

# Sequential Sampling-based Method for Two-Stage Distributionally Robust Linear Programs

November 11, 2019

In the remainder of this paper, we make the following assumptions:

- (A1) The first-stage feasible region  $\mathcal{X} := \{x : Ax \geq b, x \geq 0\}$  is a non-empty and compact set.
- (A2) There exists a  $L > -\infty$  such that  $L \leq Q(x, \tilde{\xi}) < \infty$ , almost surely.
- (A3)  $\Xi$  is a compact metric space.
- (A4) The ambiguity set  $\mathfrak{P}$  is weakly compact, and  $P\{|\tilde{\xi} - \xi| < \delta\} > 0$  for any  $\xi \in \Xi$  and  $P \in \mathfrak{P}$ .

## 1 Convergence Analysis of the Regularized Algorithm

In this section we provide the convergence result of our sequential sampling-based method. In order to facilitate the exposition of our theoretical results, we will define certain quantities. Note that these quantities are defined only for notational convenience, and are not necessarily computed during the course of the algorithm in the form represented here.

We begin by writing the approximate recourse function as:

$$Q^k(x, \xi) = \max \{ \pi^\top [r(\xi) - T(\xi)x] \mid \pi \in \Pi^k \}, \quad \forall k \geq 1. \quad (1)$$

These approximations are defined as pointwise maximum of affine functions (cuts), and therefore are piecewise affine convex functions. The following lemma captures the main properties of the sequence of approximations generated. The proof of part (a) of the Lemma is due to Lemma 1 and Theorem 2 in [1] which captures the limiting behavior of the minorants generated in 2-SD algorithm. We provide the complete proof here as certain constructions used are relevant in subsequent discussions.

**Lemma 1.** *The sequence of functions  $\{Q^k(x, \xi)\}_{k \geq 1}$  generated by the algorithm satisfies the following:*

- (a) *For a given  $(x, \xi) \in \mathcal{X} \times \Xi$ , the sequence converges uniformly to a continuous function  $g(x, \xi)$ .*
- (b) *For an infinite subsequence  $\{x^k\}_{\mathcal{K}}$  of candidate solutions generated by the algorithm such that  $\{x^k\}_{\mathcal{K}} \rightarrow \bar{x}$ , the function  $g$  satisfies  $g(\bar{x}, \xi) = Q(\bar{x}, \xi)$  for all  $\xi \in \Xi$ .*
- (c) *If for each  $x \in \mathcal{X}$ ,  $\mathbb{E}_P[Q(x, \tilde{\xi})]$  is continuous with respect to  $P \in \mathfrak{P}$ , then for a given sequence  $\{(x^k, P^k)\}_{k \geq 1}$  that converges to  $(\bar{x}, \bar{P})$ ,*

$$\lim_{k \rightarrow \infty} \mathbb{E}_{P^k}[Q^k(x^k, \tilde{\xi})] = \mathbb{E}_{\bar{P}}[Q(\bar{x}, \tilde{\xi})]$$

*with probability one.*

*Proof.* For part (a), recall that  $\mathcal{X} \times \Xi$  is a compact set ((A1)) and (A3)), and  $\{Q^k\}$  is a sequence of continuous (piecewise linear and convex) functions. The construction of the set of dual vertices satisfies  $\Pi_k \subseteq \Pi_{k+1} \subseteq \Pi$  that ensures that the sequence  $\{Q^k\}$  is monotonically increasing, i.e.,  $Q^k(x, \xi) \leq Q^{k+1}(x, \xi) \leq Q(x, \xi)$  for all  $(x, \xi) \in (\mathcal{X}, \Xi)$ , and bounded by a finite function  $Q$  (due to (A2)). Therefore, this sequence pointwise converges to some function  $g(x, \xi) \leq Q(x, \xi)$ . Since the set of dual vertices  $\Pi$  is finite, the set  $\lim_{k \rightarrow \infty} \Pi_k := \bar{\Pi} (\subseteq \Pi)$  is also a finite. Clearly,

$$g(x, \xi) = \lim_{k \rightarrow \infty} Q^k(x, \xi) = \max \{ \pi^\top [r(\xi) - T(\xi)x] \mid \pi \in \bar{\Pi} \}$$

is the optimal value of a LP, and hence, is a continuous function. The compactness of  $\mathcal{X} \times \Xi$ , and continuity, monotonicity and pointwise convergence of  $\{Q^k\}$  to  $g$  guarantees that the sequence converges uniformly to  $g$  (Theorem 7.13 in [2]).

For part (b), let  $\xi$  be given. Assumptions (A3) and (A4) ensures for any  $\delta > 0$  and  $k \in \mathcal{K}$  that  $|\xi^k - \xi| < \delta$  infinitely often, with probability one. Further, since  $\{x^k\}_{\mathcal{K}} \rightarrow \bar{x}$ ,  $Q$  is a continuous function and  $\{Q^k\}$  is a uniformly convergent sequence of continuous functions there exists a further subsequence  $\{x^{k'}, \xi^{k'}\}_{\mathcal{K}'}$  and  $K < \infty$  such that

$$|Q(\bar{x}, \xi) - Q(\bar{x}, \xi^{k'})| < \epsilon/3, \quad |Q(\bar{x}, \xi^{k'}) - Q(x^{k'}, \xi^{k'})| < \epsilon/3, \quad \text{and} \quad |Q^{k'}(x^{k'}, \xi^{k'}) - Q^{k'}(x^{k'}, \xi)| < \epsilon/3$$

for all  $k' \geq K$ . Since, the recourse function is computed exactly for  $(x^{k'}, \xi^{k'})$  as input, the affine function generated in iteration  $k'$  is a supporting hyperplane to  $Q(x^{k'}, \xi^{k'})$ , i.e.,  $Q^{k'}(x^{k'}, \xi^{k'}) = Q(x^{k'}, \xi^{k'})$ . Therefore, for every  $\epsilon > 0$ , there exists a subsequence  $\{x^{k'}\}_{k' \in \mathcal{K}'}$  and  $K < \infty$  such that

$$\begin{aligned} |Q(\bar{x}, \xi) - Q^{k'}(x^{k'}, \xi)| &< |Q(\bar{x}, \xi) - Q(\bar{x}, \xi^{k'})| + |Q(\bar{x}, \xi^{k'}) - Q^{k'}(x^{k'}, \xi^{k'})| + |Q^{k'}(x^{k'}, \xi^{k'}) - Q^{k'}(x^{k'}, \xi)| \\ &= |Q(\bar{x}, \xi) - Q(\bar{x}, \xi^{k'})| + |Q(\bar{x}, \xi^{k'}) - Q(x^{k'}, \xi^{k'})| + |Q^{k'}(x^{k'}, \xi^{k'}) - Q^{k'}(x^{k'}, \xi)| < \epsilon \end{aligned}$$

for all  $k' \geq K$ . It follows that  $Q^{k'}(\bar{x}, \xi) \rightarrow g(\bar{x}, \xi) = Q(\bar{x}, \xi)$  with probability one. This completes the proof of part (a).

For part (c), note for  $x \in \mathcal{X}$  that  $\{Q^k\}$  is a sequence of measurable functions such that  $Q^k \rightarrow g$  as  $k \rightarrow \infty$ . Under assumption (A2), the sequence satisfies  $0 \leq Q^1(x, \xi) - L \leq Q^2(x, \xi) - L \leq \dots$  for  $\xi \in \Xi$ . Using the Lebesgue's monotone convergence theorem (Theorem 11.28 in [2]),

$$\mathbb{E}_P[\lim_{k \rightarrow \infty} Q^k(x, \tilde{\xi})] = \mathbb{E}_P[g(x, \tilde{\xi})]$$

for any  $P \in \mathfrak{P}$ . Specifically, we have  $\mathbb{E}_P[g(\hat{x}, \tilde{\xi})] = \mathbb{E}_P[Q(\hat{x}, \tilde{\xi})]$  using the result from part (a). Invoking the continuity of  $\mathbb{E}_P[Q(x, \cdot)]$  with respect to  $P \in \mathfrak{P}$ , we have

$$\lim_{P \rightarrow \bar{P}} \mathbb{E}_P[\lim_{k \in \mathcal{K}'} Q^k(\hat{x}, \tilde{\xi})] = \mathbb{E}_{\bar{P}}[Q(\hat{x}, \tilde{\xi})].$$

This completes the proof.  $\square$

Using the approximate recourse function first-stage objective function approximation in (??) can be restated as

$$f^k(x) = \max_{P \in \mathfrak{P}^k} \{ F^k(x, P) := c^\top x + \mathbb{E}_P[Q^k(x, \tilde{\xi})] \}. \quad (2)$$

We further define  $\phi = \min_{x \in \mathcal{X}} F(x)$  and  $\phi^k = \min_{x \in \mathcal{X}} F^k(x)$ .

Recall that, the candidate solution is given by

$$x^k \in \arg \min \{ f^{k-1}(x) + \frac{\rho}{2} \|x - \hat{x}^{k-1}\|^2 \}. \quad (3)$$

The extremal distribution associated with the candidate solution  $x^k$  will be denoted as  $P^k \in \mathfrak{P}^k$ . The incumbent solution is updated if the following inequality is satisfied

$$f^k(x^k) - f^k(\hat{x}^{k-1}) < \gamma[f^{k-1}(x^k) - f^{k-1}(\hat{x}^{k-1})] \quad (4)$$

for  $\gamma > 0$ . Let  $\mathcal{K} := \{k_1, k_2, \dots, k_n, \dots\}$  denote the set of iterations at which the incumbent solution was updated. The optimality of  $x^k$  implies that

$$f^{k-1}(x^k) + \frac{\rho}{2}\|x^k - \hat{x}^{k-1}\|^2 \leq f^{k-1}(\hat{x}^{k-1}).$$

Therefore,  $\theta^k := f^{k-1}(x^k) - f^{k-1}(\hat{x}^{k-1}) \leq 0$ .

**Lemma 2.** *The sequence of candidate solutions generated by the algorithm satisfies*

$$f(x^*) \leq \liminf_{k \rightarrow \infty} f^{k-1}(x^k),$$

with probability one.

*Proof.* Suppose the hypothesis is not true, and for  $\epsilon > 0$  we have

$$f(x^*) - \epsilon > \liminf_{k \rightarrow \infty} f^{k-1}(x^k).$$

Therefore, there exists a subsequence  $\{x^i\}_{i \in \mathcal{I}}$ , such that

$$f(x^*) - \epsilon > f^{i-1}(x^i) \quad \forall i \in \mathcal{I}. \quad (5)$$

Since  $\mathcal{X}$  is a compact set, there is a further subsequence indexed by  $\mathcal{I}' \subset \mathcal{I}$  such that  $\{x^i\}_{i \in \mathcal{I}'} \rightarrow \bar{x} \in \mathcal{X}$  and  $\{P^i\}_{i \in \mathcal{I}'} \rightarrow \bar{P}$ . Over this subsequence we have

$$\begin{aligned} f^{i-1}(x^i) &= c^\top x^i + \mathbb{E}_{P^i}[Q^{i-1}(x^i, \tilde{\xi})] \\ &\geq c^\top x^i + \mathbb{E}_P[Q^{i-1}(x^i, \tilde{\xi})] \quad \forall i \in \mathcal{I}', P \in \mathfrak{P}^i \\ \Rightarrow f(x^*) - \epsilon &> \lim_{i \in \mathcal{I}'} c^\top x^i + \mathbb{E}_P[Q^{i-1}(x^i, \tilde{\xi})]. \end{aligned}$$

The first inequality is due to the extremal nature of probability distribution  $P^i$ . Using lemma ??, we have

$$f(x^*) - \epsilon > c^\top \bar{x} + \mathbb{E}_{\bar{P}}[Q(\bar{x}, \tilde{\xi})]$$

which contradicts the hypothesis that  $x^*$  is an optimal solution.  $\square$

**Lemma 3.** *There exists a subsequence of iterations, denoted as  $\mathcal{K}^*$ , such that  $\lim_{k \in \mathcal{K}^*} \theta^k = 0$ .*

*Proof.* We will consider two cases depending on whether the set  $|\mathcal{K}|$  is finite or not. First, suppose that  $|\mathcal{K}|$  is not finite. By the incumbent update rule

$$f^{k_n}(x^{k_n}) - f^{k_n}(\hat{x}^{k_n-1}) < \gamma[f^{k_n-1}(x^{k_n}) - f^{k_n-1}(\hat{x}^{k_n-1})] = \gamma\theta^{k_n} \leq 0 \quad \forall k_n \in \mathcal{K}.$$

Since  $x^{k_n} = \hat{x}^{k_n}$  and  $\hat{x}^{k_n-1} = \hat{x}^{k_n-1}$ , we have

$$f^{k_n}(\hat{x}^{k_n}) - f^{k_n}(\hat{x}^{k_n-1}) \leq \gamma\theta^{k_n} \leq 0.$$

This implies that  $\limsup_{n \rightarrow \infty} \theta^{k_n} \leq 0$ . The left-hand side of the above inequality captures the improvement in the objective function value at the current incumbent solution over the previous

incumbent solution. Using the above, we can write the average improvement attained over  $n$  incumbent changes as

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^n \left[ f^{k_j}(\hat{x}^{k_j}) - f^{k_j}(\hat{x}^{k_{j-1}}) \right] \leq \frac{1}{n} \sum_{j=1}^n \theta^{k_j} \leq 0 \quad \forall n, \\ \Rightarrow & \frac{1}{n} \left[ \left( f^{k_n}(\hat{x}^{k_n}) - f^{k_1}(\hat{x}^{k_0}) \right) + \sum_{j=1}^{n-1} \left( f^{k_j}(\hat{x}^{k_j}) - f^{k_{j-1}}(\hat{x}^{k_j}) \right) \right] \leq \frac{1}{n} \sum_{j=1}^n \theta^{k_j} \leq 0 \quad \forall n. \end{aligned}$$

Under the assumption that the dual feasible region is non-empty and bounded (this is ensured by relatively complete recourse), the first term in the brackets reduces to zero as  $n \rightarrow \infty$ . The second term converges to zero, with probability one, if we have uniform convergence. Thus,

$$\lim_{n \rightarrow \infty} \frac{1}{m} \sum_{j=1}^n \theta^{k_n} = 0$$

with probability one. Further,

$$\lim_{n \rightarrow \infty} \frac{1}{m} \sum_{j=1}^n \theta^{k_n} \leq \limsup_{n \rightarrow \infty} \theta^{k_n} \leq 0.$$

Thus, there exists a subsequence indexed by the set  $\mathcal{K}^*$  such that  $\lim_{k \in \mathcal{K}^*} \theta^k = 0$ , with probability one.

Now if  $|\mathcal{K}|$  is finite, then there exists  $\hat{x}$  and  $K < \infty$  such that for all  $k \geq K$ , we have  $\hat{x}^k = \hat{x}$ . Further, since the incumbent is not updated in iterations  $k \geq K$  we must have

$$f^k(x^k) - f^k(\hat{x}) \geq \gamma[f^{k-1}(x^k) - f^{k-1}(\hat{x})].$$

If  $\lim_{k \in \mathcal{K}^*} x^k = \bar{x}$ ,  $\lim_{k \in \mathcal{K}^*} f^k(x^k) = f$  and  $\lim_{k \in \mathcal{K}^*} f^k(\hat{x}) = f(\hat{x})$ , then we have

$$\begin{aligned} \lim_{k \in \mathcal{K}^*} f^k(x^k) - f^k(\hat{x}) & \geq \gamma \lim_{k \in \mathcal{K}^*} f^{k-1}(x^k) - f^{k-1}(\hat{x}) \\ & \Rightarrow f(\hat{x}) - f(\bar{x}) \geq \gamma(f(\hat{x}) - f(\bar{x})) \\ & = \lim_{k \in \mathcal{K}^*} \gamma \theta^k. \end{aligned}$$

Since  $\gamma \in (0, 1)$  and  $\theta^k \leq 0$ , we have  $\lim_{k \in \mathcal{K}^*} \theta^k = 0$  with probability one.  $\square$

**Theorem 4.** Let  $\{x^k\}_{k=1}^\infty$  and  $\{\hat{x}^k\}_{k \in \mathcal{K}}$  be the sequence candidate and incumbent solutions generated by the algorithm. There exists a subsequence of iterations, indexed by  $\mathcal{K}^*$ , such that every accumulation point of  $\{\hat{x}^k\}_{k \in \mathcal{K}^*}$  is an optimal solution with probability one.

*Proof.* Note that,

$$\begin{aligned} \theta^k & = f^{k-1}(x^{k_n}) - f^{k-1}(\hat{x}^{k-1}) \leq f^{k-1}(x^*) - f^{k-1}(\hat{x}^{k-1}) \\ & \leq f(x^*) - f^{k-1}(\hat{x}^{k-1}) \quad \forall k \in \mathcal{K}^*. \end{aligned}$$

The first inequality follows from the optimality of  $x^k$ , and the second inequality is due to the fact that  $f^k(x) \leq f(x)$  for all  $k \geq 1$ .

Using the result of Lemma 1, there exists a subsequence indexed by  $\mathcal{K}^*$  such that  $\bar{x}$  is the accumulation point of  $\{\hat{x}^k\}_{k \in \mathcal{K}^*}$ , applying limit over the infinite set  $\mathcal{K}^*$  we have

$$0 = \lim_{k \in \mathcal{K}^*} \theta^k \leq \lim_{k \in \mathcal{K}^*} f(x^*) - f^{k-1}(\bar{x}^{k-1}) = f(x^*) - f(\bar{x}).$$

The above implies that  $f(\bar{x}) \leq f(x^*)$  which leads us to conclude that  $x^*$  is an optimal solution with probability one. The second equality requires the following property of the approximation sequence:  $\lim_{k \in \mathcal{K}^*} f^{k-1}(\hat{x}^{k-1}) = f(\bar{x})$ .  $\square$

If incumbent changes finitely often then there exists a  $K < \infty$  such that  $\hat{x}^k = \hat{x}$  for all  $k \geq K$ . In this case, for  $k > K$

$$f^{k-1}(x^k) \leq f^{k-1}(x^k) + \frac{\sigma}{2} \|x^k - \hat{x}^{k-1}\|^2 \leq f^{k-1}(\hat{x}^{k-1}) \leq f(\hat{x}^{k-1}).$$

Therefore,  $\limsup_{k \rightarrow \infty} f^{k-1}(x^k) \leq f(\hat{x})$ . The last inequality in the above follows from the fact that  $f^k(x) \leq f(x)$  for all  $k \geq 1$ . Incomplete.

## References

- [1] J. L. Hige and S Sen. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16(3):650–669, 1991.
- [2] W. Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.