

Article

An Approximate Algorithm for Sparse Distributionally Robust Optimization

Ruyu Wang ¹, Yaozhong Hu ², Cong Liu ¹ and Quanwei Gao ^{3,*} 

¹ College of Science, Northwest A&F University, Xianyang 712100, China

² Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

³ College of Information Engineering, Northwest A&F University, Xianyang 712100, China

* Correspondence: qwgao@nwfau.edu.cn

Abstract

In this paper, we propose a sparse distributionally robust optimization (DRO) model incorporating the Conditional Value-at-Risk (CVaR) measure to control tail risks in uncertain environments. The model utilizes sparsity to reduce transaction costs and enhance operational efficiency. We reformulate the problem as a Min-Max-Min optimization and convert it into an equivalent non-smooth minimization problem. To address this computational challenge, we develop an approximate discretization (AD) scheme for the underlying continuous random vector and prove its convergence to the original non-smooth formulation under mild conditions. The resulting problem can be efficiently solved using a subgradient method. While our analysis focuses on CVaR penalty, this approach is applicable to a broader class of non-smooth convex regularizers. The experimental results on the portfolio selection problem confirm the effectiveness and scalability of the proposed AD algorithm.

Keywords: approximate algorithm; sparse optimization; distributionally robust optimization (DRO); portfolio selection; Conditional Value-at-Risk (CVaR)



Academic Editor: Christos Gogos

Received: 29 June 2025

Revised: 29 July 2025

Accepted: 5 August 2025

Published: 7 August 2025

Citation: Wang, R.; Hu, Y.; Liu, C.; Gao, Q. An Approximate Algorithm for Sparse Distributionally Robust Optimization. *Information* **2025**, *16*, 676. <https://doi.org/10.3390/info16080676>

Copyright: © 2025 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/>).

1. Introduction

The problem of robust optimization has gained significant attention in various fields due to its ability to handle uncertainty in optimization problems (see [1–3]). In particular, DRO has emerged as an effective approach for modeling uncertain parameters, where the objective is to optimize a performance criterion under the worst-case distribution of uncertain data (see [4–8]). While the CVaR measure has been widely used in DRO to account for tail risk, the application of such methods often faces computational challenges, especially when dealing with non-smooth optimization problems. CVaR is widely considered to be a more effective measurement of extreme losses, particularly in the face of large market fluctuations or black swan events (e.g., [9–14]), as it better captures the potential risks of a portfolio, and it can be determined from the formula

$$\phi_{\beta}(x) = \min_{\alpha \in \mathbb{R}} \{ \alpha + (1 - \beta)^{-1} \mathbb{E}[\ell(x, \xi) - \alpha]_+ \}, \quad (1)$$

where $\ell(x, \xi)$ is a convex loss function in $x \in X$ that depends on some vector of parameters $\xi \in \Xi$.

We propose a sparse DRO model that leverages the CVaR risk measure to manage uncertainty in the decision process. By incorporating sparsity into the model, we aim to improve both the computational efficiency and practical applicability of the solution,

especially in high-dimensional settings where traditional methods may be computationally expensive. For example, traditional portfolio optimization methods often lead to a full investment solution, where all assets have non-zero weights. However, in practice, investors generally seek a sparse portfolio (e.g., [15–17]), where most asset weights are zero, to reduce transaction costs and management complexity. When considering DRO, a challenge remains as to how to ensure the robustness of the model while achieving a sparse solution. Therefore, achieving a sparse portfolio while maintaining robustness has become a core problem that needs to be addressed.

The proposed model can be formulated as a Min-Max-Min optimization problem, which is then transformed into an equivalent non-smooth minimization problem. A key contribution of this paper is the introduction of an approximate discretization scheme to solve the resulting non-smooth minimization problem. The proposed scheme involves discretizing the continuous random vector and allows for efficient computation through subgradient methods or smoothing-based algorithms. Moreover, we demonstrate that the discretization scheme converges to the equivalent non-smooth formulation under mild conditions. Although we use the CVaR penalty for illustration, the methodology extends to general non-smooth convex penalty functions, which are broadly applicable to various optimization problems that arise in practice. The proposed approximate discretization (AD) algorithm in Algorithm 1 belongs to the class of data-driven optimization frameworks, which have gained increasing attention due to their flexibility and scalability in handling complex and uncertain environments. Unlike physics-based computational approaches [18–20], data-driven approaches rely on observed or simulated data to model uncertainties and guide decision-making [4].

In Section 2, we propose the sparse DRO with CVaR (SDRPC) model and establish an equivalent tractable reformulation for SDRPC using the Lagrangian dual problem. Meanwhile, we prove that the dual problem is convex and the set of optimal solutions is nonempty. The objective of the dual problem includes the maximization of an infinite number of convex functions. We propose a discretization scheme for this problem and study its convergence in Section 3.

Throughout the paper, we use the following notation: We use $\mathbb{S}_+^d$ to denote a cone of symmetric positive semi-definite matrices, and $\Delta_d = \left\{x \in \mathbb{R}^d \mid \sum_{i=1}^d x_i = 1, x \geq 0\right\}$. For a number $r \in \mathbb{R}$, we denote $[r]_+ = \max\{r, 0\}$. Given two square matrices M and N , we write $M \preceq N$ to indicate $N - M$ being positive semi-definite. The notation $\langle M, N \rangle = \sum_{i,j} M_{i,j} N_{i,j}$. For the sake of simplicity, we denote $v = (x, \alpha, q, \Lambda)$ and $\mathcal{V} = \Delta_d \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{S}_+^m$.

2. The Sparse DRO with CVaR (SDRPC) Model

In order to find a sparse and robust optimal solution, we propose the SDRPC model as follows:

$$(\text{Min} - \text{Max} - \text{Min problem}) \quad \min_{x \in \Delta_d} \max_{P \in \mathcal{P}} \mathbb{E}_P[\mathbf{F}(x, \xi)] + \tau_1 \|x\|_1 + \tau_2 \phi_\beta(x), \quad (2)$$

where $\tau_1, \tau_2 > 0$ are given parameters, $\mathbf{F}(x, \xi)$ is convex and continuous, and the CVaR $\phi_\beta(x)$ is defined in (1). Note that ξ is a random variable with distribution $P \in \mathcal{P}$. The general formulation of $\mathbf{F}(x, \xi)$ makes the model scalable and adaptable to a wide range of economic problems. Whether applied to portfolio optimization, production planning, or supply chain management, the framework can accommodate large, complex systems and diverse economic environments, making it a versatile tool in practical economic decision-making. Although problem (2) appears to be a typical minimax formulation at first glance—minimizing over the decision variable $x \in \Delta_d$ and maximizing over the

worst-case distribution $P \in \mathcal{P}$ —it further involves an inner minimization over the scalar threshold $\alpha \in \mathbb{R}$ due to the definition of the CVaR term in (1). Therefore, problem (2) can be regarded as a Min-Max-Min problem.

To describe the set of probability measures $\mathcal{P}$, let us denote by $\hat{\mu}$ and $\hat{\Sigma}$ the reference values of the mean vector and covariance matrix of the historical data $(\hat{\xi}^1, \dots, \hat{\xi}^N)$. We assume that $\hat{\Sigma}$ is a symmetric and positive definite matrix. The ambiguity set $\mathcal{P}$ in (2) is constructed by moment constraints, as intensively studied by Delage and Ye in [4] and by Xu et al. [21]:

$$\mathcal{P} = \left\{ P \in \mathcal{M} \mid \begin{array}{l} (\mathbb{E}_P[\xi] - \hat{\mu})^\top \hat{\Sigma}^{-1} (\mathbb{E}_P[\xi] - \hat{\mu}) \leq \kappa_1 \\ \mathbb{E}_P[(\xi - \hat{\mu})(\xi - \hat{\mu})^\top] \preceq \kappa_2 \hat{\Sigma} \end{array} \right\}, \quad (3)$$

where $\kappa_1 \geq 0, \kappa_2 \geq 1$ are two given numbers and $\mathcal{M}$ is the convex set of all probability measures in the measurable space $(\Xi, \mathbb{B})$, with $\Xi \subseteq \mathbb{R}^m$ denoting a convex compact set known to contain the support of P and $\mathbb{B}$ denoting the Borel σ -algebra on Ξ . It is easy to observe that the first constraint in (3), through the Schur complement theorem (Section A.5.5 of [22]), can be equivalently written as

$$\mathbb{E}_P \begin{bmatrix} -\hat{\Sigma} & \hat{\mu} - \xi \\ (\hat{\mu} - \xi)^\top & -\kappa_1 \end{bmatrix} \preceq 0.$$

We denote

$$\begin{aligned} h_1(v) &:= \langle \kappa_2 \hat{\Sigma}, \Lambda \rangle + (\Lambda^\top \hat{\mu} + q)^\top \hat{\mu} + \tau_1 \|x\|_1 + \tau_2 \alpha, \\ h_2(v, \xi) &:= \mathbf{F}(x, \xi) - (\Lambda^\top \xi + q)^\top \xi + \frac{\tau_2}{1 - \beta} [\ell(x, \xi) - \alpha]_+, \\ \hat{K}(x, \alpha, \xi) &:= \mathbf{F}(x, \xi) + \frac{\tau_2}{1 - \beta} [\ell(x, \xi) - \alpha]_+ + \tau_1 \|x\|_1 + \tau_2 \alpha. \end{aligned} \quad (4)$$

Below we will reformulate the “Min-Max-Min” problem (2) as the following non-smooth SDP:

$$\min_{v \in \mathcal{V}} \{ \varphi(v) := \max_{\xi \in \Xi} h(v, \xi) \}, \quad \text{where} \quad h(v, \xi) := h_1(v) + \sqrt{\kappa_1} \|\hat{\Sigma}^{\frac{1}{2}}(q + 2\Lambda \hat{\mu})\| + h_2(v, \xi). \quad (5)$$

Theorem 1 (Non-smooth SDP reformulation of SDRPC). *Consider the SDRPC problem (2) and its associated non-smooth SDP problem (5). Then the optimal values for (2) and (5) are equal, in the sense that x is an optimal solution of (2) if and only if there exists $v \in \mathcal{V}$, which is an optimal solution of (5).*

Proof. Based on (4), problem (2) is equivalent to the following problem:

$$\min_{x \in \Delta_d} \max_{P \in \mathcal{P}} \min_{\alpha \in \mathbb{R}} \int_{\Xi} \hat{K}(x, \alpha, \xi) P(d\xi). \quad (6)$$

Since Ξ is a compact set in (3), according to Section 2.2 in [21] and the boundedness of probability measures on a compact set, the set of all probability measures on $(\Xi, \mathbb{B})$ is compact under the topology of weak convergence. Then the minimax theorem, Theorem 4.2 in [23], is applicable to the $\max_{P \in \mathcal{P}} \min_{\alpha \in \mathbb{R}}$ expression in (6). This is due to the convexity of the function $\hat{K}(x, \alpha, P)$ with respect to α (stemming from the convexity of linear and plus functions) and the concavity (indeed, linearity) with respect to P , along with the compactness of $\mathcal{P}$. Then, interchanging the $\max$ and $\min$ operators in (6) leads to an equivalent form:

$$\min_{x \in \Delta_d} \min_{\alpha \in \mathbb{R}} \max_{P \in \mathcal{P}} \int_{\Xi} \hat{K}(x, \alpha, \xi) P(d\xi). \quad (7)$$

We start by transforming the inner maximization problem of (7):

$$\begin{aligned} \max_{P \in \mathcal{M}} \quad & \mathbb{E}_P[\hat{K}(x, \alpha, \xi)] \\ \text{s.t.} \quad & \mathbb{E}_P \begin{bmatrix} -\hat{\Sigma} & \hat{\mu} - \xi \\ (\hat{\mu} - \xi)^\top & -\kappa_1 \end{bmatrix} \preceq 0, \\ & \mathbb{E}_P[(\xi - \hat{\mu})(\xi - \hat{\mu})^\top] \preceq \kappa_2 \hat{\Sigma}, \\ & \mathbb{E}_P[1] = 1, \end{aligned} \quad (8)$$

which uses the ambiguity set (3). From the definition (4) of $\hat{K}(x, \alpha, \xi)$, the Lagrange function of (8) is

$$\begin{aligned} \int \mathbf{F}(x, \xi) - (\Lambda^\top \xi - 2\hat{q})^\top \xi + 2\hat{\mu}^\top \Lambda \xi + \frac{\tau_2}{1-\beta} [\ell(x, \xi) - \alpha]_+ - r \, dP(\xi) \\ + \langle \kappa_2 \hat{\Sigma}, \Lambda \rangle - (\Lambda^\top \hat{\mu} + 2\hat{q})^\top \hat{\mu} + \tau_1 \|x\|_1 + \tau_2 \alpha + r + \langle \hat{\Sigma}, Q \rangle + \kappa_1 s. \end{aligned}$$

Here, the parameters $r \in \mathbb{R}$ and $\Lambda \in \mathbb{R}^{m \times m}$ are the dual variables for the last and second constraints of (8), respectively. And $Q \in \mathbb{R}^{m \times m}$, $\hat{q} \in \mathbb{R}^m$ and $s \in \mathbb{R}$ together form a matrix consisting of the dual variables for the first constraint of (8). Then the Lagrange dual variable of (8) takes the following form:

$$\min_{r, \Lambda, Q, \hat{q}, s} \quad \langle \kappa_2 \hat{\Sigma}, \Lambda \rangle - (\Lambda^\top \hat{\mu} + 2\hat{q})^\top \hat{\mu} + \tau_1 \|x\|_1 + \tau_2 \alpha + r + \langle \hat{\Sigma}, Q \rangle + \kappa_1 s \quad (9)$$

$$\text{s.t.} \quad \mathbf{F}(x, \xi) - (\Lambda^\top \xi - 2\hat{q})^\top \xi + 2\hat{\mu}^\top \Lambda \xi + \frac{\tau_2}{1-\beta} [\ell(x, \xi) - \alpha]_+ - r \leq 0, \quad \forall \xi \in \Xi,$$

$$r \in \mathbb{R}, \quad \Lambda \succeq 0, \quad \begin{bmatrix} Q & \hat{q} \\ \hat{q}^\top & s \end{bmatrix} \succeq 0. \quad (10)$$

Similarly to the procedure used in the proof of Lemma 1 in [4], we can simplify the dual problem by analytically solving for the variables $(Q, \hat{q}, s)$ while keeping (Λ, r) fixed. For the sake of completeness, we briefly outline these steps below. In view of the semi-definite constraint (10), we consider two cases for the variable s^* : either $s^* = 0$ or $s^* > 0$. Let us first consider the case of $s^* = 0$. In this scenario, if $\hat{q}^* \neq 0$, this would lead to $\hat{q}^{*\top} \hat{q}^* > 0$ and

$$\begin{bmatrix} \hat{q}^* \\ y \end{bmatrix}^\top \begin{bmatrix} Q^* & \hat{q}^* \\ \hat{q}^{*\top} & s^* \end{bmatrix} \begin{bmatrix} \hat{q}^* \\ y \end{bmatrix} = \hat{q}^{*\top} Q^* \hat{q}^* - 2\hat{q}^{*\top} \hat{q}^* y < 0, \quad \text{for } y > \frac{\hat{q}^{*\top} Q^* \hat{q}^*}{2\hat{q}^{*\top} \hat{q}^*},$$

which contradict Equation (10). Therefore, it must be the case that $\hat{q}^* = 0$. Furthermore, given that $\hat{\Sigma} \succ 0$, $Q \succeq 0$ and the objective is to minimize $\langle \hat{\Sigma}, Q \rangle$, we can conclude that $Q^* = 0$. Let us consider the case when $s^* > 0$. According to the Schur complement of Section A.5.5 in [22], (10) is equivalent to $Q \succeq \frac{1}{s} \hat{q} \hat{q}^\top$. Since $\hat{\Sigma} \succ 0$ and the objective is to minimize $\langle \hat{\Sigma}, Q \rangle$, we can deduce that $Q^* = \frac{1}{s} \hat{q} \hat{q}^\top$, $s^* = \arg \min_{s>0} \frac{1}{s} \hat{q}^\top \hat{\Sigma} \hat{q} + \kappa_1 s = \frac{\|\hat{\Sigma}^{\frac{1}{2}} \hat{q}\|}{\sqrt{\kappa_1}}$, and $\langle \hat{\Sigma}, Q^* \rangle + \kappa_1 s^* = \sqrt{\kappa_1} \|\hat{\Sigma}^{1/2} \hat{q}\|$. For the above two cases, after substituting $q = -2(\hat{q} + \Lambda \hat{\mu})$, the Lagrange dual Formulas (9) and (10) of (8) simplify to

$$\begin{aligned} \min_{r, \Lambda, q} \quad & r + \langle \kappa_2 \hat{\Sigma}, \Lambda \rangle + (\Lambda^\top \hat{\mu} + q)^\top \hat{\mu} + \tau_1 \|x\|_1 + \tau_2 \alpha + \sqrt{\kappa_1} \|\hat{\Sigma}^{1/2} (q + 2\Lambda \hat{\mu})\| \\ \text{s.t.} \quad & \mathbf{F}(x, \xi) - (\Lambda^\top \xi + q)^\top \xi + \frac{\tau_2}{1-\beta} [\ell(x, \xi) - \alpha]_+ - r \leq 0, \quad \forall \xi \in \Xi, \\ & r \in \mathbb{R}, \quad q \in \mathbb{R}^m, \quad \Lambda \in \mathbb{S}_+^m. \end{aligned} \quad (11)$$

Following Shapiro's duality theory for moment problems, Proposition 3.4 in [24] and Equation (2.3) in [21], the Slater-type condition of (8) can be written as

$$(1, 0, 0) \in \text{int} \left\{ \left(\langle P, 1 \rangle, \left\langle P, \begin{pmatrix} -\hat{\Sigma} & \hat{\mu} - \xi \\ (\hat{\mu} - \xi)^\top & -\kappa_1 \end{pmatrix} \right\rangle, \left\langle P, (\xi - \hat{\mu})(\xi - \hat{\mu})^\top - \kappa_2 \hat{\Sigma} \right\rangle \right) + \{0\} \times \mathbb{S}_+^{m+1} \times \mathbb{S}_+^m | P \in \mathcal{M} \right\}, \quad (12)$$

where $\langle P, \Psi(\xi) \rangle = \int_{\Xi} \Psi(\xi) P(d\xi)$. From Example 2.3 in [21], the moment constraints (3) satisfy the Slater-type condition (12). Then the equivalence between problems (8) and (11) holds. Since the support set Ξ is compact, we know that the optimal value of (8) is finite. According to Proposition 3.4 in [24], if the common optimal value of the primal problem and the dual problem is finite, then the set of optimal solutions to the dual problem is nonempty. Consequently, we deduce that the set of optimal solutions to the dual problem (11) is nonempty and bounded.

Given any fixed x and α , we obtain the equivalence between (8) and its dual problem (11) from the paragraph above. We now return to discuss semi-infinite programming (11). The main difficulty in solving a semi-infinite programming problem comes from infinitely many constraints since there are infinitely many values of ξ in the sample space Ξ . We rewrite $h_2(v, \xi) \leq r, \forall \xi \in \Xi$, for $\max_{\xi \in \Xi} \{h_2(v, \xi)\} \leq r$. Then the Lagrange dual problem (11) is equivalent to

$$\begin{aligned} \min_{q, \Lambda} \quad & \max_{\xi \in \Xi} h(v, \xi) \\ \text{s.t.} \quad & q \in \mathbb{R}^m, \quad \Lambda \in \mathbb{S}_+^m. \end{aligned}$$

We use the fact that Min-Min operators can be performed jointly. This leads to an equivalent formulation (5) for SDRPC. $\square$

Using similar arguments to those used for tractable reformulation in Lemma 1 of [4] by Delage and Ye, we equivalently transform (8) to a semi-infinite problem (11) through the Lagrange dual equation and then use a strategy that further simplifies this dual problem by solving analytically for the $(m+1)^2$ dual variables corresponding to the first constraint in (8) using an auxiliary vector q of m variables while keeping the other dual variables. In fact, this strategy reduces $m^2 + m + 1$ unknown variables, at the expense of adding a non-smooth term $\|\cdot\|$ to the objective function.

So far, we have transformed the SDRPC index tracking model (2) and equivalently reformulated it as the SDP in (5), as shown in Theorem 1. Since Δ_d and $\mathcal{P}$ are bounded and the objective function in (2) is continuous, we know that the solution set of the original problem (2) is nonempty and bounded. Using Theorem 1, the existence of optimal solutions to the SDP (5) is guaranteed.

3. The Discretization Scheme

For the purpose of computation, we provide a discretization model for the equivalent tractable reformulation. We show the existence of solutions for the discretization model under mild conditions. The convergence results of the optimal values and solutions of the discretization scheme to those of the original equivalent reformulation are also given under mild assumptions.

In what follows, we consider the discrete approximation of (5). Our first step is to develop a discrete approximation of the continuous support set Ξ . Let $\Xi_{[N]} = \{\xi^1, \dots, \xi^N\}$

be the independent and identically distributed (i.i.d.) samples of ξ drawn by Monte Carlo sampling from the set Ξ . We consider the discretization scheme of (5):

$$\min_{\nu \in \mathcal{V}} \{\varphi^N(\nu) := \max_{\xi \in \Xi_{[N]}} h(\nu, \xi)\}, \quad (13)$$

where h is defined in (5). Then the corresponding approximation to $\phi_\beta(x)$ in (1) is

$$\phi_\beta^N(x) = \min_{\alpha \in \mathbb{R}} \left\{ \alpha + \frac{1}{(1-\beta)N} \sum_{i=1}^N [\ell(x, \xi^i) - \alpha]_+ \right\}. \quad (14)$$

Let $\mathbb{A}$ be a compact set consisting of the values of α for which the minimum in $\phi_\beta^N(x)$ is attained. We respectively denote the optimal values of (13) and (5) as $\hat{\vartheta}_N$ and $\hat{\vartheta}$. In Theorem 2, we state the convergence of problem (13) to problem (5) in terms of the optimal value.

Theorem 2. *The non-smooth model in (13) is convex and the solution set of (13) is nonempty. Suppose that the optimal value of (8) is finite, and let $\xi^1, \dots, \xi^N$ be i.i.d. samples of ξ , with a continuous probability distribution P over Ξ such that*

$$P\left(\|\xi - \xi^0\| < \delta\right) \geq C_2 \delta^{\gamma_2} \quad (15)$$

for any fixed point $\xi^0 \in \Xi$ and $\delta \in (0, \delta_0)$, where C_2, γ_2 and δ_0 are some positive constants. When N is sufficiently large, for any positive number ε , there exist positive constants $\hat{C}(\varepsilon)$ and $\hat{\beta}(\varepsilon)$ such that

$$\text{Prob}(|\hat{\vartheta}_N - \hat{\vartheta}| \geq \varepsilon) \leq \hat{C}(\varepsilon) e^{-\hat{\beta}(\varepsilon)N}. \quad (16)$$

Proof. For any $\lambda \in (0, 1)$, $q^1, q^{(2)} \in \mathbb{R}^m$, and $\Lambda^{(1)}, \Lambda^{(2)} \in \mathbb{S}_+^m$, we have, through direct computation and the convexity of $\|\cdot\|$, that

$$\begin{aligned} & \sqrt{\kappa_1} \left\| \hat{\Sigma}^{\frac{1}{2}} \left((\lambda q^{(1)} + (1-\lambda)q^{(2)}) + 2(\lambda_1 \Lambda^{(1)} + (1-\lambda)\Lambda^{(2)})\hat{\mu} \right) \right\| \\ & \leq \lambda \sqrt{\kappa_1} \left\| \hat{\Sigma}^{\frac{1}{2}} (q^1 + 2\Lambda^{(1)}\hat{\mu}) \right\| + (1-\lambda) \sqrt{\kappa_1} \left\| \hat{\Sigma}^{\frac{1}{2}} (q^{(2)} + 2\Lambda^{(2)}\hat{\mu}) \right\|. \end{aligned}$$

Therefore, using the definition of a convex function, we can show without difficulty that $h(\nu, \xi)$ in (5) with respect to ν is a convex function. From Proposition 1.38 in [25], the maximum of a finite number of convex functions is also convex. In other words, $\varphi^N(\nu)$ is convex in ν . Moreover, the feasible set $\mathcal{V}$ is a convex set. Hence, the non-smooth model (13) is convex.

From (14), we see that if the sampling generates a collection of vectors $\xi^1, \dots, \xi^N$, then the maximum with respect to α in (14) is achieved at a finite α . In other words, we may restrict the maximum with respect to α to be taken within a closed interval $[-c, c]$ for some sufficiently large positive number c ; see [9,21]. Let $\mathbb{A}$ be a compact set consisting of the values of α for which the minimum in $\phi_\beta^N(x)$ is attained. From the compactness of Δ_d and $\Xi_{[N]}$, there exist $x^* \in \Delta_d$ and $\xi^* \in \Xi_{[N]}$ as the optimal solutions of (13). Then, the solution set of (13) is nonempty. We notice the following points:

(i) Denote by

$$M_\nu(t) := \mathbb{E} \left[e^{t(h(\nu, \xi) - \mathbb{E}[h(\nu, \xi)])} \right]$$

the moment-generating function of the random variable $h(\nu, \xi) - \mathbb{E}[h(\nu, \xi)]$. Since Ξ is a compact set and from Section 3.1 in [21], for each $\nu \in \mathcal{V}$, $\sup_{\xi \in \Xi} h(\nu, \xi) < \infty$ and the moment-generating function $M_\nu(t)$ is finitely valued for all t in the neighborhood of zero.

(ii) Through the continuity of the function h , we can establish the existence of a nonnegative measurable function $\kappa : \Xi \rightarrow \mathbb{R}_+$ and a constant $\gamma > 0$ such that for all $\xi \in \Xi$,

$$|h(\nu', \xi) - h(\nu'', \xi)| \leq \kappa(\xi) \|\nu' - \nu''\|^\gamma, \quad \forall \nu', \nu'' \in \mathcal{V}.$$

(iii) Considering the boundedness of the support set Ξ and Section 5 in [26], we can conclude that the moment-generating function $M_\kappa(t)$ of $\kappa(\xi)$ is finite for all t in the neighborhood of zero.

The above facts, (i)–(iii), along with (15) and the continuity of $h(\nu, \cdot)$ over Ξ enable us to conclude that the relationship between the optimal values of (13) and (5), i.e., (16), holds true, as indicated by Lemma 3.1 (i) in [21]. $\square$

Remark 1. From Proposition 1 in [12], condition (15) is very weak: it can be guaranteed whenever the density of ξ is bounded below by a positive real-valued analytic function. It is easily seen to hold when the density is bounded away from zero around ξ^0 , and so it is an effective condition.

4. The Approximate Discretization Algorithm

The approximate discretization scheme in Section 3 can be solved using either a subgradient algorithm [27–29] or a smoothing projected-gradient algorithm [30–32]. In this section, we use the subgradient algorithm to solve (13), i.e., the discretization scheme of the equivalent reformulation. To do this, the subgradient algorithm uses the simple iteration

$$\nu_{k+1} = P_{\mathcal{V}}(\nu_k - \alpha_k g_k). \quad (17)$$

Here ν_k is the k -th iterate, g_k is any subgradient of φ^N at ν_k , $P_{\mathcal{V}}$ is the projected operator, and $\alpha_k > 0$ is the k -th step size.

Now we are ready to provide a basic scheme for the approximate discretization (AD) algorithm in Algorithm 1.

Algorithm 1 The approximate discretization (AD) algorithm for DRO

- Step 0.1. Set $\mathbf{F}(x, \xi)$ in (2) and $\ell(x, \xi)$ in (1).
- Step 0.2. Interchange the $\max_{P \in \mathcal{P}}$ and $\min_{\alpha \in \mathbb{R}}$ operators in (6), leading to an equivalent form (7).
- Step 0.3. Calculate the Lagrange dual variable (11) of the inner maximization problem (8) and obtain the non-smooth SDP (5).
- Step 0.4. Set $\Xi_{[N]} = \{\xi^1, \dots, \xi^N\}$ as i.i.d. samples of ξ drawn by Monte Carlo sampling from the set Ξ .
- Step 0.5. Set the discretization scheme of (5), i.e., (13).
- Step 0.6. Set the initial point $\nu_0 \in \mathcal{V}$ and the step size $\{\alpha_k > 0\}$ for any $k \geq 1$.
- Step 1. For $k \geq 1$, call the subgradient algorithm (17) m times to obtain

$$\nu_{k+1} = P_{\mathcal{V}}(\nu_k - \alpha_k g_k), \quad k = 1, \dots, m-1.$$

Output: ν_m

Compared to the smoothed projected-gradient method, the subgradient method offers several distinct advantages. First, it eliminates the need to update a sequence of smoothing parameters, thereby reducing the number of parameters in the algorithm. Second, for complex non-smooth functions, constructing a suitable smoothed approximation often results in intricate formulations, leading to significantly increased computational costs

in the smoothed projected-gradient method. In contrast, the subgradient method avoids such overhead. Third, the subgradient method allows for a variety of step size strategies, including both adaptive step size rules and fixed step size schemes, providing greater flexibility in practical implementation.

Remark 2. In this paper, we focus on a flexible framework by requiring the loss function $\ell(x, \xi)$ in CVaR formulation to be convex, but without imposing any specific form or structure on $\ell(x, \xi)$. This relaxed assumption allows us to consider a wide range of loss functions, making our model more adaptable to various applications where the exact form of the loss function may not be known in advance. Similarly, we do not place any restrictive conditions on the convex continuous function $\mathbf{F}(x, \xi)$ in the optimization problem (2). By allowing this general form of $\mathbf{F}(x, \xi)$, we further extend the applicability of our SDRPC model. This flexibility ensures that the model can accommodate various types of objective functions and constraints that may arise in different settings, such as portfolio optimization, resource allocation, and risk management.

Remark 3. Although the proposed AD algorithm is developed and illustrated in the context of the CVaR penalty, its applicability is not limited to this specific regularizer. In fact, the reformulation and discretization techniques presented in this work are valid for a broad class of non-smooth convex regularization terms, such as the ℓ_2 -norm, quantile loss, and Huber loss. As long as the involved penalty term remains convex (possibly non-smooth), the equivalence transformation and convergence guarantees established in this paper remain applicable. This generality significantly improves the scope and versatility of the AD algorithm.

5. Applications and Numerical Results

5.1. Applications

The SDRPC model (2) is broadly applicable across various domains due to the versatility in selecting $\mathbf{F}(x, \xi)$ and $\ell(x, \xi)$ in $\phi_\beta(x)$ (1). Below, we illustrate three practical examples that arise from different choices of these functions.

Example 1 (Lasso sparse index tracking problem). Index tracking aims to replicate the index of a financial market by constructing a portfolio consisting of assets in that market that minimizes the tracking error, which measures how closely the portfolio mimics the performance of the benchmark. Let $\xi_B = (\xi_{B,1}, \dots, \xi_{B,d}) \in \mathbb{R}^d$ be the observed historical individual return vector of the d assets and $\xi_a \in \mathbb{R}$ be the observed corresponding random market index return for $j = 1, \dots, N$. We denote this by $\xi = (\xi_B^T, \xi_a)^T \in \mathbb{R}^{d+1}$. Let $x = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ be the tracking portfolio, with x_i being the investment weight in the i th component stock. The risk-averse variant of the Lasso sparse index tracking problem in [33] can be formulated as

$$\min_{x \in \Delta_d} \max_{P \in \mathcal{P}} \mathbb{E}_P \left[\left(\xi_a - \xi_B^T x \right)^2 \right] + \tau_1 \|x\|_1 + \tau_2 \phi_\beta(x),$$

where $\ell(x, \xi) = -\xi_B^T x$ in $\phi_\beta(x)$, τ_1 and τ_2 are given regularization parameters, and the ℓ_1 -norm penalty aims to enhance the sparsity of the portfolio. Since short selling is prohibited, nonnegativity constraints are enforced on the decision variables.

Example 2 (Multiproduct newsvendor problem [34]). Assume that a newsvendor trades in $i = 1, \dots, d$ products. Before observing the uncertain demands ξ_i , the newsvendor orders x_i units of product i at the wholesale price c_i . Once ξ_i is observed, she can sell the quantity $\min(x_i, \xi_i)$ at the retail price v_i . Any unsold stock $(x_i - \xi_i)_+$ is cleared at the salvage price g_i , and any unsatisfied demand $(\xi_i - x_i)_+$ is lost. We study the risk-averse variant of the multiproduct newsvendor problem with $\mathbf{F}(x, \xi) = U(\ell(x, \xi))$, where $U(y) := e^{y/10}$ is an exponential disutility function and

$$\ell(x, \xi) = (c - v)^T x + (v - g)^T (x - \xi)_+.$$

Example 3 (Portfolio optimization problem). We consider a portfolio optimization problem [21] where the investor makes an optimal decision using historical return rates from the National Association of Securities Dealers Automated Quotations (NASDAQ) index: <https://cn.investing.com>. To simplify the discussions, we ignore the transaction fee; therefore, the total value of the portfolio is $F(x, \xi) = -\xi^T x$ in (2) and $\ell(x, \xi) = -\xi^T x$ in (1).

5.2. Numerical Results

We consider the portfolio optimization problem in Example 3, where the investor makes an optimal decision using a historical return rate of $d = 40$ between January 2005 and July 2023 from NASDAQ, which contains $N_{\text{tol}} = 4675$ samples; that is, $q = 4675$ in (13). We denote the daily return rates of n assets on the i -th day as $\xi_{B,i} = (\xi_{B_{1,i}}, \xi_{B_{2,i}}, \dots, \xi_{B_{n,i}})^T$, where $\xi_{B_{j,i}}$ represents the natural logarithm of the closing price divided by the opening price of the j -th asset on the i -th day for $i = 1, \dots, N_{\text{tol}}$. Meanwhile, we set $\kappa_1 = 0.1$ and $\kappa_2 = 1.1$ in (3). All experiments are performed in Windows 11 on an AMD Ryzen 9 7900X 12-Core CPU at 4.70 GHz with 32 GB of RAM using MATLAB R2024b.

We systematically investigate the impact of parameters τ_1 and τ_2 on the performance of the proposed model (2) by conducting experiments over a grid of values $\{10^{-3}, 10^{-2}, 10^{-1}\}$. Table 1 indicates that increasing either τ_1 or τ_2 leads to a significant rise in the objective value (Obj). This trend suggests that both regularization terms, while enhancing sparsity and controlling risk, exert a suppressive effect on the optimal solution—especially when their magnitudes are large, resulting in strong penalization.

Table 1. Obj and CPU time under different parameter settings.

τ_1	τ_2	Obj	CPU Time
10^{-3}	10^{-3}	0.2082	59.9984
10^{-3}	10^{-2}	0.3031	74.0531
10^{-3}	10^{-1}	0.5183	78.4453
10^{-2}	10^{-3}	0.2882	68.1875
10^{-2}	10^{-2}	0.3799	76.6968
10^{-2}	10^{-1}	0.5275	79.2125
10^{-1}	10^{-3}	0.4013	69.6593
10^{-1}	10^{-2}	0.4225	77.2062
10^{-1}	10^{-1}	0.6086	80.0687

The minimum Obj is achieved at $\tau_1 = 10^{-3}$ and $\tau_2 = 10^{-3}$, while the maximum value occurs at $\tau_1 = 10^{-1}$ and $\tau_2 = 10^{-1}$, with an increase of nearly threefold in the objective function. This substantial degradation in performance confirms that the overly aggressive enforcement of sparsity and CVaR risk control can severely affect solution quality. This trend is visually corroborated by the bivariate heat map in Figure 1. Each grid corresponds to a specific (τ_1, τ_2) combination, with its associated Obj mapped to a color using a continuous colormap. The colormap is chosen such that blue corresponds to the lowest (best) Obj, while yellow represents the highest (worse) Obj.

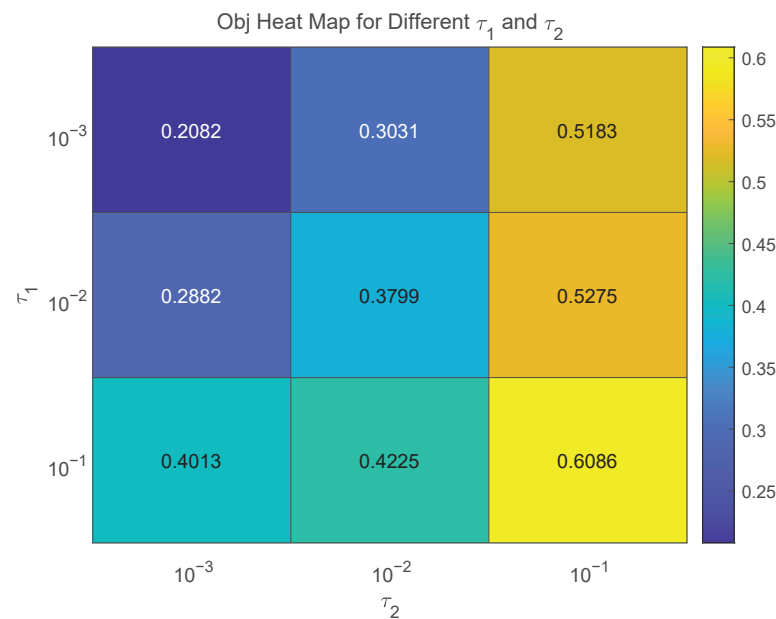


Figure 1. The objective value heat map for different τ_1 and τ_2 .

A closer examination of Table 1 reveals that the effects of τ_1 and τ_2 are not symmetric. When τ_1 is held fixed, increasing τ_2 causes more drastic changes in the objective function, suggesting that the CVaR-related regularization term (τ_2) has a more sensitive and dominant influence on the optimization outcome in this context. To further analyze the effect of τ_1 , Figure 2 presents the objective trends as τ_1 varies while τ_2 is fixed.

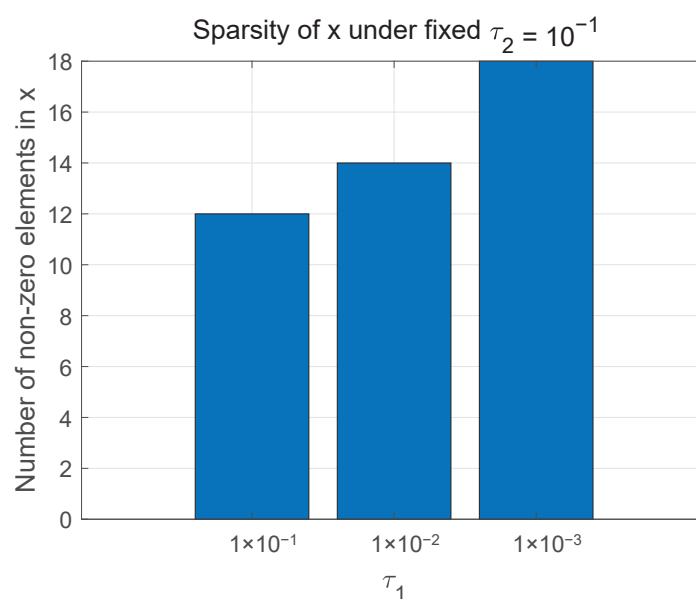


Figure 2. Sparsity of solution vector x under different parameter settings.

Moreover, Figure 3 depicts the convergence trajectory of Obj with respect to CPU time under the settings $\tau_1 = 10^{-3}$ and $\tau_2 = 10^{-3}$. The curve indicates that the AD algorithm converges rapidly, with Obj reaching a relatively stable value within approximately 1.6 s. The algorithm continues running up to 60 s only due to the preset stopping criterion of 1000 iterations. Similar convergence behavior is observed under other parameter configurations, further confirming the stability and computational efficiency of our AD algorithm.

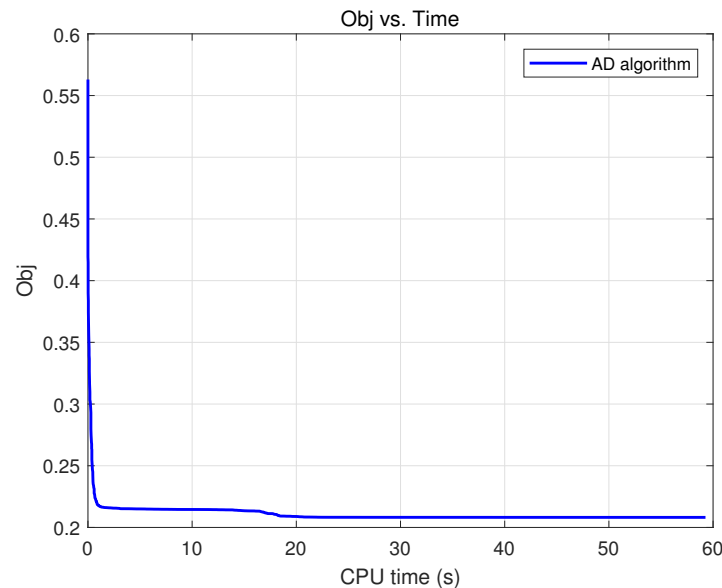


Figure 3. Obj vs. CPU time under $\tau_1 = 10^{-3}$ and $\tau_2 = 10^{-3}$.

6. Conclusions

In this paper, we propose a sparse model that combines DRO and the CVaR penalty, where the CVaR penalty can be replaced by other convex functions. We transform the model to an equivalent non-smooth positive semi-definite type of programming, where the non-smoothness arises from the maximum of numerous non-smooth functions. An approximate discretization algorithm of the non-smooth positive semi-definite programming is provided, which is convergent under mild conditions. This discretization scheme of the equivalent reformulation can be solved directly by a subgradient method or a smoothing projected-gradient method. We propose an approximate discretization (AD) algorithm by combining a low-complexity subgradient method with the preceding transformation. Furthermore, numerical experiments on a portfolio optimization problem validate the effectiveness and efficiency of the AD algorithm in real-world scenarios.

Author Contributions: Conceptualization, R.W., Q.G., C.L. and Y.H.; methodology, R.W., Q.G., C.L. and Y.H.; writing—original draft preparation, R.W., Q.G. and C.L.; writing—review and editing, Q.G. and Y.H.; supervision, Q.G. and Y.H.; funding acquisition, R.W., Q.G. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Scientific Startup Foundation for Doctors of Northwest A&F University (Z1090324139, Z1090125002), the Natural Sciences and Engineering Research Council of Canada (RGPIN 2024-05941), and a centennial fund from the University of Alberta.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Beyer, H.G.; Sendhoff, B. Robust optimization—A comprehensive survey. *Comput. Methods Appl. Mech. Eng.* **2007**, *196*, 3190–3218. [\[CrossRef\]](#)
2. Bertsimas, D.; Brown, D.B.; Caramanis, C. Theory and applications of robust optimization. *SIAM Rev.* **2011**, *53*, 464–501. [\[CrossRef\]](#)

3. Gabrel, V.; Murat, C.; Thiele, A. Recent advances in robust optimization: An overview. *Eur. J. Oper. Res.* **2014**, *235*, 471–483. [[CrossRef](#)]
4. Delage, E.; Ye, Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **2010**, *58*, 595–612. [[CrossRef](#)]
5. Rahimian, H.; Mehrotra, S. Distributionally robust optimization: A review. *arXiv* **2019**, arXiv:1908.05659. [[CrossRef](#)]
6. Levy, D.; Carmon, Y.; Duchi, J.C.; Sidford, A. Large-scale methods for distributionally robust optimization. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 8847–8860.
7. Shapiro, A.; Zhou, E.; Lin, Y. Bayesian distributionally robust optimization. *SIAM J. Optim.* **2023**, *33*, 1279–1304. [[CrossRef](#)]
8. Fan, Z.; Ji, R.; Lejeune, M.A. Distributionally robust portfolio optimization under marginal and copula ambiguity. *J. Optim. Theory Appl.* **2024**, *203*, 2870–2907. [[CrossRef](#)]
9. Rockafellar, R.T.; Uryasev, S. Optimization of Conditional Value-at-Risk. *J. Risk. Res.* **2000**, *2*, 21–42. [[CrossRef](#)]
10. Noyan, N. Risk-averse two-stage stochastic programming with an application to disaster management. *Comput. Oper. Res.* **2012**, *39*, 541–559. [[CrossRef](#)]
11. Arpón, S.; Homem-de Mello, T.; Pagnoncelli, B. Scenario reduction for stochastic programs with Conditional Value-at-Risk. *Math. Program.* **2018**, *170*, 327–356. [[CrossRef](#)]
12. Anderson, E.; Xu, H.; Zhang, D. Varying confidence levels for CVaR risk measures and minimax limits. *Math. Program.* **2020**, *180*, 327–370. [[CrossRef](#)]
13. Behera, J.; Pasayat, A.K.; Behera, H.; Kumar, P. Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105843. [[CrossRef](#)]
14. Yang, C.; Wu, Z.; Li, X.; Fars, A. Risk-constrained stochastic scheduling for energy hub: Integrating renewables, demand response, and electric vehicles. *Energy* **2024**, *288*, 129680. [[CrossRef](#)]
15. Brodie, J.; Daubechies, I.; De Mol, C.; Giannone, D.; Loris, I. Sparse and stable Markowitz portfolios. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12267–12272. [[CrossRef](#)] [[PubMed](#)]
16. Fastrich, B.; Paterlini, S.; Winker, P. Constructing optimal sparse portfolios using regularization methods. *Comput. Manag. Sci.* **2015**, *12*, 417–434. [[CrossRef](#)]
17. Dai, Z.; Wen, F. Some improved sparse and stable portfolio optimization problems. *Financ. Res. Lett.* **2018**, *27*, 46–52. [[CrossRef](#)]
18. Chai, B.; Eisenbart, B.; Nikzad, M.; Fox, B.; Blythe, A.; Blanchard, P.; Dahl, J. Simulation-based optimisation for injection configuration design of liquid composite moulding processes: A review. *Compos. Part A Appl. Sci. Manuf.* **2021**, *149*, 106540. [[CrossRef](#)]
19. Wijaya, W.; Bickerton, S.; Kelly, P. Meso-scale compaction simulation of multi-layer 2D textile reinforcements: A Kirchhoff-based large-strain non-linear elastic constitutive tow model. *Compos. Part A Appl. Sci. Manuf.* **2020**, *137*, 106017. [[CrossRef](#)]
20. Ali, M.A.; Irfan, M.S.; Khan, T.; Khalid, M.Y.; Umer, R. Graphene nanoparticles as data generating digital materials in industry 4.0. *Sci. Rep.* **2023**, *13*, 4945. [[CrossRef](#)]
21. Xu, H.; Liu, Y.; Sun, H. Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Math. Program.* **2018**, *169*, 489–529. [[CrossRef](#)]
22. Boyd, S.P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
23. Sion, M. On general minimax theorems. *Pac. J. Math.* **1958**, *8*, 171–176. [[CrossRef](#)]
24. Shapiro, A. On duality theory of conic linear problems. In *Semi-Infinite Programming*; Springer: New York, NY, USA, 2001; pp. 135–165.
25. Mordukhovich, B.; Nam, N.M. *An Easy Path to Convex Analysis and Applications*; Springer Nature: Berlin/Heidelberg, Germany, 2014.
26. Shapiro, A.; Xu, H. Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization* **2008**, *57*, 395–418. [[CrossRef](#)]
27. Polyak, B.T. Subgradient methods: A survey of Soviet research. In *Proceedings of the Nonsmooth Optimization: Proceedings of a IIASA Workshop, Laxenburg, Austria, 28 March–8 April 1977*; pp. 5–29.
28. Boyd, S.; Xiao, L.; Mutapcic, A. *Subgradient Methods*; Lecture Notes of EE392o; Stanford University: Stanford, CA, USA, 2004.
29. Nedic, A.; Bertsekas, D.P. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.* **2001**, *12*, 109–138. [[CrossRef](#)]
30. Chen, X. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program.* **2012**, *134*, 71–99. [[CrossRef](#)]
31. Zhang, C.; Chen, X. Smoothing projected gradient method and its application to stochastic linear complementarity problems. *SIAM J. Optim.* **2009**, *20*, 627–649. [[CrossRef](#)]
32. Zhang, C.; Chen, X. A smoothing active set method for linearly constrained non-lipschitz nonconvex optimization. *SIAM J. Optim.* **2020**, *30*, 1–30. [[CrossRef](#)]

33. Sant'Anna, L.R.; Caldeira, J.F.; Filomena, T.P. Lasso-based index tracking and statistical arbitrage long-short strategies. *N. Amer. J. Econ. Financ.* **2020**, *51*, 101055. [[CrossRef](#)]
34. Wiesemann, W.; Kuhn, D.; Sim, M. Distributionally robust convex optimization. *Oper. Res.* **2014**, *62*, 1358–1376. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.