{\sqrt{N}}\right) \le 4\exp\{1\}\exp\{-\Theta\} \tag{2.28}$$

where the constants $K_1(X)$ *and* $K_2(X)$ *are given by*

$$K_1(X) = \frac{D_X(M_2^2+2L^2)}{\sqrt{2(M_2^2+L^2)}} \text{ and } K_2(X) = \frac{D_X M_2^2}{\sqrt{2(M_2^2+L^2)}} + 2D_X M_2 + M_1,$$

with D_X *given by* (2.11).

(ii) *If Assumptions 1 and 4 hold,* (2.28) *holds with the right hand side replaced by* $(3+\exp\{1\})\exp\{-\frac{1}{4}\Theta^2\}$.

Proof. To prove (i), we shall first prove that for any $\Theta > 0$

(2.29)
$$\mathbb{P}\left(f^N - f(x_*) > \frac{D_X}{\sqrt{2(M_2^2+L^2)\,N}}\left[M_2^2 + 2L^2 + \Theta\left[M_2^2 + 2M_2\sqrt{2(M_2^2+L^2)}\right]\right]\right) \leq 2\exp\{1\}\exp\{-\Theta\},$$

where f^N is given by (2.13). Using Assumption 1, we have $\|G_\tau\|_*^2 = \|f'(x_\tau) + \Delta_\tau\|_*^2 \leq 2(\|f'(x_\tau)\|_*^2 + \|\Delta_\tau\|_*^2) \leq 2(L^2 + \|\Delta_\tau\|_*^2)$. Combined with (2.18), this implies that

(2.30)
$$\begin{array}{rcl} f^N - f(x_*) & \leq & \frac{1}{\Gamma_N}\left[\frac{D_X^2}{2} + \sum_{\tau=1}^N \gamma_\tau^2\Big(L^2 + \|\Delta_\tau\|_*^2\Big)\right] + \frac{1}{\Gamma_N}\sum_{\tau=1}^N \gamma_\tau\langle\Delta_\tau, x_* - x_\tau\rangle \\ & \leq & \frac{D_X(M_2^2+2L^2)}{\sqrt{2(M_2^2+L^2)}\sqrt{N}} + \frac{D_X M_2^2}{\sqrt{2(M_2^2+L^2)}\sqrt{N}}\mathcal{A} + \frac{2D_X M_2}{N}\mathcal{B} \end{array}$$

where

$$\mathcal{A} = \frac{1}{NM_2^2}\sum_{\tau=1}^N \|\Delta_\tau\|_*^2 \quad \text{and} \quad \mathcal{B} = \frac{1}{2D_X M_2}\sum_{\tau=1}^N \langle\Delta_\tau, x_* - x_\tau\rangle. \tag{2.31}$$

Setting $\zeta_\tau = \|\Delta_\tau\|_*^2/M_2^2$ and invoking (2.6.b), we get $\mathbb{E}\Big[\exp\{\zeta_\tau\}\Big] \leq \exp\{1\}$ for all $\tau \leq N$, whence, due to the convexity of the exponent,

$$\mathbb{E}\Big[\exp\{\mathcal{A}\}\Big] = \mathbb{E}\Big[\exp\{\frac{1}{N}\sum_{\tau=1}^N \zeta_\tau\}\Big] \leq \frac{1}{N}\sum_{\tau=1}^N \mathbb{E}\Big[\exp\{\zeta_\tau\}\Big] \leq \exp\{1\}$$

as well. As a result,

$$\forall\,\Theta > 0 : \mathbb{P}\Big(\mathcal{A} > \Theta\Big) \leq \exp\{-\Theta\}\mathbb{E}\Big[\exp\{\mathcal{A}\}\Big] \leq \exp\{1-\Theta\}. \tag{2.32}$$

Now let us set $\eta_\tau = \frac{1}{2D_X M_2}\langle\Delta_\tau, x_* - x_\tau\rangle$, so that $\mathcal{B} = \sum_{\tau=1}^N \eta_\tau$. Denoting by $\mathbb{E}_{|\tau-1}$ the conditional, $\xi^{\tau-1}$ being fixed, expectation, we have

$$\mathbb{E}_{|\tau-1}\Big[\eta_\tau\Big] = 0 \quad \text{and} \quad \mathbb{E}_{|\tau-1}\Big[\exp\{\eta_\tau^2\}\Big] \leq \exp\{1\},$$

where the first relation is due to $\mathbb{E}_{|\tau-1}\Big[\Delta_\tau\Big] = 0$ combined with the fact that $x_* - x_\tau$ is a deterministic function of $\xi^{\tau-1}$, and the second relation is due to (2.6.b) combined with the fact that $\|x_* - x_\tau\| \leq 2D_X$. Using Lemma 2.2, we obtain for any $\Theta > 0$

$$\mathbb{P}\Big(\mathcal{B} > \Theta\sqrt{N}\Big) \leq \exp\{-\Theta^2/4\}. \tag{2.33}$$

Combining (2.30), (2.32), and (2.33), we obtain for every $\Theta > 0$

$$\begin{array}{l}\mathbb{P}\left(f^N - f(x_*) > \frac{D_X(M_2^2+2L^2)}{\sqrt{2(M_2^2+L^2)\,N}} + \frac{\Theta}{\sqrt{N}}\left[\frac{D_X M_2^2}{\sqrt{2(M_2^2+L^2)}} + 2D_X M_2\right]\right)\\ \leq \exp\{1-\Theta\} + \exp\{-\Theta^2/4\} \leq 2\exp\{1\}\exp\{-\Theta\},\end{array} \tag{2.34}$$

which is (2.29).

Next,

$$g^N - f^N = \frac{M_1}{N}\left[\sum_{\tau=1}^N \chi_\tau\right],\ \chi_\tau = \frac{\delta_\tau}{M_1}.$$

Observing that χ_τ is a deterministic function of ξ^τ and that

$$\mathbb{E}_{|\tau-1}\Big[\chi_\tau\Big] = 0 \quad \text{and} \quad \mathbb{E}_{|\tau-1}\Big[\exp\{\chi_\tau^2\}\Big] \leq \exp\{1\},\, 1 \leq \tau \leq N$$

(we have used (2.6.a)), we can use once again Lemma 2.2 to obtain for all $\Theta > 0$:

$$\mathbb{P}\left(g^N - f^N > \Theta\frac{M_1}{\sqrt{N}}\right) \leq \exp\{-\Theta^2/4\}$$

and

$$\mathbb{P}\left(g^N - f^N < -\Theta\frac{M_1}{\sqrt{N}}\right) \leq \exp\{-\Theta^2/4\}.$$

Thus,

$$\forall\,\Theta > 0 : \mathbb{P}\left(\Big|g^N - f^N\Big| > \Theta\frac{M_1}{\sqrt{N}}\right) \leq 2\exp\{-\Theta^2/4\},$$

which, combined with (2.29) implies (2.28), i.e., item (i) of the lemma.

Finally, under Assumption 4, we have $\mathbb{P}\Big(\mathcal{A} > 1\Big) = 0$, which combines with (2.32) to imply that

$$\forall \Theta > 0 : \mathbb{P}\Big(\mathcal{A} > \Theta\Big) \leq \exp\{1-\Theta^2\},$$

meaning that the right hand side in (2.29) can be replaced with $\exp\{1-\Theta^2\} + \exp\{-\Theta^2/4\}$, which proves item (ii). □

2.2. **Stochastic Mirror Descent algorithm.** The algorithm to be described is given by a *proximal setup*, that is, by a norm $\|\cdot\|$ on E and a *distance-generating function* $\omega(x) : X \to \mathbb{R}$. This function should

- be convex and continuous on X,
- admit a continuous on $X^o = \{x \in X : \partial\omega(x) \neq \emptyset\}$ selection $\omega'(x)$ of subgradients, and
- be compatible with $\|\cdot\|$, meaning that $\omega(\cdot)$ is strongly convex, modulus $\mu(\omega) > 0$, with respect to the norm $\|\cdot\|$:

$$\langle\omega'(x) - \omega'(y), x - y\rangle \geq \mu(\omega)\|x-y\|^2 \ \forall x, y \in X^o.$$

The proximal setup induces the following entities:

(1) the *ω-center of* X given by $x_\omega = \operatorname{argmin}_{x\in X}\omega(x) \in X^o$;
(2) the *Bregman distance* or prox-function

$$V_x(y) = \omega(y) - \omega(x) - \langle\omega'(x), y - x\rangle \geq \frac{\mu(\omega)}{2}\|x-y\|^2, \tag{2.35}$$

for $x \in X^o$, $y \in X$ (the concluding inequality is due to the strong convexity of ω);

(3) the *ω-radius of X* defined as

$$D_{\omega,X} = \sqrt{2[\max_{x \in X} \omega(x) - \min_{x \in X} \omega(x)]}.$$

It is easily seen that $\langle \omega'(x_\omega), x - x_\omega \rangle \geq 0$ for all $x \in X$, whence

$$\forall x \in X : \frac{\mu(\omega)}{2}\|x - x_\omega\|^2 \leq V_{x_\omega}(x) = \omega(x) - \omega(x_\omega) - \underbrace{\langle \omega'(x_\omega), x - x_\omega \rangle}_{\geq 0} \leq \omega(x) - \omega(x_\omega) \leq \frac{1}{2} D^2_{\omega,X}, \tag{2.36}$$

and

$$\forall x \in X : \|x - x_\omega\| \leq \frac{D_{\omega,X}}{\sqrt{\mu(\omega)}}. \tag{2.37}$$

(4) *The proximal mapping*

$$\text{Prox}_x(\zeta) = \text{argmin}_{\,y \in X}\{\omega(y) + \langle \zeta - \omega'(x), y \rangle\} \quad [x \in X^o, \zeta \in E]. \tag{2.38}$$

This mapping clearly takes its values in X^o.

Taking $x_+ = \text{Prox}_x(\zeta)$, the optimality conditions for the optimization problem $\min_{y \in X}\{\omega(y) + \langle \zeta - \omega'(x), y \rangle\}$ in which x_+ is the optimal solution read

$$\forall y \in X : \langle \omega'(x_+) + \zeta - \omega'(x), y - x_+ \rangle \geq 0.$$

Rearranging the terms, some simple arithmetics shows that this condition can be written equivalently as

$$x_+ = \text{Prox}_x(\zeta) \Rightarrow \langle \zeta, x_+ - y \rangle \leq V_x(y) - V_{x_+}(y) - V_x(x_+) \;\forall y \in X. \tag{2.39}$$

Algorithm 2: Stochastic Mirror Descent.

Initialization. Take $x_1 = x_\omega$. Fix the number of iterations $N-1$ and positive deterministic stepsizes $\gamma_1, \ldots, \gamma_N$.

Loop. For $t = 1, \ldots, N-1$, compute

$$x_{t+1} \;=\; \text{Prox}_{x_t}(\gamma_t G(x_t, \xi_t)). \tag{2.40}$$

Outputs:

$$x^N = \frac{1}{\Gamma_N}\sum_{\tau=1}^N \gamma_\tau x_\tau \text{ and } g^N = \frac{1}{\Gamma_N}\left[\sum_{\tau=1}^N \gamma_\tau g(x_\tau, \xi_\tau)\right] \text{ with}$$

$$\Gamma_N = \sum_{\tau=1}^N \gamma_\tau.$$

When $\omega(x) = \frac{1}{2}\|x\|_2^2$ and the scalar product $\langle \cdot, \cdot \rangle$ is given by $\langle x, y \rangle = x^\top y$ then $\text{Prox}_x(\zeta) = \Pi_X(x - \zeta)$ and in this particular case, the Mirror Descent algorithm is the RSA algorithm given by the recurrence (2.8).

In what follows, we provide confidence intervals for the optimal value of (1.1) on the basis of the points generated by the SMD algorithm, thus extending Proposition 2.3. We first need a technical lemma:

Lemma 2.4. *Let $e_1, ..., e_N$ be a sequence of vectors from E, $\gamma_1, ..., \gamma_N$ be nonnegative reals, and let $u_1, ..., u_N \in X$ be given by the recurrence*

$$\begin{array}{rcl} u_1 & = & x_\omega \\ u_{\tau+1} & = & \mathrm{Prox}_{u_\tau}(\gamma_\tau e_\tau),\ 1 \le \tau \le N-1. \end{array}$$

Then

$$\forall y \in X : \sum_{\tau=1}^N \gamma_\tau \langle e_\tau, u_\tau - y\rangle \le \frac{1}{2} D_{\omega,X}^2 + \frac{1}{2\mu(\omega)} \sum_{\tau=1}^N \gamma_\tau^2 \|e_\tau\|_*^2. \tag{2.41}$$

Proof. Invoking (2.39), we get

$$\forall y \in X : \langle \gamma_\tau e_\tau, u_{\tau+1} - y\rangle \le V_{u_\tau}(y) - V_{u_{\tau+1}}(y) - V_{u_\tau}(u_{\tau+1}),$$

whence

$$\begin{array}{rcl} \forall y \in X : & & \\ \gamma_\tau \langle e_\tau, u_\tau - y\rangle & \le & V_{u_\tau}(y) - V_{u_{\tau+1}}(y) + \Big[\gamma_\tau \langle e_\tau, u_\tau - u_{\tau+1}\rangle - V_{u_\tau}(u_{\tau+1})\Big] \\ & \le & V_{u_\tau}(y) - V_{u_{\tau+1}}(y) + \Big[\gamma_\tau \|e_\tau\|_* \|u_\tau - u_{\tau+1}\| - \frac{\mu(\omega)}{2}\|u_\tau - u_{\tau+1}\|^2\Big] \\ & \le & V_{u_\tau}(y) - V_{u_{\tau+1}}(y) + \dfrac{\gamma_\tau^2 \|e_\tau\|_*^2}{2\mu(\omega)}, \end{array}$$

where we have used (2.35) for the second inequality. Summing up the resulting inequalities over $\tau = 1, ..., N$, and taking into account that $V_{u_{N+1}}(y) \ge 0$ by (2.35) and $V_{u_1}(y) \le \frac{1}{2} D_{\omega,X}^2$ by (2.36) (recall that $u_1 = x_\omega$), we arrive at (2.41). □

Applying Lemma 2.4 to $e_\tau = G(x_\tau, \xi_\tau)$ and in relation (2.41) specifying y as a minimizer x_* of f over X, we get:

$$\sum_{\tau=1}^N \gamma_\tau \langle G(x_\tau, \xi_\tau), x_\tau - x_*\rangle \le \frac{1}{2} D_{\omega,X}^2 + \frac{1}{2\mu(\omega)} \sum_{\tau=1}^N \gamma_\tau^2 \|G(x_\tau, \xi_\tau)\|_*^2.$$

Using notation (2.16) of the previous section, the above inequality can be rewritten

$$\sum_{\tau=1}^N \gamma_\tau \langle f'(x_\tau), x_\tau - x_*\rangle \le \frac{D_{\omega,X}^2}{2} + \frac{1}{2\mu(\omega)} \sum_{\tau=1}^N \gamma_\tau^2 \|G_\tau\|_*^2 + \sum_{\tau=1}^N \gamma_\tau \langle \Delta_\tau, x_* - x_\tau\rangle. \tag{2.42}$$

We mentioned that when $\omega(x) = \frac{1}{2}\|x\|_2^2$, the SMD algorithm is the RSA algorithm of the previous section. In that case, $\mu(\omega) = 1$, $\|\cdot\| = \|\cdot\|_2$, $\|\cdot\|_* = \|\cdot\|_2$, and (2.42) is obtained from inequality (2.17) of the previous section for the RSA algorithm substituting D_X by $D_{\omega,X}$ (note that when choosing $x_1 = x_\omega$ for the RSA algorithm, we have $D_X \le D_{\omega,X}$ so for the RSA algorithm (2.17) gives a tighter upper bound). We can now extend the results of Lemma 2.1 and Proposition 2.3 to the SMD algorithm:

Lemma 2.5. *Let Assumptions 1 and 2 hold and assume that the number of iterations $N-1$ of the SMD algorithm is fixed in advance with stepsizes given by*

$$\gamma_\tau = \frac{D_{\omega,X}\sqrt{\mu(\omega)}}{\sqrt{2(M_2^2 + L^2)}\sqrt{N}},\ \tau = 1, \ldots, N. \tag{2.43}$$

Consider the approximation $g^N = \frac{1}{\Gamma_N}\left[\sum_{\tau=1}^N \gamma_\tau g(x_\tau, \xi_\tau)\right]$ *of* $f(x_*)$*. Then*

$$\mathbb{E}\left[\left|g^N - f(x_*)\right|\right] \leq \frac{M_1 + \frac{D_{\omega,X}}{\sqrt{\mu(\omega)}}\sqrt{2(M_2^2+L^2)}}{\sqrt{N}}. \tag{2.44}$$

Proof. It suffices to follow the proof of Lemma 2.1, starting from inequality (2.17) which needs to be replaced by (2.42) for the Mirror Descent algorithm. □

Proposition 2.6. *Assume that the number of iterations* $N-1$ *of the SMD algorithm is fixed in advance with stepsizes given by* (2.43)*. Consider the approximation* $g^N = \frac{1}{\Gamma_N}\left[\sum_{\tau=1}^N \gamma_\tau g(x_\tau, \xi_\tau)\right]$ *of* $f(x_*)$*. Then,*

(i) *if Assumptions 1 and 3 hold, for any* $\Theta > 0$*, we have*

$$\mathbb{P}\left(\left|g^N - f(x_*)\right| > \frac{K_1(X) + \Theta K_2(X)}{\sqrt{N}}\right) \leq 4\exp\{1\}\exp\{-\Theta\} \tag{2.45}$$

where the constants $K_1(X)$ *and* $K_2(X)$ *are given by*

$$K_1(X) = \frac{D_{\omega,X}(M_2^2+2L^2)}{\sqrt{2(M_2^2+L^2)\mu(\omega)}} \text{ and } K_2(X) = \frac{D_{\omega,X}M_2^2}{\sqrt{2(M_2^2+L^2)\mu(\omega)}} + \frac{2D_{\omega,X}M_2}{\sqrt{\mu(\omega)}} + M_1.$$

(ii) *If Assumptions 1 and 4 hold, then* (2.45) *holds with the right hand side replaced by* $(3+\exp\{1\})\exp\{-\frac{1}{4}\Theta^2\}$*.*

Proof. It suffices to follow the proof of Proposition 2.3, knowing that inequality (2.17) needs to be replaced by (2.42) for the Mirror Descent algorithm. In particular, recalling that (2.37) holds, inequality (2.30) becomes

$$f^N - f(x_*) \quad \leq \quad \frac{D_{\omega,X}(M_2^2+2L^2)}{\sqrt{2(M_2^2+L^2)\mu(\omega)N}} + \frac{D_{\omega,X}M_2^2}{\sqrt{2(M_2^2+L^2)\mu(\omega)\,N}}\mathcal{A} + \frac{2D_{\omega,X}M_2}{\sqrt{\mu(\omega)N}}\mathcal{B}$$

now with

$$\mathcal{A} = \frac{1}{NM_2^2}\sum_{\tau=1}^N \|\Delta_\tau\|_*^2 \quad \text{and} \quad \mathcal{B} = \frac{\sqrt{\mu(\omega)}}{2D_{\omega,X}M_2}\sum_{\tau=1}^N \langle \Delta_\tau, x_* - x_\tau\rangle.$$

□

In the case when f is uniformly convex with convexity parameters ρ and $\mu(f)$, (1.1) has a unique optimal solution x_* and we can additionally bound from above $\mathbb{E}[\|x^N - x_*\|^\rho]$ by an $O(1/\sqrt{N})$ upper bound.

We recall that f is uniformly convex on X with convexity parameters $\rho \geq 2$ and $\mu(f) > 0$ if for all $t \in [0,1]$ and for all $x, y \in X$,

$$f(tx+(1-t)y) \leq tf(x)+(1-t)f(y)-\frac{\mu(f)}{2}t(1-t)(t^{\rho-1}+(1-t)^{\rho-1})\|x-y\|^\rho. \tag{2.46}$$

If a uniformly convex function f is subdifferentiable at x, then

$$\forall y \in X,\ f(y) \geq f(x) + \langle f'(x), y-x\rangle + \frac{\mu(f)}{2}\|y-x\|^\rho$$

and if f is subdifferentiable at two points $x, y \in X$, then

$$\langle f'(y) - f'(x), y-x\rangle \geq \mu(f)\|y-x\|^\rho.$$

Note that if $g(\cdot,\xi)$ is uniformly convex for every ξ then $f(x)=\mathbb{E}\Big[g(x,\xi)\Big]$ is uniformly convex with the same convexity parameters.

Lemma 2.7. *Let Assumptions 1 and 2 hold and assume that the number of iterations $N-1$ of the SMD algorithm is fixed in advance with stepsizes given by* (2.43). *Consider the approximation* $g^N=\frac{1}{\Gamma_N}\left[\sum_{\tau=1}^{N}\gamma_\tau g(x_\tau,\xi_\tau)\right]$ *of $f(x_*)$ and assume that f is uniformly convex. Then* (2.44) *holds and*

$$\mathbb{E}\left[\|x^N-x_*\|^\rho\right]\le \frac{D_{\omega,X}\sqrt{2(M_2^2+L^2)}}{\mu(f)\sqrt{\mu(\omega)}\sqrt{N}}. \tag{2.47}$$

Proof. For every $\tau=1,\dots,N$, since $x_\tau\in X$, the first order optimality conditions give

$$\langle f'(x_*),x_\tau-x_*\rangle\ge 0.$$

Using this inequality and the fact that f is uniformly convex yields

$$\mu(f)\|x_\tau-x_*\|^\rho\le\langle f'(x_\tau)-f'(x_*),x_\tau-x_*\rangle\le\langle f'(x_\tau),x_\tau-x_*\rangle. \tag{2.48}$$

Next, note that since $\rho\ge 2$, the function $\|x\|^\rho$ from E to $\mathbb{R}_+$ is convex as a composition of the convex monotone function x^ρ from $\mathbb{R}_+$ to $\mathbb{R}_+$ and of the convex function $\|x\|$ from E to $\mathbb{R}_+$. It follows that

$$\begin{array}{rcl}\|x^N-x_*\|^\rho & = & \|\frac{1}{\Gamma_N}\sum_{\tau=1}^{N}\gamma_\tau(x_\tau-x_*)\|^\rho\le\sum_{\tau=1}^{N}\frac{\gamma_\tau}{\Gamma_N}\|x_\tau-x_*\|^\rho\\ & \le & \frac{1}{\mu(f)}\sum_{\tau=1}^{N}\frac{\gamma_\tau}{\Gamma_N}\langle f'(x_\tau),x_\tau-x_*\rangle \text{ using (2.48).}\end{array} \tag{2.49}$$

Finally, we prove (2.47) using the above inequality and (2.42), and following the proof of Lemma 2.1. □

2.3. **Multi-step Stochastic Mirror Descent.** The analysis of the SMD algorithm of the previous section was done taking $x_1=x_\omega$ as a starting point. In the case when f is uniformly convex, Algorithm 3 below is a multi-step version of the Stochastic mirror Descent algorithm starting from an arbitrary point $y_1=x_1\in X$. A similar multi-step algorithm was presented in [5] for the *method of dual averaging.* The proofs of this section are adaptations of the proofs of [5] to our setting. We assume in this section that f is uniformly convex, i.e., satisfies (2.46). For multi-step Algorithm 3, at step t, Algorithm 2 is run for N_t-1 iterations starting from y_t instead of x_ω with steps that are constant along these iterations but that are decreasing with the algorithm step t. The output y_{t+1} of step t is the initial point for the next run of Algorithm 2, at step $t+1$. To describe Algorithm 3, it is convenient to introduce

(1) $x^N(x,\gamma)$: the approximate solution provided by Algorithm 2 run for $N-1$ iterations with constant step γ and using $x_1=x$ instead of $x_1=x_\omega$ as a starting point;
(2) $g^N(x,\gamma)$: the approximation of the optimal value of (1.1) computed as in (2.9) where the points $x_1,\dots,x_N$ are generated by Algorithm 2 run for $N-1$ iterations with constant step γ and using $x_1=x$ instead of $x_1=x_\omega$ as a starting point.

In Proposition 2.8, we provide an upper bound for the mean error on the optimal value that is divided by two at each step. We will assume that the prox-function is quadratically growing:

Assumption 5. There exists $0 < M(\omega) < +\infty$ such that

$$V_x(y) \leq \frac{1}{2} M(\omega) \|x - y\|^2 \text{ for all } x, y \in X. \tag{2.50}$$

Recall that the RSA algorithm amounts to taking $V_x(y) = \frac{1}{2}\|x-y\|^2$ with $\|\cdot\| = \|\cdot\|_2$ and thus (2.50) holds in this case with $M(\omega) = 1$.

Algorithm 3: multi-step Stochastic Mirror Descent.

Initialization. Take $y_1 = x_1 \in X$. Fix the number of steps m.

Loop. For $t = 1, \ldots, m$,

1) Compute

$$N_t = \left\lceil \frac{2^{3+\frac{2(t-1)(\rho-1)}{\rho}} (L^2 + M_2^2) M(\omega)}{\mu^2(f)\mu(\omega) D_X^{2(\rho-1)}} \right\rceil \tag{2.51}$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x.

2) Compute $\gamma^t = \dfrac{D_X}{2^{\frac{t-1}{\rho}}\sqrt{N_t}} \sqrt{\dfrac{M(\omega)\mu(\omega)}{2(L^2 + M_2^2)}}$.
3) Run Algorithm 2 (Stochastic Mirror Descent) for $N_t - 1$ iterations, starting from y_t instead of x_ω, to compute $y_{t+1} = x^{N_t}(y_t, \gamma^t)$ obtained using iterations (2.40) with constant step γ^t at each iteration.

Outputs: $y_{m+1} = x^{N_m}(y_m, \gamma^m)$ and $g^{N_m}(y_m, \gamma^m)$.

If for Algorithm 2 (SMD algorithm), the initialization phase consists of taking an arbitrary point x_1 in X instead of x_ω, analogues of Lemmas 2.5, 2.7, and Proposition 2.6 can be obtained using Assumption 5 and replacing (2.42) by the relation

$$\sum_{\tau=1}^{N} \gamma_\tau \langle f'(x_\tau), x_\tau - x_* \rangle \leq \frac{M(\omega)}{2} \|x_1 - x_*\|^2 + \frac{1}{2\mu(\omega)} \sum_{\tau=1}^{N} \gamma_\tau^2 \|G_\tau\|_*^2 + \sum_{\tau=1}^{N} \gamma_\tau \langle \Delta_\tau, x_* - x_\tau \rangle. \tag{2.52}$$

Proposition 2.8. *Let y_{m+1} be the solution generated by Algorithm 3 after m steps. Assume that f is uniformly convex and that Assumptions 1, 2, and 5 hold. Then*

$$\mathbb{E}\Big[\|y_{m+1} - x_*\|\Big] \leq \frac{D_X}{2^{m/\rho}}, \quad \mathbb{E}\Big[\Big|f(y_{m+1}) - f(x_*)\Big|\Big] \leq \mu(f) \frac{D_X^\rho}{2^m}, \tag{2.53}$$

$$\mathbb{E}\Big[\Big|g^{N_m}(y_m, \gamma^m) - f(x_*)\Big] \leq \mu(f) \frac{D_X^\rho}{2^m} + \frac{M_1}{\sqrt{N_m}}. \tag{2.54}$$

Proof. We prove by induction that $\mathbb{E}\Big[\|y_k - x_*\|\Big] \leq D_k := \frac{D_X}{2^{(k-1)/\rho}}$ for $k = 1, \ldots, m+1$. For $k = 1$, the inequality holds. Assume that it holds for some $k < m+1$. Using

(2.52) and following the proof of Lemmas 2.5 and 2.7, we obtain

$$
\begin{array}{lrcl}
(2.55) & \mathbb{E}\Big[\|x^{N_k}(y_k,\gamma^k)-x_*\|^\rho\Big] & \leq & \dfrac{D_k}{\mu(f)\sqrt{N_k}}\sqrt{\dfrac{2(L^2+M_2^2)M(\omega)}{\mu(\omega)}},\\
 & \mathbb{E}\Big[f(x^{N_k}(y_k,\gamma^k))-f(x_*)\Big] & = & \mathbb{E}\Big[f(y_{k+1})-f(x_*)\Big]\\
(2.56) & & \leq & \dfrac{D_k}{\sqrt{N_k}}\sqrt{\dfrac{2(L^2+M_2^2)M(\omega)}{\mu(\omega)}}.
\end{array}
$$

Plugging

$$N_k \geq 8\,\frac{2^{\frac{2(k-1)(\rho-1)}{\rho}}(L^2+M_2^2)M(\omega)}{\mu^2(f)\mu(\omega)D_X^{2(\rho-1)}} = \frac{8(L^2+M_2^2)M(\omega)}{\mu^2(f)\mu(\omega)D_k^{2(\rho-1)}}$$

into (2.55) gives

$$\mathbb{E}\Big[\|y_{k+1}-x_*\|^\rho\Big] = \mathbb{E}\Big[\|x^{N_k}(y_k,\gamma^k)-x_*\|^\rho\Big] \leq D_k\frac{D_k^{\rho-1}}{2} = D_{k+1}^\rho.$$

Since for $\rho \geq 2$, the function x^ρ is convex, using Jensen inequality we conclude that $\mathbb{E}\Big[\|y_{k+1}-x_*\|\Big] \leq D_{k+1}$ which achieves the induction. Next, using (2.56), we obtain $\mathbb{E}\Big[\Big|f(y_{k+1})-f(x_*)\Big|\Big] \leq \mu(f)D_{k+1}^\rho$. Finally, we prove (2.54) using (2.53) and following the end of the proof of Lemma 2.1. □

Corollary 2.9. *Let y_{m+1} be the solution generated by Algorithm 3 after m steps. Assume that f is uniformly convex and that Assumptions 1, 2, and 5 hold. Then for any $\Theta > 0$, $\mathbb{P}\Big(\|y_{m+1}-x_*\|^\rho > 2^{-\frac{m}{2}}\Theta\Big) \leq \frac{D_X^\rho}{\Theta}2^{-\frac{m}{2}}$.*

If at most N calls to the oracle are allowed, Algorithm 3 becomes Algorithm 4.

Algorithm 4: multi-step Stochastic Mirror Descent with no more than N calls to the oracle.

Initialization. Take $y_1 = x_1 \in X$, set $\texttt{Steps} = 1$, $\texttt{Nb}_{\texttt{Call}} = N_1$ and fix the maximal number of calls N to the oracle.

Loop. While $\texttt{Nb}_{\texttt{Call}} \leq N$,

1) Compute $\gamma^{\texttt{Steps}} = \dfrac{D_X}{2^{\frac{\texttt{Steps}-1}{\rho}}\sqrt{N_{\texttt{Steps}}}}\sqrt{\dfrac{M(\omega)\mu(\omega)}{2(L^2+M_2^2)}}$ with $N_{\texttt{Steps}}$ given by (2.51).
2) Run Algorithm 2 (Stochastic Mirror Descent) for $N_{\texttt{Steps}} - 1$ iterations with $N_{\texttt{Steps}}$ given by (2.51), starting from $y_{\texttt{Steps}}$ instead of x_ω, to compute $y_{\texttt{Steps}+1} = x^{N_{\texttt{Steps}}}(y_{\texttt{Steps}}, \gamma^{\texttt{Steps}})$ obtained using iterations (2.40) with constant step $\gamma^{\texttt{Steps}}$ at each iteration.
3) $\texttt{Steps} \leftarrow \texttt{Steps} + 1$, $\texttt{Nb}_{\texttt{Call}} \leftarrow \texttt{Nb}_{\texttt{Call}} + \texttt{N}_{\texttt{Steps}}$.
   End while

Outputs: $y_{\texttt{Steps}} = x^{N_{\texttt{Steps}-1}}(y_{\texttt{Steps}-1}, \gamma^{\texttt{Steps}-1})$ and $g^{N_{\texttt{Steps}-1}}(y_{\texttt{Steps}-1}, \gamma^{\texttt{Steps}-1})$.

Proposition 2.10. *Let $y_{\texttt{Steps}+1}$ be the solution generated by Algorithm 4. Assume that f is uniformly convex and that N is sufficiently large, namely that*

$$N > 1 + \frac{2(2^\beta+1)}{\beta \ln 2} \ln\left(1 + \frac{(2^\beta - 1)}{A(f,\omega)} N\right), \tag{2.57}$$

where $A(f,\omega) = \frac{8(L^2+M_2^2)M(\omega)}{\mu^2(f)\mu(\omega)D_X^{2(\rho-1)}}$ and where $1 \le \beta = 2\frac{\rho-1}{\rho} < 2$. If Assumptions 1, 2, and 5 hold then

$$\begin{array}{lll}\mathbb{E}\Big[\|y_{\texttt{Steps}+1} - x_*\|^\rho\Big] & \le & D_X^\rho \left[\dfrac{2^{\beta+1}A(f,\omega)}{(2^\beta-1)(N-1)+2A(f,\omega)}\right]^{1/\beta}, \\ \mathbb{E}\Big[\Big|f(y_{\texttt{Steps}+1}) - f(x_*)\Big|\Big] & \le & \mu(f) D_X^\rho \left[\dfrac{2^{\beta+1}A(f,\omega)}{(2^\beta-1)(N-1)+2A(f,\omega)}\right]^{1/\beta},\end{array}$$

and $\mathbb{E}\Big[\Big|g^{N_{\texttt{Steps}}}(y_{\texttt{Steps}}, \gamma^{\texttt{Steps}}) - f(x_*)\Big|\Big]$ is bounded from above by

$$\mu(f) D_X^\rho \left[\frac{2^{\beta+1}A(f,\omega)}{(2^\beta-1)(N-1)+2A(f,\omega)}\right]^{1/\beta} + \frac{M_1}{\sqrt{N_{\texttt{Steps}}}}.$$

Proof. In the proof of Proposition (2.8), we have shown that

$$\mathbb{E}\Big[\|y_{\texttt{Steps}+1} - x_*\|^\rho\Big] \le \frac{D_X^\rho}{2^{\texttt{Steps}}} \quad \text{and} \quad \mathbb{E}\Big[\Big|f(y_{\texttt{Steps}+1}) - f(x_*)\Big|\Big] \le \mu(f)\frac{D_X^\rho}{2^{\texttt{Steps}}}. \tag{2.58}$$

Denoting for short $A(f,\omega)$ by A, we will show that

$$\frac{1}{2^{\texttt{Steps}}} \le \left[\frac{2^{\beta+1}A}{(2^\beta-1)(N-1)+2A}\right]^{1/\beta}, \tag{2.59}$$

which, plugged into (2.58), will prove the proposition. Let us check that (2.59) indeed holds. By definition of N_t and of the number of steps of Algorithm 4, we have

$$\texttt{Steps} + 1 + \frac{2^{\beta\,(\texttt{Steps}+1)} - 1}{2^\beta - 1} A = \sum_{t=1}^{\texttt{Steps}+1} (1 + 2^{(t-1)\beta} A) > \sum_{t=1}^{\texttt{Steps}+1} N_t > N$$

which can be written

$$\frac{2^{\beta\,\texttt{Steps}}}{2^\beta - 1} A > \frac{1}{2^\beta}\left(N - \texttt{Steps} - 1 + \frac{A}{2^\beta - 1}\right), \tag{2.60}$$

and

$$N \ge \sum_{t=1}^{\texttt{Steps}} N_t \ge \sum_{t=1}^{\texttt{Steps}} 2^{(t-1)\beta} A = \frac{2^{\beta\,\texttt{Steps}} - 1}{2^\beta - 1} A. \tag{2.61}$$

From (2.61), we obtain an upper bound on the number of steps:

$$\texttt{Steps} \le \frac{\ln\left(1 + \frac{(2^\beta-1)}{A} N\right)}{\beta \ln 2}. \tag{2.62}$$

Combining (2.60), (2.61), and (2.62) gives

$$\begin{aligned}
&\frac{-A}{2^\beta-1}+\frac{1}{2^\beta}\left(N-1+\frac{A}{2^\beta-1}\right)\le\frac{\texttt{Steps}}{2^\beta}+\sum_{t=1}^{\texttt{Steps}}N_t\\
&\le\frac{\texttt{Steps}}{2^\beta}+\sum_{t=1}^{\texttt{Steps}}(1+2^{(t-1)\beta}A)\le\texttt{Steps}(1+\frac{1}{2^\beta})+\frac{2^{\beta\,\texttt{Steps}}-1}{2^\beta-1}A\\
&\le\frac{\ln\left(1+\frac{(2^\beta-1)}{A}N\right)}{\beta\ln 2}(1+\frac{1}{2^\beta})+\frac{2^{\beta\,\texttt{Steps}}-1}{2^\beta-1}A. \qquad (2.63)
\end{aligned}$$

Plugging (2.64) into (2.63) and rearranging the terms gives (2.59). □

Proposition (2.10) gives an $O(1/N^{\rho/2(\rho-1)})$ upper bound for $\mathbb{E}\Big[\Big|f(y_{\texttt{Steps}+1})-f(x_*)\Big|\Big]$, which is tighter, since $1\le\rho/2(\rho-1)<2$, than the upper bounds obtained in the previous sections in the convex case.

Finally, we provide a confidence interval for the optimal value of (1.1), obtained using the following multi-step modified version of Algorithm 3 (a confidence interval can also be obtained for the optimal value of (1.1) using a similar modified version of Algorithm 4):

Algorithm 3': variant of Algorithm 3.

Algorithm 3 with the following modification: for each step t, when Algorithm 2 is run for N_t-1 iterations, the proximal mapping used in (2.40) is now defined replacing in (2.38) the set X by $X\cap B(y_t,\frac{D_X}{2^{(t-1)/\rho}})$.

Proposition 2.11. *Let y_{m+1} be the solution generated by Algorithm 3'. Assume that f is uniformly convex, fix $\Theta>0$, and assume that N_k is sufficiently large for $k=1,\dots,m$, namely that*

$$N_k\ge 2^{\left[2k-2\frac{(k-1)}{\rho}\right]}\Big(K_1(X)+\Theta K_2(X)\Big)^2 \qquad (2.64)$$

with

$$\begin{aligned}
K_1(X)&=\sqrt{\frac{M(\omega)}{2\mu(\omega)(L^2+M_2^2)}}\left(\frac{2L^2+M_2^2}{\mu(f)D_X^{\rho-1}}\right) \text{ and}\\
K_2(X)&=\left(M_2^2\sqrt{\frac{M(\omega)}{2\mu(\omega)(L^2+M_2^2)}}+2M_2\right)\frac{1}{\mu(f)D_X^{\rho-1}}.
\end{aligned}$$

Then if Assumptions 1, 2, and 5 hold, we have

$$\mathbb{P}\left(\left|g^{N_m}(y_m,\gamma^m)-f(x_*)\right|>\frac{\mu(f)D_X^\rho}{2^{m-1}}+\Theta\frac{M_1}{\sqrt{N_m}}\right)\le 2m\exp\{1-\Theta\}+2\exp\{-\tfrac{1}{4}\Theta^2\}.$$

Proof. Let us fix $\Theta>0$. Denoting by $x_\tau,\tau=1,\dots,N_m$, the points generated at the m-th step of the algorithm and setting $f^{N_m}=\frac{1}{N_m}\sum_{\tau=1}^{N_m}f(x_\tau)$, following the proof of Proposition 2.3, we have $\mathbb{P}\left(\left|g^{N_m}(y_m,\gamma^m)-f^{N_m}\right|>\Theta\frac{M_1}{\sqrt{N_m}}\right)\le 2\exp\{-\frac{1}{4}\Theta^2\}$. We now show that

$$\mathbb{P}\left(\left|f^{N_m}-f(x_*)\right|>\frac{\mu(f)D_X^\rho}{2^{m-1}}\right)\le 2m\exp\{1-\Theta\}, \qquad (2.65)$$

which will achieve the proof of the proposition. The proof is by induction on the number of steps of the algorithm. The induction hypothesis is that for some step $k\in\{1,\dots,m\}$, $\forall\ell=1,\dots,k$, there is a set S_ℓ of probability 1 if $\ell=1$ and at least $1-2\exp\{1-\Theta\}$ otherwise such that on $\cap_{\ell=1}^{k} S_\ell$, we have $\|y_k - x_*\| \leq D_k = \frac{D_X}{2^{(k-1)/\rho}}$. For $k=1$, the result holds. Assume now the induction hypothesis for some $k \in \{1,\ldots,m\}$. We intend to show that (2.65) holds with m substituted by k and that there is a set S_{k+1} of probability at least $1-2\exp\{1-\Theta\}$ such that on $\cap_{\ell=1}^{k+1} S_\ell$, we have $\|y_{k+1}-x_*\| \leq D_{k+1} = \frac{D_X}{2^{k/\rho}}$. Denoting now by $x_\tau, \tau=1,\ldots,N_k$, the points generated at the k-th step of the algorithm, using (2.52) and the fact that $\|G_\tau\|_*^2 \leq 2(L^2 + \|\Delta_\tau\|_*^2)$, we have for $f^{N_k} - f(x_*)$ the upper bound

(2.66)

$$\begin{array}{l}\frac{1}{N_k \gamma^k}\left[\frac{M(\omega)}{2}\|y_k - x_*\|^2 + \frac{1}{\mu(\omega)}\sum_{\tau=1}^{N} \gamma_\tau^2 (L^2 + \|\Delta_\tau\|_*^2) + \sum_{\tau=1}^{N} \gamma_\tau \langle \Delta_\tau, x_* - x_\tau\rangle\right] \\ \leq U_k := \frac{M(\omega) D_k^2}{2 N_k \gamma^k} + \frac{L^2 \gamma^k}{\mu(\omega)} + \frac{\gamma^k M_2^2}{\mu(\omega)}\mathcal{A}_k + \frac{2 D_k M_2}{N_k}\mathcal{B}_k \text{ on } \cap_{\ell=1}^{k} S_\ell\end{array}$$

where

$$\mathcal{A}_k = \frac{1}{N_k M_2^2}\sum_{\tau=1}^{N_k} \|\Delta_\tau\|_*^2 \text{ and } \mathcal{B}_k = \frac{1}{2 D_k M_2}\sum_{\tau=1}^{N_k}\langle \Delta_\tau, x_* - x_\tau\rangle.$$

Observe that on $\cap_{\ell=1}^{k} S_\ell$, we have $\|x_* - y_k\| \leq D_k$ and by definition of x_τ, we have $\|y_k - x_\tau\| \leq D_k$ for $\tau = 1,\ldots,N_k$. It follows that we can follow the proof of Proposition 2.3 to show that for any $\Theta > 0$,

$$\mathbb{P}\Big(\mathcal{A}_k > \Theta\Big) \leq \exp\{1-\Theta\} \text{ and } \mathbb{P}\Big(\mathcal{B}_k > \Theta\sqrt{N_k}\Big) \leq \exp\{-\frac{1}{4}\Theta^2\}.$$

Thus there is a set S_{k+1} of probability at least $1-2\exp\{1-\Theta\}$ such that on S_{k+1}, we have $\mathcal{A}_k \leq \Theta$ and $\mathcal{B}_k \leq \Theta\sqrt{N_k}$. Next, on $\cap_{\ell=1}^{k+1} S_\ell$, plugging into (2.66) the upper bounds Θ and $\Theta\sqrt{N_k}$ for respectively $\mathcal{A}_k$ and $\mathcal{B}_k$, using the definition of γ^k, and the lower bound (2.64) on N_k, we obtain for $f^{N_k} - f(x_*)$ the upper bound $\frac{\mu(f) D_X^\rho}{2^k} = \mu(f) D_{k+1}^\rho$. Observing that $\mathbb{P}(\cap_{\ell=1}^{k+1} S_\ell) \geq 1 - 2k\exp\{1-\Theta\}$, we have shown (2.65) with step m substituted by step k. Finally, using (2.49), we have on $\cap_{\ell=1}^{k+1} S_\ell$ for $\|y_{k+1} - x_*\|^\rho$ the upper bound $\frac{U_k}{\mu(f)}$ where U_k is defined in (2.66). Since we have just shown that on $\cap_{\ell=1}^{k+1} S_\ell$, U_k is bounded from above by $\mu(f) D_{k+1}^\rho$, this achieves the induction step. □

2.4. **Application to two stage stochastic convex programs.** Consider the case when (1.1) is a two-stage risk-neutral stochastic convex program, i.e., $\mathcal{R}_i = \mathbb{E}$ is the expectation, x is the first stage decision variable, $f(x) = f_1(x) + \mathbb{E}[\mathcal{Q}(x,\xi)]$ where $\mathcal{Q}(x,\xi)$ is the second stage cost given by

$$\mathcal{Q}(x,\xi) = \left\{\begin{array}{l}\min_y \; f_2(x,y,\xi)\\ y \in \mathcal{S}(x,\xi) = \{y : h(x,y,\xi) \leq 0, Ax + By = b(\xi)\}\end{array}\right. \quad (2.67)$$

for some function h taking values in $\mathbb{R}^m$ and some random vector $\xi \in L_p(\Omega,\mathcal{F},\mathbb{P})$ with $p \geq 2$ and support Θ. We make the following assumptions:

- (H0) $X \subset \mathbb{R}^p$ is a nonempty, compact, and convex set;
- (H1) f_1 is convex, proper, and lower semicontinuous;
- (H2) for every $x \in X$ and $y \in \mathbb{R}^q$ the function $f_2(x,y,\cdot)$ is measurable and for every $\xi \in \Theta$, the function $f_2(\cdot,\cdot,\xi)$ is differentiable, convex, proper, and lower semicontinuous;
- (H3) for every $\xi \in \Theta$, the function $h(\cdot,\cdot,\xi)$ is convex and differentiable;
- (H4) for every $x \in X$ and for every $\xi \in \Theta$ the set $\mathcal{S}(x,\xi)$ is compact and there exists $y_{x,\xi} \in \mathcal{S}(x,\xi)$ such that $h(x, y_{x,\xi}, \xi) < 0$.

With the notation of Section 1, we have $f(x) = \mathbb{E}[g(x,\xi)]$ where $g(x,\xi) = f_1(x) + \mathcal{Q}(x,\xi)$. Assumptions (H1), (H2), and (H3) imply the convexity of f. Assumptions (H2), and (H4) imply that for every $x \in X$ and $\xi \in \Theta$, the second stage cost $\mathcal{Q}(x,\xi)$ is finite which implies the finiteness of $\delta(x,\xi)$. Relations (2.5)-(a), (2.6)-(a), and (2.7)-(a) in respectively Assumptions 2, 3, and 4 are thus satisfied. Assumptions (H2), (H3), and (H4) imply that for every $x \in X$ and $\xi \in \Theta$, the function $x \to \mathcal{Q}(x,\xi)$ is subdifferentiable on X with bounded subgradients at any $x \in X$. For fixed $x \in X$ and $\xi \in \Theta$, let $y(x,\xi)$ be an optimal solution of (2.67) and let $(\lambda(x,\xi), \mu(x,\xi))$ be an optimal dual solution for (2.67) where $\lambda(x,\xi)$ and $\mu(x,\xi)$ are optimal Lagrange multipliers for respectively the equality and inequality constraints. Then for any $x \in X$ and $\xi \in \Theta$, denoting by $I(x,y,\xi) := \{i \in \{1,\ldots,m\} : h_i(x,y,\xi) = 0\}$ the set of active inequality constraints at y for problem (2.67),

$$S(x,\xi) = \nabla_x f_2(x, y(x,\xi), \xi) + A^\top \lambda(x,\xi) + \sum_{i \in I(x,y(x,\xi),\xi)} \mu_i(x,\xi) \nabla_x h(x, y(x,\xi), \xi)$$

belongs to the subdifferential $\partial_x \mathcal{Q}(x,\xi)$ and is bounded. As a result, for any $x \in X$, denoting by $s_1(x)$ an arbitrary element from $\partial f_1(x)$, $f'(x) := \mathbb{E}[G(x,\xi)]$ is a subgradient of f at x with $G(x,\xi) = s_1(x) + S(x,\xi)$ and $\|G(x,\xi)\|_*$ bounded for any $x \in X$ and $\xi \in \Theta$. It follows that Assumption 1 is satisfied as well as Relations (2.5)-(b), (2.6)-(b), and (2.7)-(b) in respectively Assumptions 2, 3, and 4.

We can thus apply the developments of Sections 2.1 and 2.2 to solve a two-stage stochastic risk-neutral convex optimization problem and to obtain a confidence interval on its optimal value when assumptions (H0), (H1), (H2), (H3), and (H4) are satisfied.

If, additionally, f_1 is uniformly convex on X and if for every $\xi \in \Theta$ the function $f_2(\cdot,\cdot,\xi)$ is uniformly convex, then f is uniformly convex on X and Section (2.3) can be used to solve the corresponding stochastic program (1.1) and to obtain a confidence interval on its optimal value.

3. Quality of approximate solutions in the risk-averse case

Consider problem (1.1) with $\mathcal{R} = \mathcal{R}_i$ an extended polyhedral risk measure:

Definition 3.1. [3] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $h(z) = (h_1(z), \ldots, h_{n_{2,2}}(z))^\top$ for given functions $h_1, \ldots, h_{n_{2,2}} : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to L_{p'}(\Omega, \mathcal{F}, \mathbb{P})$ with $2 \leq p' \leq p$. A risk measure $\mathcal{R}$ on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [2, \infty)$ is called* extended polyhedral *if there exist matrices $A_1, A_2, B_{2,0}, B_{2,1}$, and vectors a_1, a_2, c_1, c_2 such that for every random variable $Z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$*

$$\mathcal{R}(Z) = \left\{ \begin{array}{l} \inf\, c_1^\top y_1 + \mathbb{E}[c_2^\top y_2] \\ y_1 \in \mathbb{R}^{k_1},\ y_2 \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{k_2}), \\ A_1 y_1 \leq a_1,\ A_2 y_2 \leq a_2\ a.s., \\ B_{2,1} y_1 + B_{2,0} y_2 = h(Z)\ a.s. \end{array} \right. \tag{3.68}$$

In what follows, we make the following assumption on h in (3.68):

(H0') The function $h(z)$ is affine: $h(z) = z b_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$.

Representation (3.68) can alternatively be written

$$\mathcal{R}(Z) = \left\{ \begin{array}{l} \inf_{y_1}\ c_1^\top y_1 + \mathbb{E}[\mathcal{Q}(y_1, Z)] \\ A_1 y_1 \leq a_1, \end{array} \right. \tag{3.69}$$

where the recourse function $\mathcal{Q}(y_1, z)$ is given by

$$\mathcal{Q}(y_1, z) = \left\{ \begin{array}{l} \inf_{y_2} \, c_2^\top y_2 \\ A_2 y_2 \leq a_2 \\ B_{2,0} y_2 = z b_2 + \tilde{b}_2 - B_{2,1} y_1. \end{array} \right. \tag{3.70}$$

In other words, $\mathcal{R}(Z)$ is the optimal value of a two-stage stochastic program where Z appears in the right-hand side of the second stage problem. We can re-write (1.1) as

$$\left\{ \begin{array}{l} \inf\limits_{y_1, x} \, c_1^\top y_1 + \mathbb{E}\Big[\mathcal{Q}\Big(y_1, g(x, \xi)\Big)\Big] \\ A_1 y_1 \leq a_1, \; x \in X, \end{array} \right. \tag{3.71}$$

with $\mathcal{Q}(\cdot, \cdot)$ given by (3.70). This problem is of the form (1.1) with $\mathcal{R}_i$ the expectation and with $x, g(x, \xi)$, and X respectively replaced by $\tilde{x} = (y_1; x)$, $\tilde{g}(\tilde{x}, \xi) = c_1^\top y_1 + \mathcal{Q}\Big(y_1, g(x, \xi)\Big)$, and $\tilde{X} = \{\tilde{x} = (y_1; x) : x \in X, A_1 y_1 \leq a_1\}$. To guarantee the convexity of the objective function in (3.71) and Assumptions 1-4, we make the following assumptions on $\mathcal{R}$ and g:

(H1') Complete recourse: $Y_1 := \{y_1 : A_1 y_1 \leq a_1\}$ is nonempty and bounded and $\{B_{2,0} y_2 : A_2 y_2 \leq a_2\} = \mathbb{R}^{n_{2,2}}$.

(H2') The feasible set

$$\mathcal{D} = \{\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^{n_{2,2}} \times \mathbb{R}^{n_{2,1}} \; : \; \lambda_2 \leq 0, \;\; B_{2,0}^\top \lambda_1 + A_2^\top \lambda_2 = c_2\} \tag{3.72}$$

of the dual of the second stage problem (3.70) is nonempty.

(H3') The set $\mathcal{D}$ given by (3.72) is bounded.

(H4') For the set $\mathcal{D}$ given by (3.72), we have that $\mathcal{D} \subseteq \{-b_2\}^* \times \mathbb{R}^{n_{2,1}}$.

(H5') For every ξ, the function $g(\cdot, \xi)$ is convex and lower semicontinuous on X and the subdifferential $\partial_x g(x, \xi)$ is bounded for every $x \in X$.

If X is closed, bounded, and convex, (H1') implies that $\tilde{X}$ is also closed, bounded and convex. Moreover, we can show that assumptions (H1'), (H2'), (H3'), (H4'), and (H5') imply that the objective function in (3.71) is convex and has bounded subgradients:

Lemma 3.2. *Consider the objective function* $f(\tilde{x}) = c_1^\top y_1 + \mathbb{E}\Big[\mathcal{Q}\Big(y_1, g(x, \xi)\Big)\Big]$ *of* (3.71) *in variable* $\tilde{x} = (y_1; x)$. *Assume that (H1'), (H2'), (H3'), (H4'), and (H5') hold. Then*

(i) $\mathcal{Q}\Big(y_1, g(x, \tilde{\xi})\Big)$ *is finite for every* $\tilde{\xi}$ *and every* $\tilde{x} \in \tilde{X}$;

(ii) *for every* $\tilde{\xi}$, *the function* $\tilde{x} \rightarrow \mathcal{Q}_{\tilde{\xi}}(\tilde{x}) = \mathcal{Q}\Big(y_1, g(x, \tilde{\xi})\Big)$ *is convex and has bounded subgradients on* $\tilde{X}$;

(iii) f *is convex and has bounded subgradients on* $\tilde{X}$.

Proof. Since (H1') holds, for every $y_1 \in Y_1$ and every $z \in \mathbb{R}$, the feasible set of problem (3.70) which defines $\mathcal{Q}(y, z)$ is nonempty. Due to (H2'), the feasible set of the dual of this problem is nonempty too. It follows that both the primal and the dual have the same finite optimal value (this shows item (i)) and by duality we can express $\mathcal{Q}(y_1, z)$ as the optimal value of the dual problem:

$$\mathcal{Q}(y_1, z) = \max_{(\lambda_1, \lambda_2) \in \mathcal{D}} \lambda_1^\top (z b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2 \tag{3.73}$$

with $\mathcal{D}$ given by (3.72). Next, observe that $\mathcal{Q}(y_1, \cdot)$ is monotone:

$$\forall y_1 \in Y_1,\ \forall z_1, z_2 \in \mathbb{R},\ z_1 \geq z_2 \Rightarrow \mathcal{Q}(y_1, z_1) \geq \mathcal{Q}(y_1, z_2). \tag{3.74}$$

Indeed, if $z_1 \geq z_2$, for every $(\lambda_1, \lambda_2) \in \mathcal{D}$, since (H4') holds, we have $\lambda_1^\top b_2 \geq 0$ and

$$\lambda_1^\top (z_1 b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2 \geq \lambda_1^\top (z_2 b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2$$

for every $y_1 \in Y_1$. Taking the maximum when $(\lambda_1, \lambda_2) \in \mathcal{D}$ in each side of the previous inequality gives $\mathcal{Q}(y_1, z_1) \geq \mathcal{Q}(y_1, z_2)$. Now take a realization $\tilde{\xi}$ of ξ and $\tilde{x} = (y_1; x), \tilde{x}_0 = (y_1^0; x_0) \in \tilde{X}$. Using the convexity of $g(\cdot, \tilde{\xi})$, we have

$$g(x, \tilde{\xi}) \geq g(x_0, \tilde{\xi}) + g_x'(x_0, \tilde{\xi})^\top (x - x_0)$$

where $g_x'(x_0, \tilde{\xi})$ is a subgradient of $g(\cdot, \tilde{\xi})$ at x_0. Combining this inequality and (3.74) gives

$$\mathcal{Q}_{\tilde{\xi}}(\tilde{x}) = \mathcal{Q}\Big(y_1, g(x, \tilde{\xi})\Big) \geq \mathcal{Q}\Big(y_1, g(x_0, \tilde{\xi}) + g_x'(x_0, \tilde{\xi})^\top (x - x_0)\Big)$$

for every $y_1 \in Y_1$. Next, we have that $\mathcal{Q}(y_1, z)$ is convex and its subdifferential is given by

$$\partial \mathcal{Q}(y_1, z) = \left\{ \begin{pmatrix} B_{2,1}^\top \lambda_1 \\ \lambda_1^\top b_2 \end{pmatrix} : (\lambda_1, \lambda_2) \in \mathbb{R}^{n_{2,2}} \times \mathbb{R}^{n_{2,1}} \in \mathcal{D}^*_{y,z} \right\}$$

where $\mathcal{D}^*_{y,z}$ is the set of optimal solutions to the dual problem (3.73). Denoting by $(\lambda_1(y, z), \lambda_2(y, z))$ an optimal solution to (3.73), we then have

$$\mathcal{Q}_{\tilde{\xi}}(\tilde{x}) = \mathcal{Q}\Big(y_1, g(x, \tilde{\xi})\Big) \geq \mathcal{Q}_{\tilde{\xi}}(\tilde{x}_0) + \begin{pmatrix} B_{2,1}^\top \lambda_1(y_1^0, g(x_0, \tilde{\xi})) \\ \lambda_1(y_1^0, g(x_0, \tilde{\xi}))^\top b_2 g_x'(x_0, \tilde{\xi}) \end{pmatrix}^\top \Big(\tilde{x} - \tilde{x}_0\Big).$$

It follows that for every $\tilde{\xi}$, $\mathcal{Q}_{\tilde{\xi}}(\cdot)$ is convex and its subdifferential is given by

$$\partial \mathcal{Q}_{\tilde{\xi}}(y_1^0, x_0) = \left\{ \begin{pmatrix} B_{2,1}^\top \lambda_1 \\ \lambda_1^\top b_2 g_x'(x_0, \tilde{\xi}) \end{pmatrix} : (\lambda_1, \lambda_2) \in \mathbb{R}^{n_{2,2}} \times \mathbb{R}^{n_{2,1}} \in \mathcal{D}^*_{y_1^0,\, g(x_0, \tilde{\xi})} \right\}.$$

Since $\mathcal{D}^*_{y_1^0,\, g(x_0, \tilde{\xi})}$ is a subset of the bounded set $\mathcal{D}$ and since (H5') holds, all subgradients of $\mathcal{Q}_{\tilde{\xi}}(\cdot)$ are bounded for every $\tilde{\xi}$: we have proved (ii). Item (iii) is an immediate consequence of (ii). □

It follows from Lemma 3.2-(iii) that Assumption 1 is satisfied. With the notation of Section 2 and of Lemma 3.2, we have $\delta(\tilde{x}, \xi) = \mathcal{Q}_{\xi}(\tilde{x}) - \mathbb{E}[\mathcal{Q}_{\xi}(\tilde{x})]$, which is finite for every ξ and $\tilde{x} \in \tilde{X}$ using Lemma 3.2-(i). It follows that relations (2.5)-(a), (2.6)-(a), and (2.7)-(a) respectively in Assumptions 2, 3, and 4 are satisfied. Finally Lemma 3.2-(ii) shows that relations (2.5)-(b), (2.6)-(b), and (2.7)-(b) respectively in Assumptions 2, 3, and 4 are also satisfied. This shows that we can use the developments of Sections 2.1, and 2.2 to solve the two-stage stochastic risk-averse convex optimization problem (1.1) and to obtain a confidence interval on its optimal value when $\mathcal{R}_i$ is an extended polyhedral risk measure and when assumptions (H0'), (H1'), (H2'), (H3'), (H4'), and (H5') are satisfied.

Risk-averse stochastic programs expressed in terms of extended polyhedral risk measures (EPRM) share many properties with risk-neutral stochastic programs. Moreover, many popular risk measures can be written as extended polyhedral risk measures satisfying assumptions (H0'), (H1'), (H2'), (H3'), and (H4'). Examples of such risk measures are the CVaR, some spectral risk measures, the optimized certainty equivalent and the expected utility with piecewise affine utility function.

We refer to Examples 2.16 and 2.17 in [3] for a discussion on these examples. Conditions ensuring that an EPRM is convex, coherent or consistent with second order stochastic dominance are given in [3]. Multiperiod versions of these risk measures are also defined in [3]. In this context, a convenient property of the corresponding risk-averse program is that we can write dynamic programming equations and solve it, in the case when the problem is convex, by decomposition using for instance SDDP [7]; see [3] for more details and examples of multiperiod EPRM. EPRM are an extension of the polyhedral risk measures introduced in [2] where the reader will find additional examples of (extended) polyhedral risk measures.

4. Testing hypotheses on the optimal value of several stochastic programs

Let us now go back to our motivating hypotheses testing problems: given m samples from the distributions of m random vectors $\xi_1, \ldots, \xi_m \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, we want to study tests (1.2)-(a), (b), (c) where ρ_i is given by (1.1) with $\mathcal{R}_i$ an extended polyhedral risk measure. Let assumptions (H1'), (H2'), (H3'), (H4'), and (H5') defined in Section 3 be satisfied. Let us start with test (1.2)-(a):

$$H_0 : \; \rho_1(\xi_1) = \rho_2(\xi_2) = \ldots = \rho_m(\xi_m) \;\text{ against }\; \overline{H_0}. \tag{4.75}$$

For every $i = 1, \ldots, m$, using N_i calls to the stochastic oracle, we have shown in Sections 2 and 3 how to obtain a confidence interval for $\rho_i(\xi_i)$: for any $\Theta > 0$,

$$\mathbb{P}\left(\left|g^{N_i} - \rho_i(\xi_i)\right| > h^{N_i}(\Theta)\right) \leq \varepsilon(\Theta)$$

for some estimator g^{N_i} of $\rho_i(\xi_i)$ and known functions $h^{N_i}(\Theta)$ and $\varepsilon(\Theta)$ that depend on the algorithm used to estimate $\rho_i(\xi_i)$. For instance, since Assumptions 1 and 4 hold, for the Stochastic Mirror Descent algorithm we have

$$h^{N_i}(\Theta) = \frac{K_1(X) + \Theta K_2(X)}{\sqrt{N_i}} \text{ and } \varepsilon(\Theta) = (3 + \exp\{1\}) \exp\{-\frac{1}{4}\Theta^2\}.$$

where the constants $K_1(X)$ and $K_2(X)$ are given by

$$K_1(X) = \frac{D_{\omega,X}(M_2^2 + 2L^2)}{\sqrt{2(M_2^2 + L^2)\mu(\omega)}} \text{ and } K_2(X) = \frac{D_{\omega,X} M_2^2}{\sqrt{2(M_2^2 + L^2)\mu(\omega)}} + \frac{2D_{\omega,X} M_2}{\sqrt{\mu(\omega)}} + M_1.$$

In all cases, $\varepsilon(\Theta)$ is a nonnegative, decreasing function with $\lim_{\Theta \to +\infty} \varepsilon(\Theta) = 0$. Let $0 < \alpha < 1$ and let us fix Θ sufficiently large to have $\varepsilon(\Theta) \leq \frac{\alpha}{m}$. We define for test (4.75) the rejection region W_a to be the set of samples such that the realizations of the corresponding confidence intervals have no intersection, i.e.,

$$\begin{array}{lll} W_a & = & \left\{ \xi \; : \; \bigcap\limits_{i}^{m} \left[g^{N_i} - h^{N_i}(\Theta), g^{N_i} + h^{N_i}(\Theta) \right] = \emptyset \right\} \\ & = & \left\{ \xi \; : \; \max\limits_{i=1,\ldots,m} \left[g^{N_i} - h^{N_i}(\Theta) \right] > \min\limits_{i=1,\ldots,m} \left[g^{N_i} + h^{N_i}(\Theta) \right] \right\}. \end{array}$$

If H_0 holds, denoting $\rho_0 = \rho_1(\xi_1) = \ldots = \rho_m(\xi_m)$, we have

$$\begin{array}{l}
\mathbb{P}\Big(\max\limits_{i=1,\ldots,m} \Big[g^{N_i} - h^{N_i}(\Theta)\Big] > \min\limits_{i=1,\ldots,m} \Big[g^{N_i} + h^{N_i}(\Theta)\Big]\Big) \\
= \mathbb{P}\left(\max\limits_{i=1,\ldots,m} \Big[g^{N_i} - h^{N_i}(\Theta) - \rho_0\Big] + \max\limits_{i=1,\ldots,m} \Big[\rho_0 - g^{N_i} - h^{N_i}(\Theta)\Big] > 0\right) \\
\leq \sum_{i=1}^m \Big[\mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) - \rho_i(\xi_i) > 0\Big) + \mathbb{P}\Big(\rho_i(\xi_i) - g^{N_i} - h^{N_i}(\Theta) > 0\Big)\Big] \\
\leq m\varepsilon(\Theta) \leq \alpha
\end{array}$$

and W_a is a rejection region for (4.75) for some significance level not larger that α. Moreover, as stated in the following lemma, if H_0 does not hold and if two optimal values are sufficiently distant then the probability to accept H_0 will be small:

Lemma 4.1. *Consider test (4.75) with rejection region W_a. If for some $i, j \in \{1, \ldots, m\}$ with $i \neq j$, we have $\rho_i(\xi_i) > \rho_j(\xi_j) + 2\Big(h^{N_i}(\Theta) + h^{N_j}(\Theta)\Big)$ then the probability to accept H_0 is not larger than $\frac{2\alpha}{m}$.*

Proof. We first check that

$$\left.\begin{array}{ll}
\rho_i(\xi_i) > \rho_j(\xi_j) + 2\Big(h^{N_i}(\Theta) + h^{N_j}(\Theta)\Big) & (a) \\
g^{N_j} - h^{N_j}(\Theta) \leq \rho_j(\xi_j) & (b) \\
\rho_i(\xi_i) \leq g^{N_i} + h^{N_i}(\Theta) & (c)
\end{array}\right\} \Rightarrow g^{N_j} + h^{N_j}(\Theta) < g^{N_i} - h^{N_i}(\Theta). \tag{4.76}$$

Indeed, if (4.76)-(a),(b), and (c) hold then

$$g^{N_j} + h^{N_j}(\Theta) \overset{(4.76)-(b)}{\leq} \rho_j(\xi_j) + 2h^{N_j}(\Theta) \overset{(4.76)-(a)}{<} \rho_i(\xi_i) - 2h^{N_i}(\Theta) \overset{(4.76)-(c)}{\leq} g^{N_i} - h^{N_i}(\Theta).$$

Assume now that $\rho_i(\xi_i) > \rho_j(\xi_j) + 2\Big(h^{N_i}(\Theta) + h^{N_j}(\Theta)\Big)$. We have $\mathbb{P}\Big(g^{N_j} - h^{N_j}(\Theta) \leq \rho_j(\xi_j)\Big) \geq 1 - \frac{\alpha}{m}$ and $\mathbb{P}\Big(g^{N_i} + h^{N_i}(\Theta) \geq \rho_i(\xi_i)\Big) \geq 1 - \frac{\alpha}{m}$. Next, since $g^{N_j} + h^{N_j}(\Theta) < g^{N_i} - h^{N_i}(\Theta)$ implies that H_0 is rejected, we get

$$\begin{array}{rcl}
\mathbb{P}\Big(\text{reject } H_0\Big) & \geq & \mathbb{P}\Big(g^{N_j} + h^{N_j}(\Theta) < g^{N_i} - h^{N_i}(\Theta)\Big) \\
& \overset{(4.76)}{\geq} & \mathbb{P}\Big(\Big\{g^{N_j} - h^{N_j}(\Theta) \leq \rho_j(\xi_j)\Big\} \bigcap \Big\{\rho_i(\xi_i) \leq g^{N_i} + h^{N_i}(\Theta)\Big\}\Big) \\
& \geq & \mathbb{P}\Big(g^{N_j} - h^{N_j}(\Theta) \leq \rho_j(\xi_j)\Big) + \mathbb{P}\Big(\rho_i(\xi_i) \leq g^{N_i} + h^{N_i}(\Theta)\Big) - 1 \\
& \geq & 1 - \frac{2\alpha}{m}
\end{array}$$

which achieves the proof of the lemma. □

Consider now test (1.2)-(b):

$$H_0^i : \rho_i(\xi_i) \leq \rho_j(\xi_j) \text{ for } 1 \leq j \neq i \leq m \text{ against } \overline{H_0^i}. \tag{4.77}$$

Let $0 < \alpha < 1$ and fix Θ sufficiently large to have $\varepsilon(\Theta) \leq \frac{\alpha}{2(m-1)}$. We define the rejection region

$$W_b = \Big\{\xi \,:\, \exists 1 \leq j \neq i \leq m \text{ such that } g^{N_i} - h^{N_i}(\Theta) > g^{N_j} + h^{N_j}(\Theta)\Big\}.$$

If H_0 holds, we have

$$\begin{array}{l}\mathbb{P}\Big(\exists\, 1 \leq j \neq i \leq m \,:\, g^{N_i} - h^{N_i}(\Theta) > g^{N_j} + h^{N_j}(\Theta)\Big]\Big)\\ \leq \displaystyle\sum_{1 \leq j \neq i \leq m} \mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) > g^{N_j} + h^{N_j}(\Theta)\Big)\\ \leq \displaystyle\sum_{1 \leq j \neq i \leq m} \Big(\mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) - \rho_i(\xi_i) + \rho_j(\xi_j) - g^{N_j} - h^{N_j}(\Theta) > 0\Big)\Big)\\ \leq \displaystyle\sum_{1 \leq j \neq i \leq m} \Big(\mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) - \rho_i(\xi_i) > 0\Big) + \mathbb{P}\Big(\rho_j(\xi_j) - g^{N_j} - h^{N_j}(\Theta) > 0\Big)\Big)\\ \leq 2(m-1)\varepsilon(\Theta) \leq \alpha\end{array}$$

and W_b is a rejection region for (4.77) for some significance level not larger that α. We also have an analog of Lemma 4.1:

Lemma 4.2. *Consider test* (4.77) *with rejection region* W_b. *If for some* $j \in \{1, \ldots, m\}$ *with* $i \neq j$, *we have* $\rho_i(\xi_i) > \rho_j(\xi_j) + 2\Big(h^{N_i}(\Theta) + h^{N_j}(\Theta)\Big)$ *then the probability to accept* H_0 *is not larger than* $\frac{\alpha}{m-1}$.

Proof. The proof is analog to the proof of Lemma 4.1. □

Finally, consider test (1.2)-(c):

$$H_0 : \rho_1(\xi_1) \leq \rho_2(\xi_2) \leq \ldots \leq \rho_m(\xi_m) \text{ against } \overline{H_0}. \tag{4.78}$$

Let $0 < \alpha < 1$ and fix Θ sufficiently large to have $\varepsilon(\Theta) \leq \frac{\alpha}{2(m-1)}$. We define the rejection region

$$W_c = \left\{\xi \,:\, \exists i \in \{1, \ldots, m-1\} \text{ such that } g^{N_i} - h^{N_i}(\Theta) > g^{N_{i+1}} + h^{N_{i+1}}(\Theta)\right\}.$$

If H_0 holds, we have

$$\begin{array}{l}\mathbb{P}\Big(\exists i \in \{1, \ldots, m-1\} \,:\, g^{N_i} - h^{N_i}(\Theta) > g^{N_{i+1}} + h^{N_{i+1}}(\Theta)\Big]\Big)\\ \leq \displaystyle\sum_{i=1}^{m-1} \mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) > g^{N_{i+1}} + h^{N_{i+1}}(\Theta)\Big)\\ \leq \displaystyle\sum_{i=1}^{m-1} \mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) - \rho_i(\xi_i) + \rho_{i+1}(\xi_{i+1}) - g^{N_{i+1}} - h^{N_{i+1}}(\Theta) > 0\Big)\\ \leq \displaystyle\sum_{i=1}^{m-1} \Big(\mathbb{P}\Big(g^{N_i} - h^{N_i}(\Theta) - \rho_i(\xi_i) > 0\Big) + \mathbb{P}\Big(\rho_{i+1}(\xi_{i+1}) - g^{N_{i+1}} - h^{N_{i+1}}(\Theta) > 0\Big)\Big)\\ \leq 2(m-1)\varepsilon(\Theta) \leq \alpha\end{array}$$

and W_c is a rejection region for (4.78) for some significance level not larger that α. As for test (4.75), we can bound from above the type II error under some assumptions:

Lemma 4.3. *Consider test* (4.78) *with rejection region* W_c. *If for some* $i \in \{1, \ldots, m-1\}$ *we have* $\rho_i(\xi_i) > \rho_{i+1}(\xi_{i+1}) + 2\Big(h^{N_i}(\Theta) + h^{N_{i+1}}(\Theta)\Big)$ *then the probability to accept* H_0 *is not larger than* $\frac{\alpha}{m-1}$.

Proof. The proof is analog to the proof of Lemma 4.1. □

5. Application to the testing of hypotheses on several risk measure values and future work

Let $\mathcal{R}$ be an extended polyhedral risk measure (EPRM). Since the value of an EPRM is the optimal value of a linear two-stage risk-neutral stochastic optimization problem, if risk measure $\mathcal{R}$ satisfies Assumptions (H1'), (H2'),(H3'), and (H4') given in Section 3, we can use the developments of Section 2 to obtain a confidence interval for the risk measure value $\mathcal{R}(\xi)$ on the basis of a sample $\tilde{\xi}_1, \ldots, \tilde{\xi}_m$, of ξ. Moreover, the tests

$$\begin{array}{l}(a)\ H_0: \ \mathcal{R}(\xi_1) = \mathcal{R}(\xi_2) = \ldots = \mathcal{R}(\xi_m) \ \text{ against } \overline{H_0},\\ (b)\ H_0^i: \ \mathcal{R}(\xi_i) \leq \mathcal{R}(\xi_j), \ \ 1 \leq j \neq i \leq m \ \text{ against } \overline{H_0^i},\\ (c)\ H_0: \ \mathcal{R}(\xi_1) \leq \mathcal{R}(\xi_2) \leq \ldots \leq \mathcal{R}(\xi_m) \ \text{ against } \overline{H_0},\end{array} \tag{5.79}$$

for random one dimensional variables $\xi_1, \ldots, \xi_m$, are special cases of tests (1.2)-(a), (b), (c). These tests are of interest if we wish to compare the risk of m one dimensional distributions $\xi_1, \ldots, \xi_m$ and choose the one with smallest risk when the risk is assessed using risk measure $\mathcal{R}$. Under Assumptions (H1'), (H2'),(H3'), and (H4') given in Section 3, the rejection regions for these tests are given in Section 4.

Finally, we mention two possible extensions of this work: (i) take a coherent risk measure for $\mathcal{R}_i$ in (1.1) and (ii) on the basis of samples of stochastic processes (X_t) and (Y_t) and given conditional risk mappings $\mathcal{R}_2, \ldots, \mathcal{R}_T$, study the test

$$\begin{array}{ll}H_0: & \mathcal{R}_2\Big[X_2 + \ldots + \mathcal{R}_{T-1}\Big[X_{T-1} + \mathcal{R}_T\Big[X_T\Big]\Big]\Big] \quad \text{against } \overline{H_0}.\\ & = \mathcal{R}_2\Big[Y_2 + \ldots + \mathcal{R}_{T-1}\Big[Y_{T-1} + \mathcal{R}_T\Big[Y_T\Big]\Big]\Big]\end{array}$$

Acknowledgments. The author would like to thank Arkadi Nemirovski, Anatoli Iouditski, and Alexander Shapiro for helpful discussions. The author's research was partially supported by an FGV grant, CNPq grant 307287/2013-0, and PRONEX-Optimization.

References

[1] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[2] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16:69–95, 2005.

[3] V. Guigues and W. Römisch. Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures. *SIAM Journal on Optimization*, 22:286–312, 2012.

[4] B.L. Jones, M.L. Puri, and R. Zitikis. Testing hypotheses about the equality of several risk measure values with applications in insurance. *Mathematics and Economics*, 38:253–270, 2006.

[5] A. Juditsky and Y. Nesterov. Primal-dual subgradient methods for minimizing uniformly convex functions. *Arxiv (2010).*

[6] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *Siam Journal on Optimization*, 19:1574–1609, 2009.

[7] M.V.F. Pereira and L.M.V.G Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

Vincent Guigues: FGV/EMAp, 190 Praia de Botafogo, Rio de Janeiro, Brazil, vguigues@fgv.br, UFRJ, Escola Politécnica, Departamento de Engenharia Industrial, Ilha do Fundão, CT, Bloco F, Rio de Janeiro, Brazil