ROBUST SOLUTIONS TO A SYSTEM OF STOCHASTIC VERTICAL LINEAR COMPLEMENTARITY PROBLEMS AND APPLICATIONS IN PORTFOLIO SELECTIONS

XIAO ZHA

PhD

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University
Department of Applied Mathematics

Robust Solutions to a System of Stochastic Vertical Linear Complementarity Problems and Applications in Portfolio Selections

Xiao Zha

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

October, 2025

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

_____Xiao Zha_____(Name of student)

To my beloved parents and all those who have made this beautiful journey unforgettable.

Abstract

The stochastic vertical linear complementarity problem (SVLCP) provides a mathematical model encompassing the optimality conditions of stochastic linear and quadratic programs, as well as stochastic bimatrix games. Various formulations and variants of the SVLCP have been proposed and analyzed in the literature. This thesis introduces a new stochastic minimization formulation designed to obtain robust solutions for systems of SVLCPs. The proposed formulation minimizes a risk function subject to stochastic vertical linear complementarity constraints. The content is mainly divided into the following two aspects.

The first part of this thesis addresses the case where the underlying probability space consists of finitely many scenarios. We reformulate the proposed formulation with a finite support set as a linearly constrained piecewise smooth minimization problem by a penalty method. We prove the existence of exact penalty parameters regarding global and local minimizers. We define a smoothing function of the piecewise smooth objective function and show that the smoothing function satisfies the Kurdyka-Łojasiewicz (KL) property. Moreover, we propose a smoothing block coordinate descent (SBCD) algorithm, and prove that the sequence generated by the algorithm globally converges to an ϵ-Clarke stationary point of the penalty problem by the KL property for any $\epsilon > 0$. Finally, we apply our formulation and algorithm to portfolio selection problems with real data. Numerical results demonstrate the robustness of our approach.

The second part of the thesis investigates the sample average approximation (SAA) of the stochastic minimization formulation. We begin by introducing an implicit reformulation of the original problem, which is then approximated through the SAA scheme. The convergence properties of this approximation are analyzed in two folds. First, as the sample size tends to infinity, the optimal value of the SAA problem converges to the counterpart in the true problem. Second, as the sample size approaches infinity, an optimal solution of the SAA problem becomes an optimal solution of the true problem with probability one (w.p. 1). Moreover, we prove the uniform exponential convergence of the objective values and further show that this result implies exponential convergence of the SAA solutions toward the solution set of the original problem. Finally, the theoretical findings are complemented by extensive numerical experiments on portfolio selection problems, which confirm the convergence behavior of the SAA method.

This thesis contains research results of the following paper which was written during the period of my Ph.D study at the Department of Applied Mathematics, The Hong Kong Polytechnic University:

- Zha X, Allen-Zhao Z, Chen X (2024) Robust solutions to a system of stochastic vertical linear complementarity problems. *submitted to Mathematics of Operations Research (Minor Revision).*

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Xiaojun Chen, for introducing me to the filed of mathematical optimization and for her invaluable guidance and immeasurable support during my PhD journey, not only cultivating my research taste, but also reshaping my life perspectives fundamentally. She has consistently motivated and encouraged me to strive for higher goals and has been an unwavering source of knowledge and inspiration.

I am also thankful to my collaborator, Prof. Zhihua Zhao, for introducing me to the portfolio selection problem, as well as for his invaluable advice and insightful discussions. I would also like to thank Prof. Wei Bian for the fruitful discussion on the KL property which significantly improved the work.

Furthermore, I am truly grateful to be part of Prof. Chen's research group and would like to extend my sincere thanks to my academic brothers and sisters: Prof. Chao Zhang, Prof. Congpei An, Prof. Wei Bian, Prof. Xin Liu, Prof. Yanfang Zhang, Prof. Hailin Sun, Prof. Zaikun Zhang, Prof. Xiao Wang, Dr. Yang Zhou, Dr. Lei Yang, Dr. Bo Wen, Dr. Hong Wang, Dr. Chao Li, Dr. Fang He, Dr. Xiaozhou Wang, Dr. Jianfeng Luo, Dr. Wei Liu, Dr. Shisen Liu, Dr. Guan Wang, Dr. Lei Wang, Dr. Xingbang Cui, Dr. You Zhao, Dr. Yong Zhao, Dr. Lin Chen, Dr. Jian Guo, Dr. Fan Wu, Dr. Fei Fang, Dr. Kaixin Gao, Dr. Shijie Yu, Dr Zicheng Qiu, Dr. Yue Wang, Dr. Yifan He, Miss. Xin Qu, Miss. Lingzi Jin, Miss. Yixuan Zhang, Mr. Zhouxing Luo, Mr. Wentao Ma, Miss. Yiyang Li, Mr. Junchao Duan. Specifically, I

am deeply grateful for all the invaluable suggestions and support that Dr. Lei Yang and Dr. Jie Jiang provided during the early days of my PhD studies. I also greatly appreciate all the helpful discussions with Dr. Lin Chen and Dr. Xingbang Cui. My sincere gratitude also goes to Dr Lei Wang for carefully proofreading this thesis and providing valuable feedbacks.

Last but not least, I am thankful to my parents and beloved wife Yuxin for their unconditional love and unwavering support throughout my PhD journey.

Contents

List of Figures

List of Tables

List of Notation

$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of n-dimensional real vectors
$\mathbb{R}^{m\times n}$	the set of $m \times n$ real matrices
$[m]$	for a given $m \in \mathbb{N}$, $[m] := \{1, 2, \ldots, m\}$
$\lceil a \rceil$	the smallest integer greater than or equal to $a \in \mathbb{R}$
$x^\top$	the transpose of a matrix/vector x
$x \perp y$	the inner product of vectors x, y is equal to 0
$\|x\|$	the ℓ_2-norm of a vector x
$\|A\|$	the ℓ_2-norm of a matrix A
$\mathrm{dist}(x, \Omega)$	the distance from a vector x to the set Ω, i.e., $\mathrm{dist}(x, \Omega) := \inf\{\|x - y\| : y \in \Omega\}$
$\mathcal{B}_\gamma(a)$	closed ball of radius $\gamma > 0$ centered at a point $a \in \mathbb{R}^k$
$\mathbb{D}(\mathcal{X}, \mathcal{Y})$	the deviation of the set $\mathcal{X}$ from the set $\mathcal{Y}$, i.e., $\mathbb{D}(\mathcal{X}, \mathcal{Y}) := \sup_{x\in\mathcal{X}} \mathrm{dist}(x, \mathcal{Y})$
$\min(x, y)$	the vector with its i-th component being $\min(x_i, y_i)$
e	a vector in proper dimension with all elements equal to 1
$\nabla f(x)$	the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$

$\widehat{\partial}\,\varphi(x)$	the regular subdifferential of a function $\varphi : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$
$\bar{\partial}\,\varphi(x)$	the limiting subdifferential of a function $\varphi : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$
$\partial\,\varphi(x)$	the Clarke subdifferential of a function $\varphi : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$
$(X, \mathcal{F}, \mathbb{P})$	a probability space, where X is a sample space, $\mathcal{F}$ is a σ-algebra, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure
ξ	a random variable defined in the probability space $(X, \mathcal{F}, \mathbb{P})$
$\mathbb{E}[\xi]$	the expectation of ξ
i.i.d.	independent and identically distributed
a.e.	almost everywhere
w.p.1	with probability 1

Chapter 1

Introduction

The first chapter gives an introduction to our model and presents a literature review on related topics.

1.1 Background

Let $\xi : \Omega \to \mathbb{R}^l$ be a random variable defined in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a support set $\Xi \subset \mathbb{R}^l$ and $\mathcal{M} : \Xi \to \mathbb{R}^m$ be the space of measurable functions. In this thesis, we consider the following stochastic minimization problem: find $(x, u(\cdot)) \in \mathbb{R}^n \times \mathcal{M}$ which solves

$$\begin{array}{ll} \min & \mathbb{E}[\|A(\xi)x - u(\xi)\|^2] + \dfrac{\lambda}{2}\|x\|^2 \\ \text{s.t.} & 0 \le Bx + b \perp Hx + h \ge 0, \\ & 0 \le D(\xi)u(\xi) + c(\xi) \perp M(\xi)u(\xi) + q(\xi) \ge 0 \ \text{ for a.e. } \xi \in \Xi, \end{array} \tag{1.1.1}$$

where $\mathbb{E}[\cdot]$ denotes the expected value over Ξ, $\lambda > 0$ is a regularization parameter, $B, H \in \mathbb{R}^{n_1 \times n}, b, h \in \mathbb{R}^{n_1}$, $A(\cdot) : \Xi \to \mathbb{R}^{m \times n}$, $D(\cdot), M(\cdot) : \Xi \to \mathbb{R}^{m_1 \times m}$ and $c(\cdot), q(\cdot) : \Xi \to \mathbb{R}^{m_1}$ are matrix and vector valued mappings, respectively; $y \perp z$ means $y^\top z = 0$ for any $y, z \in \mathbb{R}^d$; the abbreviation "a.e." stands for "almost everywhere". The constraints in problem (1.1.1) is a system of stochastic vertical linear complementarity problems. Obviously, the vertical linear complementarity problem (VLCP) is a

generalization of the linear complementarity problem (LCP). For VLCP and LCP, see the references [15, 18, 21, 22, 44]

1.2 Literature review

Problem (1.1.1) can be viewed as a stochastic mathematical program with equilibrium constraints (SMPEC), which has been studied for two decades [16, 31, 38, 42].

In [38], Patriksson and Wynter first introduced a clear and rigorous formulation of SMPECs, combining stochastic optimization with equilibrium constraints. This extends the classical framework of mathematical programs with equilibrium constraints (MPECs) [33] to incorporate stochastic elements, making them applicable to real-world problems involving uncertainty. They also discussed the existence of solutions to SMPECs and provide sufficient conditions under which solutions exist. Then, in [13], Christiansen et al. proposed an SMPEC model to address a class of stochastic bilevel programming problems in structural optimization. Their approach captures uncertainties such as material properties and loading conditions within a hierarchical decision-making framework, where the upper-level represents the designer's objectives and the lower-level models system responses or constraints. Through numerical examples, they demonstrate that the SMPEC model enables robust and efficient structural designs under uncertainty. In [42], Shapiro and Xu considered the following SMPEC

$$\begin{array}{ll} \min & \mathbb{E}[\Phi(x, u(\xi), \xi)] \\ \text{s.t.} & x \in \mathcal{X}, \\ & 0 \in \Psi(x, u(\xi), \xi) + \mathcal{N}_{\mathcal{C}}(u(\xi)) \ \text{ for a.e. } \xi \in \Xi, \end{array} \tag{1.2.1}$$

where $\mathcal{X} \subset \mathbb{R}^n$ is a nonempty set, $\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}$ and $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^m$, $\mathcal{C} \subset \mathbb{R}^m$ is a nonempty convex closed set and $\mathcal{N}_{\mathcal{C}}(u(\xi))$ denotes the normal cone to $\mathcal{C}$ at $u(\xi)$. Shapiro and Xu [42] discussed the sample average approximation (SAA) method applied to (1.2.1) with a focus on the case where the variational constraint

has a unique solution for every $x \in \mathcal{X}$ and almost every realization of $\xi \in \Xi$. In [16], Cui et al. established complexity bounds of implicit smoothing first/zeroth-order methods under assumption that $\Psi(x, \cdot, \xi)$ is strongly monotone which can ensure the stochastic variational constraint has a unique solution for every $x \in \mathcal{X}$ and almost every realization of $\xi \in \Xi$. In [31], Lin et al. considered a mathematical program with linear complementarity constraints (SMPLCC) where the variational constraint in (1.2.1) is the linear complementarity constraint as follows

$$0 \leq u(\xi) \perp M(\xi)u(\xi) + q(\xi) + N(\xi)x + z(\xi) \geq 0,\ z(\xi) \geq 0 \ \text{ for a.e. } \xi \in \Xi, \quad (1.2.2)$$

where $N(\xi) \in \mathbb{R}^{m_1 \times n}$ and $z(\xi) \in \mathbb{R}^{m_1}$. Under Assumption that $M(\xi)$ is a R_0 matrix for every $\xi \in \Xi$, Lin et al. [31] established convergence of a combined smoothing implicit programming and penalty method for the problem with a finite set Ξ. From [15, Proposition 3.9.23], such assumption implies that the solution set of the linear complementarity problem (LCP) in (1.2.2) is bounded for any $x \in \mathcal{X}$ and $\xi \in \Xi$.

To the best of our knowledge, the relatively complete recourse condition has been assumed in most existing literature on SMPEC, which means that the "wait and see" solution $u(\xi)$ of the stochastic variational constraint of (1.2.1) exists for every $x \in \mathcal{X}$ and almost every $\xi \in \Xi$. However, in practice, relatively complete recourse condition may fail in many real-world applications. As an example given in [12], when considering to make a decision on building a power station for providing electrical power to satisfy the demand, it could be practically impossible to make sure that the uncertain demand will be satisfied under any possible circumstances. Other example is from forecast of future traffic state for public transport project design and evaluation based on robust Wardrop's user equilibrium assignment [53]. In a global traffic network, some local traffic networks contain uncertainties in demand and road capacity and have their own equilibrium state $u(\xi)$ between the travel demand and supply at any scenario ξ. The forecast of traffic state of global networks is based on

the expected valued of Wardrop's user equilibrium assignment which provides a set $\mathcal{X}$ of "here and now" solutions. It is possible that some local networks do not have equilibrium state with additional recourse $N(\xi)x$ in their demand for any $x \in \mathcal{X}$ and $\xi \in \Xi$.

To find a robust solution without assuming the relatively complete recourse condition for SMPEC, we propose the stochastic minimization model (1.1.1). For the traffic equilibrium problems, a "here and now" solution x^* is a solution of expected value formulation of the stochastic LCP

$$0 \leq x \perp \mathbb{E}[\tilde{H}(\xi)x + \tilde{h}(\xi)] \geq 0,$$

where $\tilde{H}(\xi) \in \mathbb{R}^{n\times n}$ and $\tilde{h}(\xi) \in \mathbb{R}^n$ for $\xi \in \Xi$. A "wait and see" solution $u(\xi)$ is a solution of the stochastic LCP

$$0 \leq u(\xi) \perp M(\xi)u(\xi) + q(\xi) \geq 0 \quad \text{for a.e. } \xi \in \Xi.$$

The objective is to minimize the expected distance between the sub-vector of x^* at local networks and $u(\xi)$. For such case, in problem (1.1.1), $n_1 = n, m_1 = m, n > m$, $B = I_n, D(\xi) = I_m, b = 0, c(\xi) = 0$, $H = \mathbb{E}(\tilde{H}(\xi)), h = \mathbb{E}(\tilde{h}(\xi))$ and $A(\xi) \equiv A$ is a projection from a global variable to a local variable.

Problem (1.1.1) is a generalization of SMPLCC in the sense that the linear complementarity constraints are vertical linear complementarity constraints (VLCC) with $n_1 \leq n, m_1 \leq m, m \leq n$. Note that the stochastic VLCC is independent of the "here and now" variable x, and thus the relatively complete recourse condition is not necessary for the study of problem (1.1.1). The robustness of the "here and now" solution is achieved by minimizing the objective function.

1.3 Summary of contributions of the thesis

The contributions of this thesis are summarized as follows.

- In the first part, to find approximate solutions of problem (1.1.1), we consider the probability space based on finitely many scenarios $\Xi = \{\xi_1, ..., \xi_N\}$, where each $\xi_i \in \mathbb{R}^l$ has a known probability $p_i > 0$ and these probabilities sum up to one.

  - We propose a stochastic minimization model (1.1.1) to find a robust solution of a system of stochastic vertical linear complementarity problems (SVLCPs). We present a penalty reformulation (3.0.2) of the discrete problem (3.0.1). We show that the solution set of (3.0.1) is not empty and bounded, and there is a penalty parameter $\rho^* > 0$ such that problem (3.0.1) and problem (3.0.2) with any $\rho > \rho^*$ have same global minimizers. Moreover, any local minimizer of problem (3.0.1) is a local minimizer of problem (3.0.2).

  - Using the block structure of the constraints in problem (3.0.2), we develop a smoothing block coordinate descent (SBCD) algorithm for (3.0.2). At each step of the algorithm, we can solve N convex quadratic programs in parallel. We prove that the smoothing function $\tilde{\phi}$ satisfies the KL property and the sequence generated by the algorithm globally converges to an ϵ-Clarke stationary point of the penalty problem (3.0.2) by the KL property for any $\epsilon > 0$.

  - To verify the efficiency of our approach, we applied (3.0.2) and the SBCD algorithm to robust portfolios selection problems under uncertain environment. The "here and now" solution is expected valued formulation

of stochastic LCP for global portfolios, while the "wait and see" solution is from stochastic LCP formulation of local portfolios in unstable assets, as the most active part of the global portfolios. Consequently, we obtain favorable out-of-sample performances in both single-asset and industry stress tests.

- In the second part, we consider the sample average approximation (SAA) method for solving problem (1.1.1).
  - First, we present an implicit reformulation for problem (1.1.1) which eliminates the term $u(\xi)$ and therefore is of help for convergence analysis. Then we approximate it with the SAA method.
  - Next, we conduct a convergence analysis for the SAA problem. Specifically, we show that, as the sample size $N \to \infty$, the optimal value and optimal solutions of the SAA problem converge to their counterparts of the true problem with probability one.
  - Next, we establish the uniform exponential convergence of the function values and, furthermore, translate this convergence to the exponential convergence of an optimal solution of the SAA problem converging to the set of optimal solutions of the true problem.
  - Finally, we also conduct extensive numerical experiments on the aforementioned portfolio selection problems to show the convergence of the SAA method.

1.4 Organization of the thesis

The thesis is organized as follows.

- In Chapter 1, we introduce the background and provide a literature review on the stochastic mathematical programming with equilibrium constraints and the relatively complete recourse condition.

- In Chapter 2, we first give basic notation and concepts. Then, we give a comprehensive introduction to the KL property and KL functions, especially the semialgebraic functions.

- In Chapter 3, we consider the case of a discrete finite support set Ξ. Specifically, we first propose an exact penalization problem which partially penalizes the constraints and then design a smoothing block coordinate descent (SBCD) algorithm for solving the penalty problem. Then, we show the convergence analysis of our SBCD algorithm. Finally, we conduct extensive numerical experiments for the portfolio selection problem.

- In Chapter 4, we consider the SAA problem and its convergence analysis. Specifically, we first give an implicit reformulation of the original problem and then approximate it with the SAA method. Then, we present convergence analysis on the SAA problem and show its exponential convergence rate. Finally, numerical experiments do support the convergence analysis.

- In Chapter 5, we summarize the main results of our work and list some possible future works.

Chapter 2

Preliminary

2.1 Notation

Throughout the thesis, we use the following notation. We use $\|\cdot\|$ to denote the ℓ_2-norm for both vectors and matrices. Let $[N]$ be the index set $\{1, \cdots, N\}$. For any vectors $x, y \in \mathbb{R}^n$, $\min(x, y) = (\min(x_1, y_1), \cdots, \min(x_n, y_n))^\top$. For a closed set $\mathcal{X} \subseteq \mathbb{R}^n$, denote its convex hull by $\mathrm{conv}(\mathcal{X})$, and the distance of $x \in \mathbb{R}^n$ to $\mathcal{X}$ by

$$\mathrm{dist}(x, \mathcal{X}) = \inf_{y \in \mathcal{X}} \|x - y\|.$$

For two closed sets $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$, let $\mathbb{D}(\mathcal{Y}, \mathcal{X}) := \sup_{y \in \mathcal{Y}} \mathrm{dist}(y, \mathcal{X})$ denote the deviation of $\mathcal{Y}$ from the set $\mathcal{X}$. Furthermore, the normal cone of a closed and convex set $\mathcal{X}$ at $x \in \mathcal{X}$ is denoted by

$$\mathcal{N}_{\mathcal{X}}(x) = \{v \in \mathbb{R}^n : \langle v, z - x \rangle \leq 0, \text{ for all } z \in \mathcal{X}\}.$$

Denote by $\mathcal{B}_\gamma(a)$ the k-dimensional closed ball of radius $\gamma > 0$ centered at a point $a \in \mathbb{R}^k$, i.e., $\mathcal{B}_\gamma(a) = \{x \in \mathbb{R}^k : \|x - a\| \leq \gamma\}$. We use $\mathcal{B}_\gamma^k$ to denote $\mathcal{B}_\gamma(0)$ for $0 \in \mathbb{R}^k$.

For a proper lower semicontinuos function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the domain is defined by $\mathrm{dom}\, \varphi = \{y : \varphi(y) < \infty\}$. For each $y \in \mathrm{dom}\, \varphi$, the regular subdifferential [39, Definition 8.3(a)] is defined by

$$\widehat{\partial}\, \varphi(y) := \left\{ v \in \mathbb{R}^n : \liminf_{z \to y, z \neq y} \frac{\varphi(z) - \varphi(y) - \langle v, z - y \rangle}{\|z - y\|} \geq 0 \right\}.$$

The limiting subdifferential [39, Definition 8.3(b)] at $y \in \operatorname{dom} \varphi$ is then given by

$$\bar{\partial}\, \varphi(y) := \left\{ v \in \mathbb{R}^n : \exists y_t \to y, \varphi(y_t) \to \varphi(y), v_t \in \widehat{\partial}\, \varphi(y) \to v \right\}.$$

For a locally Lipschitz continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$, the Clarke subdifferential [14, section 1.2] at $y^* \in \mathbb{R}^n$ is defined by

$$\partial \varphi(y^*) = \operatorname{conv} \left\{ \lim_{y \to y^*} \nabla \varphi(y) : \varphi \text{ is differentiable at } y \right\}.$$

We say that $x^* \in \mathcal{X}$ is an ϵ-Clarke stationary point of φ over a nonempty closed convex set $\mathcal{X}$ if

$$\operatorname{dist}\left(0, \partial \varphi(x^*) + \mathcal{N}_{\mathcal{X}}(x^*)\right) \leq \epsilon$$

for $\epsilon > 0$.

2.2 Kurdyka-Łojasiewicz property

First, recall the definition of Kurdyka-Łojasiewicz (KL) property.

Definition 2.1 (KL property and KL function)**.** *We say that a proper closed function $g : \mathbb{R}^n \to (-\infty, \infty]$ has the KL property at $\bar{w} \in \operatorname{dom} \bar{\partial} g$ if there exist a neighborhood $\mathcal{N}$ of $\bar{w}$, $\alpha \in (0, \infty]$ and a continuous concave function $\psi : [0, \alpha) \to \mathbb{R}_+$ with $\psi(0) = 0$ such that the following conditions hold.*

(i) ψ is continuously differentiable on $(0, \alpha)$ with $\psi' > 0$ over $(0, \alpha)$.

(ii) For $\forall w \in \mathcal{N}$ with $g(\bar{w}) < g(w) < g(\bar{w}) + \alpha$, one has

$$\psi'(g(w) - g(\bar{w}))\, \operatorname{dist}(0, \bar{\partial} g(w)) \geq 1.$$

Furthermore, if g satisfies the KL property at every point in $\operatorname{dom} \bar{\partial} g$, then it is called a KL function.

Definition 2.2 (KL exponent)**.** *For a proper closed function g satisfying the KL property at $\bar{w} \in dom\,\partial f$, if the corresponding function ψ can be chosen as $\psi(s) = \bar{c}s^{1-\gamma}$ for some $\bar{c} > 0$ and $\alpha \in [0, 1)$, i.e., there exist $c, \epsilon > 0$ and $\alpha \in (0, \infty]$ so that*

$$\operatorname{dist}(0, \bar{\partial} g(w)) \geq c(g(w) - g(\bar{w}))^{\gamma}.$$

where $\|w - \bar{w}\| \leq \epsilon$ and $g(\bar{w}) < g(w) < g(\bar{w}) + \alpha$, then we say that g has the KL property at $\bar{w}$ with an exponent of γ. If f is a KL function and has the same exponent γ at any $\bar{w} \in dom\,\partial f$, then we say that g is a KL function with an exponent of γ.

Remark 2.1. *A proper closed function f satisfies the KL property with any exponent $\gamma \in [0, 1]$ at any non-critical point $\bar{x} \in dom\,\partial f$ [29, Lemma 2.1]. Thus, to show that a proper closed function f is a KL function, it suffices to show that f satisfies the KL property at critical points.*

The KL property is a fundamental concept in optimization and variational analysis, especially in the study of nonsmooth and nonconvex problems. It provides a unifying framework to analyze the convergence behavior of iterative algorithms for a broad class of optimization problems. This property traced its origins to the work of Łojasiewicz in 1963 [32], who introduced a gradient inequality for analytic functions, and was later extended by Kurdyka in 1998 [25] to more general settings.

Originally introduced for real analytic functions, Łojasiewicz [32] showed that every critical point of a real analytic function f, satisfies a gradient inequality

$$\|\nabla f(x)\| \geq C|f(x) - f(x^*)|^{\gamma}, \tag{2.2.1}$$

where x^* is a critical point, $C > 0$, and $\gamma \in [0, 1)$ is a constant depending on the function and the critical point. This inequality became a cornerstone in proving the local convergence of gradient-based method near critical points for analytic functions. Then, Kurdyka [25] extended the gradient inequality to the class of

differentiable functions definable in an o-minimal structure. Subsequently, in [6], Bolte et al. explored and proposed the KL property for nonsmooth functions, including those encountered in modern optimization. The KL property has been instrumental in analyzing the convergence of various first-order methods including proximal algorithm [1, 24], difference-of-convex algorithm [45], proximal alternating linearized minimization algorithm [7], Douglas-Rachford splitting algorithm [28], alternating direction method of multipliers [23, 27, 50, 51], block prox-linear method [49], and random reshuffling method [30].

Furthermore, the KL exponent plays a key role in understanding the convergence rate of optimization algorithms, particularly in problems involving nonsmooth and nonconvex functions. For a certain optimization problem and a specific algorithm applied, we first construct a suitable potential function which is supposed to satisfy the KL property. Assume that $\{x^k\}$ is a bounded sequence generated by the algorithm. Furthermore, assume that somehow one can determine the KL exponent. Then, the following results on the convergence rate hold:

- If $\gamma = 0$, then $\{x^k\}$ converges finitely;
- If $\gamma \in (0, \frac{1}{2}]$, then $\{x^k\}$ converges locally linearly;
- If $\gamma \in (\frac{1}{2}, 1]$, then $\{x^k\}$ converges locally sublinearly.

Consequently, the KL exponent has garnered significant attention in recent research. We refer interested readers to [29, 36, 37, 52] for the calculus rules for computing explicitly the KL exponent.

Finally, we focus on semialgebraic functions, a large class of functions which are the KL functions. Recall that a set $\mathcal{C} \subset \mathbb{R}^n$ is called semialgebraic if it can be represented as a finite union of sets of the form

$$\{x \in \mathbb{R}^n : p_i(x) = 0, q_i(x) > 0, i = 1, ..., N\},$$

where p_i, q_i are real polynomial functions. A function is called semialgebraic if its graph is a semialgebraic set. Semialgebraic function satisfies the KL property [5, 6] with exponent $\gamma \in [0, 1)$. Some basic rules and properties for semialgebraic sets and functions are summarized as below.

1. For a semialgebraic set $\mathcal{C}$, its closure $\mathrm{cl}(\mathcal{C})$ is also semialgebraic.

2. Indicator functions of semialgebraic sets are semialgebraic.

3. Finite sums and products of semialgebraic functions are semialgebraic.

4. Scalar products are semialgebraic.

5. The composition of semialgebraic functions or mappings are semialgebraic.

6. A parametrized supremum (infimum) function $f(x) = \sup_{y \in \mathcal{Y}} g(x, y)$ (respectively, $f(x) = \inf_{y \in \mathcal{Y}} g(x, y)$) is semialgebraic if $\mathcal{Y}$ and g are semialgebraic.

Specifically, note that the ℓ_2-norm $\|x\|$ is semialgebraic [4].

Chapter 3

A smoothing block coordinate descent algorithm

In this chapter, to find approximate solutions of problem (1.1.1), we consider the probability space based on finitely many scenarios $\Xi = \{\xi_1, ..., \xi_N\}$, where each $\xi_i \in \mathbb{R}^l$ has a known probability $p_i > 0$ and these probabilities sum up to 1. Specifically, we first propose an exact penalization which partially penalize the constraints and then design a smoothing block coordinate descent (SBCD) algorithm for solving the penalty problem. Then, we show the convergence analysis of our SBCD algorithm. Finally, we conduct extensive numerical experiments for the portfolio selection problem.

Let $A_i = A(\xi_i), D_i = D(\xi_i), c_i = c(\xi_i), M_i = M(\xi_i), q_i = q(\xi_i)$ for $\xi_i \in \Xi$ and $i \in [N]$. Specifically, we suppose that the following assumption holds.

Assumption 3.1. *The sets*

$$\mathcal{X} := \{x \in \mathbb{R}^n : 0 \leq Bx + b \perp Hx + h \geq 0\}$$

$$\mathcal{U}_i := \{u_i \in \mathbb{R}^m : 0 \leq D_i u_i + c_i \perp M_i u_i + q_i \geq 0\}, \quad \forall\, i \in [N]$$

are not empty.

Some sufficient conditions for the solvability of vertical linear complementarity problems can be found in [21]. Moreover, some recent results on stochastic complementarity problems can be found in [43].

Then, the discrete form of problem (1.1.1) has the following form

$$\begin{array}{ll} \min\limits_{x,U} & f(x,U) := \sum\limits_{i=1}^{N} p_i \left\|A_i x - u_i\right\|^2 + \frac{\lambda}{2}\|x\|^2 \\ \text{s.t.} & x \in \mathcal{X},\ u_i \in \mathcal{U}_i,\ i \in [N], \end{array} \tag{3.0.1}$$

where $U = (u_1^\top, \cdots, u_N^\top)^\top \in \mathbb{R}^{mN}$. Let $\mathcal{S} = \prod\limits_{i=0}^{N} \mathcal{S}_i \subseteq \mathbb{R}^\nu$ with $\nu = n + mN$ and

$$\begin{array}{rcl} \mathcal{S}_0 & := & \{x \in \mathbb{R}^n : Bx + b \geq 0,\ Hx + h \geq 0\}; \\ \mathcal{S}_i & := & \{u_i \in \mathbb{R}^m : D_i u_i + c_i \geq 0,\ M_i u_i + q_i \geq 0\},\ \forall i \in [N]. \end{array}$$

We consider the following piecewise smooth penalty [34] optimization problem with linear constraints for (3.0.1)

$$\min_{(x;U)\in\mathcal{S}} \ \phi(x,U) := f(x,U) + \rho\left(\theta(x) + \sum_{i=1}^{N} r_i(u_i)\right) \tag{3.0.2}$$

with a penalty parameter $\rho > 0$, and

$$\begin{array}{rcl} \theta(x) & := & e^\top(Bx + b) - e^\top(Bx + b - Hx - h)_+, \\ r_i(u_i) & := & e^\top(D_i u_i + c_i) - e^\top(D_i u_i + c_i - M_i u_i - q_i)_+,\ \forall i \in [N], \end{array}$$

where e is the all-one vector with appropriate dimension and

$$(s)_+ = (\max(s_1, 0), \ldots, \max(s_k, 0))^\top$$

for a k-dimensional vector s.

We use the following smoothing function with $\mu > 0$ to approximate $(s)_+$:

$$\tilde{t}(s_i, \mu) = \frac{1}{2}\left(s_i + \sqrt{s_i^2 + 4\mu^2}\right), \quad i = 1, \ldots, k, \tag{3.0.3}$$

and define a smoothing function of $\phi(x,U)$ as follows,

$$\tilde{\phi}(x, U, \mu) := f(x,U) + \rho\tilde{\theta}(x,\mu) + \rho\sum_{i=1}^{N} \tilde{r}_i(u_i, \mu),$$

where $\tilde{\theta}$ and $\tilde{r}_i$ are smoothing functions for θ and r_i respectively.

3.1 Exact penalty

In this section, we show that (3.0.2) is an exact penalty reformulation of (3.0.1) whenever the penalty parameter ρ exceeds a certain positive number.

Let $z = [x^\top, U^\top]^\top$ and

$$\mathcal{C} = \left\{ z \in \mathbb{R}^\nu \,:\, \theta(x) + \sum_{i=1}^{N} r_i(u_i) \le 0 \right\}.$$

The feasible set $\mathcal{F} := \mathcal{X} \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_N$ of problem (3.0.1) is the intersection of two sets $\mathcal{S}$ and $\mathcal{C}$, namely, $\mathcal{F} = \mathcal{S} \bigcap \mathcal{C}$. By Assumption 3.1, there is $z^0 \in \mathcal{F}$. Let

$$\gamma = \sqrt{f(z^0)} \left(\sqrt{\frac{2}{\lambda}} + \sum_{i=1}^{N} \sqrt{\frac{1}{p_i}} + \|A_i\| \sqrt{\frac{2}{\lambda}} \right).$$

Lemma 3.1. *The solution set of problem* (3.0.1) *is not empty and contained in* $\mathcal{B}_\gamma^\nu$.

Proof. Since the objective function f is continuous and the feasible set $\mathcal{F}$ is not empty, the solution set of problem (3.0.1) is not empty and contained in the level set $\mathcal{D}(f) = \{z \in \mathbb{R}^\nu : f(z) \le f(z^0)\}$.

Now we show $\mathcal{D}(f) \subseteq \mathcal{B}_\gamma^\nu$. Note that $z \in \mathcal{D}(f)$ implies that $\frac{\lambda}{2}\|x\|^2 \le f(z^0)$, i.e., $\|x\| \le \sqrt{\frac{2f(z^0)}{\lambda}}$. Moreover, $z \in \mathcal{D}(f)$ also implies that $p_i \|A_i x - u_i\|^2 \le f(z^0)$, i.e., $\|A_i x - u_i\| \le \sqrt{\frac{f(z^0)}{p_i}}$, which implies

$$\sum_{i=1}^{N} \|u_i\| \le \sum_{i=1}^{N} \|u_i - A_i x\| + \|A_i x\| \le \sqrt{f(z^0)} \sum_{i=1}^{N} \left(\sqrt{\frac{1}{p_i}} + \|A_i\| \sqrt{\frac{2}{\lambda}} \right).$$

Hence, we obtain

$$\|z\| \le \|x\| + \sum_{i=1}^{N} \|u_i\| \le \sqrt{\frac{2f(z^0)}{\lambda}} + \sqrt{f(z^0)} \sum_{i=1}^{N} \left(\sqrt{\frac{1}{p_i}} + \|A_i\| \sqrt{\frac{2}{\lambda}} \right).$$

We complete the proof. □

Let $\mathcal{K} = \{z \in \mathbb{R}^\nu : \|z\|_\infty \leq \gamma\}$. Obviously, $\mathcal{S} \cap \mathcal{K}$ is a nonempty bounded polyhedral convex set and $\mathcal{S} \cap \mathcal{B}_\gamma^\nu \subseteq \mathcal{S} \cap \mathcal{K}$. Note that θ and $r_i, i \in [N]$ are piecewise linear concave functions. From Theorem 5 in [26], we obtain the following lemma.

Lemma 3.2. *There is a $\tau > 0$ such that*

$$\operatorname{dist}(z, \mathcal{F} \cap \mathcal{K}) \leq \tau \left(\theta(x) + \sum_{i=1}^{N} r_i(u_i) \right)$$

for all $z \in \mathcal{S} \cap \mathcal{K}$.

We are now ready to show the equivalence between (3.0.1) and (3.0.2) regarding global minimizers.

Theorem 3.1. *The following statements hold.*

(i) There is a $\rho^* > 0$ such that any global minimizer z^* of (3.0.1) *is a global minimizer z^* of* (3.0.2) *whenever $\rho \geq \rho^*$.*

(ii) If z^* is a global minimizer of (3.0.2) *for some $\rho \geq \rho^*$, then z^* is a global minimizer of* (3.0.1).

Proof. The objective function f of problem (3.0.1) is continuously differentiable. Let L_f be a Lipschitz constant of f over the compact set $\mathcal{S} \cap \mathcal{K}$.

From Lemma 3.1 and $\mathcal{B}_\gamma^\nu \subseteq \mathcal{K}$, the set of global minimizers of problem (3.0.1) and the set of global minimizers of the following problem

$$\min\ f(z), \quad \text{s.t.} \quad z \in \mathcal{C} \cap \mathcal{S} \cap \mathcal{K} \tag{3.1.1}$$

are equal.

From Lemma 3.2 and Lemma 3.1 in [11], the following problem

$$\begin{array}{ll} \min & f(z) + \rho \left(\theta(x) + \sum\limits_{i=1}^{N} r_i(u_i) \right) \\ \text{s.t.} & z \in \mathcal{S} \cap \mathcal{K} \end{array} \tag{3.1.2}$$

with some $\rho \geq \rho^* := \tau L_f$ is an exact penalty reformulation of (3.1.1), where τ is the error bound parameter in Lemma 3.2. Therefore, any global minimizer of (3.0.1) is a global minimizer of (3.1.2), whenever $\rho \geq \rho^*$, and any global minimizer of (3.1.2) for some $\rho \geq \rho^*$ is a global minimizer of (3.0.1).

From $z^0 \in \mathcal{F}$, we have $\theta(x^0) + \sum_{i=1}^{N} r_i(u_i^0) = 0$ and thus the level set of the objective function of (3.1.2) is contained in the level set of the objective function of (3.0.1) with $f(z^0)$. Hence

$$\left\{ z \in \mathbb{R}^\nu \,:\, f(z) + \rho(\theta(x) + \sum_{i=1}^{N} r_i(u_i)) \leq f(z^0) \right\} \subseteq \mathcal{D}(f) \subseteq \mathcal{B}_\gamma^\nu \subseteq \mathcal{K},$$

which implies that the set of global minimizers of problem (3.1.2) and the set of global minimizers of the following problem

$$\min_{z \in \mathcal{S}} \; f(z) + \rho \left(\theta(x) + \sum_{i=1}^{N} r_i(u_i) \right),$$

that is problem (3.0.2), are equal.

We complete the proof. □

We next present a result on the local minimizers of problems (3.0.1) and (3.0.2), which can be proved using the similar proof of Theorem 3.1.

Corollary 3.1. *Suppose that z^* is a local minimizer of* (3.0.1)*. Then there exists a $\rho^* > 0$ such that z^* is a local minimizer of* (3.0.2) *whenever $\rho > \rho^*$.*

Proof. Suppose z^* is a local minimizer of (3.0.1). Then there is $\gamma_1 > 0$ such that z^* is a global minimizer of (3.0.1) in $\mathcal{F} \cap \mathcal{K}_1$, where $\mathcal{K}_1 := \{z \in \mathbb{R}^\nu : \|z - z^*\|_\infty \leq \gamma_1\}$. Similar to Lemma 3.2, there is a $\tau_1 > 0$ such that

$$\operatorname{dist}(z, \mathcal{F} \cap \mathcal{K}_1) \leq \tau_1 \left(\theta(x) + \sum_{i=1}^{N} r_i(u_i) \right)$$

for all $z \in \mathcal{S} \cap \mathcal{K}_1$. Then, following the proof of Theorem 3.1, we can show that there exists a $\rho^* > 0$ such that z^* is a local minimizer of (3.0.2) whenever $\rho > \rho^*$. □

Remark 3.1. *The penalty parameter ρ depends on the error bound parameter τ and the Lipschitz constant L_f. It is extremely hard to estimate the threshold value ρ^*. Therefore, in practice, we just set ρ to be large enough according to the residual of our final solution. Trivially, we can show that*

$$\phi(z) = f(z) + \rho_0\theta(x) + \rho_1 \sum_{i=1}^{N} r_i(u_i) \tag{3.1.3}$$

is also an exact penalty reformulation of (3.0.1) *whenever* $\min(\rho_0, \rho_1) \geq \tau L_f$. *Therefore, we can adjust the penalty parameters in a more flexible way.*

3.2 Algorithm design and analysis

In this section, we define a smoothing function of ϕ and prove that it satisfies the Kurdyka-Łojasiewicz (KL) property. Moreover we present a smoothing block coordinate descent (SBCD) algorithm for the exact penalty problem (3.0.2) of problem (3.0.1) based on the block structures of its constraints. We prove that the sequence generated by the SBCD algorithm globally converges to an ϵ-Clarke stationary point of problem (3.0.2) by using the KL property for any $\epsilon > 0$.

3.2.1 Algorithm description

In this subsection, we present the SBCD algorithm which combines block coordinate descent (BCD) methods and smoothing methods. First, we use smoothing functions $\tilde{\theta}(\cdot, \mu)$ and $\tilde{r}_i(\cdot, \mu)$ to approximate $\theta(\cdot)$ and $r_i(\cdot), i \in [N]$ where $\mu > 0$ is the smoothing parameter. Then, we use the BCD method to solve $N + 1$ smooth strongly convex problems for a fixed $\mu > 0$. Moreover, we update the smoothing parameter based on the reduction of the smoothing objective function values.

We illustrate how to obtain a smoothing function $\tilde{\theta}$ of $\theta(\cdot)$ and give its relevant properties. Similar definitions and properties hold for each $r_i(\cdot)$ and $\tilde{r}_i$, $i \in [N]$. Before processing, we give the definition of the smoothing function.

Definition 3.1. *We call $\tilde{\theta} : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ a smoothing function of the piecewise linear concave function θ in* (3.0.2)*, if $\tilde{\theta}(\cdot, \mu)$ is continuously differentiable in $\mathbb{R}^n$ for any fixed $\mu > 0$ and $\lim_{x \to \bar{x}, \mu \downarrow 0} \tilde{\theta}(x, \mu) = \theta(\bar{x}), \forall \bar{x} \in \mathcal{S}_0$.*

We use the smoothing function in (3.0.3) to construct a smoothing function of the plus function $(s)_+$ and define a smoothing function of θ as follows

$$\tilde{\theta}(x, \mu) = e^\top (Bx + b) - \sum_{j=1}^{n_1} \tilde{t} \left([Bx + b - Hx - h]_j, \mu \right). \tag{3.2.1}$$

Similarly,

$$\tilde{r}_i(u_i, \mu) = e^\top (D_i u_i + c_i) - \sum_{j=1}^{m_1} \tilde{t} \left([D_i u_i + c_i - M_i u_i - q_i]_j, \mu \right). \tag{3.2.2}$$

From [9, 10], we have the following proposition.

Proposition 3.1. *The function $\tilde{\theta}$* (3.2.1) *with $\tilde{t}$ defined in* (3.0.3) *is a smoothing function of θ and has the following properties.*

(i) $\tilde{\theta}(x, \cdot)$ is strictly decreasing for any fixed $x \in \mathbb{R}^n$;

(ii) $\tilde{\theta}(\cdot, \mu)$ is concave for any fixed $\mu > 0$;

(iii) $\mathrm{conv}\{\lim_{\mu \downarrow 0} \nabla \tilde{\theta}(\bar{x}, \mu)\} = \partial \theta(\bar{x}), \forall \bar{x} \in \mathcal{S}_0$;

(iv) $\tilde{\theta}(x, \cdot)$ is Lipschitz continuous with the Lipschitz constant $\kappa_0 = n_1$, namely, $|\tilde{\theta}(x, \mu_1) - \tilde{\theta}(x, \mu_2)| \leq \kappa_0 |\mu_1 - \mu_2|, \quad \forall x \in \mathbb{R}^n$;

(v) $\nabla\tilde{\theta}(\cdot,\mu)$ *is Lipschitz continuous with the Lipschitz constant* $\frac{L_0}{\mu}$ *and* $L_0 \geq n_1(\|B\|_1 + \|H\|_1)$ *for any fixed* $\mu > 0$.

For (i), (ii) and (iv), see [9, Proposition 2.1]. For (iii), see [10, Theorem 1]. Finally, (v) can be shown by definition (3.2.1) with (3.0.3).

Similarly, we can easily show that the function $\tilde{r}_i$ (3.2.2) is a smoothing function of $r_i,\ \forall i \in [N]$, with properties in Proposition 3.1. We use $\frac{L_i}{\mu}$ to denote the Lipschitz constant $\nabla\tilde{r}_i(\cdot,\mu)$ with respect to u_i, and $\kappa_i = m_1$ the Lipschitz constant for each $\tilde{r}_i$ with respect to μ. By Definition 3.1 and Proposition 3.1, we have

$$\begin{aligned} 0 \leq \theta(x) - \tilde{\theta}(x,\mu) &\leq \kappa_0\mu \quad \forall x \in \mathbb{R}^n; \\ 0 \leq r_i(u_i) - \tilde{r}_i(u_i,\mu) &\leq \kappa_i\mu, \quad \forall u_i \in \mathbb{R}^m,\ \forall i \in [N]. \end{aligned}$$

When it is clear from context, we simply denote the gradient of $\tilde{\theta}(\cdot,\mu)$ and $\tilde{r}_i(\cdot,\mu)$ at x and u_i as $\nabla\tilde{\theta}(x,\mu)$ and $\nabla\tilde{r}_i(u_i,\mu)$, respectively.

Now we present the SBCD algorithm for problem (3.0.2).

Algorithm 1 The SBCD algorithm for (3.0.2)

Require: Initialize $(x^0, U^0) \in \mathcal{F}$, $\mu_0 = \mu_{-1} > 0$, $\epsilon > 0$, $\beta > 0$, $\sigma \in (0,1)$, and set $k = 0$.

Ensure: (x^k, U^k).

1: **while** a termination criterion is not met **do**

2: Update x:

$$x^{k+1} = \arg\min_{x \in \mathcal{S}_0} f(x, U^k) + \rho \left(\nabla \tilde{\theta}(x^k, \mu_k)\right)^\top x + \frac{\beta}{2\mu_k} \left\|x - x^k\right\|^2. \tag{3.2.3}$$

3: Update u_i for all $i \in [N]$:

$$u_i^{k+1} = \arg\min_{u_i \in \mathcal{S}_i} p_i \left\|A_i x^{k+1} - u_i\right\|^2 + \rho \left(\nabla \tilde{r}_i(u_i^k, \mu_k)\right)^\top u_i + \frac{\beta}{2\mu_k} \left\|u_i - u_i^k\right\|^2. \tag{3.2.4}$$

4: If

$$\tilde{\phi}(x^{k+1}, U^{k+1}, \mu_k) + \kappa\mu_k - \tilde{\phi}(x^k, U^k, \mu_{k-1}) - \kappa\mu_{k-1} \leq -\frac{\epsilon^2}{2\beta}\mu_k, \tag{3.2.5}$$

set $\mu_{k+1} = \mu_k$; Otherwise, set

$$\mu_{k+1} = \sigma\mu_k. \tag{3.2.6}$$

5: Let $k \leftarrow k + 1$ and go to **step 1**.

6: **end while**

Remark 3.2. *At each step of Algorithm 1, we apply the SBCD method [48, 49] to the smoothing problem*

$$\min_{(x;U) \in \mathcal{S}} \tilde{\phi}(x, U, \mu_k) \tag{3.2.7}$$

and update the variables (x, U) in the Gauss-Seidel fashion. The subproblems (3.2.3) *and* (3.2.4) *are strongly convex quadratic programs and can be efficiently solved by many software packages, e.g., MOSEK, IPOPT, OSQP,* ***quadprog*** *in Matlab. Moreover, we update the smoothing parameter μ by monitoring the reduction of the value of the smoothing function with the smoothing parameter and the reducing scheme* (3.2.6).

The following proposition on gradient consistence of $\tilde{\phi}$ will be used in the conver-

gence analysis of the SBCD algorithm in the next subsection.

Proposition 3.2. *For any $(\bar{x}, \bar{U}) \in \mathcal{S}$, we have*

$$\left\{ \lim_{(x,U)\to(\bar{x},\bar{U}),\mu\downarrow 0} \nabla\tilde{\phi}(x, U, \mu) \right\} \subseteq \partial\phi(\bar{x}, \bar{U}). \tag{3.2.8}$$

Proof. Note that the functions $-\theta, -r_i, \forall i \in [N]$ are piecewise linear convex and therefore regular. By [8, Corollary 1], we have

$$\begin{aligned} &\left\{ \lim_{x\to\bar{x},\mu\downarrow 0} \nabla\tilde{\theta}(x, \mu) \right\} \subseteq \partial\theta(\bar{x}), \\ &\left\{ \lim_{u_i\to\bar{u}_i,\mu\downarrow 0} \nabla\tilde{r}_i(u_i, \mu) \right\} \subseteq \partial r_i(\bar{u}_i),\ i \in [N]. \end{aligned} \tag{3.2.9}$$

Since f is continuously differentiable, and variables of θ and r_i for $i \in [N]$ are separable, we obtain (3.2.8) from (3.2.9) and [14, Corollary 2, p. 39]. □

3.2.2 Convergence analysis

In this subsection, we show the convergence analysis for the SBCD algorithm. Throughout this subsection, let $\{x^{k+1}, U^{k+1}, \mu_k\}$ be the sequences generated by Algorithm 1. For simplicity, let $z^k := [(x^k)^\top, (U^k)^\top]^\top$ when it is appropriate.

First, we estimate the decrease of the objective function in (3.2.7) as the number of iterations increases.

Lemma 3.3. *For any $k \in \mathbb{N}$, there holds*

$$\tilde{\phi}(z^k, \mu_k) - \tilde{\phi}(z^{k+1}, \mu_k) \geq \frac{\beta}{2\mu_k} \left\| z^{k+1} - z^k \right\|^2. \tag{3.2.10}$$

Proof. First, it follows from Proposition 3.1-(ii) that $\tilde{\theta}(\cdot, \mu)$ is concave, namely, it holds that

$$\tilde{\theta}(x^k, \mu_k) - \tilde{\theta}(x^{k+1}, \mu_k) \geq \left\langle \nabla\tilde{\theta}(x^k, \mu_k), x^k - x^{k+1} \right\rangle. \tag{3.2.11}$$

Since x^{k+1} is the minimizer of (3.2.3), we have

$$f(x^k, U^k) - f(x^{k+1}, U^k) \geq \rho \left\langle \nabla\tilde{\theta}(x^k, \mu_k), x^{k+1} - x^k \right\rangle + \frac{\beta}{2\mu_k} \left\| x^{k+1} - x^k \right\|^2. \quad (3.2.12)$$

Together with (3.2.11) and (3.2.12), we have

$$\begin{aligned} & \tilde{\phi}(x^k, U^k, \mu_k) - \tilde{\phi}(x^{k+1}, U^k, \mu_k) \\ = \;& f(x^k, U^k) + \rho\tilde{\theta}(x^k, \mu_k) - f(x^{k+1}, U^k) - \rho\tilde{\theta}(x^{k+1}, \mu_k) \\ \geq \;& \rho \left\langle \nabla\tilde{\theta}(x^k, \mu_k), x^{k+1} - x^k \right\rangle + \frac{\beta}{2\mu_k} \left\| x^{k+1} - x^k \right\|^2 \\ & + \rho \left(\tilde{\theta}(x^k, \mu_k) - \tilde{\theta}(x^{k+1}, \mu_k) \right) \\ \geq \;& \frac{\beta}{2\mu_k} \left\| x^{k+1} - x^k \right\|^2. \end{aligned} \quad (3.2.13)$$

Similarly, by the concavity of $\tilde{r}_i$ and the fact that u_i^{k+1} is the minimizer of (3.2.4), we can show that for $\forall i \in [N]$

$$\begin{aligned} & p_i \left\| A_i x^{k+1} - u_i^k \right\|^2 + \rho\tilde{r}_i(u_i^k, \mu_k) - \left[p_i \left\| A_i x^{k+1} - u_i^{k+1} \right\|^2 + \rho\tilde{r}_i(u_i^{k+1}, \mu_k) \right] \\ & \geq \frac{\beta}{2\mu_k} \left\| u_i^{k+1} - u_i^k \right\|^2. \end{aligned} \quad (3.2.14)$$

Summing up inequalities (3.2.14) for all $i \in [N]$, we have

$$\tilde{\phi}(x^{k+1}, U^k, \mu_k) - \tilde{\phi}(x^{k+1}, U^{k+1}, \mu_k) \geq \frac{\beta}{2\mu_k} \left\| U^{k+1} - U^k \right\|^2. \quad (3.2.15)$$

Finally, sums up inequalities (3.2.15) and (3.2.13). Then, the statement is established.

□

Next we show the convergence of the sequence $\{\phi(z^k)\}$ of the objective function values in (3.0.2), by utilizing the following auxiliary function

$$\mathcal{E}(z, \mu) := \tilde{\phi}(z, \mu) + \kappa\mu$$

with $\kappa := \rho(n_1 + Nm_1)$. From (iv) of Proposition 3.1, we have

$$\left|\tilde{\phi}(z, \mu_1) - \tilde{\phi}(z, \mu_2)\right| \leq \kappa|\mu_1 - \mu_2| \tag{3.2.16}$$

for any $z \in \mathbb{R}^{n+mN}$ and $\mu_1 > 0, \mu_2 > 0$.

Lemma 3.4. *The following statements hold.*

(i) The sequence of smoothing parameters converges to 0, i.e., $\lim_{k\to\infty} \mu_k = 0$*;*

(ii) The sequence $\{\mathcal{E}(z^{k+1}, \mu_k)\}$ *is convergent;*

(iii) The sequence $\{z^k\}$ *is bounded.*

Proof. (i) It is not hard to observe that the sequence $\{\mu_k\}$ is nonincreasing and bounded, and therefore converges. Next, we prove the statement by contradiction, i.e., assume that $\lim\limits_{k\to\infty} \mu_k = \hat{\mu} > 0$. It implies that (3.2.6) happens at most finite times. Then, there must exist a large enough $K \in \mathbb{N}$ such that $\mu_k \equiv \hat{\mu}$ for any $k \geq K$, and moreover,

$$\mathcal{E}(z^{k+1}, \mu_k) - \mathcal{E}(z^k, \mu_{k-1}) \leq -\frac{\epsilon^2}{2\beta}\hat{\mu},$$

which implies that $\lim\limits_{k\to\infty} \mathcal{E}(z^{k+1}, \mu_k) = -\infty$. However, by (i) of Proposition 3.1, we have

$$\mathcal{E}(z^{k+1}, \mu_k) \geq \phi(z^{k+1}) \geq 0. \tag{3.2.17}$$

Hence, the assumption leads to a contradiction. This finishes the proof of (i).

(ii) By (3.2.17), we know that the sequence $\{\mathcal{E}(z^{k+1}, \mu_k)\}$ is lower bounded. Next, we show that it is monotonically nonincreasing.

Recall that from Lemma 3.3 we have

$$\tilde{\phi}(z^k, \mu_k) \geq \tilde{\phi}(z^{k+1}, \mu_k) + \frac{\beta}{2\mu_k}\|z^{k+1} - z^k\|^2. \tag{3.2.18}$$

By (i) and (vi) of Proposition 3.1, we have

$$0 \leq \tilde{\phi}(z^k, \mu_k) - \tilde{\phi}(z^k, \mu_{k-1}) \leq \kappa(\mu_{k-1} - \mu_k). \tag{3.2.19}$$

Together with (3.2.18) and (3.2.19), we have

$$\mathcal{E}(z^k, \mu_{k-1}) - \mathcal{E}(z^{k+1}, \mu_k) \geq \frac{\beta}{2\mu_k}\|z^{k+1} - z^k\|^2, \tag{3.2.20}$$

which implies that the statement (ii) is true.

(iii) It follows from (3.2.17) and the nonincreasing of $\{\mathcal{E}(z^{k+1}, \mu_k)\}$ that we have

$$\phi(z^{k+1}) \leq \mathcal{E}(z^{k+1}, \mu_k) \leq \mathcal{E}(z^1, \mu_0).$$

Therefore, by $z^k \in \mathcal{S}$ and

$$\{z \in \mathbb{R}^\nu \,:\, \phi(z) \leq \mathcal{E}(z^1, \mu_0)\} \subset \{z \in \mathbb{R}^\nu \,:\, f(z) \leq \mathcal{E}(z^1, \mu_0)\},$$

we conclude that $\{z^k\}$ is bounded from the boundedness of the level set of f. See the proof of Lemma 3.1. □

Theorem 3.2. *The sequence $\{\phi(z^k)\}$ generated by the SBCD algorithm is convergent.*

Proof. By (i) of Proposition 3.1, for any fixed $k \geq 1$, we have

$$0 \leq \phi(z^{k+1}) - \tilde{\phi}(z^{k+1}, \mu_k) \leq \kappa\mu_k,$$

which implies that

$$\mathcal{E}(z^{k+1}, \mu_k) - \kappa\mu_k \leq \phi(z^{k+1}) \leq \mathcal{E}(z^{k+1}, \mu_k).$$

It then follows from Lemma 3.4 that we have

$$\lim_{k\to\infty} \mathcal{E}(z^{k+1}, \mu_k) = \lim_{k\to\infty} \phi(z^{k+1}).$$

This finishes the proof. □

Next, we show that there exists a subsequence of $\{z^k\}$ such that its any accumulation point is an ϵ-Clarke stationary point of (3.0.2), which lays a foundation for the sequence convergence of $\{z^k\}$.

Lemma 3.5. *Let $\mathcal{N}^s := \{k : \mu_{k+1} \neq \mu_k\}$. For any $\epsilon > 0$, every accumulation point of $\{z^k : k \in \mathcal{N}^s\}$ is an ϵ-Clarke stationary point of* (3.0.2).

Proof. By Lemma 3.4, $\{z^k : k \in \mathcal{N}^s\}$ is bounded. Let $z^* := [(x^*)^\top, (U^*)^\top]^\top$ be an accumulation point of $\{z^k : k \in \mathcal{N}^s\}$ with a convergent subsequence $\{z^k\}_{k\in\mathfrak{K}}$ and $\mathfrak{K} \subset \mathcal{N}^s$. Since (3.2.5) fails for $k \in \mathcal{N}^s$, we have

$$\tilde{\phi}(z^{k+1}, \mu_k) + \kappa\mu_k - \tilde{\phi}(z^k, \mu_{k-1}) - \kappa\mu_{k-1} > -\frac{\epsilon^2}{2\beta}\mu_k.$$

Considering (3.2.20), we then have $\frac{\beta}{2\mu_k}\|z^{k+1} - z^k\|^2 \leq \frac{\epsilon^2}{2\beta}\mu_k$ for $k \in \mathfrak{K}$, which together with Lemma 3.4 -(i) implies that

$$\frac{\beta\|z^{k+1} - z^k\|}{\mu_k} \leq \epsilon \quad \text{for } k \in \mathfrak{K}, \quad \text{and} \quad \lim_{\substack{k\in\mathfrak{K}\\ k\to\infty}} z^{k+1} = z^*. \tag{3.2.21}$$

Then, according to the first-order necessary optimality condition of subproblem (3.2.3), we have

$$-\left[\nabla_x f(x^{k+1}, U^k) + \rho\nabla\tilde{\theta}(x^k, \mu_k) + \frac{\beta}{\mu_k}(x^{k+1} - x^k)\right] \in \mathcal{N}_{\mathcal{S}_0}(x^{k+1}),\ \forall k \in \mathfrak{K}. \tag{3.2.22}$$

Similarly, for subproblem (3.2.4) with any $i \in [N]$, we have that for $\forall k \in \mathfrak{K}$

$$-\left[2p_i(u_i^{k+1} - A_i x^{k+1}) + \rho\nabla\tilde{r}_i(u_i^k, \mu_k) + \frac{\beta}{\mu_k}(u_i^{k+1} - u_i^k)\right] \in \mathcal{N}_{\mathcal{S}_i}(u_i^{k+1}). \tag{3.2.23}$$

Together with (3.2.22) and (3.2.23), we obtain

$$-G_k + \frac{\beta}{\mu_k}(z^k - z^{k+1}) \in \mathcal{N}_{\mathcal{S}}(z^{k+1}),\ \forall k \in \mathfrak{K}, \tag{3.2.24}$$

with

$$G_k = \begin{bmatrix} \nabla_x f(x^{k+1}, U^k) + \rho \nabla \tilde{\theta}(x^k, \mu_k) \\ 2p_1(u_1^{k+1} - A_i x^{k+1}) + \rho \nabla \tilde{r}_1(u_1^k, \mu_k) \\ \vdots \\ 2p_N(u_N^{k+1} - A_i x^{k+1}) + \rho \nabla \tilde{r}_N(u_N^k, \mu_k) \end{bmatrix}.$$

By (3.2.21) and the gradient consistency of the smoothing function $\tilde{\phi}$ in Proposition 3.2, we have

$$\lim_{\substack{k \in \mathfrak{K} \\ k \to \infty}} G_k \in \partial \phi(z^*).$$

Finally, using the outer semicontinuity of normal cone [39], (3.2.21), and the gradient consistency of the smoothing function $\tilde{\phi}$ in Proposition 3.2, and taking limits on both sides of (3.2.24) as $k \in \mathfrak{K}$ and $k \to \infty$, we obtain

$$\operatorname{dist}(0, \partial \phi(z^*) + \mathcal{N}_{\mathcal{S}}(z^*)) \leq \epsilon.$$

We complete the proof. □

Remark 3.3. *With inequality* (3.2.5)*, we can give Lemma 3.7 which presents an upper bound for the sum of* μ_k*. This upper bound will be used in Theorem 3.3 to prove the KL property and the whole sequence convergence. On the other hand, from the proof of Lemma 3.5, it is easy to verify that if (3.2.5) in Algorithm 1 is changed to*

$$\tilde{\phi}(x^{k+1}, U^{k+1}, \mu_k) + \kappa \mu_k - \tilde{\phi}(x^k, U^k, \mu_{k-1}) - \kappa \mu_{k-1} \leq -\frac{\mu_k^2}{2\beta}, \tag{3.2.25}$$

that is, set $\epsilon = \sqrt{\mu_k}$*, then the statement of Lemma 3.5 has the following version "Any accumulation point of* $\{z^k : k \in \mathcal{N}^s\}$ *is a Clarke stationary point of* (3.0.2)*."*

Finally, by the KL property, we show the sequence generated by the SBCD algorithm converges to an ϵ-Clarke stationary point. First, we prove that certain functions are semialgebraic and therefore KL functions.

Lemma 3.6. *The function $\tilde{\theta}$ of (x, μ) as defined in* (3.2.1) *is a KL function.*

Proof. It is trivial to show that

$$\begin{aligned}\tilde{\theta}(x,\mu) &= \frac{1}{2}\, e^\top (Bx + b + Hx + h) - \frac{1}{2}\sum_{j=1}^{n_1} \left\| \begin{bmatrix} (B_j - H_j)^\top x + b_j - h_j \\ 2\mu \end{bmatrix} \right\| \\ &= \frac{1}{2}\, e^\top (Bx + b + Hx + h) - \frac{1}{2}\sum_{j=1}^{n_1} \left\| \begin{bmatrix} B_j - H_j & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ \mu \end{bmatrix} + \begin{bmatrix} b_j - h_j \\ 0 \end{bmatrix} \right\|\end{aligned}$$

with B_j, H_j denoting the j-th row of the matrices B and H respectively, and b_j, h_j denoting the j-th element of the vectors b and h respectively. First, note that the Euclidean norm $\|\cdot\|$ is semialgebraic [4, 48]. Also it is well-known that the composition of semialgebraic functions is semialgebraic. Moreover, it is well-known that finite sums of semialgebraic functions is semialgebraic. Therefore, we conclude that the function $\tilde{\theta}$ is semialgebraic and therefore a KL function. □

Corollary 3.2. *Let $F(x, U, \mu) := \tilde{\phi}(x, U, \mu) + \kappa\mu + \delta_{\mathcal{S}_0}(x) + \sum_{i=1}^{N} \delta_{\mathcal{S}_i}(u_i)$. Then F is semialgebraic and therefore a KL function.*

Proof. In a similar way to the proof of Lemma 3.6, we can show that each function $\tilde{r}_i$ of (u_i, μ) is semialgebraic. Therefore, the function $\tilde{\phi}$ of (x, U, μ) is semialgebraic. Note that indicator function of a polyhedral set is semialgebraic. Hence, we conclude that the function F is semialgebraic and therefore a KL function. □

In what follows, let $z^* = [(x^*)^\top, (U^*)^\top]^\top$ be an arbitrary accumulation point of $\{z^k : k \in \mathcal{N}^s\}$ with a convergent subsequence $\{z^k\}_{k\in\mathfrak{K}}$ and $\mathfrak{K} \subset \mathcal{N}^s$, $w^k = [(z^k)^\top, \mu_{k-1}]^\top$ and $w^* = [(z^*)^\top, 0]^\top$. For simplicity, we set $F_k := F(w^k)$ and $F^* := F(w^*)$.

Let $L_{\nabla f}$ be the Lipschitz constant of ∇f, $\frac{L_0}{\mu}$ and $\frac{L_i}{\mu}$ be the Lipschitz constants of $\nabla\tilde{\theta}(x, \mu)$ and $\nabla\tilde{r}_i(u_i, \mu)$, with $\mu > 0$, respectively, and let $L_p = \rho \max\{L_0, L_1, \cdots, L_N\}$. First, let us give the following lemma.

Lemma 3.7. *The following statement holds:*

$$\sum_{k=0}^{\infty} \mu_k \leq \frac{\mu_0}{1-\sigma} + \frac{2\beta}{\epsilon^2}(F_0 - F^*).$$

Proof. First note that

$$\sum_{k \in \mathcal{N}^s}^{\infty} \mu_k = \sum_{j=0}^{\infty} \mu_0 \sigma^j = \frac{\mu_0}{1-\sigma}.$$

Since $\{x^k\} \subset \mathcal{S}_0$ and $\{u_i^k\} \subset \mathcal{S}_i$, we have $F_k = \mathcal{E}(z^k, \mu_{k-1})$.

When $k \notin \mathcal{N}^s$, (3.2.5) gives that

$$\frac{\epsilon^2}{2\beta}\mu_k \leq F_k - F_{k+1}.$$

According to the nonincreasing property of $\{F_k\}$ and Lemma 3.4-(ii), we conclude that

$$\sum_{k \notin \mathcal{N}^s}^{\infty} \mu_k \leq \frac{2\beta}{\epsilon^2}(F_0 - F^*).$$

Overall, we have

$$\sum_{k=0}^{\infty} \mu_k \leq \frac{\mu_0}{1-\sigma} + \frac{2\beta}{\epsilon^2}(F_0 - F^*),$$

which completes the proof. □

Lemma 3.5 shows a subsequence convergence of $\{z^k\}$ to an ϵ-Clarke stationary point of problem (3.0.2). Next, in Theorem 3.3, by the KL property together with Lemma 3.5, we can further prove that $\{z^k\}$ is a Cauchy sequence and, therefore, has whole sequence convergence to an ϵ-Clarke stationary point of problem (3.0.2).

Theorem 3.3. *For any $\epsilon > 0$, the sequence $\{z^k\}$ generated by Algorithm 1 converges to an ϵ-Clarke stationary point of problem* (3.0.2).

Proof. Since $\{x^k\} \subset \mathcal{S}_0$ and $\{u_i^k\} \subset \mathcal{S}_i$, by (3.2.20), we have

$$F_k - F_{k+1} = \mathcal{E}(z^k, \mu_{k-1}) - \mathcal{E}(z^{k+1}, \mu_k) \geq \frac{\beta}{2\mu_k}\|z^k - z^{k+1}\|^2. \tag{3.2.26}$$

Moreover, from Lemma 3.4, inequality (3.2.19) and Theorem 3.2, we have

$$F^* \leq F_k, \qquad \lim_{k\to\infty} \mu_k = 0 \quad \text{and} \quad \lim_{k\to\infty} F_k = F^*. \tag{3.2.27}$$

By Corollary 3.2, F is a KL function. Denote by $\mathcal{N}$ of the neighborhood of w^*, α and $\psi : [0, \alpha) \to \mathbb{R}_+$ the associated objects in Definition 2.1 relative to F and w^*.

Since z^* is an accumulation point of $\{z^k : k \in \mathcal{N}^s\}$, from (3.2.26) and (3.2.27), there are k_0 and ζ such that for $\forall k \geq k_0$, $\mathcal{B}_\zeta(w^*) \subset \mathcal{N}$, $F^* < F_k < F^* + \alpha$,

$$\Gamma_2\mu_{k_0}+\Gamma_1\psi(F_{k_0}-F^*)+2\sqrt{\frac{2\mu_{k_0}(F_{k_0}-F^*)}{\beta}}+\frac{2\beta\Gamma_0}{\epsilon^2}(F_{k_0}-F^*)+\|z^{k_0}-z^*\| < \zeta, \tag{3.2.28}$$

where $\Gamma_1 = \dfrac{2(L_{\nabla f}\mu_0 + \beta + L_p)}{\beta}, \Gamma_0 = \dfrac{2\kappa}{\beta\Gamma_1}, \Gamma_2 = \dfrac{\Gamma_0}{1-\sigma} + 1$ are constants. Here (3.2.28) means that, to use the KL property for proving whole sequence convergence, we need to start from an initial point w^{k_0} that is close enough to w^*.

Note that ψ in Definition 2.1 is concave and $\psi'(F_k - F^*) > 0$. The KL property and (3.2.26) yield

$$\begin{aligned} &\psi(F_k - F^*) - \psi(F_{k+1} - F^*) \\ &\geq \psi'(F_k - F^*)(F_k - F_{k+1}) \\ &\geq \frac{\psi'(F_k - F^*)\beta}{2\mu_k}\|z^k - z^{k+1}\|^2. \end{aligned} \tag{3.2.29}$$

Without loss of generality, in the rest of this proof, we set $k_0 = 0$ for simplicity of the proof.

We first claim that

$$w^k \in \mathcal{B}_\zeta(w^*), \quad \forall\ k \geq 0. \tag{3.2.30}$$

Let us check (3.2.30) for $k = 0$ and $k = 1$. In view of (3.2.28), w^0 lies in $\mathcal{B}_\zeta(w^*)$. As for w^1, (3.2.26) yields in particular

$$\frac{\beta}{2\mu_0}\|z^0 - z^1\|^2 \leq F_0 - F_1 \leq F_0 - F^*. \tag{3.2.31}$$

Hence,

$$\begin{aligned} \|w^1 - w^*\| &\leq \|z^1 - z^*\| + \mu_0 \\ &\leq \|z^1 - z^0\| + \|z^0 - z^*\| + \mu_0 \\ &\leq \sqrt{\frac{2\mu_0(F_0 - F^*)}{\beta}} + \|z^0 - z^*\| + \mu_0. \end{aligned} \tag{3.2.32}$$

Therefore, w^1 lies in $\mathcal{B}_\zeta(w^*)$ in view of (3.2.28).

Suppose $w^k \in \mathcal{B}_\zeta(w^*)$ for all $0 \leq k \leq K$ with $K \geq 1$. We next show $w^{K+1} \in \mathcal{B}_\zeta(w^*)$. Since $w^k \in \mathcal{B}_\zeta(w^*)$ and $F^* < F_k < F^* + \alpha$ for all $0 \leq k \leq K$, we have the Kurdyka-Łojasiewicz inequality

$$\psi'(F_k - F^*)\mathrm{dist}(0, \bar{\partial}F(w^k)) \geq 1 \tag{3.2.33}$$

with

$$\bar{\partial}F(w^k) = \begin{bmatrix} \nabla_x f(x^k, U^k) + \rho\nabla\tilde{\theta}(x^k, \mu_{k-1}) + \mathcal{N}_{\mathcal{S}_0}(x^k) \\ 2p_1(u_1^k - A_1x^k) + \rho\nabla\tilde{r}_1(u_1^k, \mu_{k-1}) + \mathcal{N}_{\mathcal{S}_1}(u_1^k) \\ \vdots \\ 2p_N(u_N^k - A_Nx^k) + \rho\nabla\tilde{r}_N(u_N^k, \mu_{k-1}) + \mathcal{N}_{\mathcal{S}_N}(u_N^k) \\ \kappa + \nabla_\mu\tilde{\phi}(x^k, U^k, \mu_{k-1}) \end{bmatrix}$$

which is due to the feasibility of z^k, the regularity of each set $\mathcal{S}_i$ and therefore the regularity of each indicator function. By the first-order optimality condition of subproblems (3.2.3) and (3.2.4), we have

$$\begin{aligned} -\Big[\nabla_x f(x^k, U^{k-1}) + \rho\nabla\tilde{\theta}(x^{k-1}, \mu_{k-1}) + \tfrac{\beta}{\mu_{k-1}}(x^k - x^{k-1})\Big] &\in \mathcal{N}_{\mathcal{S}_0}(x^k); \\ -\Big[2\,p_i(u_i^k - A_ix^k) + \rho\nabla\tilde{r}_i(u_i^{k-1}, \mu_{k-1}) + \tfrac{\beta}{\mu_{k-1}}(u_i^k - u_i^{k-1})\Big] &\in \mathcal{N}_{\mathcal{S}_i}(u_i^k),\ \forall i \in [N]. \end{aligned}$$

Then,

$$\begin{bmatrix} \nabla_x f(x^k, U^k) - \nabla_x f(x^k, U^{k-1}) + \rho\left(\nabla\tilde{\theta}(x^k, \mu_{k-1}) - \nabla\tilde{\theta}(x^{k-1}, \mu_{k-1})\right) - \frac{\beta}{\mu_{k-1}}\left(x^k - x^{k-1}\right) \\ \rho\left(\nabla\tilde{r}_1(u_1^k, \mu_{k-1}) - \nabla\tilde{r}_1(u_1^{k-1}, \mu_{k-1})\right) - \frac{\beta}{\mu_{k-1}}\left(u_1^k - u_1^{k-1}\right) \\ \vdots \\ \rho\left(\nabla\tilde{r}_N(u_N^k, \mu_{k-1}) - \nabla\tilde{r}_N(u_N^{k-1}, \mu_{k-1})\right) - \frac{\beta}{\mu_{k-1}}\left(u_N^k - u_N^{k-1}\right) \\ \kappa + \nabla_\mu\tilde{\phi}(x^k, U^k, \mu_{k-1}) \end{bmatrix}$$

is an element of $\bar{\partial}F(w^k)$ for $k \geq 1$. Note that $0 \leq \kappa + \nabla_\mu\tilde{\phi}(x^k, U^k, \mu_{k-1}) \leq \kappa$ from (3.2.16). Therefore, for $k \geq 1$, we have

$$\begin{aligned} & \operatorname{dist}(0, \bar{\partial}F(w^k)) \\ \leq\ & \left\|\nabla_x f(x^k, U^k) - \nabla_x f(x^k, U^{k-1})\right\| + \frac{L_p}{\mu_{k-1}}\left\|(x^k, U^k) - (x^{k-1}, U^{k-1})\right\| \\ & + \frac{\beta}{\mu_{k-1}}\left\|(x^k, U^k) - (x^{k-1}, U^{k-1})\right\| + \kappa \\ \leq\ & L_{\nabla f}\left\|U^k - U^{k-1}\right\| + \frac{L_p}{\mu_{k-1}}\left\|(x^k, U^k) - (x^{k-1}, U^{k-1})\right\| + \frac{\beta}{\mu_{k-1}}\left\|(x^k, U^k) - (x^{k-1}, U^{k-1})\right\| + \kappa \\ \leq\ & \left(L_{\nabla f} + \frac{L_p}{\mu_{k-1}} + \frac{\beta}{\mu_{k-1}}\right)\left\|z^k - z^{k-1}\right\| + \kappa. \end{aligned}$$

Therefore, we have

$$\psi'(F_k - F^*) \geq \frac{1}{\left(L_{\nabla f} + \frac{L_p}{\mu_{k-1}} + \frac{\beta}{\mu_{k-1}}\right)\left\|z^k - z^{k-1}\right\| + \kappa}.$$

Together with (3.2.26) and (3.2.29), we further have

$$\begin{aligned} & \psi(F_k - F^*) - \psi(F_{k+1} - F^*) \\ \geq\ & \frac{\frac{\beta}{\mu_k}\left\|z^k - z^{k+1}\right\|^2}{2\left[\left(L_{\nabla f} + \frac{L_p}{\mu_{k-1}} + \frac{\beta}{\mu_{k-1}}\right)\left\|z^k - z^{k-1}\right\| + \kappa\right]} \\ \geq\ & \frac{\frac{\beta}{\mu_{k-1}}\left\|z^k - z^{k+1}\right\|^2}{2\left[\left(L_{\nabla f} + \frac{L_p}{\mu_{k-1}} + \frac{\beta}{\mu_{k-1}}\right)\left\|z^k - z^{k-1}\right\| + \kappa\right]} \\ \geq\ & \frac{\left\|z^k - z^{k+1}\right\|^2}{2\left[\left(L_{\nabla f}\frac{\mu_{k-1}}{\beta} + L_p\frac{1}{\beta} + 1\right)\left\|z^k - z^{k-1}\right\| + \kappa\frac{\mu_{k-1}}{\beta}\right]} \\ \geq\ & \frac{1}{\Gamma_1} \cdot \frac{\left\|z^k - z^{k+1}\right\|^2}{\left\|z^k - z^{k-1}\right\| + \Gamma_0\mu_{k-1}}, \end{aligned}$$

with $\Gamma_1 = \frac{2(L_{\nabla f}\mu_0 + L_p + \beta)}{\beta}, \Gamma_0 = \frac{2\kappa}{\beta\Gamma_1}$.

Hence, for $1 \leq k \leq K$

$$\left(\|z^k - z^{k-1}\| + \Gamma_0\mu_{k-1}\right)^{\frac{1}{2}} \left(\Gamma_1(\psi(F_k - F^*) - \psi(F_{k+1} - F^*))\right)^{\frac{1}{2}} \geq \|z^k - z^{k+1}\|.$$

Together with $ab \leq (a^2 + b^2)/2$, we have

$$\|z^k - z^{k-1}\| + \Gamma_0\mu_{k-1} + \Gamma_1 \left(\psi(F_k - F^*) - \psi(F_{k+1} - F^*)\right) \geq 2\|z^k - z^{k+1}\|. \quad (3.2.34)$$

Since (3.2.34) holds for all $1 \leq k \leq K$, let us sum over k

$$\begin{array}{rl} & \|z^1 - z^0\| + \Gamma_0 \sum_{k=1}^K \mu_{k-1} + \Gamma_1(\psi(F_1 - F^*) - \psi(F_{K+1} - F^*)) \\ \geq & \sum_{k=1}^K \left\|z^k - z^{k+1}\right\| + \left\|z^{K+1} - z^K\right\|. \end{array}$$

Hence, in view of the monotonicity of ψ and F_k

$$\sum_{k=1}^K \left\|z^k - z^{k+1}\right\| \leq \left\|z^1 - z^0\right\| + \Gamma_1\psi(F_0 - F^*) + \Gamma_0 \sum_{k=1}^K \mu_{k-1}.$$

Recall Lemma 3.7, we have

$$\sum_{k=1}^{\infty} \mu_{k-1} \leq \frac{\mu_0}{1-\sigma} + \frac{2\beta}{\epsilon^2}(F_0 - F^*).$$

We finally get

$$\begin{array}{rl} & \left\|w^{K+1} - w^*\right\| \\ \leq & \|z^{K+1} - z^*\| + \mu_K \\ \leq & \|z^* - z^1\| + \sum_{k=1}^K \left\|z^k - z^{k+1}\right\| + \mu_K \\ \leq & \|z^* - z^1\| + \|z^1 - z^0\| + \Gamma_1\psi(F_0 - F^*) + \Gamma_2\mu_0 \\ \leq & 2\|z^1 - z^0\| + \|z^0 - z^*\| + \Gamma_1\psi(F_0 - F^*) + \Gamma_2\mu_0 \\ \leq & 2\sqrt{\frac{2\mu_0(F_0 - F^*)}{\beta}} + \frac{2\beta\Gamma_0}{\epsilon^2}(F_0 - F^*) + \|z^0 - z^*\| + \Gamma_1\psi(F_0 - F^*) + \Gamma_2\mu_0, \end{array}$$

with $\Gamma_2 = \dfrac{\Gamma_0}{1-\sigma} + 1$, which shows $w^{K+1} \in \mathcal{B}_\zeta(w^*)$. Here the last inequality is due to (3.2.31).

Hence we complete the proof of (3.2.30).

Therefore, the inequality (3.2.34) indeed holds for $\forall k \geq 1$; let us sum it for k running from some K_0 to some $K > K_0$

$$\sum_{k=K_0}^{K} \left\|z^k - z^{k+1}\right\| \leq \left\|z^{K_0} - z^{K_0-1}\right\| + \Gamma_1\psi(F_{K_0} - F^*) + \Gamma_2\mu_{K_0} + \frac{2\beta\Gamma_0}{\epsilon^2}(F_{K_0} - F^*).$$

Let $K \to \infty$, together with (3.2.26), we have

$$\begin{aligned}
& \textstyle\sum_{k=K_0}^{\infty} \left\|w^k - w^{k+1}\right\| \\
\leq\ & \textstyle\sum_{k=K_0}^{\infty} \left\|z^k - z^{k+1}\right\| + \mu_{k-1} - \mu_k \\
\leq\ & \sqrt{\frac{2\mu_{K_0-1}}{\beta}}\sqrt{F_{K_0-1} - F^*} + \Gamma_1\psi(F_{K_0} - F^*) + \Gamma_2\mu_{K_0} + \frac{2\beta\Gamma_0}{\epsilon^2}(F_{K_0} - F^*) + \mu_{K_0-1} \\
\leq\ & \sqrt{\frac{2\mu_{K_0-1}}{\beta}}\sqrt{F_{K_0-1} - F^*} + \Gamma_1\psi(F_{K_0-1} - F^*) + (\Gamma_2+1)\mu_{K_0-1} + \frac{2\beta\Gamma_0}{\epsilon^2}(F_{K_0-1} - F^*),
\end{aligned}$$

which implies that $\{z^k\}$ is a Cauchy sequence and therefore convergent. In view of Lemma 3.5, we conclude that its limit is an ϵ-Clarke stationary point of (3.0.2). $\square$

3.3 Numerical experiments

In this section, we conduct the empirical tests by applying our proposed method to the stochastic mean-variance portfolio selection problem. Our purposes are to demonstrate the robust performance of the SBCD algorithm with a prescribed tolerance, and explore the impact of different tolerances on algorithm performance.

In the tests, our termination criterion in Algorithm 1 is set as

$$\text{number of iterations} > 1000 \quad \text{or} \quad \mu_k \leq 10^{-4}. \tag{3.3.1}$$

The parameters in the penalty problem (3.1.3) and Algorithm 1 are set as

$$\lambda = 0.001,\ \rho_0 = 100,\ \rho_1 = 200,\ \mu_0 = 10, \beta = 1,\ \sigma = 0.5. \tag{3.3.2}$$

All the computational tests were conducted on an Lenovo ThinkCentre Desktop (Intel core I7-12700 CPU, 2.1 GHz, 64 GB RAM) with using Matlab R2024b.

3.3.1 Problem statement

Based on the optimality conditions, a solution to the LCP corresponds to a solution of the convex constrained quadratic program with its associated Lagrange multipliers, which play a significant role in portfolio optimization applications. The classic Markowitz's mean-variance (MV) portfolio optimization [35] is formulated as

$$\begin{array}{ll} \min\limits_{w\in\mathbb{R}^k} & \vartheta w^\top \Sigma w - \varrho^\top w \\ \text{s.t.} & 0 \le w \le a,\ e^\top w = 1, \end{array} \tag{3.3.3}$$

where $w \in \mathbb{R}^k$ denotes the portfolio weights of assets; $\varrho \in \mathbb{R}^k$ and $\Sigma \in \mathbb{S}^k_+$ are the mean vector and the covariance matrix of returns on assets, respectively; $a \in \mathbb{R}^k_+$ is upper bounds enforced on w, satisfying $e^\top a \ge 1$; $\vartheta > 0$ is the risk aversion factor. In what follows, unless otherwise specified, we set $\vartheta = \frac{1}{2}$ and $a = 0.2 \cdot e$ in all the tests. Obviously, it is equivalent to the LCP(q, M) with

$$M = \begin{bmatrix} \Sigma & -C^\top \\ C & 0 \end{bmatrix},\ C = \begin{bmatrix} e^\top \\ -e^\top \\ -I \end{bmatrix},\ q = \begin{bmatrix} -\varrho \\ -1 \\ 1 \\ a \end{bmatrix},\ x = \begin{bmatrix} w \\ y \end{bmatrix},$$

with the dual variables $y \in \mathbb{R}^{k+2}_+$.

Without loss of generality, we assume that the financial market contains two types of assets: stable assets (e.g., LondonGold, US10Y) and unstable assets (e.g., stocks). As well known, the above MV-based LCP is of high input sensitivity, mainly derived

from multi-scenario variations of unstable assets [3]. Here, we attempt to construct a global portfolio including stable and unstable assets that has the minimum expected Euclidean distance to the "wait and see" solution set of unstable assets, in order to remain highly effective in almost all scenario variations. Therefore, a global robust decision-making model is formulated as problem (1.1.1) with $n_1 = n, m_1 = m, n > m$, $B = I_n, D(\xi) = I_m, b = 0, c(\xi) = 0$, and $A(\xi) \equiv A$ being a projected matrix from global assets to unstable assets.

In practical, we consider the following discretized form:

$$\begin{aligned} \min_{x,U} \quad & \frac{1}{N}\sum_{i=1}^{N} \|Ax - u_i\|^2 + \frac{\lambda}{2}\|x\|^2 \\ \text{s.t.} \quad & \begin{cases} x \geq 0,\ Hx + h \geq 0,\ x^\top (Hx + h) = 0 \\ u_i \geq 0,\ M_i u_i + q_i \geq 0,\ u_i^\top (M_i u_i + q_i) = 0,\ i \in [N], \end{cases} \end{aligned} \tag{3.3.4}$$

where the input matrices and vectors H, h, M_i, q_i are constructed from historical data. We will illustrate how to construct these matrices and vectors in the next subsection.

3.3.2 Experimental setup

As aforementioned, the unstable assets are from stock exchange market while the stable assets are London golden investment (LondonGold) and US Treasury 10-Year Notes (US10Y). The stock data sets used in our experiments are selected from the standard ones in OR-library (termed data 1-3) and China A-share market (termed data 4-6), respectively. For the standard data sets, weekly prices of the stocks from 1992 to 1997 of Hang Seng (Hong Kong), Standard and Poor's 100 (USA), and the Nikkei index (Japan) are used. For data 4-5, the daily closing prices from 2010 to 2017 in China A-share market are used. Initially, we collect the daily prices of 3980 stocks from 2010 to 2017 in China stock market. For each stock, the prices before IPO (initial public offering) are labeled as zero. We delete the stocks with more than 200 daily prices of zero and obtain 1100 stocks with more consistent daily prices. We

sort the 1100 stocks by their average return over 2010 to 2017, and take the top 200 stocks of the sorted as the sample, where the first half is used as data 4 and the remaining as data 5. See Table 3.1 for the description of all the data sets. For data 6, the daily closing prices from 2015 to 2024 for the CSI 300 index [1] components in China A-share market are used. Initially, we collect the daily prices of the 300 stocks from 2015 to 2024 in China stock market. Then we do the same preprocessing as in the case for data 4 and data 5 and take the top 100 stocks as data 6.

Table 3.1: Description of the six real data sets.

Data set	J	Location	Index	T	Description
data 1	31	Hong Kong	Hang Seng	291	weekly prices from 1992 to 1997
data 2	98	USA	S&P 100	291	weekly prices from 1992 to 1997
data 3	225	Japan	Nikkei 225	291	weekly prices from 1992 to 1997
data 4	100	China	A-share	1939	daily prices from 2010 to 2017
data 5	100	China	A-share	1939	daily prices from 2010 to 2017
data 6	100	China	CSI300	2431	daily prices from 2015 to 2024

For each data set, let the observations of the stock closing prices be $\{P_{j,t} : j \in [J], t \in [T]\}$. Then, we can compute the (logarithmic) returns on stocks:

$$r_{j,t} = \log \frac{P_{j,t+1}}{P_{i,t}},\ j \in [J],\ t \in [T-1].$$

As a result, we obtain a return matrix $R \in \mathbb{R}^{(T-1)\times J}$ with $R(t,j) = r_{j,t}, j \in [J], t \in [T-1]$. Each data set is partitioned into two subsets at the ratio of 9:1, termed a training set and a testing set. The training set, called in-sample set, consists of the first 9/10 of the data and is used to compute the optimal portfolio. The testing set, called out-of-sample set, contains the rest of the data and is used to test the performance of the resulting optimal portfolio.

[1] The CSI 300 is a capitalization-weighted stock market index designed to replicate the performance of the top 300 stocks traded on the Shanghai Stock Exchange and the Shenzhen Stock Exchange. The CSI 300 index is deemed the Chinese counterpart of the S&P 500 index. Source: Wikipedia

Next, we elaborate on how to construct the input data H, h, M_i, q_i with rolling window procedure.

(i) `Set up` M_i, q_i. Let $R \in \mathbb{R}^{(T-1)\times J}$ be the return matrix of the unstable assets over the period $[1, T-1]$. By the splitting ratio, we have $R_{in} \in \mathbb{R}^{T_{in}\times J}$ and $R_{out} \in \mathbb{R}^{T_{out}\times J}$, with $T_{in} = \lceil 0.9(T-1) \rceil$ and $T_{out} = (T-1) - T_{in}$, namely, R_{in} is the first T_{in} rows of R and the rest of R is R_{out}. Let W be a rolling window and ς be a step size. For $i = 1, 2, ..., \lceil \frac{T_{in}}{\varsigma} \rceil$, we take $\varrho_i \in \mathbb{R}^J$ and $\Sigma_i \in \mathbb{R}^{J\times J}$ in the following way:

$$\varrho_i = \text{mean}(R_{in}((i-1)\varsigma + 1 : (i-1)\varsigma + W, :)),$$

$$\Sigma_i = \text{cov}(R_{in}((i-1)\varsigma + 1 : (i-1)\varsigma + W, :)).$$

Then,

$$M_i = \begin{bmatrix} \Sigma_i & -C^\top \\ C & 0 \end{bmatrix},\ C = \begin{bmatrix} e^\top \\ -e^\top \\ -I \end{bmatrix},\ q_i = \begin{bmatrix} -\varrho_i \\ -1 \\ 1 \\ a \end{bmatrix}.$$

(ii) `Set up` H, h. We can also obtain a return matrix $\hat{R} \in \mathbb{R}^{(T-1)\times(J+2)}$ of all assets including stable assets, where the last two columns are the returns of LondonGold and US10Y. Then, we take $\bar{\varrho} \in \mathbb{R}^{J+2}$ and $\bar{\Sigma} \in \mathbb{R}^{(J+2)\times(J+2)}$ in the following way:

$$\bar{\varrho} = \text{mean}(\hat{R}), \quad \bar{\Sigma} = \text{cov}(\hat{R}).$$

Then, we take

$$H = \begin{bmatrix} \bar{\Sigma} & -C^\top \\ C & 0 \end{bmatrix},\ C = \begin{bmatrix} e^\top \\ -e^\top \\ -I \end{bmatrix},\ h = \begin{bmatrix} -\bar{\varrho} \\ -1 \\ 1 \\ a \end{bmatrix}.$$

For data 1-3, we use $W = 3$ and $\varsigma = 1$; for data 4-5, we take $W = 15$ and $\varsigma = 3$. Then, each window has almost the same duration either 3 trading weeks or 15 trading days.

Note that the covariance matrix Σ is positive semi-definite, and the LCP(q, M) is equivalent to the convex quadratic program (3.3.3). The LCP(q, M) has a solution when $e^\top a \geq 1$. Hence problem (3.3.4) satisfies Assumption 3.1 when $e^\top a \geq 1$.

Moreover, the out-of-sample portfolio performance is more concerned in practice because it represents the possible future returns of the portfolio in a sense. For comparison, we define the two criteria to measure the out-of-sample performance as follows:

- the annualized return (AR) of the portfolio x constructed by strategy Z:

$$\mathrm{AR}(x) = \frac{\mathrm{CR}(x)}{N_w} \times 50 \quad \text{or} \quad \frac{\mathrm{CR}(x)}{N_d} \times 250,$$

  where $\mathrm{CR}(x)$ is the out-of-sample cumulative return of x; N_w and N_d denote the number of weeks and days, respectively, in the out-of-sample set.

- the Sharpe ratio (SR) of x constructed by strategy Z:

$$\mathrm{SR}(x) = \frac{\varrho^\top x - r_f}{\sqrt{x^\top \Sigma x}},$$

  where ϱ and Σ are the out-of-sample mean vector and covariance matrix, respectively, and r_f denotes the risk free rate which is set to be $\frac{2\%}{50}$ for data 1-3 and $\frac{2\%}{250}$ for data 4-6.

The Sharpe ratio is a widely used metric in finance to measure the risk-adjusted return of an investment or portfolio. It helps investors understand how much excess return they are earning for the risk they are taking. Apparently, the greater the two criteria, the better the strategy in terms of the out-of-sample AR and SR, respectively.

3.3.3 Out-of-sample robustness

To demonstrate the robust performance of the SBCD algorithm with a prescribed tolerance, we compare the out-of-sample performances of the following four strategies:

- **SVLCP** x^*: the SVLCP model (3.3.4) solved with our proposed SBCD algorithm with $\epsilon = 10^{-3}$;
- **EV** x_{ev}: the expected valued formulation LCP(h, H) solved with the semi-smooth Newton algorithm [2];
- **Naive** x_{na}: the naive evenly weighted portfolio [17];
- **DRO** x_{dro}: inspired by [2, 20], we solve the following penalized distributionally robust optimization with the Kullback-Leibler divergence

$$\begin{aligned} \min_{x} \quad & \eta_1 g_0(x) + \sup_{p\in\Delta}\left\{\sum_{i=1}^{N} p_i g_i(x) - \eta_2 \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{\hat{p}_i}\right)\right\} \\ \text{s.t.} \quad & x \geq 0, \end{aligned} \tag{3.3.5}$$

  where

$$g_0(x) := \max_{y\geq 0}\left\{\langle Hx + h, x - y\rangle - \frac{\alpha}{2}\|y - x\|^2\right\},$$

$$g_i(x) := \max_{y\geq 0}\left\{\langle M_i(A_i x) + q_i, A_i x - y\rangle - \frac{\alpha}{2}\|y - A_i x\|^2\right\}, i \in [N],$$

  with some $\alpha > 0$, are the regularized gap functions, $\hat{p}$ denotes a nominal distribution, $\Delta := \left\{p \in \mathbb{R}^N : p \geq 0, e^\top p = 1\right\}$ and η_1, η_2 are positive constants. Specifically, we take $\hat{p} = e/N$.

In the tests, we simulate three market scenarios including no-shock, single-asset shock and group shock as follows:

[2] The solver, developed by Y. Tassa, https://www.mathworks.com/matlabcentral/fileexchange/20952-lcp-mcp-solver-newton-based.

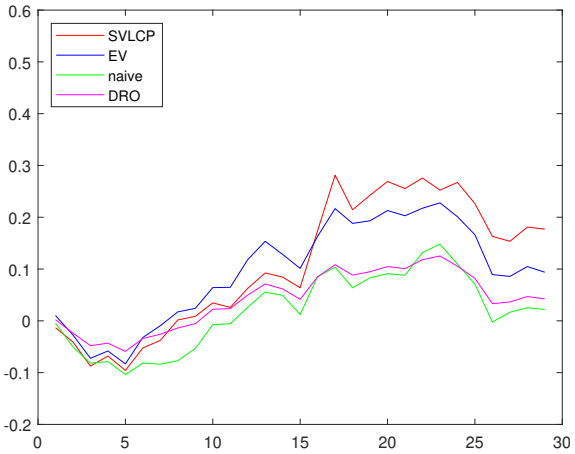
- **no-shock (NS)**: no perturbations on the out-of-sample sets.

- **single-asset shock (SS)**: randomly choose a stock and add a disturbance ($\pm 0.05*$ rand(T_{out}, 1) for weekly prices data 1-3 and $\pm 0.01*$ rand(T_{out}, 1) for daily prices data 4-6) to its returns. The sign is chosen by a Bernoulli distribution with $p = 0.5$ and stands for a positive or negative news for the specific stock.

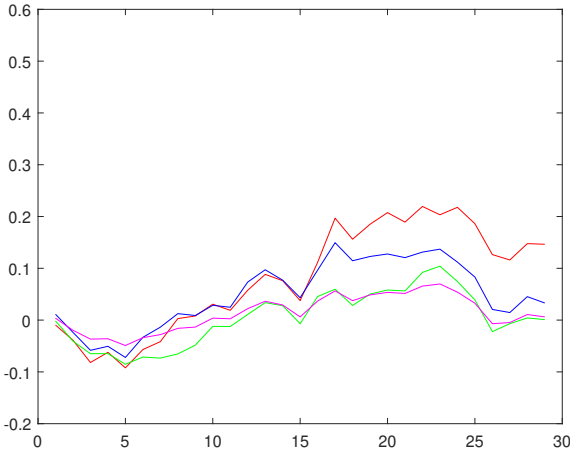
- **group shock (GS)**: cluster all the stocks into $5\lceil \frac{J}{50} \rceil$ groups or industries by using the Matlab package **spectralcluster**; and then add a disturbance of the same sign ($\pm 0.05*$ rand(T_{out}, group size) for weekly prices data 1-3 and $\pm 0.01*$ rand(T_{out}, 1) for daily prices data 4-6) to all members in a random group. The sign is chosen in the same way. This group disturbance indeed simulates a positive or negative news for the specific group or industry.

In addition, we investigate the impact of investment strategies on three types of investors with different risk-attitudes, which are characterized by various pre-allocated weights on stable assets. When the pre-allocated weights are set to $0, 0.1$ or 0.25 on both LondonGold and US10Y, we call the investors as `risk-taking (RT)`, `risk-neutral (RN)` and `risk-averse (RA)`, respectively. Accordingly, we need to modify the budgetary constraint "$e^\top w = 1$" to "$e^\top w = 1, 0.8, 0.5$" for restricting the cumulative weight of unstable assets.

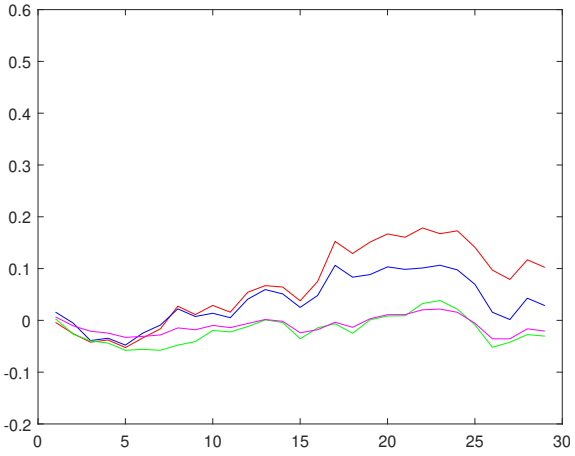
Now, we report the out-of-sample performance for the three types of investors on the six data sets. Firstly, under the **NS** scenario, we compare the out-of-sample trends of cumulative returns over time on the six data sets reported in Figures 3.1–3.6.



(a) Risk-taking

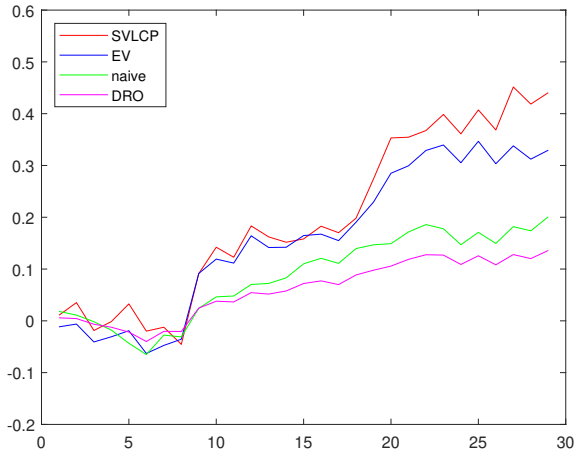


(b) Risk-neutral

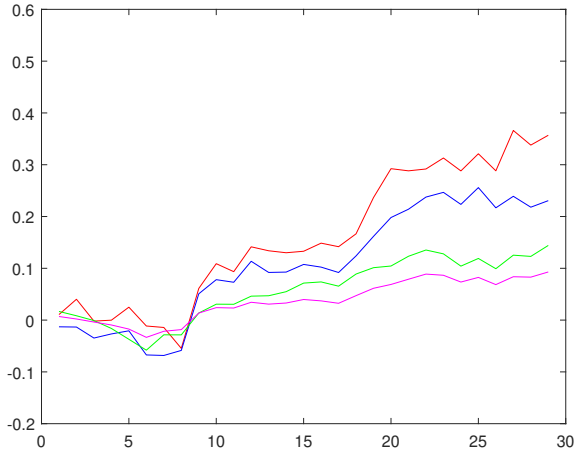


(c) Risk-averse

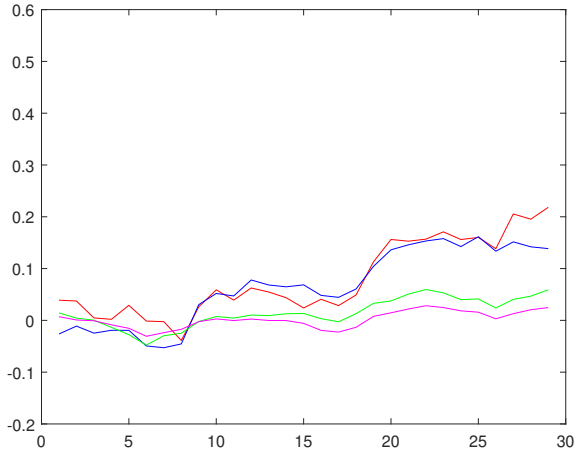
Figure 3.1: Comparisons of out-of-sample cumulative returns over time for the four strategies on data 1.



(a) Risk-taking

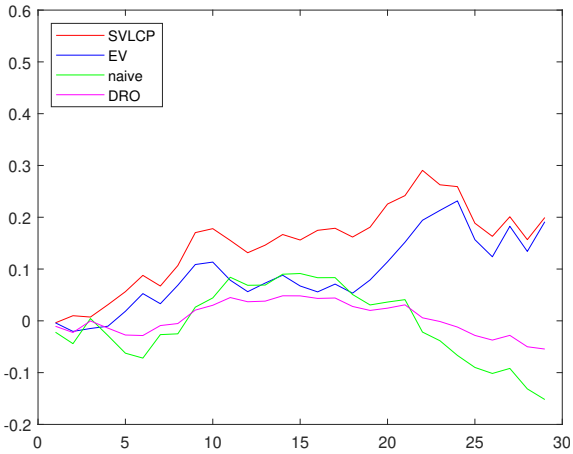


(b) Risk-neutral

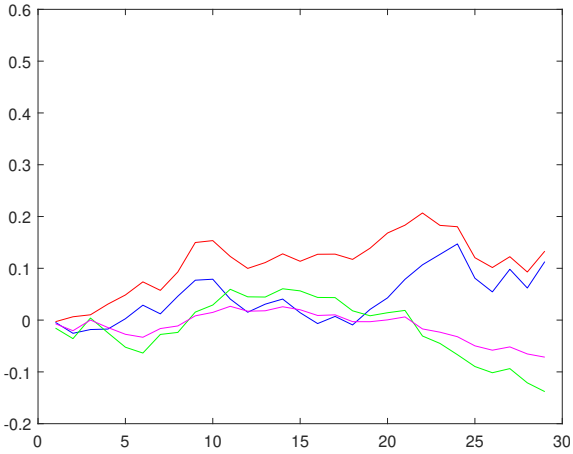


(c) Risk-averse

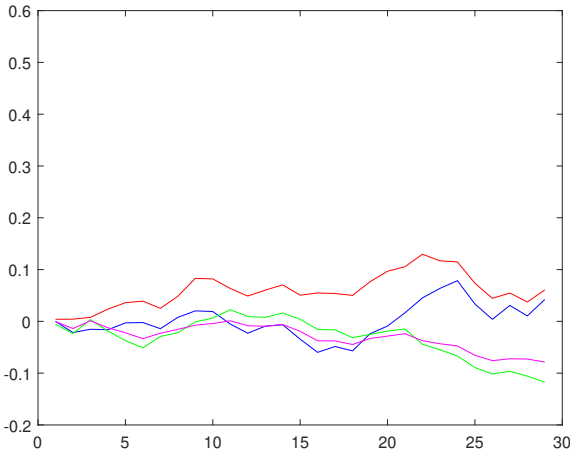
Figure 3.2: Comparisons of out-of-sample cumulative returns over time for the four strategies on data 2.



(a) Risk-taking

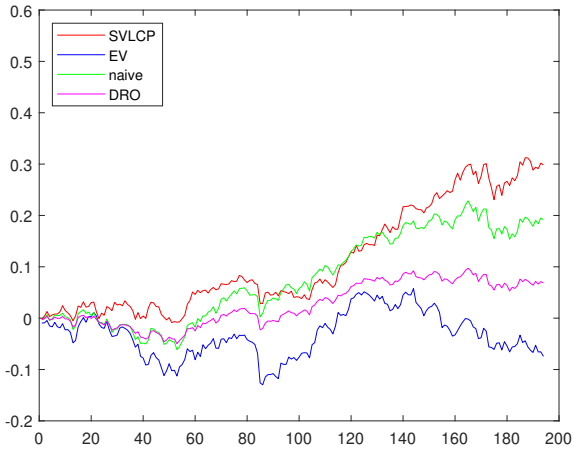


(b) Risk-neutral

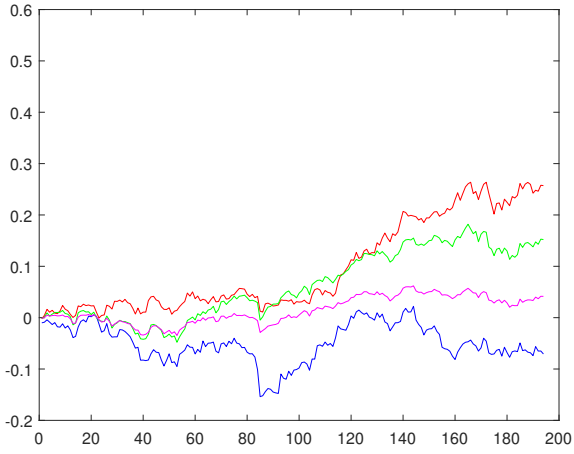


(c) Risk-averse

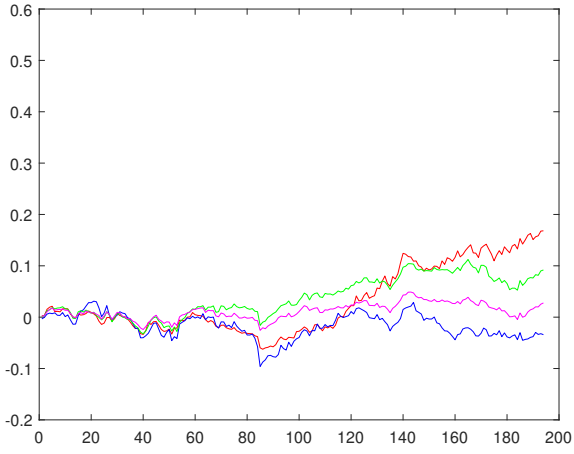
Figure 3.3: Comparisons of out-of-sample cumulative returns over time for the four strategies on data 3.



(a) Risk-taking

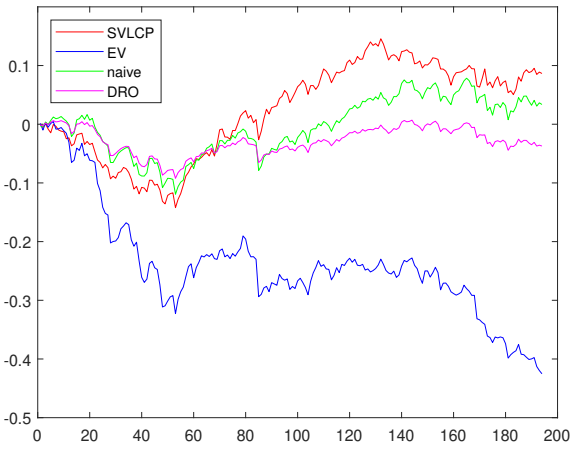


(b) Risk-neutral

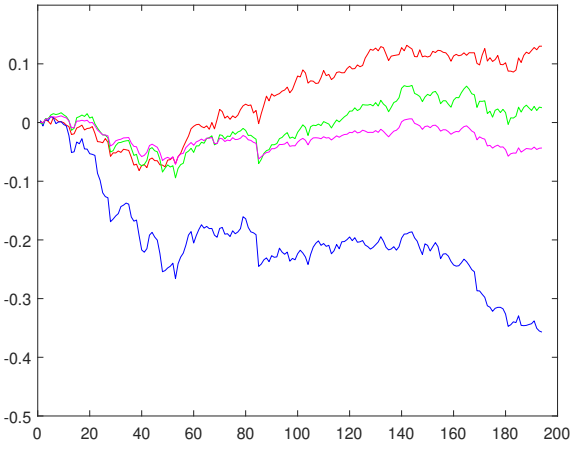


(c) Risk-averse

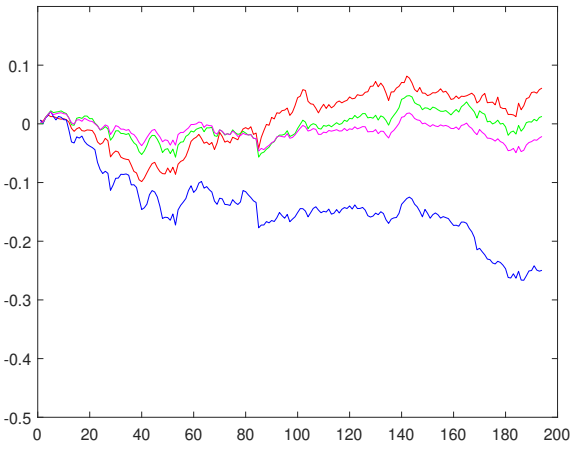
Figure 3.4: Comparisons of out-of-sample cumulative returns over time for the four strategies on data 4.



(a) Risk-taking

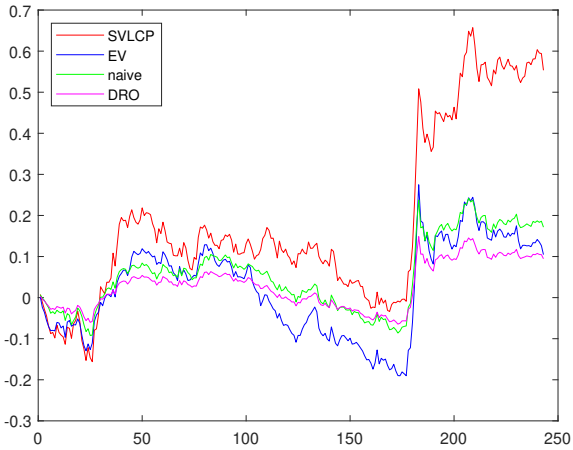


(b) Risk-neutral

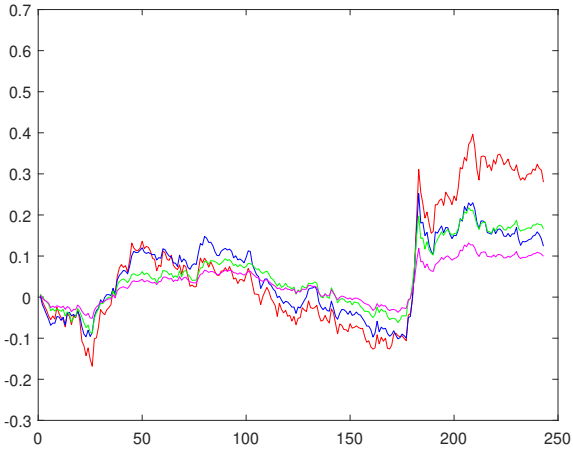


(c) Risk-averse

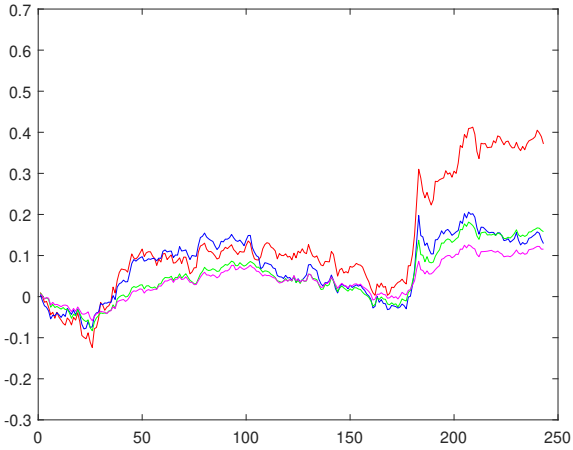
Figure 3.5: Comparisons of out-of-sample cumulative returns over time for the four strategies on data 5.



(a) Risk-taking



(b) Risk-neutral



(c) Risk-averse

Figure 3.6: Comparisons of out-of-sample cumulative returns over time for the four strategies on data 6.

Moreover, we calculate the out-of-sample values of AR and SR of the portfolios generated by the aforementioned three strategies, respectively. Numerical results are presented in the columns of "NS" of Tables 3.2–3.7. Then, under the **SS** and **GS** scenarios, we conduct 10000 random tests for verifying the robustness of portfolios, and report the means of AR/SR and also the averages of the worst 5% of AR/SR, denotes as $\text{AR}^{5\%}$ and $\text{SR}^{5\%}$, listed under the columns of "SS" and "GS" in Tables 3.2–3.7, respectively.

From the results of the **NS**, **SS** and **GS** scenarios, we have the following observations.

(i) Compared with the EV, Naive and DRO strategies, our SVLCP strategy shows advantages for investors with three different risk-attitudes in terms of the out-of-sample AR and SR. This is because our AR and SR are both higher than those of the other three strategies in all instances no matter the market scenarios. Furthermore, under the SS and GS scenarios, considering investors with three different risk-attitudes, our SVLCP strategy has the best $\text{AR}^{5\%}$ in 72.2% (26/36) instances and best $\text{SR}^{5\%}$ in 80.6% (29/36) instances which also show the robustness of our solution.

(ii) Moreover, our SVLCP strategy shows advantages throughout the process of cumulative returns than the others. This is because the cumulative returns of our strategy are mostly higher than that of the other three strategies for different investors on all the data sets.

3.3.4 The impact of tolerance

Next, we illustrate the impact of ϵ on the out-of-sample performance of our SVLCP strategy under different market scenarios. Specially, we set $\epsilon = 10^i$ $(i = -4, -3, \cdots, 0)$,

Table 3.2: Out-of-sample performances for 10000 random tests on data 1 under three market scenarios.

Scenarios	Investors	Criteria	x^*	x_{ev}	x_{na}	x_{dro}
	RT	AR	**0.305**	0.162	0.038	0.073
		SR	**0.138**	0.084	0.011	0.052
NS	RN	AR	**0.253**	0.057	0.002	0.011
		SR	**0.132**	0.025	-0.014	-0.012
	RA	AR	**0.176**	0.049	-0.053	-0.036
		SR	**0.116**	0.024	-0.082	-0.094
	RT	AR	**0.305**	0.161	0.037	0.073
		$AR^{5\%}$	**0.038**	-0.106	-0.010	-0.038
		SR	**0.138**	0.083	0.011	0.052
		$SR^{5\%}$	**0.009**	-0.075	-0.019	-0.057
SS	RN	AR	**0.251**	0.056	0.001	0.010
		$AR^{5\%}$	**-0.009**	-0.205	-0.037	-0.082
		SR	**0.131**	0.025	-0.014	-0.012
		$SR^{5\%}$	**-0.017**	-0.152	-0.044	-0.124
	RA	AR	**0.176**	0.049	-0.053	-0.036
		$AR^{5\%}$	**-0.034**	-0.153	-0.077	-0.087
		SR	**0.115**	0.023	-0.083	-0.094
		$SR^{5\%}$	**-0.040**	-0.140	-0.109	-0.179
	RT	AR	**0.304**	0.159	0.037	0.072
		$AR^{5\%}$	**-0.243**	-0.369	-0.417	-0.239
		SR	**0.137**	0.082	0.010	0.051
		$SR^{5\%}$	**-0.127**	-0.230	-0.272	-0.254
GS	RN	AR	**0.251**	0.054	0.001	0.009
		$AR^{5\%}$	-0.281	-0.239	-0.363	**-0.177**
		SR	**0.131**	0.023	-0.015	-0.013
		$SR^{5\%}$	**-0.170**	-0.176	-0.295	-0.241
	RA	AR	**0.175**	0.046	-0.053	-0.037
		$AR^{5\%}$	-0.225	-0.236	-0.280	**-0.139**
		SR	**0.114**	0.021	-0.083	-0.095
		$SR^{5\%}$	**-0.181**	-0.207	-0.340	-0.267

Table 3.3: Out-of-sample performances for 10000 random tests on data 2 under three market scenarios.

Scenarios	Investors	Criteria	x^*	x_{ev}	x_{na}	x_{dro}
	RT	AR	**0.760**	0.568	0.346	0.234
		SR	**0.330**	0.310	0.309	0.304
NS	RN	AR	**0.615**	0.398	0.249	0.160
		SR	**0.316**	0.252	0.270	0.266
	RA	AR	**0.376**	0.239	0.102	0.043
		SR	**0.246**	0.193	0.136	0.053
	RT	AR	**0.759**	0.568	0.346	0.234
		$AR^{5\%}$	**0.643**	0.453	0.331	0.209
		SR	**0.329**	0.310	0.308	0.304
		$SR^{5\%}$	0.273	0.244	**0.294**	0.267
SS	RN	AR	**0.615**	0.398	0.248	0.160
		$AR^{5\%}$	**0.513**	0.308	0.236	0.141
		SR	**0.316**	0.252	0.270	0.266
		$SR^{5\%}$	**0.262**	0.192	0.256	0.230
	RA	AR	**0.376**	0.239	0.102	0.043
		$AR^{5\%}$	**0.309**	0.184	0.094	0.033
		SR	**0.245**	0.193	0.136	0.053
		$SR^{5\%}$	**0.199**	0.145	0.123	0.030
	RT	AR	**0.763**	0.570	0.347	0.235
		$AR^{5\%}$	0.032	0.062	**0.116**	0.093
		SR	**0.331**	0.310	0.309	0.304
		$SR^{5\%}$	0.005	0.023	0.091	**0.103**
GS	RN	AR	**0.617**	0.399	0.249	0.161
		$AR^{5\%}$	**0.127**	-0.106	0.063	0.044
		SR	**0.317**	0.252	0.270	0.266
		$SR^{5\%}$	**0.056**	-0.083	0.052	0.045
	RA	AR	**0.378**	0.240	0.102	0.043
		$AR^{5\%}$	-0.105	-0.033	-0.013	**-0.012**
		SR	**0.246**	0.194	0.136	0.053
		$SR^{5\%}$	-0.085	**-0.047**	-0.056	-0.075

Table 3.4: Out-of-sample performances for 10000 random tests on data 3 under three market scenarios.

Scenarios	Investors	Criteria	x^*	x_{ev}	x_{na}	x_{dro}
	RT	AR	**0.344**	0.330	-0.262	-0.094
		SR	**0.212**	0.192	-0.204	-0.176
NS	RN	AR	**0.230**	0.195	-0.238	-0.123
		SR	**0.167**	0.122	-0.234	-0.253
	RA	AR	**0.105**	0.073	-0.202	-0.135
		SR	**0.095**	0.050	-0.303	-0.334
	RT	AR	**0.344**	0.330	-0.262	-0.094
		$AR^{5\%}$	**0.287**	0.273	-0.269	-0.103
		SR	**0.211**	0.191	-0.204	-0.176
		$SR^{5\%}$	**0.175**	0.156	-0.208	-0.190
SS	RN	AR	**0.230**	0.195	-0.238	-0.123
		$AR^{5\%}$	**0.190**	0.151	-0.243	-0.130
		SR	**0.167**	0.122	-0.235	-0.253
		$SR^{5\%}$	**0.135**	0.092	-0.239	-0.265
	RA	AR	**0.105**	0.073	-0.202	-0.135
		$AR^{5\%}$	**0.079**	0.044	-0.206	-0.139
		SR	**0.095**	0.050	-0.303	-0.334
		$SR^{5\%}$	**0.066**	0.023	-0.308	-0.342
	RT	AR	**0.345**	0.331	-0.261	-0.094
		$AR^{5\%}$	**-0.108**	-0.145	-0.369	-0.160
		SR	**0.212**	0.192	-0.203	-0.175
		$SR^{5\%}$	**-0.083**	-0.101	-0.281	-0.278
GS	RN	AR	**0.231**	0.196	-0.237	-0.123
		$AR^{5\%}$	**-0.159**	-0.197	-0.324	-0.178
		SR	**0.168**	0.123	-0.234	-0.253
		$SR^{5\%}$	**-0.140**	-0.150	-0.313	-0.349
	RA	AR	**0.106**	0.074	-0.202	-0.135
		$AR^{5\%}$	**-0.138**	-0.175	-0.256	-0.164
		SR	**0.096**	0.051	-0.303	-0.334
		$SR^{5\%}$	**-0.174**	-0.183	-0.377	-0.396

Table 3.5: Out-of-sample performances for 10000 random tests on data 4 under three market scenarios.

Scenarios	Investors	Criteria	x^*	x_{ev}	x_{na}	x_{dro}
	RT	AR	**0.385**	-0.096	0.248	0.088
		SR	**0.140**	-0.040	0.092	0.045
NS	RN	AR	**0.331**	-0.091	0.196	0.054
		SR	**0.133**	-0.041	0.088	0.028
	RA	AR	**0.217**	-0.044	0.118	0.036
		SR	**0.104**	-0.031	0.062	0.013
	RT	AR	**0.385**	-0.096	0.248	0.088
		$AR^{5\%}$	**0.261**	-0.195	0.234	0.066
		SR	**0.147**	-0.033	0.100	0.058
		$SR^{5\%}$	**0.100**	-0.067	0.095	0.043
SS	RN	AR	**0.332**	-0.091	0.196	0.053
		$AR^{5\%}$	**0.235**	-0.194	0.185	0.034
		SR	**0.142**	-0.033	0.098	0.045
		$SR^{5\%}$	**0.101**	-0.071	0.092	0.029
	RA	AR	**0.217**	-0.044	0.118	0.036
		$AR^{5\%}$	**0.153**	-0.108	0.112	0.026
		SR	**0.114**	-0.021	0.074	0.029
		$SR^{5\%}$	**0.081**	-0.052	0.070	0.021
	RT	AR	**0.387**	-0.093	0.249	0.090
		$AR^{5\%}$	**0.016**	-0.507	-0.002	-0.057
		SR	**0.148**	-0.032	0.101	0.059
		$SR^{5\%}$	**0.006**	-0.174	-0.001	-0.038
GS	RN	AR	**0.334**	-0.088	0.197	0.055
		$AR^{5\%}$	-0.038	-0.550	**-0.004**	-0.052
		SR	**0.143**	-0.032	0.098	0.045
		$SR^{5\%}$	-0.016	-0.201	**-0.002**	-0.043
	RA	AR	**0.219**	-0.042	0.119	0.036
		$AR^{5\%}$	-0.145	-0.318	**-0.006**	-0.029
		SR	**0.116**	-0.020	0.075	0.029
		$SR^{5\%}$	-0.076	-0.153	**-0.004**	-0.023

Table 3.6: Out-of-sample performances for 10000 random tests on data 5 under three market scenarios.

Scenarios	Investors	Criteria	x^*	x_{ev}	x_{na}	x_{dro}
	RT	AR	**0.111**	-0.548	0.043	-0.077
		SR	**0.035**	-0.172	0.009	-0.063
NS	RN	AR	**0.168**	-0.460	0.033	-0.056
		SR	**0.074**	-0.173	0.006	-0.059
	RA	AR	**0.079**	-0.322	0.016	-0.028
		SR	**0.030**	-0.158	-0.002	-0.037
	RT	AR	**0.111**	-0.548	0.043	-0.077
		$AR^{5\%}$	-0.014	-0.660	**0.030**	-0.101
		SR	**0.043**	-0.166	0.017	-0.050
		$SR^{5\%}$	-0.005	-0.200	**0.012**	-0.066
SS	RN	AR	**0.167**	-0.460	0.032	-0.056
		$AR^{5\%}$	**0.062**	-0.553	0.022	-0.074
		SR	**0.084**	-0.166	0.016	-0.044
		$SR^{5\%}$	**0.031**	-0.199	0.010	-0.057
	RA	AR	**0.079**	-0.322	0.016	-0.028
		$AR^{5\%}$	**0.013**	-0.382	0.010	-0.037
		SR	**0.041**	-0.148	0.010	-0.022
		$SR^{5\%}$	**0.007**	-0.176	0.006	-0.029
	RT	AR	**0.112**	-0.546	0.045	-0.076
		$AR^{5\%}$	-0.390	-1.213	-0.419	**-0.376**
		SR	**0.043**	-0.165	0.018	-0.049
		$SR^{5\%}$	**-0.150**	-0.367	-0.164	-0.245
GS	RN	AR	**0.168**	-0.458	0.034	-0.055
		$AR^{5\%}$	**-0.215**	-0.980	-0.338	-0.290
		SR	**0.084**	-0.165	0.016	-0.043
		$SR^{5\%}$	**-0.108**	-0.352	-0.162	-0.226
	RA	AR	**0.078**	-0.320	0.017	-0.027
		$AR^{5\%}$	-0.172	-0.646	-0.215	**-0.164**
		SR	**0.040**	-0.148	0.010	-0.021
		$SR^{5\%}$	**-0.088**	-0.298	-0.131	-0.129

Table 3.7: Out-of-sample performances for 10000 random tests on data 6 under three market scenarios.

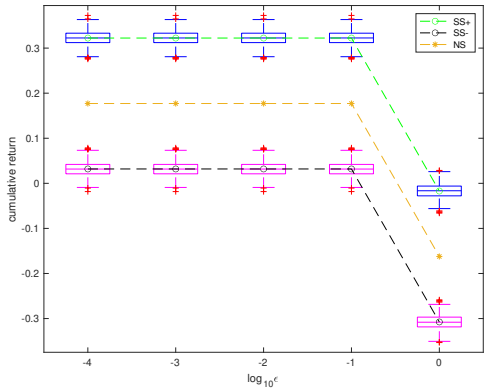
Scenarios	Investors	Criteria	x^*	x_{ev}	x_{na}	x_{dro}
	RT	AR	**0.569**	0.108	0.176	0.097
		SR	**0.080**	0.016	0.041	0.031
NS	RN	AR	**0.288**	0.128	0.171	0.102
		SR	**0.089**	0.025	0.049	0.045
	RA	AR	**0.355**	0.133	0.163	0.091
		SR	**0.108**	0.035	0.064	0.057
	RT	AR	**0.569**	0.107	0.176	0.097
		$AR^{5\%}$	**0.444**	-0.010	0.163	0.072
		SR	**0.083**	0.019	0.047	0.037
		$SR^{5\%}$	**0.064**	-0.002	0.043	0.027
SS	RN	AR	**0.379**	0.127	0.171	0.058
		$AR^{5\%}$	**0.278**	0.022	0.160	0.051
		SR	**0.094**	0.029	0.056	0.069
		$SR^{5\%}$	**0.069**	0.005	0.053	0.060
	RA	AR	**0.355**	0.133	0.163	0.091
		$AR^{5\%}$	**0.294**	0.070	0.156	0.087
		SR	**0.114**	0.041	0.073	0.072
		$SR^{5\%}$	**0.095**	0.022	0.071	0.070
	RT	AR	**0.566**	0.105	0.175	0.097
		$AR^{5\%}$	-0.181	-0.233	-0.211	**-0.105**
		SR	**0.082**	0.019	0.046	0.039
		$SR^{5\%}$	**-0.026**	-0.042	-0.056	-0.042
GS	RN	AR	**0.379**	0.126	0.170	0.057
		$AR^{5\%}$	**0.005**	-0.128	-0.139	-0.005
		SR	**0.094**	0.029	0.056	0.068
		$SR^{5\%}$	**0.001**	-0.030	-0.046	-0.005
	RA	AR	**0.355**	0.131	0.162	0.091
		$AR^{5\%}$	**0.105**	-0.098	-0.031	0.059
		SR	**0.114**	0.041	0.073	0.072
		$SR^{5\%}$	0.034	-0.030	-0.014	**0.047**

respectively, and then, when all else is the same as aforementioned, compare the out-of-sample cumulative returns on the six data sets. To avoid repetition, we take the **SS** scenario as an example; the results under the **GS** scenario are similar and therefore ignored. Under the **NS** scenario, we plot the out-of-sample cumulative returns for each i, and connect them with yellow lines. Under the **SS** scenario, we conduct 10000 positive and 10000 negative random tests for each i, respectively, and present a blue (positive) and a magenta (negative) box plots of the cumulative returns, along with the connecting lines of the mean points. The comparison results on data 1-6 are reported separately in Figures 3.7–3.12.

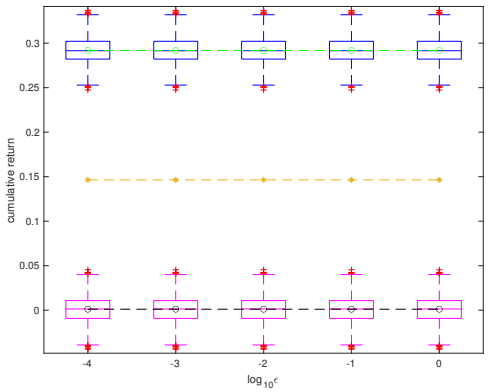
From Figures 3.7–3.12, we have the following observations.

(iii) When the tolerance ϵ is less than 10^{-2}, the out-of-sample performance of our SVLCP strategy tends to stabilize on almost all data sets regardless of the market shock. It implies that our proposed SBCD method converges consistently to an ϵ-Clarke stationary point in most instances.

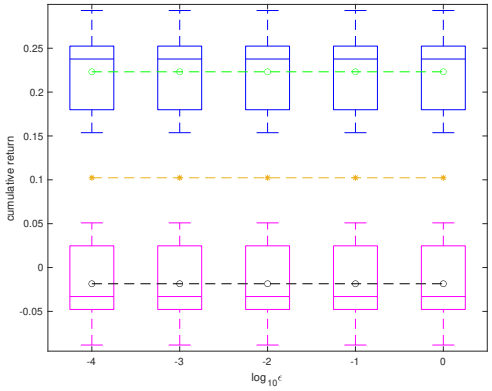
(iv) In addition, under the **SS** scenario, our SVLCP strategy obtains very concentrated out-of-sample cumulative returns on almost all data sets whether under positive or negative shock, which further demonstrates the robustness of our proposed algorithm.



(a) Risk-taking

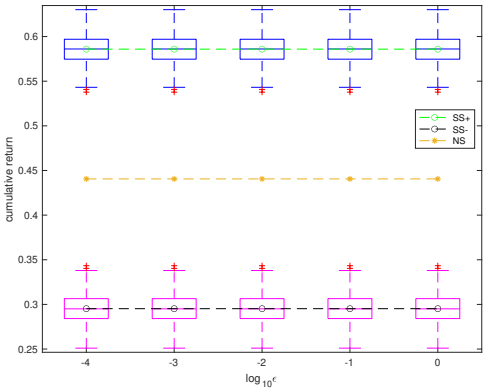


(b) Risk-neutral

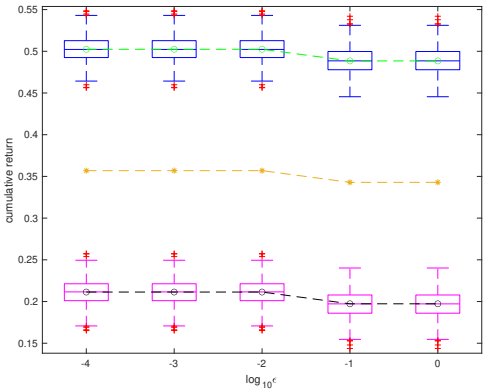


(c) Risk-averse

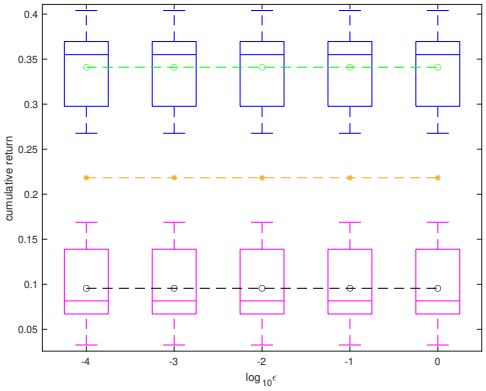
Figure 3.7: Boxplot of out-of-sample cumulative returns of random tests on data 1 for different ϵ.



(a) Risk-taking

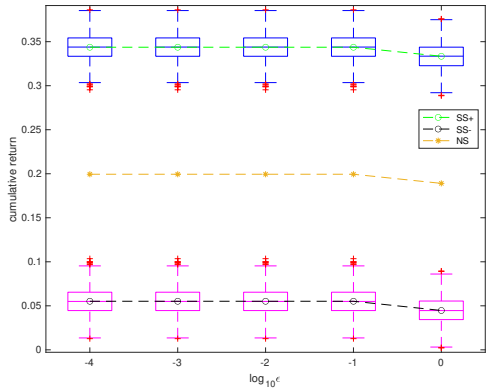


(b) Risk-neutral

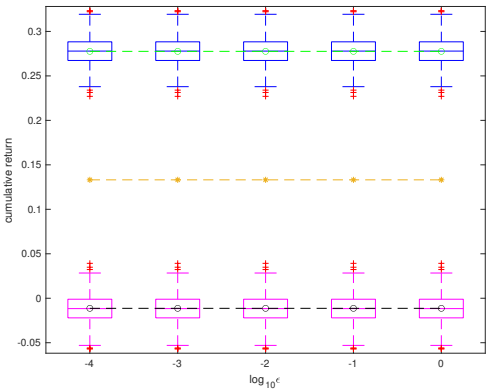


(c) Risk-averse

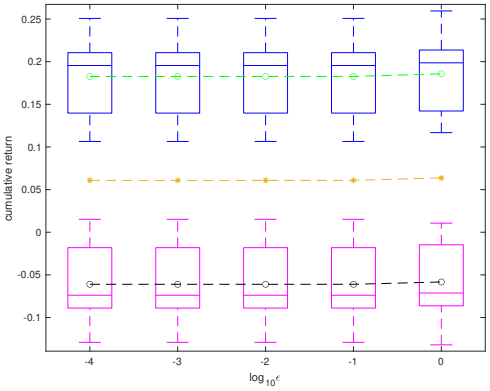
Figure 3.8: Boxplot of out-of-sample cumulative returns of random tests on data 2 for different ϵ.



(a) Risk-taking

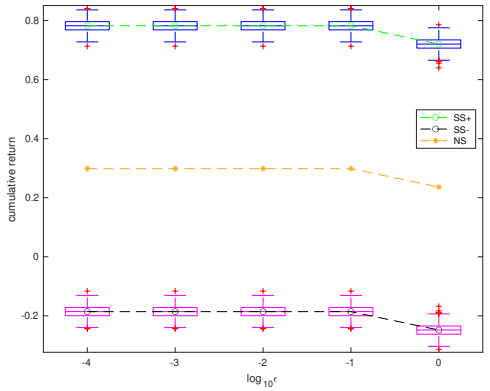


(b) Risk-neutral

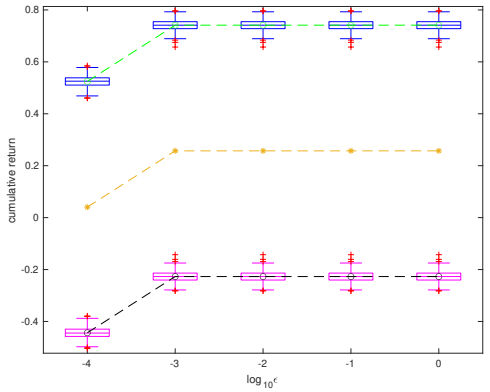


(c) Risk-averse

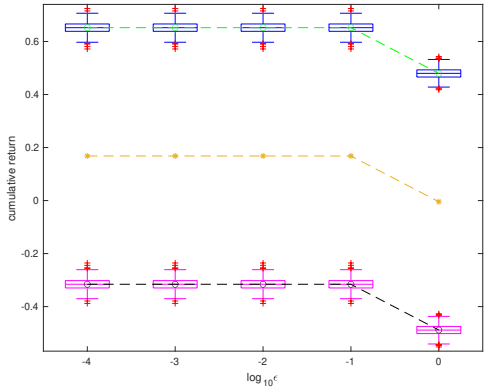
Figure 3.9: Boxplot of out-of-sample cumulative returns of random tests on data 3 for different ϵ.



(a) Risk-taking

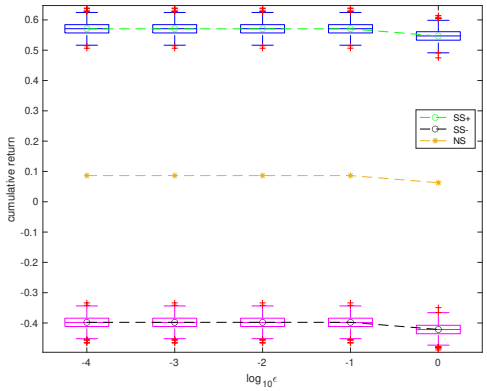


(b) Risk-neutral

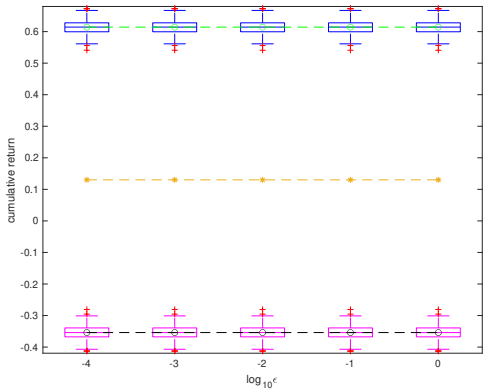


(c) Risk-averse

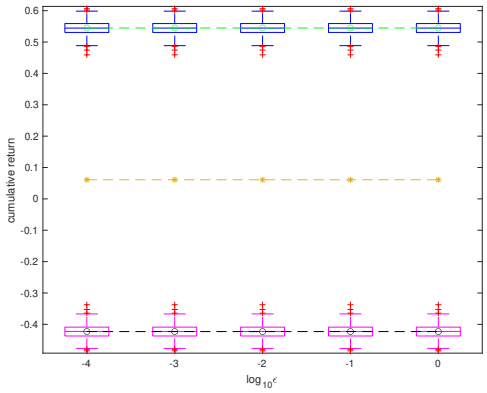
Figure 3.10: Boxplot of out-of-sample cumulative returns of random tests on data 4 for different ϵ.



(a) Risk-taking

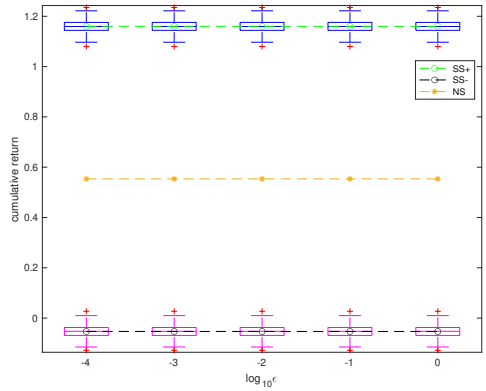


(b) Risk-neutral

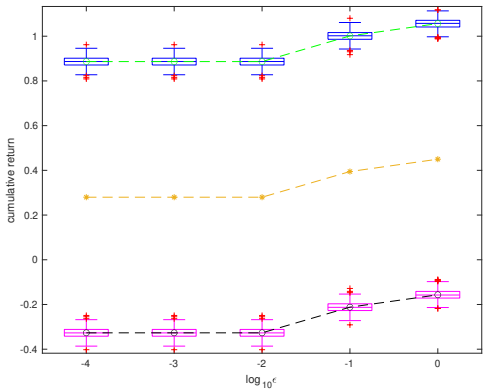


(c) Risk-averse

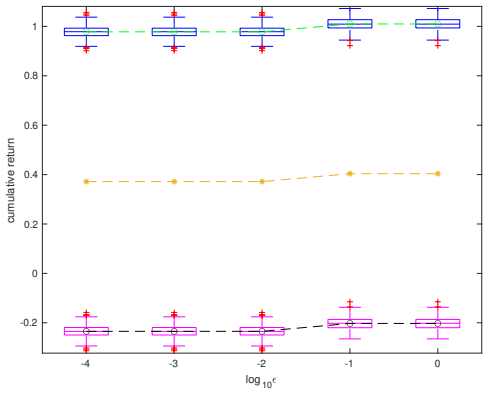
Figure 3.11: Boxplot of out-of-sample cumulative returns of random tests on data 5 for different ϵ.



(a) Risk-taking



(b) Risk-neutral

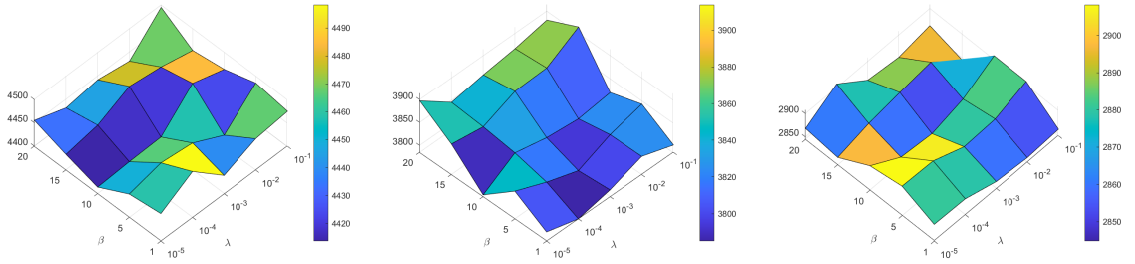


(c) Risk-averse

Figure 3.12: Boxplot of out-of-sample cumulative returns of random tests on data 6 for different ϵ.

3.3.5 Sensitivity analysis on β and λ

In this subsection, we conduct sensitivity analysis on the parameters β and λ. We plot the objective function values of (3.0.2) at ϵ-stationary solutions corresponding to different parameter settings. Specifically, we set $\beta = [0.1, 1, 5, 10, 20, 50]$ and $\lambda = [0.0001, 0.001, 0.01, 0.1, 1]$. The other parameters are just the same as (3.3.2). Here we take data 1 as an example and give the 3D plots in Figure 3.13.



(a) Risk-taking (b) Risk-neutral (c) Risk-averse

Figure 3.13: 3D plot of the objective function values (3.0.2) on data 1 for different β and λ.

Furthermore, we calculate the relative errors with taking the function value for $\beta = 10$ and $\lambda = 0.001$ as a benchmark. The relative errors fluctuate between $[-0.14\%, 1.77\%]$, $[-0.84\%, 2.52\%]$ and $[-0.30\%, 1.92\%]$, respectively, which shows that the function value does not change significantly, and therefore our model is robust.

Chapter 4

Sample average approximation method

In this chapter, we consider a sample average approximation (SAA) problem of (1.1.1) and present its convergence analysis.

The SAA method [19, 42, 46, 47] is a widely used approach for solving stochastic optimization problems, where the objective function or constraints involve uncertainty modeled by random variables. The method replaces the involved expectations or probabilistic terms in the original problem with their sample averages computed using a finite number of samples of the random variables. This transforms the stochastic problem into a deterministic optimization problem that is easier to solve.

4.1 An implicit reformulation

In this section, we give an implicit reformulation of problem (1.1.1) and then approximate it with the SAA method.

Intuitively, we could write the SAA problem of (1.1.1) as follows

$$\begin{aligned} &\min_{x,u_1,\ldots,u_N} \hat{f}_N(x) := \frac{1}{N}\sum_{i=1}^{N} \|A(\xi_i)x - u_i\|^2 + \frac{\lambda}{2}\|x\|^2 \\ &\text{s.t.} \begin{cases} 0 \le Bx + b \perp Hx + h \ge 0 \\ 0 \le C(\xi_i)u_i + d(\xi_i) \perp M(\xi_i)u_i + q(\xi_i) \ge 0, \text{ for } i \in [N]. \end{cases} \end{aligned} \tag{4.1.1}$$

However, this formulation is not appropriate for SAA convergence analysis. The increasing variables $u_1, \ldots, u_N$ make it hard to analyze the behavior of the SAA method. Therefore, we consider an implicit reformulation of problem (1.1.1) which "eliminates" the variables $u_1, \ldots, u_N$.

Alternatively, we reformulate problem (1.1.1) as following

$$\begin{aligned} \min \quad & f(x) := \mathbb{E}[F(x, \xi)] \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned} \tag{4.1.2}$$

where $F(x, \xi) := p(x, \xi) + \dfrac{\lambda}{2}\|x\|^2$ with $p(x, \xi)$ being the optimal value of

$$\begin{aligned} \min \quad & \|A(\xi)x - u(\xi)\|^2 \\ \text{s.t.} \quad & u(\xi) \in \mathcal{U}(\xi). \end{aligned} \tag{4.1.3}$$

For equivalence of formulations (1.1.1) and (4.1.2), see [40, Chapter 1, section 2.4] and [42]. Then, problem (4.1.3) is a projection of $A(\xi)x$ onto $\mathcal{U}(\xi)$.

4.2 Convergence analysis

In this section, we consider the convergence of the SAA method for problem (4.1.2). Let random sample $\xi_1, \ldots, \xi_N$ be N realizations of the random vector ξ and the expected value $\mathbb{E}[F(x, \xi)]$ is approximated by the following sample average $\dfrac{1}{N}\sum_{i=1}^{N} F(x, \xi_i)$. Hence, problem (4.1.2) is approximated by the following SAA problem

$$\begin{aligned} \min_x \quad & \hat{f}_N(x) := \frac{1}{N}\sum_{i=1}^{N} F(x, \xi_i) \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned} \tag{4.2.1}$$

where $F(x, \xi_i) = p(x, \xi_i) + \dfrac{\lambda}{2}\|x\|^2$ with $p(x, \xi_i)$ being the optimal value of

$$\begin{aligned} \min \quad & \|A(\xi_i)x - u_i\|^2 \\ \text{s.t.} \quad & u_i \in \mathcal{U}(\xi_i). \end{aligned} \tag{4.2.2}$$

Problem (4.2.1) is an implicit formulation where variables $u_i, i = 1, ..., N$ are "eliminated". This type of formulation is of help for the convergence analysis.

Next, we show a convergence analysis for the SAA problem (4.2.1). Let $\hat{\vartheta}_N$ and $\hat{\mathcal{S}}_N$ denote the optimal value and set of optimal solutions of problem (4.2.1) respectively. It is important to note that $\hat{\vartheta}_N$ and $\hat{\mathcal{S}}_N$ depend on the generated samples and should, therefore, be considered as random variables. We discuss the convergence properties of $\hat{\vartheta}_N$ and $\hat{\mathcal{S}}_N$ viewed as statistical estimators of counterparts ϑ^* and $\mathcal{S}^*$ of the true problem (4.1.2). Throughout this section, we assume that the random sample $\xi_1, \ldots, \xi_N$ is independent and identically distributed (i.i.d).

Let

$$\gamma := \sqrt{\frac{2f(x^0)}{\lambda}},$$

and

$$\mathcal{C} := \mathcal{X} \cap \mathcal{B}_\gamma^n$$

with x^0 being an arbitrary feasible point of problem (4.1.2). Then we have the following result on the boundedness of the optimal solution sets $\mathcal{S}^*$ and $\hat{\mathcal{S}}_N$.

Lemma 4.1. *The solution set of problem* (4.1.2) *is not empty and contained in* $\mathcal{C}$. *Moreover, w.p.1 for* N *large enough the set* $\hat{\mathcal{S}}_N$ *is nonempty and contained in* $\mathcal{C}$.

Proof. Note that the objective function of problem (4.1.2) is continuous and coercive. Given that the feasible set $\mathcal{X}$ is not empty, the solution set of problem (4.1.2) is not empty and contained in the level set $\mathcal{D}(f) = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$. Note that $x \in \mathcal{D}(f)$ implies that $\frac{\lambda}{2}\|x\|^2 \leq f(x^0)$, i.e., $\|x\| \leq \sqrt{\frac{2f(x^0)}{\lambda}}$. Therefore, $\mathcal{S}^* \subset \mathcal{C}$.

Together with the law of large numbers, similarly, we can show that w.p.1 for N large enough the set $\hat{\mathcal{S}}_N$ is nonempty and $\hat{\mathcal{S}}_N \subset \mathcal{B}_\gamma^n$. □

Note that $\hat{f}_N(x)$ converges to $f(x)$ w.p. 1 uniformly on $\mathcal{C}$ [41, pp. 375] if

$$\sup_{x\in\mathcal{C}} \left|\hat{f}_N(x) - f(x)\right| \to 0 \quad \text{w.p. } 1 \quad \text{as} \quad N \to \infty.$$

Lemma 4.2. *The expected value function f is finite valued and continuous on $\mathcal{C}$, and $\hat{f}_N(x)$ converges to $f(x)$ w.p. 1, as $N \to \infty$, uniformly in $x \in \mathcal{C}$.*

Proof. First, note that for any $x \in \mathcal{C}$ the function $p(\cdot, \xi)$ is continuous at x for almost every $\xi \in \Xi$.

Let $u^*(\xi)$ be an optimal solution of (4.1.3). Then $u^*(\cdot) : \Xi \to \mathbb{R}^m$ is a measurable function. Moreover,

$$\begin{aligned} F(x,\xi) &= p(x,\xi) + \frac{\lambda}{2}\|x\|^2 \\ &= \|A(\xi)x - u^*(\xi)\|^2 + \frac{\lambda}{2}\|x\|^2 \\ &\leq (\|A(\xi)\|\|x\| + \|u^*(\xi)\|)^2 + \frac{\lambda}{2}\|x\|^2 \\ &\leq (\|A(\xi)\|\gamma + \|u^*(\xi)\|)^2 + \frac{\lambda\gamma^2}{2}. \end{aligned}$$

Therefore, by [41, Theorem 7.48], together with the i.i.d assumption, we conclude that the expected value function f is finite value and continuous on $\mathcal{C}$, and $\hat{f}_N(x)$ converges to $f(x)$ w.p. 1, as $N \to \infty$, uniformly in $x \in \mathcal{C}$. □

Then, we have the following theorem which establishes the SAA convergence.

Theorem 4.1. *$\hat{\vartheta}_N \to \vartheta^*$ and $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \to 0$ w.p. 1 as $N \to \infty$.*

Proof. Due to Lemma 4.1 and Lemma 4.2, by [41, Theorem 5.3], we established the statement. □

4.3 Exponential convergence rate

In the previous section, we established the convergence of the SAA problem. Specifically, we show that the optimal value and set of optimal solutions of the SAA problem converge to the counterparts of the true problem with probability one as the sample size increases. However, an important issue remains unaddressed, which is both conceptually and computationally significant: how large should the sample size be to achieve the desired accuracy of the SAA estimators? Next, we show the uniform exponential convergence of $\hat{f}_N(x)$ to $f(x)$ on $\mathcal{X}$. Furthermore, we translate the uniform exponential convergence of functional value to the exponential convergence of solutions of the SAA problem to the set of optimal solution of the true problem.

Cramér's Large Deviation Theorem plays an important role in the following analysis. It is a fundamental result in probability theory that describes the asymptotic behavior of the probabilities of rare events for sums of i.i.d. random variables. Specifically, it provides an exponential decay rate for the probability that the sum deviates significantly from its expected value. The formal definition is given as follows [41, pp. 387-388].

Definition 4.1 (Cramér's Large Deviation Theorem)**.** *Consider an i.i.d. sequence $Z_1, \ldots, Z_N$ of replications of a real valued random variable Z, and let $Y_N := \frac{1}{N}\sum_{i=1}^{N} Z_i$ be the corresponding sample average. Suppose that Z has finite mean $\mu_Z := \mathbb{E}[Z]$. Then, for any real number $a \geq \mu_Z$, it holds that*

$$\mathbb{P}(Y_N \geq a) \leq e^{-NI(a)}, \tag{4.3.1}$$

where

$$I(y) := \sup_{t \in \mathbb{R}} \{ty - \Lambda(t)\},$$

called the rate function, is the conjugate of the logarithmic moment-generating function

$\Lambda(t) := \ln M(t)$ *and* $M(t) := \mathbb{E}[e^{tZ}]$ *is the moment-generating function (MGF) of random variable* Z.

We denote by

$$M_x(t) := \mathbb{E}\left\{e^{t[F(x,\xi)-f(x)]}\right\}$$

the moment generating function of the random variable $F(x,\xi) - f(x)$. Furthermore, let us make the following assumption.

Assumption 4.1. *For every* $x \in \mathcal{C}$ *the moment generating function* $M_x(t)$ *is finite valued for all* t *in a neighborhood of zero.*

Proposition 4.1. *Suppose Assumption 4.1 holds. Then, for any* $\epsilon > 0$*, there exist positive constants* $c(\epsilon)$ *and* $\beta(\epsilon)$*, independent of* N*, such that*

$$\mathbb{P}\left\{\sup_{x\in\mathcal{C}} |\hat{f}_N(x) - f(x)| \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.3.2}$$

Proof. This proof is primarily based on the proof of [42, Theorem 5.1].

First note that $\mathcal{C}$ is a compact subset of $\mathcal{X}$. Therefore, we can construct a ν-net to approximate $\mathcal{C}$. Specifically, for a $\nu > 0$, let $\bar{x}_1, \ldots, \bar{x}_T \in \mathcal{C}$ be such that for every $x \in \mathcal{C}$ there exists $\bar{x}_i, i \in \{1, \ldots, T\}$, such that $\|x - \bar{x}_i\| \leq \nu$, i.e. $\{\bar{x}_1, \ldots, \bar{x}_T\}$ is a ν-net in $\mathcal{C}$. We can choose this net in such a way that $T \leq [O(1)d/\nu]^n$, where $d := \sup_{x',x\in\mathcal{C}} \|x' - x\|$ is the diameter of $\mathcal{C}$ and $O(1)$ is a generic constant.

For an $x \in \mathcal{C}$, let $i(x) \in \arg\min_{1\leq i\leq T} \|x - \bar{x}_i\|$. By construction of the ν-net we have that $\|x - \bar{x}_{i(x)}\| \leq \nu$ for every $x \in \mathcal{C}$. Then

$$\begin{aligned}|\hat{f}_N(x) - f(x)| &\leq |\hat{f}_N(x) - \hat{f}_N(\bar{x}_{i(x)})| + |\hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)})| + |f(\bar{x}_{i(x)}) - f(x)| \\ &\leq 2(1 + \lambda\gamma)\nu + |\hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)})|.\end{aligned}$$

Here the second inequality is due to the Lipschitz continuity of functions $\hat{f}_N$ and f on $\mathcal{C}$.

According to the Cramér's Large Deviation Theorem, we have that for any $x \in \mathcal{C}$ and $\epsilon > 0$ it holds that

$$\mathbb{P}\left\{\hat{f}_N(x) - f(x) \geq \epsilon\right\} \leq e^{-NI_x(\epsilon)}, \tag{4.3.3}$$

where

$$I_x(y) := \sup_{t \in \mathbb{R}} \{ty - \ln M_x(t)\}$$

is the rate function of random variable $F(x, \xi) - f(x)$.

Similarly,

$$\mathbb{P}\left\{\hat{f}_N(x) - f(x) \leq -\epsilon\right\} \leq e^{-NI_x(-\epsilon)},$$

and hence

$$\mathbb{P}\left\{|\hat{f}_N(x) - f(x)| \geq \epsilon\right\} \leq e^{-NI_x(\epsilon)} + e^{-NI_x(-\epsilon)}. \tag{4.3.4}$$

By Assumption 4.1, we have that both $I_x(\epsilon)$ and $I_x(-\epsilon)$ are positive for every $x \in \mathcal{C}$.

Consider $Z_i := \hat{f}_N(\bar{x}_i) - f(\bar{x}_i), i = 1, \ldots, T$. We have that the event $\{\max_{1 \leq i \leq T} |Z_i| \geq \epsilon\}$ is equal to the union of the events $\{|Z_i| \geq \epsilon\}, i = 1, \ldots, T$, and hence

$$\mathbb{P}\left(\max_{1 \leq i \leq T} |Z_i| \geq \epsilon\right) \leq \sum_{i=1}^{T} \mathbb{P}(|Z_i| \geq \epsilon).$$

Together with (4.3.4) this implies that

$$\mathbb{P}\left(\max_{1 \leq i \leq T} |\hat{f}_N(x) - f(x)| \geq \epsilon\right) \leq 2 \sum_{i=1}^{T} \exp\left\{-N \min\{I_x(\epsilon), I_x(-\epsilon)\}\right\}.$$

Now let us take a ν-net with such ν that $2(1 + \lambda\gamma)\nu = \epsilon/2$, i.e. $\nu := \dfrac{\epsilon}{4(1 + \lambda\gamma)}$.

Hence,

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{x\in\mathcal{C}}|\hat{f}_N(x)-f(x)|\geq\epsilon\right\}\\
\leq&\mathbb{P}\left\{\max_{1\leq i\leq T}|\hat{f}_N(\bar{x}_i)-f(\bar{x}_i)|\geq\frac{\epsilon}{2}\right\} \qquad (4.3.5)\\
\leq&2\sum_{i=1}^{T}\exp\left\{-N\min\left\{I_{\bar{x}_i}\left(\frac{\epsilon}{2}\right),I_{\bar{x}_i}\left(-\frac{\epsilon}{2}\right)\right\}\right\}.
\end{aligned}$$

Since the above choice of the ν-net does not depend on the sample, and both $I_{\bar{x}_i}(\frac{\epsilon}{2})$ and $I_{\bar{x}_i}(-\frac{\epsilon}{2})$ are positive, $i=1,\ldots,T$, we obtain that (4.3.5) implies (4.3.2), and hence completes the proof. □

Next, based on the uniform exponential convergence of $\hat{f}_N(x)$ to $f(x)$, we further show an exponential convergence of an optimal solution of the SAA problem to the set of optimal solutions of the true problem. For this translation, we first introduce the following lemma which basically states that when $\sup_{x\in\mathcal{X}}|\hat{f}_N(x)-f(x)|$ is sufficiently small, the set of optimal solutions of the SAA problem is close to the set of optimal solutions of the true problem if both sets are nonempty.

Lemma 4.3. *For any $\epsilon>0$, there exists a $\delta>0$ such that if $\sup_{x\in\mathcal{X}}|\hat{f}_N(x)-f(x)|<\delta$, then $\mathbb{D}(\hat{\mathcal{S}}_N,\mathcal{S}^*)<\epsilon$.*

Proof. First, we show that if $\sup_{x\in\mathcal{X}}|\hat{f}_N(x)-f(x)|<\delta$, then $|\hat{\vartheta}_N-\vartheta^*|<\delta$.

Suppose $\hat{\vartheta}_N,\vartheta^*$ are attained at $x_1,x_2\in\mathcal{X}$ respectively. Without loss of generality, we may assume that $\hat{\vartheta}_N\geq\vartheta^*$. Then,

$$\hat{\vartheta}_N-\vartheta^*=\hat{f}_N(x_1)-f(x_2)\leq\hat{f}_N(x_2)-f(x_2)<\delta.$$

Similarly, if $\vartheta^*\geq\hat{\vartheta}_N$, we have $\vartheta^*-\hat{\vartheta}_N<\delta$.

Then, we prove that if $\sup_{x\in\mathcal{X}}|\hat{f}_N(x)-f(x)|<\delta$, then $|\hat{\vartheta}_N-\vartheta^*|<\delta$.

Let

$$R(\epsilon) := \inf_{x:d(x,S^*)\geq\epsilon} f(x) - \vartheta^*.$$

Obviously, $R(\epsilon) > 0$. Let $\delta = R(\epsilon)/3$ and $\sup_{x\in\mathcal{X}} |\hat{f}_N(x) - f(x)| < \delta$.

Then, for any point $x \in \mathcal{X}$ with $d(x, S^*) \geq \epsilon$, we have

$$\begin{aligned} |\hat{f}_N(x) - \hat{\vartheta}_N| &= |\hat{f}_N(x) - \hat{\vartheta}_N - (f(x) - \vartheta^*) + f(x) - \vartheta^*| \\ &\geq |f(x) - \vartheta^*| - |\hat{f}_N(x) - f(x)| - |\hat{\vartheta}_N - \vartheta^*| \\ &\geq 3\delta - \delta - \delta = \delta > 0. \end{aligned}$$

This indicates that $x \notin \hat{\mathcal{S}}_N$. Therefore, for any $\hat{x}^* \in \hat{\mathcal{S}}_N$, $d(\hat{x}^*, \mathcal{S}^*) \leq \epsilon$, which implies $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) < \epsilon$. □

Theorem 4.2. *Let $\hat{x}^*$ be a solution of the SAA problem* (4.2.1)*. Suppose Assumption 4.1 holds. Then, for every $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$, independent of N, such that*

$$\mathbb{P}\{\mathrm{dist}(\hat{x}^*, \mathcal{S}^*) \geq \epsilon\} \leq \hat{c}(\epsilon) e^{-N\hat{\beta}(\epsilon)} \tag{4.3.6}$$

for N sufficiently large.

Proof. Let $\epsilon > 0$ be any small positive number. By Lemma 4.3, there exists a $\delta(\epsilon) > 0$ such that if $\sup_{x\in\mathcal{X}} |\hat{f}_N(x) - f(x)| < \delta(\epsilon)$ then $\mathbb{D}(\hat{x}^*, \mathcal{S}^*) < \epsilon$.

On the other hand, under Assumption 4.1, by Proposition 4.1, we have that for the $\delta(\epsilon) > 0$ there exist positive constants $c(\delta(\epsilon))$ and $\beta(\delta(\epsilon))$ (for the simplicity of notation we write them as $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$) independent of N, such that

$$\mathbb{P}\left\{\sup_{x\in\mathcal{C}} |\hat{f}_N(x) - f(x)| \geq \delta(\epsilon)\right\} \leq \hat{c}(\epsilon) e^{-N\hat{\beta}(\epsilon)}.$$

for N sufficiently large.

Consequently, we have

$$\mathbb{P}\left\{\operatorname{dist}(\hat{x}^*, \mathcal{S}^*) \geq \epsilon\right\} \leq \mathbb{P}\left\{\sup_{x\in\mathcal{C}} |\hat{f}_N(x) - f(x)| \geq \delta(\epsilon)\right\} \leq \hat{c}(\epsilon)e^{-N\hat{\beta}(\epsilon)}$$

This completes the proof. □

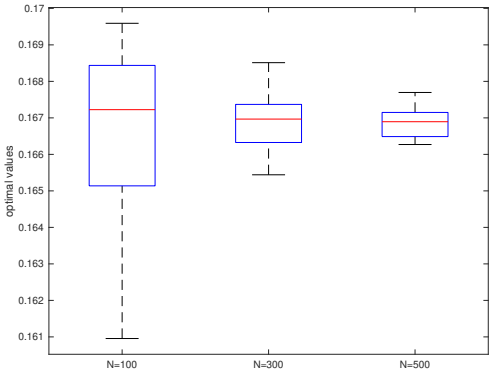
4.4 Numerical experiments

In this section, we aim to show the convergence of the SAA method. We still take the portfolio selection problem in Chapter 3 as an example. Specifically, we take the data sets 4, 5, 6 in Table 3.1 as examples. For these three data sets, by the rolling window method for generating M_i and q_i, we have $577, 577, 725$ samples in total, respectively.

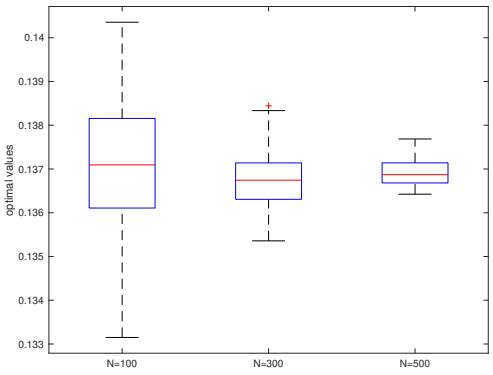
To investigate the behavior of the SAA problem, we randomly select $N = 100, 300, 500$ samples for data sets 4 and 5 and $N = 100, 400, 700$ samples for data set 6, and solve the corresponding SAA problem (4.1.1) with Algorithm 1. The parameters setting is also the same as those in Chapter 3. We repeat this process 25 times and box plot the objective value of problem (4.1.1).

All the computational tests were conducted on an Lenovo ThinkCentre Desktop (Intel core I7-12700 CPU, 2.1 GHz, 64 GB RAM) with using Matlab R2024b.

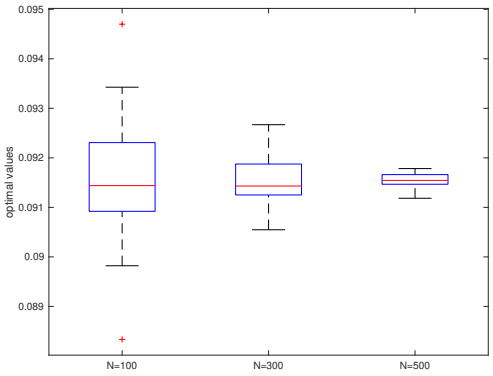
Figures 4.1–4.3 show the convergence trend of the function value as the sample size N increases.



(a) Risk-taking

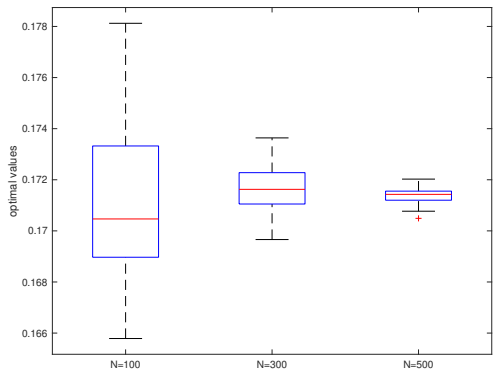


(b) Risk-neutral

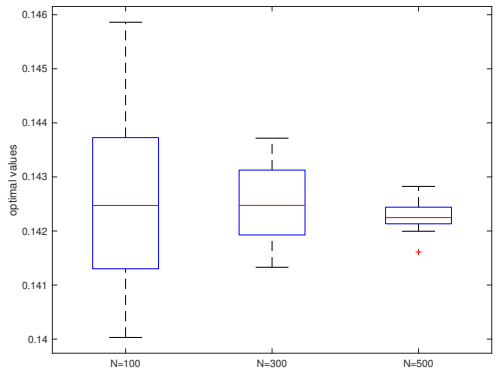


(c) Risk-averse

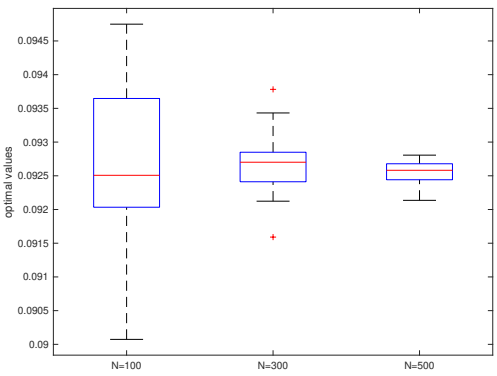
Figure 4.1: Boxplot of optimal values on data 4 for different sample size N.



(a) Risk-taking

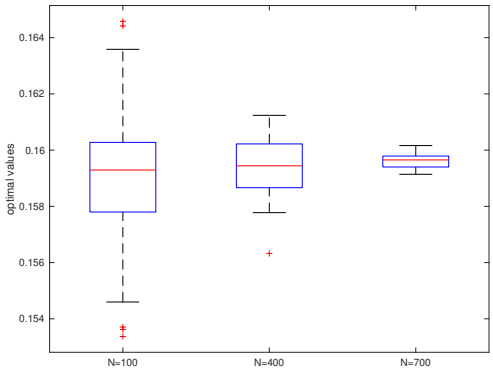


(b) Risk-neutral

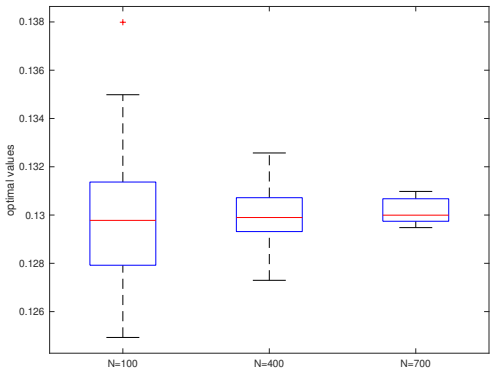


(c) Risk-averse

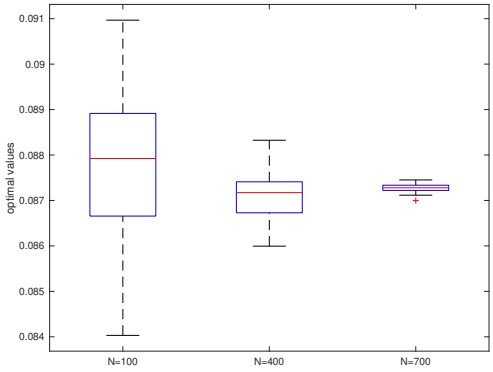
Figure 4.2: Boxplot of optimal values on data 5 for different sample size N.



(a) Risk-taking



(b) Risk-neutral



(c) Risk-averse

Figure 4.3: Boxplot of optimal values on data 6 for different sample size N.

Chapter 5

Conclusions and future work

In this chapter, we conclude the whole thesis and provide some possible remarks for the future work.

5.1 Conclusions

In this thesis, we propose stochastic optimization model (1.1.1) to find a robust solution of a system of SVLCPs and consider its applications in portfolio selection problems.

- We prove that there are penalty parameters such that problem (3.0.2) is an exact penalty problem of (1.1.1) with a finite support set regarding global minimizers and local minimizers. The objective function of problem (3.0.2) is piecewise smooth and the constraints are linear with block structure. We develop a smoothing block coordinate descent (SBCD) algorithm for solving problem (3.0.2). We prove the smoothing function satisfies the KL property and the sequence generated by the SBCD algorithm converges to an ϵ-Clarke stationary point of problem (3.0.2) by using the KL property. Our preliminary numerical experiments on portfolio selection problems with real data show the robustness of solutions obtained by our methods, comparing with the single-stage stochastic expected value formulation, the naive evenly weighted portfolio

and the portfolio arising from the distributionally robust optimization problem with the Kullback-Leibler divergence defined ambiguity set.

- We propose an implicit reformulation of the true problem and then approximate it with the SAA method. Next, we analyze the SAA convergence. Specifically, we show that, as the sample size goes to infinity, the optimal value and set of optimal solutions of the SAA problem converge to the counterparts of the true problem. Also we establish the uniform exponential convergence of the function values and translate it to the exponential convergence of an optimal solution of the SAA problem converging to the set of optimal solutions of the true problem. Finally, we also conduct extensive numerical experiments on the portfolio selection problems to show the convergence of the SAA method.

5.2 Future work

Some possible extensions for future work are listed below.

- Although it is extremely challenging, we are still interested in obtaining an explicit expression or at least an lower bound for the threshold value ρ^* in our exact penalty problem;

- We are interested in deriving the KL exponent for our problem, which will help us establish the convergence rate of our algorithm. Specifically, it might be interesting to investigate the KL exponent of a smoothing function if the original function has a KL exponent γ.

- As Remark 3.3 pointed out, if we choose the rule (3.2.25) for updating the smoothing parameter, is it possible that we can still show the whole sequence convergence with the KL property?

- A deeper discussion on the choice of smoothing function and the smoothed function satisfying the KL property would be a valuable topic for future research.
- We are also interested in stochastic approximation method and designing a stochastic algorithm for solving problem (4.1.2) directly instead of using the SAA method.

Bibliography

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, 2010.

[2] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.*, 59(2):341–357, 2013.

[3] M. J. Best and R. R. Grauer. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Rev. Financial Stud.*, 4(2):315–342, 1991.

[4] J. Bochnak, M. Coste, and M.-F. Roy. *Real Algebraic Geometry*. Springer-Verlag, Berlin, 1998.

[5] J. Bolte, A. Daniilidis, and A. Lewis. The lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.*, 17(4):1205–1223, 2007.

[6] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM J. Optim.*, 18(2):556–572, 2007.

[7] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Programming*, 146(1-2):459–494, 2014.

[8] J. V. Burke, X. Chen, and H. Sun. The subdifferential of measurable composite max integrands and smoothing approximation. *Math. Programming*, 181(2):229–264, 2020.

[9] B. Chen and X. Chen. A global and local superlinear continuation-smoothing method for p_0 and r_0 NCP or monotone NCP. *SIAM J. Optim.*, 18(2):556–572, 1999.

[10] X. Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Programming*, 134(1):71–99, 2012.

[11] X. Chen, Z. Lu, and T. K. Pong. Penalty methods for a class of non-Lipschitz optimization problems. *SIAM J. Optim.*, 26(3):1465–1492, 2016.

[12] X. Chen, H. Sun, and H. Xu. Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. *Math. Programming*, 177(1):255–289, 2019.

[13] S. Christiansen, M. Patriksson, and L. Wynter. Stochastic bilevel programming in structural optimization. *Struct. Multidiscip. Optim.*, 21(5):361–371, 2001.

[14] F. H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM, Philadephia, 1990.

[15] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. SIAM, Philadephia, 2009.

[16] S. Cui, U. V. Shanbhag, and F. Yousefian. Complexity guarantees for an implicit smoothing-enabled method for stochastic MPECs. *Math. Programming*, 198(2):1153–1225, 2023.

[17] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financial Stud.*, 22(5):1915–1953, 2009.

[18] B. C. Eaves. The linear complementarity problem. *Management Sci.*, 17(9):612–634, 1971.

[19] Y. M. Ermoliev and V. I. Norkin. Sample average approximation method for compound stochastic optimization problems. *SIAM J. Optim.*, 23(4):2231–2263, 2013.

[20] J. Gotoh, M. J. Kim, and A. E. Lim. Calibration of distributionally robust empirical optimization models. *Oper. Res.*, 69(5):1630–1650, 2021.

[21] M. S. Gowda. On the extended linear complementarity problem. *Math. Programming*, 72(1):33–50, 1996.

[22] M. S. Gowda and R. Sznajder. The generalized order linear complementarity problem. *SIAM J. Matrix Anal. Appl.*, 15(3):779–795, 1994.

[23] K. Guo, D. Han, and T.-T. Wu. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints. *International Journal of Computer Mathematics*, 94(8):1653–1669, 2017.

[24] X. Jia, C. Kanzow, and P. Mehlitz. Convergence analysis of the proximal gradient method in the presence of the Kurdyka-Lojasiewicz property without global lipschitz assumptions. *SIAM J. Optim.*, 33(4):3038–3056, 2023.

[25] K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier*, 48:769–783, 1998.

[26] H. A. Le Thi, T. Pham Dinh, and H. V. Ngai. Exact penalty and error bounds in DC programming. *J. Global Optim.*, 52(3):509–535, 2012.

[27] G. Li and T. K. Pong. Global convergence of splitting methods for nonconvex composite optimization. *SIAM J. Optim.*, 25(4):2434–2460, 2015.

[28] G. Li and T. K. Pong. Douglas-rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Math. Programming*, 159(1):371–401, 2016.

[29] G. Li and T. K. Pong. Calculus of the exponent of Kurdyka-Lojasiewicz inequality and its applications to linear convergence of first-order methods. *Found. Comput. Math.*, 18(5):1199–1232, 2018.

[30] X. Li, A. Milzarek, and J. Qiu. Convergence of random reshuffling under the kurdyka–łojasiewicz inequality. *SIAM J. Optim.*, 33(2):1092–1120, 2023.

[31] G.-H. Lin, X. Chen, and M. Fukushima. Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. *Math. Programming*, 116(1):343–368, 2009.

[32] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytique réels. *Coll. du CNRS*, pages 87–89, 1963.

[33] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, 1996.

[34] O. L. Mangasarian. Solution of general linear complementarity problems via nondifferentiable concave minimization. *Acta Math. Vietnam.*, 22:199–205, 1997.

[35] H. Markowitz. Portfolio selection. *J. Finance*, 7(1):77–91, 1952.

[36] W. Ouyang. Kurdyka-Lojasiewicz exponent via sqaure parametrization. *arXiv preprint arXiv:2506.10110*, 2025.

[37] W. Ouyang, Y. Liu, T. K. Pong, and H. Wang. Kurdyka-Lojasiewicz exponent via hadamard parametrization. *SIAM J. Optim.*, 35(1):62–91, 2025.

[38] M. Patriksson and L. Wynter. Stochastic mathematical programs with equilibrium constraints. *Oper. Res. Lett.*, 25(4):159–167, 1999.

[39] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, New York, 1998.

[40] A. Ruszczyński and A. Shapiro. Stochastic programming models. In *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 1–64. Elsevier, 2003.

[41] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.

[42] A. Shapiro and H. Xu. Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization*, 57(3):395–418, 2008.

[43] H. Sun and X. Chen. Two-stage stochastic variational inequalities: theory, algorithms and applications. *J. Oper. Res. Soc. China*, 9(1):1–32, 2021.

[44] R. Sznajder and M. S. Gowda. Generalizations of P_0- and P-properties; extended vertical and horizontal linear complementarity problems. *Linear Algebra Appl.*, 223-224:695–715, 1995.

[45] B. Wen, X. Chen, and T. K. Pong. A proximal difference-of-convex algorithm with extrapolation. *Comput. Optim. Appl.*, 69(2):297–324, 2018.

[46] H. Xu. Sample average approximation methods for a class of stochastic variational inequality problems. *Asia-Pacific Journal of Operational Research*, 27(01):103–119, 2010.

[47] H. Xu and D. Zhang. Stochastic nash equilibrium problems: sample average approximation and applications. *Comput. Optim. Appl.*, 55(3):597–645, 2013.

[48] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.*, 6(3):1758–1789, 2013.

[49] Y. Xu and W. Yin. A globally convergent algorithm for nonconvex optimization based on block coordinate update. *J. Sci. Comput.*, 72(2):700–734, 2017.

[50] L. Yang, T. K. Pong, and X. Chen. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. *SIAM J. Imaging Sci.*, 10(1):74–110, 2017.

[51] M. Yashtini. Multi-block nonconvex nonsmooth proximal admm: Convergence and rates under Kurdyka-Lojasiewicz property. *J. Global Optim.*, 190(3):966–998, 2021.

[52] P. Yu, G. Li, and T. K. Pong. Kurdyka-Lojasiewicz exponent via inf-projection. *Found. Comput. Math.*, 22(4):1171–1217, 2022.

[53] C. Zhang, X. Chen, and A. Sumalee. Robust Wardrop's user equilibrium assignment under stochastic demand and supply: Expected residual minimization approach. *Transp. Res. B Methodol.*, 45(3):534–552, 2011.