

# Article An Alternating Iteration Algorithm for a Parameter-Dependent Distributionally Robust Optimization Model

Shuang Lin<sup>1</sup>, Jie Zhang<sup>2,\*</sup> and Nan Shi<sup>2</sup>



- <sup>2</sup> School of Mathematics, Liaoning Normal University, Dalian 116029, China; shi\_nan\_lnnu@163.com
- \* Correspondence: jie\_zhang@lnnu.edu.cn

**Abstract:** Based on a successive convex programming method, an alternating iteration algorithm is proposed for solving a parameter-dependent distributionally robust optimization. Under the Slater-type condition, the convergence analysis of the algorithm is obtained. When the objective function is convex, a modified algorithm is proposed and a less-conservative solution is obtained. Lastly, some numerical tests results are illustrated to show the efficiency of the algorithm.

Keywords: distributionally robust optimization; alternating iteration algorithm; convergence analysis

**MSC:** 90C30

# 1. Introduction

In stochastic programming, the involved random variables usually satisfy certain distribution. However, in the real world, the certain distribution may be unknown or only the part of it is known. Distributionally robust optimization (DRO) method happens to be an effective way to solve such uncertain problems.

The study of the DRO method can be traced back to Scarf's early work [1], which is intended to address potential uncertainties in supply chain and inventory control. In the DRO method, historical data may not be sufficient to estimate future distribution, therefore, a larger distribution set containing the true distribution can adequately address the risk of fuzzy uncertainty sets. The DRO model has been widely used in operations research, finance and management science, see [2–6] for recent development and further research. However, most of the ambiguity set of DRO are independent of decision variable.

Recently, Zhang, Xu and Zhang [7] have proposed a parameter-dependent DRO model, where the probability of the underlying random variables depends on the decision variables and the ambiguity set is defined through parametric moment conditions with generic cone constraints. Under Slater-type conditions, the quantitative stability results are established for the parameter-dependent DRO. By recent developments from the variational theory, Royset and Wets [8] have established convergence results for approximations of a class of DRO problems with decision-dependent ambiguity sets. Their discussion covers a variety of ambiguity sets, including moment-based and stochastic-dominance-based ones. Luo and Mehrotra [9] have obtained formulations for problems that feature distributional ambiguity sets defined by decision-dependent bounds on moments. Until recently, DRO with decision-dependent ambiguity sets are mostly theoretical achievements and the algorithms for solving such DRO are not related.

In this paper, for the parameter-dependent DRO model in [7], we propose an alternating iteration algorithm for solving it and propose a less-conservative solution strategy for its special case.



Citation: Lin, S.; Zhang, J.; Shi, N. An Alternating Iteration Algorithm for a Parameter-Dependent Distributionally Robust Optimization Model. *Mathematics* **2022**, *10*, 1175. https://doi.org/10.3390/ math10071175

Academic Editor: Frank Werner

Received: 9 March 2022 Accepted: 30 March 2022 Published: 4 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). As far as we are concerned, the main contributions of this paper can be summarized as follows. Firstly, we carry out convergence analysis for alternating iteration algorithm. Under the Slater constraint qualification, we show that any cluster point of the sequence generated by the alternating iteration algorithm is an optimal solution of the parameterdependent DRO. Notice that the proof of the convergence of successive convex programming method in [10] cannot cover our convergence analysis, since the uncertain set in

ming method in [10] cannot cover our convergence analysis, since the uncertain set in Equation (1) depends on x, therefore our convergence analysis can be seen an extension of the proposition in [10]. Secondly, when the corresponding objective function is convex, a less-conservative DRO is constructed and a modified algorithm is proposed for it. At last, numerical experiments are carried out to show the efficiency of the algorithm.

The paper is organized as follows. Section 2 demonstrates the structure of the algorithm for the parameter-dependent DRO and establishes the convergence of the algorithm. In Section 3, the modified algorithm is proposed for a special case of DRO and the less-conservative solution is obtained. In Section 4, some numerical test results are illustrated to show the less conservative property of solutions obtained by the modified algorithm.

Throughout the paper, we use the following notations. By convention, we use  $\mathbf{R}^{n \times n}$  and  $S^{n \times n}$  to denote the space of all  $n \times n$  matrices and symmetric matrices respectively. For matrix  $A \in S^{n \times n}$ ,  $A \preceq 0$  means that A is a negative semidefinite symmetric matrix, ||x|| denotes the Euclidean norm of a vector x in  $\mathbf{R}^n$ . For a real-valued function  $\varphi : \mathbf{R}^n \to \mathbf{R}$ ,  $\nabla \varphi(x)$  denotes the gradient of  $\varphi$  at x.

### 2. DRO Model and Its Algorithm

Consider the following distributionally robust optimization (DRO) problem:

$$(P) \qquad \begin{array}{c} \min_{x} \quad \sup_{P \in \mathcal{P}(x)} \mathbb{E}_{P}[f(x,\xi(\omega))] \\ \text{s.t.} \quad x \in X, \end{array} \tag{1}$$

where *X* is a compact set of  $\mathbb{R}^n$ ,  $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$  is a continuously differentiability function,  $\xi : \Omega \to \Xi$  is a vector of random variables defined on probability space  $(\Omega, \mathcal{F}, P)$  with support set  $\Xi \subset \mathbb{R}^k$ , for fixed  $x \in X$ ,  $\mathcal{P}(x)$  is a set of distributions which contains the true probability distribution of random variable  $\xi$ , and  $\mathbb{E}_P[\cdot]$  denotes the expected value with respect to probability measure  $P \in \mathcal{P}(x)$ .

In this paper, we consider the case when  $\mathcal{P}(x)$  is constructed through moment condition

$$\mathcal{P}(x) := \{ P \in \mathscr{P} : \mathbb{E}_P[\Psi(x, \xi(\omega))] \in \mathcal{K} \},$$
(2)

where  $\Psi$  is a random map which consists of vectors and/or matrices with measurable random components, and  $\mathscr{P}$  denotes the set of all probability distributions/measures in the space  $(\Omega, \mathcal{F})$  and  $\mathcal{K}$  is a closed convex cone in a finite dimensional vector and/or matrix spaces. If we consider  $(\Xi, \mathscr{B})$  as a measurable space equipped with Borel sigma algebra  $\mathscr{B}$ , then  $\mathcal{P}(x)$  may be viewed as a set of probability measures defined on  $(\Xi, \mathscr{B})$  induced by the random variate  $\xi$ . To ease notation, we will use  $\xi$  to denote either the random vector  $\xi(\omega)$  or an element of  $\mathbf{R}^k$  depending on the context.

When  $\Xi$  is a finite discrete set, that is,  $\Xi = {\xi_1, \dots, \xi_N}$ , for some *N*, (2) can be written as

$$\mathcal{P}(x) = \left\{ (p_1, \cdots, p_N) : \sum_{j=1}^N p_j \Psi(x, \xi_j) \in \mathcal{K}, p_j \ge 0, \sum_{j=1}^N p_j = 1 \right\}.$$
 (3)

In this section, we consider the DRO model (1) with  $\mathcal{P}(x)$  defined by (3). In this case,

$$\mathbb{E}_P[f(x,\xi(\omega))] = \sum_{j=1}^N p_j f(x,\xi_j) \text{ and } \mathbb{E}_P[\Psi(x,\xi(\omega))] = \sum_{j=1}^N p_j \Psi(x,\xi_j).$$

In [10], a successive convex programming (SCP) method for a max–min problem with fixed compact set is proposed. However, the SCP method in [10] cannot be used to solve (1) directly, since  $\mathcal{P}(x)$  in (1) depends on x.

Based on the SCP algorithm, we propose an alternating iteration algorithm for solving (1). In the algorithm proposed, the optimal solution is obtained by alternative iteration of solutions of inner maximum problems and outer minimum problems in (1). For convenience, let

$$C = \left\{ (p_1, \cdots, p_N) \in \mathbf{R}^N : p_j \ge 0, j = 1, 2, \cdots, N, \sum_{j=1}^N p_j = 1 \right\}.$$

We know from the algorithm that if the algorithm stops in finite steps with  $C_{k+1} = C_k$ or  $v_k \le t_k$ , then  $x_k$  is an optimal solution of (1). In practice, problem (6) can be solved by its dual problem. In the case when an infinite sequence is produced, we use the following theorem to ensure the validity of the algorithm.

We introduce a notation, which is used in the proof of the convergence of the Algorithm in Table 1. Let  $P, Q \in \mathcal{P}$ , the *total variation metric* between P and Q is defined as (see, e.g., page 270 in [11]),

$$d_{TV}(P,Q) := \sup_{h \in \mathscr{H}} \left| \mathbb{E}_P[h(\xi)] - \mathbb{E}_Q[h(\xi)] \right|, \tag{4}$$

where,

$$\mathscr{H} := \left\{ h : \mathbf{R}^k \to \mathbf{R} : h \text{ is } \mathscr{B} \text{ measurable, } \sup_{\xi \in \Xi} |h(\xi)| \le 1 \right\},\tag{5}$$

Using the total variation norm, we can define the distance from a metric  $P \in \mathscr{P}$  to a metric set  $\mathcal{P} \subset \mathscr{P}$ , that is,

$$d_{TV}(Q, \mathcal{P}) := \inf_{P \in \mathcal{P}} d_{TV}(Q, P).$$

We next provide the convergence of the Algorithm in Table 1.

Table 1. The Alternating Iteration Algorithm.

1. Set k = 0 and  $C_0 = {\hat{P}}$  with  $\hat{P} \in C$  satisfying  ${x \in X : \mathbb{E}_{\hat{P}}[\Psi(x,\xi)] \in \mathcal{K}} \neq \emptyset$ .

2. Solve the outer problem

$$\min_{\substack{x,t \\ \text{s.t.}}} \quad t \\ \mathbb{E}_P[f(x,\xi(\omega))] \le t, \forall P \in \mathcal{P}_k(x) \\ x \in X,$$
(6)

and obtain the solution  $(x_k, t_k)$ , where  $\mathcal{P}_k(x) = \{P \in \mathcal{C}_k : \mathbb{E}_P[\Psi(x, \xi)] \in \mathcal{K}\}$ . 3. Solve the inner problem

$$\max_{P} \quad \mathbb{E}_{P}[f(x_{k},\xi)] \\ \text{s.t.} \quad \mathbb{E}_{P}[\Psi(x_{k},\xi)] \in \mathcal{K}, \forall P \in \mathcal{C}$$

$$(7)$$

and obtain the solution  $\hat{P}_k$  and the optimal value  $v_k$ .

4. Let  $\mathcal{C}_{k+1} = \mathcal{C}_k \cup \{\hat{P}_k\}.$ 

5. If  $C_{k+1} = C_k$  or  $v_k \le t_k$ , then a solution of (1) is found and the algorithm stops. Otherwise set k = k + 1 and goto 2.

**Theorem 1.** Let  $\{x_n\}$  be a sequence generated by Algorithm in Table 1 and  $x_0$  be a cluster point. If (a)  $(x, P) \mapsto \mathbb{E}_P[f(x, \xi)]$  and  $(x, P) \mapsto \mathbb{E}_P[\Psi(x, \xi)]$  are both continuous on  $X \times C$ , (b) for  $x \in X$ ,  $f(x, \cdot)$  and  $\Psi(x, \cdot)$  are finite valued and continuous on  $\Xi$ , (b)  $0 \in int \{\mathbb{E}_P[\Psi(x_0, \xi)] - \mathcal{K} : P \in C\}$ , then  $x_0$  is an optimal solution of problem (1).

**Proof.** Since  $C_n$  is an increasing sequence of sets and C is a compact set, we have  $\lim_{n\to\infty} C_n = cl[\bigcup_{n=1}^{\infty} C_n] := C^+$ . Since  $x_0$  is a cluster point of  $\{x_n\}$ , there exists an subsequence of  $\{x_n\}$ 

converging to  $x_0$ . Without loss of generality, for simplicity, we assume that  $x_0$  is the limit point of  $\{x_n\}$ . We know from step 2 in the algorithm that  $x_n$  is an optimal solution of

$$\min_{x} \sup_{P \in \mathcal{P}(x) \cap \mathcal{C}_{n}} \mathbb{E}_{P}[f(x, \xi(\omega))]$$
s.t.  $x \in X$ .
(8)

Let  $\hat{S}_n(x)$  and  $\hat{v}_n(x)$  denote the optimal solution set and optimal value of

$$\sup_{P\in\mathcal{P}_n(x)},\mathbb{E}_P[f(x,\xi)]$$

respectively,  $\hat{S}(x)$  and  $\hat{v}(x)$  denote the optimal solution set and optimal value of

$$\sup_{P\in\mathcal{P}(x)\cap\mathcal{C}^+}\mathbb{E}_P[f(x,\xi)],$$

respectively. Then, we have from (8) that:

$$\hat{v}_n(x_n) \le \hat{v}_n(x) \text{ for any } x \in X.$$
 (9)

We proceed the rest of the proof in three steps.

Step 1. We next show

$$\lim_{n \to \infty} \hat{v}_n(x_n) = \hat{v}(x_0). \tag{10}$$

Let  $\hat{P}_n \in \hat{S}_n(x_n)$ , by compactness of  $C^+$ ,  $\{\hat{P}_n\}$  has cluster points. We assume  $\hat{P}^*$  is a cluster point of  $\{\hat{P}_n\}$ , then there exists a subsequence  $\{n_k\} \subseteq \{n\}$  such that  $\hat{P}_{n_k}$  converges to  $\hat{P}^*$  weakly as  $k \to \infty$  and  $\hat{P}^* \in C^+$ . Under conditions (a) and (b), we have

$$\hat{v}_{n_k}(x_{n_k}) = \mathbb{E}_{\hat{P}_{n_k}}[f(x_{n_k},\xi)] \to \mathbb{E}_{\hat{P}^*}[f(x_0,\xi)] \le \hat{v}(x_0)$$

as  $k \to \infty$ . Hence, we have

$$\limsup_{n \to \infty} \vartheta_n(x_n) \le \vartheta(x_0). \tag{11}$$

Since  $\mathcal{P}_n(x_n) = \mathcal{P}(x_n) \cap \mathcal{C}_n \neq \emptyset$ , we have form condition (a) that  $\mathcal{P}(x_0) \cap \mathcal{C}^+ \neq \emptyset$ , which means that  $\hat{S}(x_0) \neq \emptyset$ . Let  $P^* \in \hat{S}(x_0)$ , we next show that there exists a sequence  $\{\hat{P}_n\}$  with  $\hat{P}_n \in \mathcal{P}_n(x_n)$  such that  $\hat{P}_n$  converges to  $P^*$  weakly as  $n \to \infty$ . Under conditions (b) and (c), we know from [Theorem 2.1] in [7] that there exist positive constants  $\gamma$  and  $\nu \in (0, 1)$  such that:

$$d_{TV}(Q, \mathcal{P}(x_n)) \leq \gamma \|x_n - x_0\|^{\nu}$$

for all  $Q \in \mathcal{P}(x_0)$  and *n* large enough, which means that for  $P^* \in \hat{S}(x_0)$ ,

$$d_{TV}(P^*, \mathcal{P}(x_n) \cap \mathcal{C}_n) \le d_{TV}(P^*, \mathcal{P}(x_n)) + d_{TV}(P^*, \mathcal{C}_n) \le \gamma \|x_n - x_0\|^{\nu} + d_{TV}(P^*, \mathcal{C}_n)$$
(12)

for *n* large enough. Let  $\hat{P}_n = \prod_{\mathcal{P}(x_n) \cap \mathcal{C}_n} (P^*)$ , then by (12), we have  $\hat{P}_n$  converges to  $P^*$  weakly as *n* converges to infinity. Consequently, under condition (b),

$$\hat{v}_n(x_n) \ge \mathbb{E}_{\hat{P}_n}[f(x_n,\xi)] \to \mathbb{E}_{P^*}[f(x_0,\xi)] = \hat{v}(x_0)$$

as  $n \to \infty$  and hence,

$$\liminf_{n \to \infty} \hat{v}(x_n) \ge \hat{v}(x_0). \tag{13}$$

Combining (11) and (13), we have  $\hat{v}(x_n)$  converges to  $\hat{v}(x_0)$  as  $n \to \infty$ .

Step 2. We next show for any fixed  $x \in X$ ,

$$\lim_{n \to \infty} \hat{v}_n(x) = \hat{v}(x). \tag{14}$$

Since  $\lim_{n\to\infty} C_n = C^+$ , we have  $\lim_{n\to\infty} \mathcal{P}(x) \cap C_n = \mathcal{P}(x) \cap C^+$ .

Then under conditions (a) and (b), similarly to the proof of step 1, we have  $\hat{v}_n(x)$  converges to  $\hat{v}(x)$  as  $n \to \infty$ .

Step 3. Combining (9), (10) and (14), we have

$$\hat{v}(x_0) \le \hat{v}(x)$$
 for any  $x \in X$ , (15)

which means that,  $x_0$  is an optimal solution of

$$\min_{x} \sup_{P \in \mathcal{P}(x) \cap \mathcal{C}^{+}} \mathbb{E}_{P}[f(x, \xi(\omega))]$$
  
s.t.  $x \in X$ . (16)

By step 3 in algorithm, we have

$$\sup_{P\in\mathcal{P}(x_n)\cap\mathcal{C}^+} \mathbb{E}_P[f(x_n,\xi)] \leq \sup_{P\in\mathcal{P}(x_n)\cap\mathcal{C}} \mathbb{E}_P[f(x_n,\xi)] = \mathbb{E}_{\hat{P}_{n+1}}[f(x_n,\xi)] \leq \sup_{P\in\mathcal{P}(x_n)\cap\mathcal{C}^+} \mathbb{E}_P[f(x_n,\xi)],$$

which means that

$$\sup_{P \in \mathcal{P}(x_n) \cap \mathcal{C}^+} \mathbb{E}_P[f(x_n, \xi)] = \sup_{P \in \mathcal{P}(x_n) \cap \mathcal{C}} \mathbb{E}_P[f(x_n, \xi)].$$

Then by the proof in step 1, letting  $n \to \infty$ , we have  $\hat{v}(x_0) = \sup_{P \in \mathcal{P}(x_0) \cap \mathcal{C}} \mathbb{E}_P[f(x_0, \xi)]$ . Consequently, by (15), we have

$$\hat{v}(x_0) \le \hat{v}(x) \le \sup_{P \in \mathcal{P}(x) \cap \mathcal{C}} \mathbb{E}_P[f(x,\xi)]$$

for all  $x \in X$ . Therefore,  $x_0$  is an optimal solution of (1).

**Remark 1.** In [10], without any constraint qualifications, the proof of the convergence of SCP method is obtained. However, in our proof, since the uncertain set in (1) depends on x, the Slater condition ensures the proof. We know from the above proof that if the uncertain set in (1) independent on x, the Slater condition can be omitted. Therefore our convergence analysis can be seen an extension of the proposition in [10].

# 3. Less Conservative Model and a Modified Algorithm

In this section, we consider a special case of (1) and provide a less-conservative model. In the case when  $\Xi = {\xi_1, \dots, \xi_N}$  and the ambiguity set is

$$\mathcal{P}(x) := \left\{ P \in \mathscr{P} : \begin{array}{c} \mathbb{E}_P[\xi - \mu_0]^T \Sigma_0^{-1} \mathbb{E}_P[\xi - \mu_0] \leq \gamma_1 \\ \mathbb{E}_P[(\xi - \mu_0)(\xi - \mu_0)^T] - \gamma_2 \Sigma_0 \leq 0 \end{array} \right\},\tag{17}$$

where  $\gamma_1$  and  $\gamma_2$  are nonnegative constants,  $\mu_0 \in \mathbf{R}^k$  and  $\Sigma_0 \in S^{k \times k}$  is positive semidefinite, the model (1) is the following problem:

$$\begin{array}{l} \min_{x \in X} \max_{(p_1, \cdots, p_N) \in \mathbf{R}^N} & \mathbb{E}_P[f(x, \xi)] \\ \text{s.t.} & \sum_{j=1}^N p_j g_1(\xi_j) \preceq 0, \\ & \sum_{j=1}^N p_j g_2(\xi_j) \preceq 0, \\ & p_j \ge 0, j = 1, \cdots, N, \\ & \sum_{j=1}^N p_j = 1. \end{array} \tag{18}$$

where

$$g_1(\xi) = \begin{bmatrix} -\Sigma_0 & \mu_0 - \zeta \\ (\mu_0 - \xi)^T & -\gamma_1 \end{bmatrix}$$

and

$$g_2(\xi) = (\xi - \mu_0)(\xi - \mu_0)^T - \gamma_2 \Sigma_0$$

The model has been investigated in [2]. As shown in [2], the constraints in (18) imply that the mean of  $\xi$  lies in an ellipsoid of size  $\gamma_1$  centered at the estimate  $\mu_0$  and the centered second moment matrix of  $\xi$  lies in a positive semidefinite cone defined with a matrix inequality.

However, in the constraints of (18), not all  $\xi_j$  lies in the ellipsoid of size  $\gamma_1$  centered at the estimate  $\mu_0$ . In practice, we may be only interested in the  $\xi_j$  which lies in the ellipsoid and omit the ones outside the ellipsoid. Consequently, we propose a less-conservative DRO model, that is

$$\begin{array}{ll} \min_{x \in X} \max_{(p_1, \cdots, p_N) \in \mathbf{R}^N} & \mathbb{E}_P[f(x, \xi)] \\ \text{s.t.} & p_j g_1(\xi_j) \preceq 0, \\ & p_j g_2(\xi_j) \preceq 0, \\ & p_j \geq 0, j = 1, \cdots, N, \\ & \sum_{j=1}^N p_j = 1. \end{array} \tag{19}$$

In the above model, if the  $\xi_j$  does not lie in an ellipsoid of size  $\gamma_1$  centered at the estimate  $\mu_0$  or does not satisfy the matrix inequality  $g_2(\xi_j) \leq 0$ , the corresponding constraints are vanished. Moreover, we can choose  $\gamma_1$  and  $\gamma_2$  such the feasible set of the inner problem is not empty, for example, for the first constraint, let  $\gamma_1 = \max\{(\xi_j - \mu_0)^T \Sigma_0^{-1} (\xi_j - \mu_0) : j = 1, 2, \dots, N\}$ . Compare with model (18), the model (19) is less conservative since the feasible set of the inner maximum problem is smaller.

Let Q be a set of probability distributions defined as

$$\mathcal{Q} = \left\{ (p_1, \cdots, p_N) \in \mathbf{R}^N : p_j g_i(\xi_j) \leq 0, \sum_{j=1}^N p_j = 1, p_j \geq 0, j = 1, \cdots, N, i = 1, 2 \right\}.$$
 (20)

Next we give a modified alternative solution algorithm for (19):

The above algorithm is based on the algorithm in Pflug and Wozabal [10] for solving a distributed robust investment problem and a cutting plane algorithm in Kelley [12] for solving convex optimization problems. A similar algorithm has been used in Xu et al. [5] to solve a different DRO model and the proof of the convergence is omitted. In the following, we provide convergence analysis of the modified alternative solution algorithm based on Theorem 1.

**Theorem 2.** Let  $\{x_n\}$  be a sequence generated by Algorithm in Table 2 and  $x_0$  be a limit point. If for each  $\xi \in \Xi$ ,  $f(\cdot, \xi)$  is continuously differentiable and convex on X, then  $x_0$  is an optimal solution of problem (19). Table 2. The Modified Alternating Iteration Algorithm.

1. Let 
$$\mathbf{P}_0 = (p_1^0, \cdots, p_N^0) \in \mathcal{Q}$$
 and  $\mathcal{Q}_0 := {\mathbf{P}_0}$  and  $x_0 \in X$ . Set  $k = 0$ .  
2. Solve the outer minimization problem  

$$\min_{\substack{x,t \\ s.t. \\ x \in X, \\ \sum_{j=1}^N p_j^k [f(x_k, \xi_j) + \nabla_x f(x_k, \xi_j)^\top (x - x_k)] \le t, \text{ for } \mathbf{P}_k = (p_1^k, \cdots, p_N^k) \in \mathcal{Q}_k$$
(21)

and obtain the solution  $(x_{k+1}, t_{k+1})$ .

3. Solve the inner maximization problem

$$\max_{\substack{(p_1,\cdots,p_N)\in\mathbf{R}^N\\ \text{s.t.}}} \sum_{j=1}^N p_j f(x_{k+1},\xi_j) \\ p_j \begin{bmatrix} -\Sigma_0 & \mu_0 - \xi\\ (\mu_0 - \xi)^\top & -\gamma_1\\ (\mu_0 - \xi)^\top & -\gamma_1 \end{bmatrix} \leq 0 \\ p_j [(\xi - \mu_0)(\xi - \mu_0)^\top] \leq \gamma_2 \Sigma_0, \\ p_j \geq 0, j = 1, \cdots, N, \\ \sum_{j=1}^N p_j = 1 \end{cases}$$

$$(22)$$

and obtain the solution  $(\mathbf{P}_{k+1}, v_{k+1})$ . 4. Let  $\mathcal{Q}_{k+1} = \mathcal{Q}_k \bigcup \{\mathbf{P}_{k+1}\}$ . If  $\mathcal{Q}_{k+1} = \mathcal{Q}_k$  or  $v_{k+1} \leq t_{k+1}$ , then stop, else let k := k+1, go to 1.

**Proof.** The proof is similar as the proof of Theorem 1. Since  $Q_n$  is an increasing sequence of sets and Q is a compact set, we have  $\lim_{n\to\infty} Q_n = \operatorname{cl}[\bigcup_{n=1}^{\infty} Q_n] := Q^+$ .

Let  $\hat{S}_n(x)$  and  $\hat{v}_n(x)$  denote the optimal solution set and optimal value of

$$\sup_{(p_1,\dots,p_N)\in\mathcal{Q}_n}\sum_{j=1}^N p_j [f(x_{n-1},\xi_j) + \nabla_x f(x_{n-1},\xi_j)^T (x-x_{n-1})]$$

respectively,  $\hat{S}(x)$  and  $\hat{v}(x)$  denote the optimal solution set and optimal value of

$$\sup_{(p_1,\cdots,p_N)\in\mathcal{Q}^+}\sum_{j=1}^N p_j f(x,\xi_j)$$

respectively. Then we have

$$\hat{v}_n(x_n) \le \hat{v}_n(x)$$
 for any  $x \in X$ . (23)

Let  $(p_1^n, \dots, p_N^n) \in \hat{S}_n(x_n)$ , by compactness of  $\mathcal{Q}^+$ ,  $\{(p_1^n, \dots, p_N^n)\}$  has cluster points. We assume  $(p_1^*, \dots, p_N^*)$  is a cluster point of  $\{(p_1^n, \dots, p_N^n)\}$ , then there exists a subsequence  $\{n_k\} \subseteq \{n\}$  such that  $(p_1^{n_k}, \dots, p_N^{n_k})$  converges to  $(p_1^*, \dots, p_N^*)$  weakly as  $k \to \infty$  and  $(p_1^*, \dots, p_N^*) \in \mathcal{Q}^+$ . Then we have

$$\begin{aligned} \hat{v}_{n_k}(x_{n_k}) &= \sum_{j=1}^N p_j^{n_k} [f(x_{n_k-1},\xi_j) + \nabla_x f(x_{n_k-1},\xi_j)^T (x_{n_k} - x_{n_k-1})] \\ &\leq \sum_{j=1}^N p_j^{n_k} f(x_{n_k},\xi_j) \to \sum_{j=1}^N p_j^* [f(x_0,\xi_j)] \le \hat{v}(x_0) \end{aligned}$$

as  $k \to \infty$ . Hence, we have

$$\limsup_{n \to \infty} \hat{v}_n(x_n) \le \hat{v}(x_0). \tag{24}$$

On the other hand, for  $(p_1^*, \dots, p_N^*) \in \hat{S}(x_0), (p_1^*, \dots, p_N^*) \in Q^+$ , which means that  $\exists (p_1^n, \dots, p_N^n) \in Q^n$  such that

$$(p_1^n,\cdots,p_N^n) \rightarrow (p_1^*,\cdots,p_N^*)$$

as  $n \to \infty$ . Therefore, we have

$$\hat{v}_n(x_n) \ge \sum_{j=1}^N p_j^n [f(x_{n-1},\xi_j) + \nabla_x f(x_{n-1},\xi_j)^T (x_n - x_{n-1})] \to \sum_{j=1}^N p_j^* f(x_0,\xi_j) = \hat{v}(x_0)$$
(25)

as  $n \to \infty$ . Combining (24) and (25), we obtain

$$\lim_{n\to\infty}\hat{v}_n(x_n)=\hat{v}(x_0).$$

The else of proof follows from the proof of Theorem 1.

**Remark 2.** Notice that the Slater condition is not used in the proof, since the uncertain set in (1) is independent on x, the Slater condition can be omitted.

#### 4. Numerical Tests

In this section, we discuss the numerical performance of proposed alternating iteration algorithm for solving (18) and (19). We do so by applying the alternating iteration algorithm to a news vender problem [4] and provide comparative analysis of the numerical results.

Suppose the company has to decide the order quantity  $x_j$  of a product to meet the demand  $\xi_j$  and the news provider trades in  $j = 1, \dots, n$  products. Before knowing the uncertain demand  $\xi_j$ , the news vender orders  $x_j$  units of product j at the wholesale price  $c_j > 0$ . Once the demand  $\xi_j$  is known, it can be quantified min $\{x_j, \xi\}$  at the retail price of  $v_j$ . Any stock that have not been sold  $(x_j - \xi_j)_+$  are cleared by the remedy price  $h_j$ . Any unsatisfied demand  $(\xi_j - x_j)_+$  is lost. The total loss of the news vendors can be described as a function of the order decision  $x := (x_1, \dots, x_n)^\top$ :

$$L(x,\xi) = c^{\top}x - v^{\top}\min(x,\xi) - h^{\top}(x-\xi)_{+} = (c-v)^{\top}x + (v-h)^{\top}(x-\xi)_{+},$$
(26)

where non-negativity and minimum operators are applied to the component method. We study the risk aversion of the news vendor problem on two models:

(H1) 
$$\min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_{P}[U(L(x,\xi))],$$
(27)

and

(H2) 
$$\min_{x \in X} \sup_{P \in \mathcal{Q}} \mathbb{E}_{P}[U(L(x,\xi))],$$
(28)

where  $U(w) := e^{w/10}$  is an exponential distribution function,

$$\mathcal{P} = \left\{ (p_1, \cdots, p_N) \in \mathbf{R}^N : \sum_{j=1}^N p_j g_i(\xi_j) \preceq 0, \sum_{j=1}^N p_j = 1, p_j \ge 0, j = 1, \cdots, N, i = 1, 2 \right\}.$$

and Q is defined as in (20). Notice that for the news vender problem, problems (18) and (19) are just (*H*1) and (*H*2) respectively.

The data are generated as follows: for *i*-th product, wholesale, retail and remedy prices are  $c_j = 0.1(5 + j - 1)$ ,  $v_j = 0.15(5 + j - 1)$  and  $h_j = 0.05(5 + j - 1)$  respectively; the product demands vector  $\xi$  is characterized by a multivariate log-normal distribution with the mean  $\mu = (\mu_1, \dots, \mu_n)$ ,  $\mu_j = 2, j = 1, \dots, n$ . In the execution of the algorithm, we use an ambiguity set Q in (20) with  $\gamma_1 = 0.1$  and  $\gamma_2 = 1.1$ . The mean and convariance matrix  $\mu_0$  and  $\Sigma_0$  are calculated to be generated through a computer. The experiments are carried out through Matlab 2016 installed on a Dell notebook computer with Windows 7 operating system and Intel Core i5 processor. The SDP subproblems in Algorithms are solved by Matlab solver "SDPT3-4.0" [13].

The computation results are shown in the Tables 3 and 4 and Figures 1–5 below sequentially. In Tables 3 and 4, we show the average cpu time (Times(s)), iteration (Iter) and optimal values (Optimal Vlue) of each test problem with different sample sizes.

From the Tables 3 and 4 and Figures 1–5, we can roughly see that problems (*H*1) and (*H*2) can be solved by the alternating iteration algorithm. We know from Figures 1–5 that when using the algorithm to solve (*H*1) and (*H*2), the number of iterations and time of solving (*H*1) are basically more than that of solving (*H*2). Moreover, the optimal values of (*H*2) is smaller than the ones of (*H*1). Since the DRO model is usually used to describe an upper bound of uncertain optimization problems, the smaller the optimal value of the DRO model, the less conservative the DRO model is. Therefore, (*H*2) is a less conservative DRO model. However, according to Figure 3, (*H*1) is more robust than (*H*2) because the curve shown by (*H*1) is more stable.

**Table 3.** The performance of (*H*1).

n	Time (s)	Iter	<b>Optimal Value</b>
2	31.234226	48	0.9669
4	44.774935	66	0.9663
6	52.961924	72	0.9616
8	52.583319	69	0.9578
10	56.231204	74	0.9659
12	65.792619	85	0.9629

Table 4. The performance of (*H*2).

n	Time (s)	Iter	<b>Optimal Value</b>
2	28.848998	48	0.9483
4	32.741311	50	0.7441
6	38.559070	57	0.7603
8	41.873316	57	0.8045
10	61.811460	76	0.8577
12	63.149645	75	0.8433



Figure 1. Comparative analysis of (*H*1) and (*H*2) on Time(s).



Figure 2. Comparative analysis of (H1) and (H2) on Iter.



Figure 3. Comparative analysis of (H1) and (H2) on Optimal Value.



Figure 4. Comparative analysis of Optimal Value from (H1).



Figure 5. Comparative analysis of Optimal Value from (H2).

The numerical results show that, in order to obtain a conservative total loss in the news vender problem, solving DRO model (*H*2) by the alternating iteration algorithm usually performs better than solving DRO model (*H*1). However, in our observations, when we only focus on the robustness, DRO model (*H*1) may be the better choice. We provide the links to the source codes as follows: https://pan.baidu.com/s/1dSmMUynZqi5LzWgn6 aUUoQ?pwd=xn44 (accessed on 25 January 2022).

#### 5. Conclusions

In this paper, we carry out convergence analysis for an alternating iteration algorithm for a distributionally robust optimization problem where the ambiguity set depends on decision variables. Convergence analysis of the alternating iteration algorithm are obtained under the Slater-type condition, which can be seen an extension of the result in [10]. When the objective function is convex, a modified alternating iteration algorithm is proposed for obtaining a less-conservative solution of DRO and the convergence analysis is established. Finally, we discuss the numerical performance of proposed alternating iteration algorithm for obtaining a conservative total loss in the news vender problem. We can undertake similar analysis when the ambiguity set in DRO is constructed in other ways such as Kullback–Leiblor divergence [14], Wasserstein metric [15,16] etc. We leave all these for future research as they are beyond the focus of this paper.

**Author Contributions:** Conceptualization, S.L. and J.Z.; methodology, S.L.; software, N.S.; validation, J.Z., S.L. and N.S.; formal analysis, S.L.; investigation, J.Z.; writing—original draft preparation, S.L.; writing—review and editing, J.Z.; visualization, N.S.; supervision, J.Z.; project administration, J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China under Project Grant Nos. 12171219 and 61877032, the Liaoning Revitalization Talents Program No. XLYC2007113, Scientific Research Fund of Liaoning Provincial Education Department under Project No. LJKZ0961.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

## References

- 1. Scarf, H. A min-max solution of an inventory problem. In *Studies in the Mathematical Theory of Inventory and Production;* Arrow, K.S., Karlin, S., Scarf, H.E., Eds.; Stanford University Press: Stanford, CA, USA, 1958; pp. 201–209.
- 2. Ye, Y.; Delage, E. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **2010**, *58*, 595–612.
- 3. Goh, J.; Sim, M. Distributionally robust optimization and its tractable approximations. Oper. Res. 2010, 58, 902–917. [CrossRef]
- 4. Wiesemann, W.; Kuhn, D.; Sim, M. Distributionally robust convex optimization. Oper. Res. 2014, 62, 1358–1376. [CrossRef]
- 5. Xu, H.; Liu, Y.C.; Sun, H.L. Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Math. Program.* **2017**, *130*, 1–22. [CrossRef]
- Shapiro, A. On duality theory of conic linear problems. In *Semi-Infinite Programming*; Goberna, M.A., Lopez, M.A., Eds.; Springer: Boston, MA, USA, 2001; pp. 135–165.
- Zhang, J.; Xu, H.; Zhang, L. Quantitative stability analysis for distributionally robust optimization with moment constraints. SIAM J. Optim. 2017, 26, 1855–1882. [CrossRef]
- Royset, J.O.; Wets, R.J.-B. Variational theory for optimization under stochastic ambiguity. SIAM J. Optim. 2017, 27, 1118–1149. [CrossRef]
- 9. Luo, F.; Mehrotra, S. Distributionally robust optimization with decision dependent ambiguity sets. *Tech. Rep.* **2018**, *14*, 2565–2594. [CrossRef]
- 10. Pflug, G.C.; Wozabal, D. Ambiguity in portfolio selection. Quantitative 2007, 7, 435–442. [CrossRef]
- 11. Athreya, K.B.; Lahiri, S.N. *Measure Theory and Probability Theory*; Springer: New York, NY, USA, 2006.
- 12. Kelley, J.E. The cutting-plane method for solving convex programs. SIAM J. Appl. Math. 1960, 8, 703–712. [CrossRef]
- Toh, K.C.; Todd, M.J.; Tütüncü, R.H. SDPT3-a Matlab software package for semidefinite programming. *Optim. Methods Softw.* 1999, 11, 545–581. [CrossRef]
- 14. Hu, Z.; Hong, L.J. Kullback-Leibler Divergence Constrained Distributionally Robust Optimization. Available online: http://www.optimization-online.org/DB\_HTML/2012/11/3677.html (accessed on 1 November 2012).
- 15. Esfahani, P.M.; Kuhn, D. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.* **2018**, 171, 115–166. [CrossRef]
- 16. Zhao, C.; Guan, Y. Data-driven risk-averse stochastic optimization with wasserstein metric. Oper. Res. Lett. 2018, 46, 262. [CrossRef]