Smooth minimization of two-stage stochastic linear programs

Shabbir Ahmed
School of Industrial & Systems Engineering,
Georgia Institute of Technology, 765 Ferst Drive, Atlanta, GA 30332.

February 3, 2006

Abstract

This note presents an application of the smooth optimization technique of Nesterov [7] for solving two-stage stochastic linear programs. It is shown that the original $O(\frac{1}{\epsilon})$ bound of [7] on the number of main iterations required to obtain an ϵ-optimal solution is retained.

1 Introduction

We consider two-stage stochastic linear problems with fixed recourse of the form

$$\min\Big\{f(x) = c^\top x + \mathbb{E}[Q(x,\boldsymbol{\xi})] : x \in X\Big\}, \tag{1}$$

with

$$Q(x,\xi) = \min\Big\{q^\top y :\ Wy = h + Tx,\ y \geq 0\Big\}. \tag{2}$$

Here $x \in \mathbb{R}^{n_1}$ and $y \in \mathbb{R}^{n_2}$ are the first and second-stage decision vectors, respectively, X is a non-empty, compact, convex set of feasible first-stage decisions, and $\boldsymbol{\xi} = (\boldsymbol{q}, \boldsymbol{h}, \boldsymbol{T})$ is a random vector with a known distribution and support Ξ. Given a first stage decision x and a realization $\xi = (q,h,T) \in \mathbb{R}^{n_2+m_2+m_2\times n_1}$ of $\boldsymbol{\xi}$, the second-stage decision is given by a solution of the linear program (2). We assume that (2) is feasible for all values of the right-hand-side (complete recourse), i.e., the positive hull of W: $\text{pos}W = \mathbb{R}^{m_2}$. Then by invoking strong duality, we can write

$$Q(x,\xi) = \max\Big\{\pi^\top(h + Tx) :\ W^\top \pi \leq q\Big\}. \tag{3}$$

Throughout the rest of this paper we use (3) as the definition of $Q(x,\xi)$. The set of dual-feasible solutions is $\Pi(q) = \{\pi \in \mathbb{R}^{m_2} : W^\top \pi \leq q\}$. We assume that $\Pi(q)$ is non-empty for all $q \in \Xi$. We also assume that $\mathbb{E}\|\boldsymbol{\xi}\|^2 < \infty$. Here $\|\cdot\|$ denotes the standard ℓ_2-norm for vectors, and the spectral norm for matrices. In the above setting, the expectation function in problem (1) is well-defined and finite, and consequently, (1) is guaranteed to have an optimal solution (cf. [8, Ch.2]).

It is well-known that, for any ξ, in general the value function $Q(x,\xi)$ is a non-smooth (albeit convex) function of x. Consequently, for general distributions of ξ, (1) is a non-smooth convex program. A variety of non-smooth minimization techniques have been successfully adapted for solving (1) (cf. [8, Ch.3]). Recently, Nesterov [7] proposed a smooth minimization framework for approximately solving a class of non-smooth problems and showed that the number of major iterations required to obtain ϵ-optimal solution is $O(\frac{1}{\epsilon})$. In this paper, we apply Nesterov's framework to the two-stage stochastic linear program (1). As in [7], the key idea is to construct a suitable smooth approximation of the value function $Q(x,\xi)$, and hence (1), and solve it using an efficient smooth minimization scheme [6]. Smooth approximations of $Q(x,\xi)$ have been investigated in [1, 3, 4]. Such approximations have typically been solved using Newton-type methods [2, 3, 4]. However, to our knowledge, no explicit complexity analysis for obtaining ϵ-optimal solutions of (1) using these methods exist. We show that by adapting the smooth optimization framework of [7] to a smooth approximation of two-stage stochastic linear programs proposed by Chen [3], the original $O(\frac{1}{\epsilon})$ bound of [7] on the number of main iterations required to obtain an ϵ-optimal solution is retained.

2 Summary of Nesterov's framework

Consider the problem

$$\min\{f(x) :\ x \in X\}, \tag{4}$$

where $X \subseteq \mathbb{R}^n$ is non-empty, compact, and convex; and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex and differentiable with Lipschitz continuous gradients, i.e.,

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\| \ \ \forall\, x_1, x_2 \in X.$$

Nesterov [6, 7] proposes the following efficient scheme for (4).

Smooth minimization scheme.

0. Choose $x_0 \in X$. Set $k = 0$.

1. Compute $f(x_k)$ and $\nabla f(x_k)$.

2. Compute $u_k = \text{argmin}\{(\nabla f(x_k))^\top (u - x_k) + \frac{L}{2}\|u - x_k\|^2 :\ u \in X\}$.

3. Compute $v_k = \text{argmin}\{\sum_{i=0}^{k} \frac{i+1}{2}[f(x_i) + (\nabla f(x_i))^\top (v - x_i)] + \frac{L}{2}\|v - x_0\|^2 :\ v \in X\}$.

4. Set $x_{k+1} = \frac{k+1}{k+3}u_k + \frac{2}{k+3}v_k$. Increment $k \leftarrow k+1$ and go to step 1.

Theorem 2.1 [7] *After k iterations of the smooth minimization scheme, we have a solution $u_k \in X$ satisfying*

$$0 \leq f(u_k) - f(x^*) \leq \frac{2L\|x^* - x_0\|^2}{k^2}, \tag{5}$$

where x^* is an optimal solution of (4).

Now, consider the following class of problems

$$\min\Big\{f(x) = \widehat{f}(x) + \max\{\pi^\top Ax - \widehat{\phi}(\pi) :\ \pi \in \Pi\} :\ x \in X\Big\}, \tag{6}$$

where $\widehat{f} : \mathbb{R}^n \mapsto \mathbb{R}$ is convex and differentiable with Lipschitz continuous gradients with constant $M \geq 0$; $\Pi \subset \mathbb{R}^m$ is non-empty, compact, and convex; $\widehat{\phi} : \mathbb{R}^m \mapsto \mathbb{R}$ is convex; and $A \in \mathbb{R}^{m\times n}$. The function f is clearly convex, but in general, non-smooth. Nesterov [7] considers the following smooth approximation of f. Let $\mu > 0$ be a "smoothness" parameter, and define

$$f_\mu(x) = \widehat{f}(x) + \max\{\pi^\top Ax - \widehat{\phi}(\pi) - \frac{\mu}{2}\|\pi\|^2 :\ \pi \in \Pi\} \tag{7}$$

Theorem 2.2 [7]

(i) f_μ is convex and continuously differentiable with gradient

$$\nabla f_\mu(x) = \nabla \widehat{f}(x) + A^\top \pi(x)$$

where $\pi(x) \in \Pi$ is the (unique) optimal solution of $\max\{\pi^\top Ax - \widehat{\phi}(\pi) - \frac{\mu}{2}\|\pi\|^2 :\ \pi \in \Pi\}$.

(ii) $\nabla f_\mu(x)$ is Lipschitz continuous with the constant

$$L_\mu = M + \frac{\|A\|^2}{\mu}.$$

(iii) f_μ *is a uniform approximation of* f*, i.e.,*

$$f_\mu(x) \le f(x) \le f_\mu(x) + \mu D(\Pi) \ \ \forall\, x$$

where $D_2 \ge \max\{\frac{1}{2}\|\pi\|^2 : \ \pi \in \Pi\}$.

Theorem 2.3 [7] *Let* $D_1 \ge \max\{\frac{1}{2}\|x - x_0\|^2 : \ x \in X\}$ *and* x^* *be an optimal solution of (6). Given* $\epsilon > 0$*, choose*

$$N \ge \frac{4\|A\|}{\epsilon}\sqrt{D_1 D_2} + \sqrt{\frac{MD_1}{\epsilon}}, \quad \text{and} \quad \mu = \frac{2\|A\|}{N+1}\sqrt{\frac{D_1}{D_2}}.$$

Then, after N *iterations of the Smooth minimization scheme on* $\min\{f_\mu(x) : \ x \in X\}$*, the solution* $u_N \in X$ *satisfies*

$$0 \le f(u_N) - f(x^*) \le \epsilon.$$

3 Smooth minimization of two-stage stochastic linear programs

Given smoothness parameter $\mu > 0$, consider the following function

$$f_\mu(x) = c^\top x + \mathbb{E}[Q_\mu(x, \boldsymbol{\xi})], \tag{8}$$

where

$$Q_\mu(x,\xi) = \max\left\{\pi^\top(h + Tx) - \frac{\mu}{2}\pi^\top\pi : \ \pi \in \Pi(q)\right\}. \tag{9}$$

For discrete distributions of $\boldsymbol{\xi}$, Chen [2, 3] showed that f_μ is differentiable with a Lipschitz continuous gradient, and, for μ small enough, is a uniform approximation of f. Next, we present these results for general distributions, and provide estimates of the Lipschitz constant and the approximation error of f_μ for any μ.

Theorem 3.1 *Given* $\mu > 0$*, the function* f_μ *is finite valued for all* $x \in X$*; it is convex and continuously differentiable with gradient*

$$\nabla f_\mu(x) = c + \mathbb{E}[\boldsymbol{T}^\top \pi(x, \boldsymbol{\xi})], \tag{10}$$

where $\pi(x,\xi)$ *is the (unique) optimal solution of (9) corresponding to* x *and* ξ*. The gradient* $\nabla f_\mu(x)$ *is Lipschitz continuous with the constant*

$$L_\mu = \frac{\mathbb{E}\|\boldsymbol{T}\|^2}{\mu}.$$

Proof: Since $\Pi(q)$ is non-empty with probability one and $\pi^\top\pi \ge 0$, we have for any x

$$-\infty < Q_\mu(x,\xi) \le Q(x,\xi) \ \ \forall\, \xi \in \Xi.$$

Thus, the integrability of $Q_\mu(x,\xi)$ follows from the integrability of $Q(x,\xi)$ for all $x \in X$. Consequently, f_μ is finite valued for all $x \in X$. From Theorem 2.2 we have that, for all $\xi \in \Xi$, $Q_\mu(\cdot,\xi)$ is convex, and continuously differentiable with gradient

$$\nabla Q_\mu(x,\xi) = T^\top \pi(x,\xi),$$

and $\nabla Q_\mu(\cdot,\xi)$ is Lipschitz continuous with constant $\|T\|^2/\mu$. Since $\mathbb{E}[Q(\cdot,\boldsymbol{\xi})]$ is Lipschitz continuous (cf. [8, Ch.2]), so is $\mathbb{E}[Q_\mu(\cdot,\boldsymbol{\xi})]$, and we have that

$$\nabla \mathbb{E}[Q_\mu(x,\boldsymbol{\xi})] = \mathbb{E}[\nabla Q_\mu(x,\boldsymbol{\xi})].$$

The remaining claims then follow. □

Lemma 1 *Let $g : \mathbb{R}^n \mapsto \mathbb{R}$ be convex and positively homogenous of degree $k > 0$ and let $A_i \subseteq \mathbb{R}^n$ be non-empty for all $i = 1, \ldots, m$. Then*

$$\max\{g(x) : x \in \sum_{i=1}^m A_i\} \le m^{k-1} \sum_{i=1}^m \max\{g(x) : x \in A_i\}.$$

Proof: We have

$$\begin{aligned}
\max\{g(x) : x \in \sum_{i=1}^m A_i\} &= m^k \max\{g(\frac{1}{m}x) : x \in \sum_{i=1}^m A_i\} \\
&= m^k \max\{g(\frac{1}{m}\sum_{i=1}^m y_i) : y_i \in A_i, \quad i = 1, \ldots, m\} \\
&\le m^k \max\{\sum_{i=1}^m \frac{1}{m} g(y_i) : y_i \in A_i, \quad i = 1, \ldots, m\} \\
&\le m^{k-1} \sum_{i=1}^m \max\{g(y_i) : y_i \in A_i\},
\end{aligned}$$

where the first step follows from positive homogeniety and the third step from convexity. □

Theorem 3.2 *There exists constants $\Gamma > 0$ depending on W such that*

$$f_\mu(x) \le f(x) \le f_\mu(x) + \mu\Gamma\mathbb{E}\|\boldsymbol{q}\|, \quad \forall\ x.$$

Proof: Note that for any x and ξ,

$$Q_\mu(x,\xi) \le Q(x,\xi) \le Q_\mu(x,\xi) + \frac{\mu}{2}\max\left\{\pi^\top\pi :\ \pi \in \Pi(q)\right\}.$$

By Hoffman's Lemma [5], there exists a constant $\Gamma > 0$ depending on W, such that $\Pi(q) \subseteq \Pi(0) + \Gamma\|q\|\mathbb{B}$ where $\mathbb{B}$ is the unit ball in $\mathbb{R}^{m2}$. Note that since $\text{pos}W = \mathbb{R}^{m_2}$, $\Pi(0) = \{0\}$. Thus

$$\begin{aligned}
\max\left\{\pi^\top\pi :\ \pi \in \Pi(q)\right\} &\le \max\left\{\pi^\top\pi :\ \pi \in \Pi(0) + \Gamma\|q\|\mathbb{B}\right\} \\
&\le 2\underbrace{\max\{\pi^\top\pi :\ \pi \in \Pi(0)\}}_{=0} + 2\Gamma\|q\|\underbrace{\max\{\pi^\top\pi :\ \pi \in \mathbb{B}\}}_{=1},
\end{aligned}$$

where the second step follows from invoking Lemma 1 with $m = k = 2$. Taking expectations, we have

$$\mathbb{E}[Q_\mu(x,\boldsymbol{\xi})] \le \mathbb{E}[Q(x,\boldsymbol{\xi})] \le \mathbb{E}[Q_\mu(x,\boldsymbol{\xi})] + \mu\Gamma\mathbb{E}\|\boldsymbol{q}\|,$$

and hence the desired result. □

Theorem 3.3 *Given $\epsilon > 0$, choose*

$$N \ge \left\lceil 8\sqrt{\mathbb{E}\|\boldsymbol{T}\|^2\mathbb{E}\|\boldsymbol{q}\|D_1\Gamma} \cdot \frac{1}{\epsilon}\right\rceil \quad \text{and} \quad \mu = \frac{1}{N}\sqrt{\frac{2\mathbb{E}\|\boldsymbol{T}\|^2 D_1}{\Gamma\mathbb{E}\|\boldsymbol{q}\|}} \tag{11}$$

Then after N iterations of the smooth minimization algorithm on (8), we have a solution $u_k \in X$ satisfying $f(u_k) - f(x^*) \le \epsilon$, where x^* is an optimal solution of (1).

Proof: From Theorem 2.1, we have that after N iterations of the smooth minimization algorithm on (8), the solution u_N satisfies

$$f_\mu(u_k) - f_\mu(x_\mu) \le \frac{2L_\mu D_1}{N^2} \le \frac{2\mathbb{E}\|\boldsymbol{T}\|^2 D_1}{\mu N^2},$$

where x_μ is an optimal solution of $\min\{f_\mu(x) : \; x \in X\}$. We have, by optimality of x_μ, $f_\mu(x_\mu) \leq f_\mu(x^*)$; and from Theorem 3.2, $f_\mu(x^*) \leq f(x^*)$ and $f(u_N) \leq f_\mu(u_N) + \mu\Gamma\mathbb{E}\|\boldsymbol{q}\|$. Thus

$$f(u_k) - f(x^*) \leq f_\mu(u_N) - f_\mu(x_\mu) + \mu\Gamma\mathbb{E}\|\boldsymbol{q}\| \leq \frac{2\mathbb{E}\|\boldsymbol{T}\|^2 D_1}{\mu N^2} + \mu\Gamma\mathbb{E}\|\boldsymbol{q}\|.$$

Minimizing with respect to μ, we get the expression for μ in (11). The expression for N is then obtained by solving for

$$\frac{2\mathbb{E}\|\boldsymbol{T}\|^2 D_1}{\mu N^2} + \mu\Gamma\mathbb{E}\|\boldsymbol{q}\| \leq \epsilon.$$

□

An important consequence of the above result is that the number of main iterations is independent of the number of possible realizations of the underlying random parameters (scenarios). Of course the complexity of evaluating of f_μ and its gradient grows linearly with the number of scenarios. As a result the overall complexity of the proposed approach grows only linearly with the number of scenarios.

References

[1] J. R. Birge, S. M. Pollock and L. Qi. A quadratic recourse function for the two-stage stochastic program. In A. Eberhard (ed.), *Progress in optimization*, Dordrecht, Kluwer, 109–121, 2000.

[2] X. Chen, L. Qi and R. Womersley. Newton's method for quadratic stochastic programs with recourse. *Journal of Computational and Applied Mathematics*, 60:29–46, 1995.

[3] X. Chen. A parallel BFGS-SQP method for stochastic linear programs. In R. L. May and A. K. Easton (eds.), *Computational Techniques and Applications*, World Scientific, 67–74, 1995.

[4] X. Chen. Newton-type methods for stochastic programming. *Mathematical and Computer Modelling*, 31:89–98, 2000.

[5] A. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards, Section B. Math. Sci.*, 49:263–265, 1952.

[6] Yu. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(\frac{1}{k^2})$. *Doklady AN SSSR* (translated as Soviet Math. Docl.), 269(3):543–547, 1983.

[7] Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103:127–152, 2005.

[8] A. Ruszczyński and A. Shapiro (eds.). *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 2003.