Operations Research Letters 45 (2017) 625–630

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Primal–dual hybrid gradient method for distributionally robust optimization problems

Yongchao Liu [a], Xiaoming Yuan [b], Shangzhi Zeng [b], Jin Zhang [b,*]

[a] School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, China
[b] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

ARTICLE INFO

Article history:
Received 5 February 2017
Received in revised form 2 October 2017
Accepted 3 October 2017
Available online 5 November 2017

Keywords:
Distributionally robust optimization
Discretization method
Primal–dual hybrid gradient
Moment conditions
Wasserstein metric

ABSTRACT

We focus on the discretization approach to distributionally robust optimization (DRO) problems and propose a numerical scheme originated from the primal–dual hybrid gradient (PDHG) method that recently has been well studied in convex optimization area. Specifically, we consider the cases where the ambiguity set of the discretized DRO model is defined through the moment condition and Wasserstein metric, respectively. Moreover, we apply the PDHG to a portfolio selection problem modelled by DRO and verify its efficiency.

1. Introduction

Distributionally robust optimization (DRO) can accommodate a vast amount of noisy and incomplete data while it truthfully captures the decision maker's attitude towards both risk and ambiguity. The study of DRO traces back to the earlier work by Scarf [20] which is motivated to address incomplete information on the underlying uncertainty in supply chain and inventory control problems. Over the past few years, it has gained substantial popularity through further contributions by, e.g., Bertsimas and Popescu [2], Delage and Ye [4], Mehrotra and Papp [13], Wiesemann et al. [22,23] to just mention a few.

Different from robust optimization problems, the functional variables in DRO problems induce more challenges on designing implementable and efficient numerical schemes. In the past decade, authors have proposed various techniques to tackle different DRO problems, such as the one-stage problems, multistage problems and chance-constrained problems, see, e.g., [4,5,11,23–25]. Most of the existing works are focused on the dual approach whose framework can be summarized as the following stages: consider the Lagrange dual of the inner max problem, then reformulate the min-max problem as a min-min (combining the min-min by min) problem with semi-infinite constraints, and finally recast the semi-infinite constraints as a linear semidefinite constraint by S-Lemma or dual method again. Wiesemann et al. [24] provide a unified framework of the SDP reformulation for DRO problems where the ambiguity set is constructed through some probabilistic and moment constraints.

Another important approach pioneered by Pflug and Wozabal [16] is to discretize the ambiguity set of DRO problems and then solve the discretized min-max optimization problem directly as a saddle-point problem in the deterministic optimization context. More recently, Xu et al. [26] propose two schemes to discretize DRO problem with moment ambiguity sets, one of which is for the dualized DRO problems and the other is directly through its ambiguity set.

In this paper, we follow the discretization approach studied in [12,16,26] to solve the DRO problem directly

$$\min_{x\in X}\max_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)], \tag{1.1}$$

where X is a compact convex set of $\mathbb{R}^n$, $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is a continuous function and for each fixed $\xi \in \Xi$, $f(\cdot,\xi)$ is convex in x, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a vector of random variables defined on measurable space $(\Omega, \mathcal{F})$ equipped with sigma algebra $\mathcal{F}$, $\mathcal{P} \subseteq \mathscr{P}(\Xi)$ is a convex set of probability distributions and $\mathscr{P}(\Xi)$ denotes the set of all probability measures on compact set Ξ.

It is known that solving a DRO problem amounts to finding a saddle point of a min-max problem and the main challenge lies in the fact that the inner maximization problem has functional variables. On the other hand, it is noticed that if the constraint set for the distribution is discrete, then the DRO problem (1.1) can be recast as a minimax problem in a finite Euclidean space which can be solved by well-studied numerical schemes in the context of saddle-point problems. Following this thought, we suggest applying the

* Corresponding author.
E-mail addresses: lyc@dlut.edu.cn (Y. Liu), xmyuan@hkbu.edu.hk (X. Yuan), 15484203@life.hkbu.edu.hk (S. Zeng), zhangjin@hkbu.edu.hk (J. Zhang).

https://doi.org/10.1016/j.orl.2017.10.001

discretization technique to approximate the DRO problem (1.1) in a finite Euclidean Space, and then consider the lifting technique to further reformulate the discretized DRO problem as a saddle-point problem with certain separable structure. Then, we implement the primal–dual hybrid gradient (PDGH), which traces back to [1] and has gained popularity particularly in the image processing area recently since the work [29] and then [3,9,10,17,27], to the reformulated saddle-point problem.

Throughout this paper, we use the following notation. Let $d(x, A) := \inf_{x' \in A} \|x - x'\|$ the distance from a point x to the set A. For two sets $\mathcal{C}$ and $\mathcal{A}$, $\mathbb{D}(\mathcal{C}, \mathcal{A}) := \sup_{x \in \mathcal{C}} d(x, \mathcal{A})$, denotes the deviation of $\mathcal{C}$ from $\mathcal{A}$ and $\mathbb{H}(\mathcal{C}, \mathcal{A}) := \max\{\mathbb{D}(\mathcal{C}, \mathcal{A}), \mathbb{D}(\mathcal{A}, \mathcal{C})\}$ denotes the Hausdorff distance between $\mathcal{A}$ and $\mathcal{C}$. Finally, for a sequence of subsets $\{\mathcal{C}_k\}$, we follow the notation [19] by using $\limsup_{k \to +\infty} \mathcal{C}_k$ to denote its outer limit, that is, $\limsup_{k \to +\infty} \mathcal{C}_k = \{x : \liminf_{k \to +\infty} d(x, \mathcal{C}_k) = 0\}$.

All the proofs are relegated to the appendix in the electronic companion (available at http://www.optimization-online.org/DB_HTML/2017/10/6238.html).

2. Description of the algorithm

In this section, we describe the discretization approach to the DRO problem (1.1) and then specify the implementation of the PDHG to the saddle-point problem reformulated by the discretized DRO problem. Results in this section will be frequently used throughout this note.

2.1. Discretization approach to DRO problems

The discretization approach means the DRO problem (1.1) is approximated by a min-max point problem in a finite Euclidean space, with the ambiguity set $\mathcal{P}$ replaced by a set of discrete distributions. This kind of research is in line with the standard approach in stochastic programming [14]. To streamline the idea of the discretization approach, let Ξ^N be a discrete subset of Ξ and $\mathscr{P}(\Xi^N)$ denote the set of all probability distributions with support set contained in Ξ^N. By restricting the ambiguity set $\mathcal{P}$ on $\mathscr{P}(\Xi^N)$, we have an approximation problem of (1.1):

$$\min_{x \in X} \max_{P \in \mathcal{P}_N} \mathbb{E}_P[f(x, \xi)], \tag{2.1}$$

where $\mathcal{P}_N := \mathcal{P} \cap \mathscr{P}(\Xi^N)$. Compared to problem (1.1), the problem (2.1) is a standard min-max problem in a finite dimensional space and hence usually is easier to be tackled.

We first study some conditions under which it becomes reasonable to approximate the true problem (1.1) via the discretized problem (2.1).

Theorem 2.1. *Let (x_N, P_N) be a solution point of the discretized DRO problem (2.1). Suppose that $\{\mathcal{P}_N\}$ converges to $\mathcal{P}$ weakly. Then any accumulation point of the sequence $\{(x_N, P_N)\}$ is a solution point of the true DRO problem (1.1).*

2.2. PDHG for saddle-point problems

Saddle-point problems arise in a wide range of areas; and they are mathematical models of some very important applications in scientific computing, economics, game theory, and so on. The literature is too voluminous to list and we just mention very few works that are the most relevant to the application to the specific saddle-point problem (2.1). For our purpose, it suffices to discuss the specific saddle-point problem:

$$\min_{s \in S} \max_{w \in W} \langle s, w \rangle. \tag{2.2}$$

Here we focus on the case that $S \subset \mathbb{R}^m$ and $W \subset \mathbb{R}^n$ are compact convex sets, which ensure problem (2.2) has a saddle-point [18, Corollary 37.6.2]. We shall specify the sets S and W later for different ways of forming the ambiguity set of distributions for the discretized DRO problem (2.1).

For the development on numerical schemes for various saddle-point problems, there is a vast set of literature. Among them are primal–dual type methods which originate from the so-called Uzawa method in [1] and have been well studied in various contexts since the work [29]. For simplicity, we just mention the PDHG method proposed in [3] which was further explained in [10] as an application of the proximal point algorithm. Other variants of the PDHG method in, e.g., [8,10,17], are also applicable, but we do not discuss them in this short paper. More precisely, if the PDHG method in [3] is applied to the saddle-point problem (2.2), the iteration scheme reads as the following.

Algorithm 2.1 *PDHG method for problem. (2.2)*

Require: $s_0 \in \mathbb{R}^m$, $w_0 \in \mathbb{R}^n$, $\epsilon > 0$, $\tau > 0$, $\sigma > 0$ *and* $\sigma\tau < 1$ **for** $k = 0, 1, 2, \ldots$ **do**

$$\begin{cases} \hat{s}_{k+1} = s_k - \tau\, w_k \\ s_{k+1} = \arg\min_{s \in S} \{\frac{1}{2\tau}\|s - \hat{s}_{k+1}\|^2\} \\ \bar{s}_{k+1} = 2s_{k+1} - s_k \\ \hat{w}_{k+1} = w_k + \sigma \bar{s}_{k+1} \\ w_{k+1} = \arg\min_{w \in W} \{\frac{1}{2\sigma}\|w - \hat{w}_{k+1}\|^2\} \end{cases}$$

if $\max(\|s_{k+1} - s_k\|, \|w_{k+1} - w_k\|) \le \epsilon$ **then**
Break.
end
end

Certainly, Algorithm 2.1 generates an iterative sequence and thus by implementing Algorithm 2.1 we can only numerically obtain an approximate solution to the saddle-point problem (2.2) subject to certain tolerance. This means only an approximate solution to the discretized DRO problem (2.1) can be obtained numerically via certain numerical schemes. To investigate the approximation of the discretized DRO problem (2.1) to the true DRO problem (1.1) via implementing Algorithm 2.1 , we introduce the concept of an ϵ_N-solution point as follows: (x_N, P_N) is said to be an ϵ_N-solution point of (2.1) if it satisfies

$$\begin{aligned} &\max_{P \in \mathcal{P}_N} \langle P, f(x_N, \xi)\rangle - \epsilon_N \le \langle P_N, f(x_N, \xi)\rangle \\ &\le \min_{x \in X} \langle P_N, f(x, \xi)\rangle + \epsilon_N, \end{aligned} \tag{2.3}$$

where $\epsilon_N > 0$ denotes the tolerance which can be well controlled by a numerical scheme with provable convergence such as Algorithm 2.1 . Indeed, Theorem 2.1 is sill true if $\epsilon_N \downarrow 0$.

Theorem 2.2. *Let $\epsilon_N \downarrow 0$ and (x_N, P_N) be an ϵ_N-solution point of the discretized DRO problem (2.1). Suppose that $\{\mathcal{P}_N\}$ converges to $\mathcal{P}$ weakly. Then any accumulation point of the sequence $\{(x_N, P_N)\}$ is a solution point of the true DRO problem (1.1).*

Note that the convergence of Algorithm 2.1 (see, e.g., [3,10]) ensures that an approximate solution to the discretized DRO problem (2.1) with an accuracy of ϵ_N satisfying $\epsilon_N \downarrow 0$ can be obtained. Hence, based on Theorem 2.2, the approximation of the discretized DRO problem (2.1) via numerically implementing Algorithm 2.1 to the true DRO problem (1.1) is justified, in terms of the optimality of both the solution points and objective function values.

3. DRO with moment ambiguity set

There are different approaches to forming the ambiguity set of distributions for the DRO problem (1.1); the moment condition is one of the most popular ways. In this section, we specify the ambiguity set in the discretized DRO problem (2.1) with the moment condition and propose a reformulation of the discretized problem that turns out to fit the saddle-point problem (2.2). The rationale of using the reformulated discretized model is also justified. More specifically, we consider the problem (1.1) where $\mathcal{P}$ is constructed as follows:

$$\mathcal{P} := \left\{P \in \mathscr{P}(\Xi)\, \mathbb{E}_P[\psi_i(\xi)] \le 0, \text{ for } i = 1, \dots, k\right\}, \tag{3.1}$$

where $\psi_i : \Xi \to \mathbb{R}^{n_i}$, $i = 1, \dots, k$, is a vector with measurable random components, and $\mathscr{P}(\Xi)$ denotes the set of all probability distributions over space $(\Xi, \mathcal{F})$. Then we may rewrite problem (1.1) as:

$$\begin{aligned} \min_{x\in X} \max_{P\in\mathscr{P}(\Xi)} \quad & \langle P, f(x,\xi)\rangle \\ \text{s.t.} \quad & \langle P, \psi_i(\xi)\rangle \le 0, i = 1, \dots, k. \end{aligned} \tag{3.2}$$

Let us denote $\Xi^N := \{\hat{\xi}_1, \dots, \hat{\xi}_N\}$ and restrict the ambiguity set $\mathcal{P}$ in (3.1) to $\mathscr{P}(\Xi^N)$, that is $\mathcal{P}_N := \mathcal{P} \cap \mathscr{P}(\Xi^N)$. Consequently, the discrete approximation problem is:

$$\begin{aligned} \min_{x\in X} \max_{p\ge 0} \quad & \langle p, F(x)\rangle \\ \text{s.t.} \quad & \langle p, \mathbf{1}\rangle = 1, \\ & \langle p, \Psi_i\rangle \le 0, \quad i = 1, \dots, k, \end{aligned} \tag{3.3}$$

where $\mathbf{1}$ denotes the vector with each component being 1, and

$$F(x) = (f(x, \hat{\xi}_1), \dots, f(x, \hat{\xi}_N))^T, \tag{3.4}$$

$$\Psi_i = (\psi_i(\hat{\xi}_1), \dots, \psi_i(\hat{\xi}_N))^T. \tag{3.5}$$

If F is a linear function, then the objective function in problem (3.3) is bilinear and problem (3.3) fits the saddle-point problem (2.2) and Algorithm 2.1 is applicable. For other cases, we use the lifting technique to recast the objective function in problem (3.3): by introducing an auxiliary variable $t := (t_1, \dots, t_N)^T$, the problem (3.3) can be rewritten as

$$\begin{aligned} \min_{x\in X, t} \max_{P\in\mathbb{R}^N_+} \quad & \langle p, t\rangle \\ \text{s.t.} \quad & \langle P, \mathbf{1}\rangle = 1, \\ & \langle P, \Psi_i\rangle \le 0, \ i = 1, \dots, k, \\ & f(x, \hat{\xi}_i) \le t_i, \ i = 1, \dots, N. \end{aligned} \tag{3.6}$$

As for any fixed $\xi \in \Xi$, $f(\cdot, \xi)$ is a convex function, problem (3.6) turns out to be a special case of the saddle-point problem (2.2) with

$$S := \left\{(x, t) : \begin{cases} x \in X, & |t_i| \le t_{\max} \\ f(x, \hat{\xi}_i) \le t_i, i = 1, \dots, N, \end{cases}\right\};$$

$$W := \left\{P \in \mathbb{R}^N_+ : \begin{cases} \langle P, \mathbf{1}\rangle = 1; \\ \langle P, \Psi_i\rangle \le 0, i = 1, \dots, k \end{cases}\right\},$$

thus Algorithm 2.1 is applicable, where $t_{\max} := \max_{x\in X, \xi\in\Xi} |f(x, \xi)|$.

The next theorem justifies that it is reasonable to solve the reformulated problem (3.6) to pursue a solution point of the problem (3.3).

Theorem 3.1. *Let (x^*, t^*, P^*) be a solution to problem (3.6). Then (x^*, P^*) is a solution point to problem (3.3). Conversely, let (x^*, P^*) be a solution to problem (3.3). Then (x^*, t^*, P^*) with $t^* := F(x^*)$ is a solution to problem (3.6).*

As presented in Theorem 2.1, if the approximate ambiguity set $\{\mathcal{P}_N\}$ converges to $\mathcal{P}$ weakly, a solution point to the approximation problem (3.3) converges to a solution point to the true DRO problem (1.1). The next proposition provides a sufficient condition to ensure the convergence of the ambiguity set $\{\mathcal{P}_N\}$ to $\mathcal{P}$ as N tends to infinity.

Proposition 3.1 *([26, Corollary 4.1]). Assume: (a) $\mathbb{H}(\Xi^N, \Xi) \to 0$ as N tends to infinity, (b) the Slater condition holds, that is, there exists $P_0 \in \mathscr{P}(\Xi)$ such that $\langle P_0, \psi_i(\xi)\rangle < 0$, $i = 1, \dots, k$. Then $\{\mathcal{P}_N\}$ converges to $\mathcal{P}$ weakly.*

Together with Theorems 2.1 and 3.1, Proposition 3.1 ensures that the sequence of optimal solutions to the problem (3.6) converges to an optimal solution to the DRO problem (3.2) as N tends to infinity. Indeed, we may present the quantitative convergence of optimal values and optimal solutions by employing the stability results in [12, Theorem 13].

Theorem 3.2. *Let ϑ and ϑ_N denote the optimal values of problems (3.6) and (3.2), (x_N, t_N, P_N) and (x^*, P^*) be the corresponding optimal solutions. Assume that (a) $\mathcal{P}$ satisfies the Slater condition; (b) for each fixed x, there exists a positive constant L independent of x such that $|f(x, \xi') - f(x, \xi'')| \le L\|\xi' - \xi''\|$. Then, there exists a positive constant C_1 such that $|\vartheta - \vartheta_N| \le C_1\mathbb{H}(\Xi^N, \Xi)$. If additionally $\mathbb{E}_{P^*}[f(\cdot, \xi)]$ satisfies the growth condition at point x^*, that is, there exists a positive constant r such that*

$$\mathbb{E}_{P^*}[f(x, \xi)] - \mathbb{E}_{P^*}[f(x^*, \xi)] \ge r\|x - x^*\|, \forall x \in X,$$

then there exists a positive constant C_2 such that $\|x^* - x_N\| \le C_2\mathbb{H}(\Xi^N, \Xi)$.

4. DRO with distance ambiguity set

Another popular way to characterize the ambiguity of a DRO problem is by a set of distributions that are sufficiently close to a given nominal distribution according to some distance defined on probability space. Particularly, we consider the Wasserstein metric, which is defined through a distance function between two probability distributions in a given compact supporting space. More specifically, given two probability distributions P and Q with support sets Ξ and $\hat{\Xi}$ respectively, the Wasserstein metric is defined as

$$\mathsf{dl}_{W(P,Q)} := \inf_{\pi}\{d(\xi, \hat{\xi}), \pi(\xi) = P, \pi(\hat{\xi}) = Q\},$$

where the infimum is taken over all joint distributions π with marginal P and Q.

We refer to, e.g., [5,6,28], for some DRO problems whose ambiguity sets are defined through the Wasserstein metric. Particularly, [5,6,28] consider the DRO problem (1.1) with ambiguity set:

$$\mathcal{P}_W := \{Q \in \mathscr{P}(\Xi) : \mathsf{dl}_{W(P,P_0)} \le c\},$$

where P_0 is a nominal probability distribution and c is a small positive number representing the robustness of the ambiguity set. Of course, with the growth of c, $\mathcal{P}_W$ becomes bigger and has a higher probability to contain the true distribution. When the nominal distribution is in form of the empirical distribution determined by direct observations of data, we may choose the parameter c based on some statistical methods:

$$P(\mathsf{dl}_W(P, P_N) \le c) \ge 1 - \exp(-\frac{c^2}{2B^2}N), \tag{4.1}$$

where N is the number of historical data and B is the diameter of Ξ. See [28, Proposition 1] for details.

Similarly, we propose a discrete approximation of the ambiguity set $\mathcal{P}_W$ as follows:

$$\mathcal{P}_W^N := \{Q \in \mathscr{P}(\Xi^N) : \mathsf{dl}_W(P, P_0) \le c\}.$$

Suppose that the nominal probability is the empirical probability based on independent and identically distributed sample

$\xi_1, \ldots, \xi_M$. Then, with the ambiguity set discretized by the mentioned Wasserstein metric, the true DRO problem (1.1) can be recast as:

$$\begin{aligned} \min_{x\in X} \max_{q\geq 0,\pi\geq 0} & \sum_{i=1}^{N} q_i f(x,\hat{\xi}_i) \\ \text{s.t.} & \sum_{i=1}^{N} \pi_{i,j} = p_j, j=1,\ldots,M \\ & \sum_{j=1}^{M} \pi_{i,j} = q_i, i=1,\ldots,N \\ & \sum_{i=1}^{N}\sum_{j=1}^{M} \pi_{i,j} d(\hat{\xi}_i, \xi_j) \leq c, \end{aligned} \tag{4.2}$$

where $\pi := (\pi_{1,1}, \ldots, \pi_{N,M})$ is the joint distribution in the space $(\Xi^N, \mathscr{B}) \times (\Xi^M, \mathscr{B})$ and $\Xi^M := \{\xi_1, \ldots, \xi_M\}$. Similar to problem (3.6), we introduce an auxiliary variable $t := (t_1, \ldots, t_N)^T$ and then reformulate (4.2) as

$$\begin{aligned} \min_{x\in X,t} \max_{q\geq 0,\pi\geq 0} & \sum_{i=1}^{N} q_i t_i \\ \text{s.t.} & \sum_{i=1}^{N} \pi_{i,j} = p_j, j=1,\ldots,M \\ & \sum_{j=1}^{M} \pi_{i,j} = q_i, i=1,\ldots,N \\ & \sum_{i=1}^{N}\sum_{j=1}^{M} \pi_{i,j} d(\hat{\xi}_i, \xi_j) \leq c \\ & f(x,\hat{\xi}_i) \leq t_i, i=1,\ldots,N, \end{aligned} \tag{4.3}$$

which fits the saddle-point problem (2.2) with

$$S := \left\{ (x,t) : \begin{cases} x\in X, & |t_i| \leq t_{\max} \\ f(x,\hat{\xi}_i) \leq t_i, & i=1,\ldots,N, \end{cases} \right\};$$

$$W := \left\{ (q,\pi) : \begin{cases} \sum_{i=1}^{N} \pi_{i,j} = p_j, j=1,\ldots,M \\ \sum_{j=1}^{M} \pi_{i,j} = q_i, i=1,\ldots,N \\ \sum_{i=1}^{N}\sum_{j=1}^{M} \pi_{i,j} d(\hat{\xi}_i, \xi_j) \leq c \\ (q,\pi) \in \mathbb{R}_+^N \times \mathbb{R}_+^{N\times M} \end{cases} \right\};$$

and thus Algorithm 2.1 can be applied directly.

The following theorem presents the convergence of the discretized DRO problem (4.3) to the true DRO problem (1.1) as N tends to infinity.

Theorem 4.1. *Let ϑ and ϑ_N denote the optimal values of problems (4.3) and (1.1), $(x_N, t_N; q_N, \pi_N)$ and (x^*, P^*) be the corresponding optimal solutions. Assume that for each fixed x, there exists a positive constant L independent of x such that $|f(x,\xi') - f(x,\xi'')| \leq L\|\xi' - \xi''\|$. Then, there exists a positive constant C_1 such that $|\vartheta - \vartheta_N| \leq C_1\mathbb{H}(\Xi^N, \Xi)$. If additionally $\mathbb{E}_{P^*}[f(\cdot,\xi)]$ satisfies the growth condition at point x^*, that is, there exists a positive constant r such that*

$$\mathbb{E}_{P^*}[f(x,\xi)] - \mathbb{E}_{P^*}[f(x^*,\xi)] \geq r\|x - x^*\|, \forall x \in X,$$

then there exists a positive constant C_2 such that $\|x^* - x_N\| \leq C_2\mathbb{H}(\Xi^N, \Xi)$.

5. Numerical results

In this section, we consider the DRO formulation of a portfolio optimization problem and implement Algorithm 2.1 to the discretized reformulation of the DRO model studied in the previous sections. Some preliminary numerical results are reported to show the efficiency of Algorithm 2.1 for solving the resulting saddle-point reformulations of the discretized DRO problems.

We consider the portfolio optimization problem, in which one is interested in maximizing the expected utility obtained from the single-step return of his investment portfolio. We consider the case where there is no trading fee, that is, given that k investment options are available, the expected utility is defined as:

$$f(x,\xi) := r_1x_1 + \cdots + r_kx_k, \tag{5.4}$$

where r_i is the random return of asset i. In the robust optimization approach to this problem, one defines a distributional set based on the sample to contain the true distribution. We consider the cases where the ambiguity set is defined through moment conditions and Wasserstein metric respectively. Here the moment-condition type of ambiguity set is defined as:

$$\mathcal{P} := \left\{ P \in \mathscr{P} : \begin{array}{l} |\mathbb{E}[\xi] - \mu_0| \leq c_1 \\ \|\mathbb{E}_P[(\xi-\mu_0)(\xi-\mu_0)^T] - \Sigma_0\|_F \leq c_2 \end{array} \right\},$$

where μ_0 and Σ_0 are sample mean and sample covariance, c_1 and c_2 are nonnegative constants. Based on [21, Theorem 3 and Corollary 6] and Bonferroni's inequality, if we choose $c_1 = \frac{\rho}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{1}{\delta}}\right)$ and $c_2 = \frac{\rho}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{1}{\delta}}\right)$, then $\mathcal{P}$ includes the true distribution with a probability of $1 - 2\delta$. For the Wasserstein metric type ambiguity set, we choose the parameter c by statistics (4.1).

We collect the following four stocks: Aberdeen Asset Management PLC, Admiral Group PLC, AMEC PLC, Anglo American PLC, PL (http://finance.google.com) (from 19th Dec 2012 to 15th Nov 2013) with a total of 230 datasets. Similar to the work [4], to ensure that the sample is independent and it follows the same distribution, we use 30 days from the most recent history to assign the portfolio. We have carried out out-of-sample tests with a rolling window of 30 days: use the first 30 data to construct the ambiguity set $\mathcal{P}$ and calculate the optimal portfolio strategy for the 31-th day and then move on a rolling basis.

For numerical experiments, we choose the robust parameters such that the true probability is contained in the ambiguity set with a probability of 99% and compare our model with the stochastic programming model, that is, taking the empirical distribution as the true distribution. We implement Algorithm 2.1 on MATLAB 2014 installed in a PC with Windows 7 operating system. We use CVX (version 1.22) developed by Grant and Boyd [7] to solve the optimization problem in Steps 2 and 4 of Algorithm 2.1. Since condition $\sigma\tau < 1$ guarantees the convergence of PDHG method (Algorithm 2.1), we set the parameters σ and τ as 9.9 and 0.1 respectively.

Table 1 summarizes the daily returns generated by the portfolio models, where "L", "H" and "A" denote respectively the lowest, highest and average return. We record the number of days when the overall portfolio return rate falls below 1 and exceeds (or equals to) 1, denoted by "Down" and "Up". We can see that, compared to 90 times in SP model, there are 102 times when the return rate exceeds 1 in the DRO models. The DRO models and average strategy $(1/n)$ achieve comparable average daily return and display stable performances within a narrow range between the best and worst return curves. Fig. 1 depicts the evolution of wealth over

Table 1
Daily return.

Model	L	H	A	Down	Up
$1/n$	0.9735	1.0246	0.9997	103	97
Wasserstein	0.9733	1.0262	0.9999	98	102
Moment	0.9732	1.0262	0.9996	98	102
SP	0.9414	1.0473	0.9970	110	90

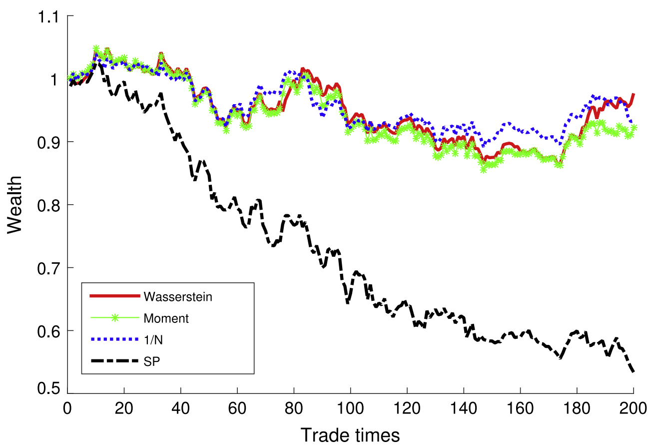

Fig. 1. Wealth evolution with the trading times.

200 trading days when managing a portfolio of four assets on a daily basis with different models. The figure indicates that all wealth lines have the tendency to going down and the wealth curves of DRO models and $1/n$ investment strategy outperform SP model. Compared to the moment type ambiguity set, distance type ambiguity set (Wasserstein metric) displays higher average daily return, wealth at the end of horizon and generate more stable daily return over the time horizon. Our experiments also verify the theory in [15] that the average investment strategy is an optimal strategy when there is only few historical data.

In the previous test, the objective function is linear, which means the decision maker is risk-neutral. We now study the following risk-averse variant of the portfolio optimization problem by considering the exponential utility function [26,24]: $U(f(x, \xi))$, where $U(y) =: e^{y/4}$ and f is defined in (5.4).

Figs. 2–3 compare the DRO models with linear and nonlinear objective functions (DRO-L and DRO-N for short) on the evolution of wealth over 200 trading days when managing a portfolio of four assets. From the two figures, we can see that the DRO-N is more stable than the DRO-L albeit it does not necessarily achieve best return in every experiment. Moreover, the DRO-N is insensitive to the type of the ambiguity sets as the two wealth curves returned by the DRO-N with moment type and distance type ambiguity sets are almost same. Fig. 4 shows the results of DRO-N, SP model with nonlinear objective functions and the $1/n$ investment strategy. The figure indicates that all wealth lines have the tendency to going down and the wealth curves of DRO models and $1/n$ investment strategy outperform SP model, but the difference of the wealth curve between the DRO model and the $1/n$ investment strategy are smaller than Fig. 1.

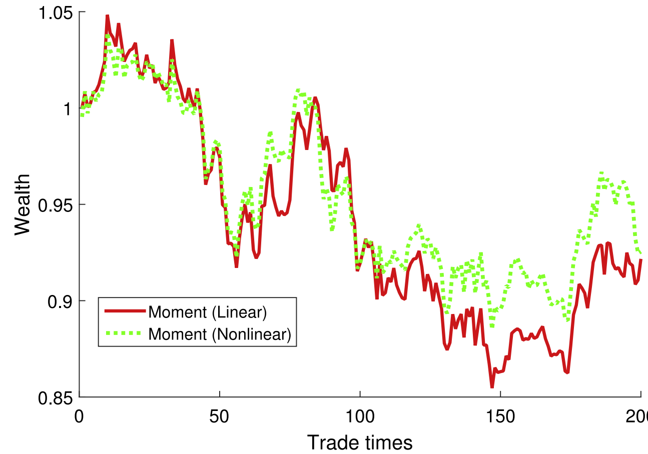

Fig. 2. Moment: Linear vs Nonlinear.

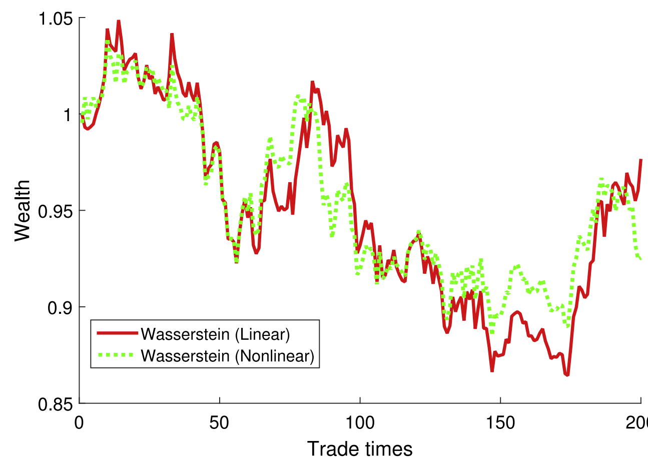

Fig. 3. Wasserstein: Linear vs Nonlinear.

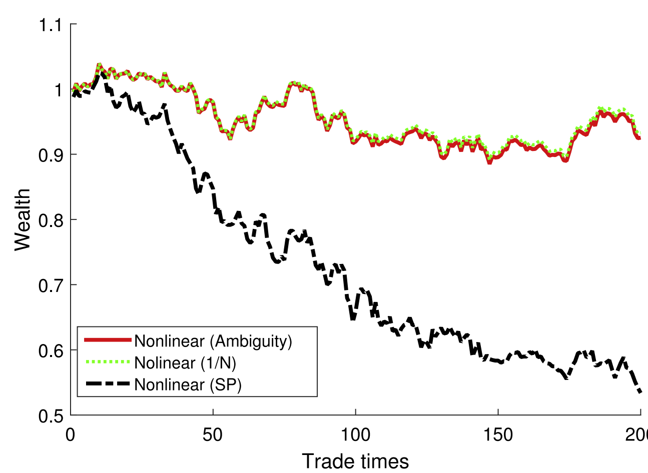

Fig. 4. Wasserstein: Linear vs Nonlinear.

6. Conclusions

Motivated by the recent research on discrete approximation method [12,16,26] to distributionally robust optimization (DRO) problem, we employ the primal–dual hybrid gradient (PDHG) method to solve the DRO problem where the ambiguity set is defined through moment condition and/or Wasserstein metric. As the PDHG method is more efficient for saddle point problem with bilinear objective function, the lifting technique is used to recast the nonlinear objective function. The preliminary numerical test on portfolio selection optimization problem demonstrates the applicability of the numerical approaches.

Acknowledgements

We would like to thank the editor for organizing an effective review and two anonymous referees for insightful comments and constructive suggestions which help us significantly consolidate the paper. The research is supported by the NSFC grants #11571056, #11771255 and #11601458; and the General Research Fund from Hong Kong Research Grants Council: HKBU12300515.

References

[1] K.J. Arrow, L. Hurwicz, H. Uzawa, Studies in Linear and Non-Linear Programming, with Contributions By H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Science, Vol. II, Stanford University Press, Stanford, Calif, 1958.

[2] D. Bertsimas, I. Popescu, Optimal inequalities in probability theory: a convex optimization approach, SIAM J. Optim. 15 (2005) 780–804.

[3] A. Chambolle, T. Pock, A first-order primal–dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vision 40 (2011) 120–145.

[4] E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, Oper. Res. 58 (2010) 592–612.

[5] P.M. Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, Math. Program. (2017). http://dx.doi.org/10.1007/s10107-017-1172-1. (in press).

[6] R. Gao, A.J. Kleywegt, Distributionally robust stochastic optimization with Wasserstein distance,manucript, 2016, http://www.optimization-online.org/DB_FILE/2016/04/5396.pdf.

[7] M. Grant, S. Boyd, CVX, for convex optimization http://www.stanfordedu/boyd/.

[8] B.S. He, F. Ma, X.M. Yuan, An algorithmic framework of generalized primal–dual hybrid gradient methods for saddle point problems, J. Math. Imaging Vision 58 (2017) 279–293.

[9] B.S. He, Y.F. You, X.M. Yuan, On the convergence of primal–dual hybrid gradient algorithm, SIAM J. Imaging Sci. 7 (2014) 2526–2537.

[10] B.S. He, X.M. Yuan, Convergence analysis of primal–dual algorithms for a saddle-point problem: From contraction perspective, SIAM J. Imaging Sci. 5 (2012) 119–149.

[11] R. Jiang, Y. Guan, Data-driven chance constrained stochastic program, Math. Program. 158 (2016) 291–327.

[12] Y. Liu, A. Pichler, H. Xu, Discete approximation and quantification in distributionally robust optimization, Math. Oper. Res., (in press).

[13] S. Mehrotra, D. Papp, A cutting surface algorithm for semiinfinite convex programming with an application to moment robust optimization, SIAM J. Optim. 24 (2014) 1670–1697.

[14] G.Ch. Pflug, A. Pichler, Multistage Stochastic Optimization, in: Springer Series in Operations Research and Financial Engineering, Springer, 2014.

[15] G.Ch. Pflug, A. Pichler, D. Wozabal, The 1/N investment strategy is optimal under high model ambiguity, J. Bank. Financ. 36 (2012) 410–417.

[16] G.Ch. Pflug, D. Wozabal, Ambiguity in portfolio selection, Quant. Finance 7 (2007) 435–442.

[17] T. Pock, A. Chambolle, Diagonal preconditioning for first order primal–dual algorithms in convex optimization, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV, Barcelona, Spain, 2011, pp. 1762–1769.

[18] R.T. Rockafellar, Convex Analysis, Princeton university press, 2015.

[19] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer, Berlin, 1998 (3rd printing 2009).

[20] H. Scarf, in: K.S. Arrow, S. Karlin, H.E. Scarf (Eds.), A Min-Max Solution of an Inventory Problem, in: Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958, pp. 201–209.

[21] J. Shawe-Taylor, N. Cristianini, Estimating the moments of a random vector with applications, in: Proc. GRETSI 2003 Conference, 2003, pp. 47–52.

[22] W. Wiesemann, D. Kuhn, B. Rustem, Robust resource allocations in temporal networks, Math. Program. 135 (2012) 437–471.

[23] W. Wiesemann, D. Kuhn, B. Rustem, Robust Markov decision process, Math. Oper. Res. 38 (2013) 153–183.

[24] W. Wiesemann, D. Kuhn, M. Sim, Distributionally robust convex optimization, Oper. Res. 62 (2014) 1358–1376.

[25] L. Xin, D.A. Goldberg, A. Shapiro, Distributionally robust multistage inventory models with moment constraints, 2013, http://www2.isye.gatech.edu/dgoldberg9/ftp/Goldberg_Shapiro_Xin_Submitted_Robust_Multistage_Inventory_Moment_Constraints.pdf..

[26] H. Xu, Y. Liu, H. Sun, Distributionally robust ptimization with matrix moment constraints: lagrange duality and cutting plane methods, Math. Program. (2017). http://dx.doi.org/10.1007/s10107-017-1143-6. (in press).

[27] X. Zhang, M. Burger, S. Osher, A unified primal–dual algorithm framework based on Bregman iteration, J. Sci. Comput. 46 (2011) 20–46.

[28] C. Zhao, Y. Guan, Data-driven risk-averse stochastic optimization with Wasserstein metric, manuscript, 2015, available at http://www.optimization-online.org/DB_FILE/2015/05/4902.pdf.

[29] M. Zhu, T.F. Chan, An efficient primal–dual hybrid gradient algorithm for total variation image restoration, CAM Report 08-34, UCLA, Los Angeles, CA 2008.