The Hong Kong Polytechnic University
Department of Applied Mathematics

Distributionally Robust Stochastic Variational Inequalities and Applications

Yanfang Zhang

A thesis submitted in partial fulfilment of
the requirements for the degree of Doctor of Philosophy

July 2012

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgment is made in the text.

Yanfang Zhang

To my family

Abstract

This thesis focuses on the stochastic variational inequality (VI). The stochastic VI has been used widely in engineering and economics as an effective mathematical model for a number of equilibrium problems involving uncertain data.

For a class of stochastic VIs, we present a new residual function defined by the gap function in Chapter 2. The expected residual minimization (ERM) formulation is a nonsmooth optimization problem with linear constraints. We prove the Lipschitz continuity and semismoothness of the objective function and the existence of minimizers of the ERM formulation. We show various desirable properties of the here and now solution, which is a minimizer of the ERM formulation.

In Chapter 3, we propose a globally convergent (a.s.) smoothing sample average approximation (SSAA) method for finding a minimizer of the ERM formulation. We show that the SSAA problems of the ERM formulation have minimizers in a compact set, and any cluster point of minimizers (stationary points) of the SSAA problems is a minimizer (a stationary point) of the ERM formulation (a.s.) as the sample size $N \to \infty$ and the smoothing parameter $\mu \downarrow 0$.

We discuss the ERM formulation for the stochastic linear VI in Chapter 4, which is convex under some mild conditions. We apply the Moreau-Yosida regularization to present an equivalent smooth convex minimization problem. To have the convexity of the sample average approximation (SAA) problems of the ERM formulation, we adopt the Tikhonov regularization. We show that any cluster point of minimizers of the Tikhonov regularized SAA problems is a minimizer of the ERM formulation as the

sample size $N \to \infty$ and the Tikhonov regularization parameter $\varepsilon \to 0$. Moreover, we prove that the minimizer is the least l_2-norm solution of the ERM formulation. We also prove the semismoothness of the gradients of the Moreau-Yosida and Tikhonov regularized SAA problems.

In Chapter 5, we discuss the distributionally robust stochastic linear VI based on the ERM formulation. We introduce the CVaR formulation defined by the ERM formulation and establish the relationship between the CVaR formulation and the ERM formulation. For a wide range of cases, we show that the two formulations have the same minimizers. Moreover, we derive the gradient consistency for the smoothing CVaR formulation. We employ the sublinear expectation to consider the distributionally robust CVaR formulation for the stochastic linear VI, and prove the existence of minimizers of the robust CVaR formulation.

We provide applications arising from traffic flow problems for stochastic VI in Chapter 6. We show the conditions and assumptions imposed in this thesis hold in such applications. Moreover, numerical results illustrate that the solutions, efficiently generated by the ERM formulation, have desirable properties.

This thesis is based on the following papers written during the period of my study at the Department of Applied Mathematics, The Hong Kong Polytechnic University as a graduate student:

1. Y. Zhang, "Numerical investigation of deterministic formulations for stochastic complementarity problems", Pac. J. Optim., 7 (2011), pp. 317-337.
2. X. Chen, R. J-B. Wets, and Y. Zhang, "Stochastic variational inequalities: residual minimization smoothing/sample average approximations", SIAM J. Optim., 22 (2012), pp. 649-673.
3. Y. Zhang and X. Chen, "Moreau-Yosida regularization for stochastic linear variational inequalities", to appear in J. Optim. Theory Appl.

Acknowledgments

Over the past three years, my supervisor Professor Xiaojun Chen has perfected a good role of advisor, mentor and friend. First and foremost, I would like to express my appreciation to her for her valuable advice, kind support and patient guidance. Her insightful and expansive knowledge of optimization constantly makes me amazing, and her encouragement and suggestions for my research are crucial. In fact, the subject for this thesis is following her advice, and she led me to learn more on this subject. My interests for doing research on this subject are raised by her. She helped me to overcome the difficulties I faced during the past days. Especially, I wish to owe my special thanks to her for her careful reading and making modifications on the draft manuscripts of this thesis. It has been valuable and pleasure for me to study under her supervision, and I shall remain appreciative for all that she has done for me.

I would like to thank Professor Roger J-B Wets (University of California, Davis) for his help and valuable suggestions on my research. During his visit here, I learnt a lot of mathematical knowledge from him for the face to face discussing. I would also like to thank Professor Yinyu Ye (Stanford University, USA) for his helpful discussion and comments on my research. I wish to show my thanks to Professor Naihua Xiu (Beijing Jiaotong University, China) for his great encouragement and insightful guidance. I am very grateful to thank Dr. Chao Zhang (Beijing Jiaotong University, China) for her helpful advices on my research.

I would also like to thank all my teachers in The Hong Kong Polytechnic University, especially to Professor Xiaoqi Yang for his encouragement and attention to my thesis.

I wish to show my thanks to the staff of the Department of Applied Mathematics at The Hong Kong Polytechnic University for their help during my study here.

Several people had an enormous help on me over the past years. To Dr. Congpei An, Dr. Wei Bian, Dr. Weijun Zhou and Mr. Yang Zhou, I am grateful to thank them for their help and valuable comments on my research. The colleagues and friends in DE 405 and DE 407 (Department of Applied Mathematics) make my study and life interesting and colorful here, and I wish to pay special thanks to them.

I would like to acknowledge The Hong Kong Polytechnic University for offering me financial support over the past three years. The support made me to have various chances to communicate and exchange ideas with researchers. I like the research environment and the excellent facilities of The Hong Kong Polytechnic University and shall always treasure them.

Last but not the least, I would like to thank my parents. They always provide me with unconditional and consistent support for my study. I can not imagine having results without their love, guidance and support.

Contents

List of Figures

List of Tables

Index of Notation

Spaces and Orthants

R: the real numbers

$\overline{R}$: the real numbers plus $\pm\infty$

R_+: the right half line

R_{++}: the set of positive numbers

R^n: the real n-dimensional space

R^n_+: the nonnegative orthant of R^n

R^n_{++}: the positive orthant of R^n

$R^{m\times n}$: the space of $m \times n$ real matrices

Matrices

A: $\equiv (a_{ij})$, a matrix with entries a_{ij}

A^T: the transpose of a matrix A

A_K: the submatrix of A with column-index in the index set $K \subseteq \{1, \cdots, n\}$ of cardinality $|K|$

I: an identity matrix

Vectors

x^T: the transpose of a vector x with components x_i

$||x||$: the l_2-norm of $x \in R^n$

e: a vector whose elements are all 1

Chapter 1

Preview and Introduction

1.1 The stochastic variational inequalities

The variational inequality (VI), the special case of which is the complementarity problem (CP), has a wide applicability across many fields. For instance, it can be used as an effective model for a number of equilibrium problems in engineering and economics. The VI provides a bridge for the study of optimization and equilibrium problems. Mathematically, the VIs arise from the constrained optimization and the Karush-Kuhn-Tucker (KKT) system. The study of the finite-dimensional VI and CP began in the mid-1960s, and had many fundamental results and articles in 1970s [22, 26].

The classic VI is the problem of finding $x \in X$ that satisfies the inclusion $-F(x) \in N_X(x)$ denoted by VI(X, F), also written as,

$$\text{find}\, x \in X \ \text{ such that } \ (u - x)^T F(x) \geq 0, \quad \forall\, u \in X;$$

here $F : R^n \to R^n$ is a continuous function, $X \subseteq R^n$ a (nonempty) closed, convex set and $N_X(x)$ is the normal cone to X at x.

It turns out that the classic VI or CP can be formulated as an equivalent formulation in terms of systems of equations or optimization problems. The VI is casted via a deterministic minimization problem by using a residual function for the VI.

Definition 1.1.1 [26] *A residual function for the VI(X, F) on a (closed) set $D \supseteq X$ is a nonnegative function $f : D \to R_+$ such that $f(x) = 0$ if and only if $x \in D$ solves the VI(X, F).*

Many residual functions for the VI have been extensively studied. Specifically, the gap function, as a basis of some residual functions for the VI, is given by:

$$g(x) = \max_{y \in X} F(x)^T (x - y), \qquad x \in D \supseteq X.$$

Based on the gap function, Fukushima in [29] introduced the regularized gap function as

$$g_c(x) = \max_{y \in X} \{F(x)^T (x - y) - \frac{c}{2}(x - y)^T G(x - y)\},$$

where c is a positive number, G is a symmetric positive definite matrix and $x \in D$.

Peng in [49] gave another residual function, the D-gap function, as

$$g_{ab}(x) = g_a(x) - g_b(x),$$

where $b > a > 0$ and $x \in R^n$.

In this thesis, our interest is the VI in a stochastic environment, which we consider it as the stochastic VI. The stochastic VI is a natural extension of VI, and it has been studied in [2, 17, 30, 31, 35, 40, 57, 63]. However, most of these articles focus on the feasible sets of the stochastic VI are deterministic. In this thesis, we consider the case where both the function and the feasible set of the stochastic VI have uncertainties.

Consider the stochastic VI where $F : \Xi \times R^n \to R^n$ is continuously differentiable in x for every $\xi \in \Xi \subseteq R^L$ and measurable in ξ for every $x \in R^n$ and

$$X_\xi = \{\, x \mid Ax = b_\xi, \quad x \geq 0\}$$

with a given matrix $A \in R^{m\times n}$ and a random vector b_ξ taking values in R^m. If $X_\xi = R^n_+$, the stochastic VI simplifies to a stochastic complementarity problem (CP).

When the function $F(\xi, x)$ is affine for almost every $\xi \in \Xi$, i.e.

$$F(\xi, x) = M_\xi x + q_\xi, \qquad a.s.,$$

the above stochastic VI reduces to the stochastic linear VI.

In this thesis, we suppose that for any $\xi \in \Xi$, the feasible set X_ξ is nonempty and the matrix A has full-row rank. In some applications, A is an incidence matrix whose entries are either 0 or 1 but the function F and the vector b depend on stochastic parameters, e.g., traffic equilibrium problems, Nash-Cournot production/distribution problems, etc.

A good formulation of the VI, in a stochastic environment, when either F, or X, or both, depend on stochastic parameters is not straightforward. Even, when just F involves stochastic parameters, say ξ, one might be led to consider a variety of formulations: find $x \in X$ such that

$$\text{prob}\big\{ -F(\xi, x) \in N_X(x)\big\} \geq \alpha, \quad \text{or} \quad -F(\hat{\xi}, x) \in N_X(x)$$

$$\text{or still} \quad E[-F(\xi, x)] \in N_X(x), \tag{1.1.1}$$

where $\alpha \in (0, 1]$, $\hat{\xi}$ stands for a guess of the future and $E[\cdot]$ denotes the expected value over $\Xi \subseteq R^L$, a set representing future states of knowledge. The last two formulations are essentially deterministic variational inequalities, the only issues being how to calculate $E[-F(\xi, x)]$ for the last one and having an undeniable capability to know the future for the second one; one might consider setting $\hat{\xi} = E[\xi]$ but that has been discredited repeatedly including in this thesis. The first formulation with $\alpha = 1$ could be converted to a large VI, involving an infinite number of inequalities when ξ is continuously distributed, that only exceptionally would have a solution. When $\alpha \in (0, 1)$, the problem takes on the form of a 'chance constraint' and would actually be quite challenging to come to grips with theoretically and computationally and this, in addition to having to validate the choice of the α. When, also the set X depends on ξ, a meaning can still be attached to the first two of these formulations but the comments made earlier about such formulations remain valid, even more so. When seeking to mimic the third formulation one runs quickly into difficulties when trying to justify replacing X_ξ by its expectation or try to compute $E[N_{X_\xi}(x) + F(\xi, x)]$.

The following two deterministic formulations have been studied for the stochastic VI when the feasible set is fixed.

- Expected Value (EV) formulation [30, 31, 35, 57, 63]: find $x \in X$ such that

$$(y - x)^T E[F(\xi, x)] \geq 0, \qquad \forall\, y \in X. \tag{1.1.2}$$

- Expected Residual Minimization (ERM) formulation [2, 16, 20, 27, 39, 40, 68, 69]:

$$\min_{x \in X} E[f(\xi, x)], \tag{1.1.3}$$

where $f(\xi, \cdot) : X \to R_+$ is a residual function for the VI$(X, F(\xi, \cdot))$ for fixed $\xi \in \Xi$.

As already pointed out, the EV formulation can be viewed as a deterministic VI$(X, \bar{F})$ with the expected function $\bar{F}(x) = E[F(\xi, x)]$. Using mean values or some other estimates for the uncertain parameters in the model may lead to seriously misleading decisions. The ERM formulation proposed by Chen and Fukushima in [16] minimizes the expected values of the 'loss' for all possible scenarios due to failure of the equilibrium. Mathematical analysis and practical examples show that the ERM formulation is robust in the sense that its solution has minimum sensitivity with respect to variations in the random parameters.

For the stochastic CP, the ERM formulation defined by different residual functions has different properties, such as smoothness and boundness [16, 20, 27, 67]. For the stochastic VI, Agdeppa et al. in [2] studied the ERM formulation given by the regularized gap function and the D-gap function to illustrate the convexity of the stochastic linear VI. Chen and Lin in [17] employed the D-gap function as a risk function to define the Conditional Value-at-Risk (CVaR) [54, 55] for the stochastic VI. In this thesis, we pay our attention to the ERM formulation defined by a new residual function which is given by the gap function and the recourse variable. We find that the condition for the convexity of our new residual function for the stochastic linear VI is weaker than that is used by Agdeppa et al. in [2].

1.2 Robust optimization for the stochastic VI

The ERM formulation aims to reduce the total loss of the decisions for all scenarios, and the numerical results in Chapter 6 show that it can give a robust optimal solution. It is defined by the expected value of random variables and it is a stochastic programming. It is inevitable that we must resort to Monte Carlo approximation [12, 57, 59] to get the expected value. A difficulty thing, which is lead by the limit information of a distribution for the random variables in practice, is more challenging and needs to be considered.

In recent years, robust optimization for the stochastic programming has attracted much attention to reduce the influence of the parametric uncertainties on the mathematical optimization. The first step in this area was taken by Scarf in [58], who defined a set of probability distributions, which is assumed to include the true distribution. Recent developments in robust optimization focus on the uncertainty sets [4, 6, 7, 8, 9, 10, 11, 18, 23, 24, 25, 32, 46] to develop some new models to overcome the conservatism of the old ones.

It is worth noting that in the field of the stochastic programming, there are already abundant results for theoretical analysis and algorithm for the linear expectation; see Birge and Louveaux [12], Kall and Wallace [36], and Ruszczyński and Shapiro [57]. However, we find that there are few references concern on a robust statistics. The robust statistics produces estimators that are not unduly affected by small changes of the model assumptions, and the standard methods may be comparatively badly affected.

The sublinear expectation [34, 51, 61] is a robust statistics and it is also called the upper expectation. The sublinear expectation is used in the situations when the probability models have uncertainty. We focus on the sublinear expectation to consider distributionally robust stochastic VI.

Let Ω be a given set and $\mathcal{H}$ be a linear space of real valued functions defined on Ω. We suppose that $\mathcal{H}$ satisfies $c \in \mathcal{H}$ for each constant c and $|\mathcal{X}| \in \mathcal{H}$ if $\mathcal{X} \in \mathcal{H}$.

Definition 1.2.1 *A sublinear expectation* $\mathbb{E}$ *is a functional* $\mathbb{E} : \mathcal{H} \to R$ *satisfying*

- *Monotonicity:* $\mathbb{E}[\mathcal{X}] \geq \mathbb{E}[\mathcal{Y}]$ *if* $\mathcal{X} \geq \mathcal{Y}$.
- *Constant preserving:* $\mathbb{E}[c] = c$ *for* $c \in R$.
- *sub-additivity: For each* $\mathcal{X}, \mathcal{Y} \in \mathcal{H}$, $\mathbb{E}[\mathcal{X} + \mathcal{Y}] \leq \mathbb{E}[\mathcal{X}] + \mathbb{E}[\mathcal{Y}]$.
- *Positive homogeneity:* $\mathbb{E}[\lambda \mathcal{X}] = \lambda \mathbb{E}[\mathcal{X}]$ *for* $\lambda \geq 0$.

It is not difficult to see that if $\mathbb{E}$ is a sublinear expectation and $\rho(\mathcal{X}) := \mathbb{E}[-\mathcal{X}]$, we can gain ρ is a **coherent risk measure** [3] which is a function satisfying the following properties:

(i) Monotonicity: For all $\mathcal{X}, \mathcal{Y} \in \mathcal{H}$ with $\mathcal{X} \geq \mathcal{Y}$, we have $\rho(\mathcal{X}) \geq \rho(\mathcal{Y})$.

(ii) Translational invariance: For all $\mathcal{X} \in \mathcal{H}$ and $\forall\, c \in R$, we have $\rho(\mathcal{X}+c) = \rho(\mathcal{X}) - c$.

(iii) Sub-additivity: For all $\mathcal{X}, \mathcal{Y} \in \mathcal{H}$, we have $\rho(\mathcal{X} + \mathcal{Y}) \leq \rho(\mathcal{X}) + \rho(\mathcal{Y})$.

(iv) Positive homogeneity: For all $\mathcal{X} \in \mathcal{H}$ and $\forall\, \lambda \geq 0$, we have $\rho(\lambda \mathcal{X}) = \lambda \rho(\mathcal{X})$.

1.3 Outline of the thesis

We make a brief preview for the rest chapters. In Chapter 2, we present a residual minimization (ERM) formulation defined by a new residual function for a class of stochastic

VI. We get that the objective function of the ERM formulation is Lipschitz continuous and semismooth which can guarantee the existence of optimal solutions of the ERM formulation. We define a here and now solution with various desirable properties and it is a minimizer of the ERM formulation. The ERM formulation of the stochastic CP, which is a special case of the stochastic VI, is also studied in this chapter. We find that under some suitable condition, optimal solutions of the ERM formulation are the same as that of the EV formulation for the stochastic CP.

In Chapter 3, smoothing sample average approximation (SSAA) method is studied. We use SSAA method to find optimal solutions of the ERM formulation. We get that SSAA method is globally convergent almost surely (a.s.). We show that optimal solutions of the SSAA problem of the ERM formulation exist in a compact set, and any cluster point of optimal solutions and stationary points of the SSAA problems is an optimal solution and a stationary point of the ERM problem (a.s.) as the sample size $N \to \infty$ and the smoothing parameter $\mu \downarrow 0$.

In Chapter 4, we focus on the stochastic linear VI which has a wide range of applications. The ERM formulation given in Chapter 2 is considered for the stochastic linear VI. To ensure the convexity of the approximation we use the Tikhonov regularization. Moreover, we employ the Moreau-Yosida regularization to get a smooth and convex approximation. We derive good theoretical properties of the Moreau-Yosida regularization for the stochastic linear VI, such as the semismoothness of the gradient of the Moreau-Yosida regularization and the convergence of the optimal solutions as the sample size $N \to \infty$ and the Tikhonov regularization parameter $\varepsilon \to 0$.

We discuss the distributionally robust stochastic linear VI using the sublinear expectation in Chapter 5. We use the ERM formulation to define the CVaR formulation for the stochastic linear VI. The relationship for the ERM formulation and the CVaR formulation is studied, and optimal solutions of the two formulations are the same for some cases. For the smoothing CVaR formulation, we can obtain its gradient consistency. The existence of optimal solutions of the robust CVaR formulation is proved. For a wide range of residual functions, we can get the explicit form of the sublinear expectation, and the distributionally robust stochastic VI can be solved efficiently.

We consider the applications and numerical experiments of the stochastic VI in Chapter 6. We concentrate our attention on the traffic equilibrium for the applications. We show that conditions and assumptions imposed in this thesis hold in such applications. Moreover, the numerical results of the SSAA method and the Moreau-Yosida regularization for the stochastic linear VI illustrate that the ERM formulation has good properties, such as robustness and high probability.

Chapter 2

The ERM formulation for the stochastic VI

2.1 Introduction

A good deterministic formulation for the stochastic VI is necessary and important. Although we have introduced some deterministic formulations for the stochastic VI in Chapter 1, there is another way to formulate the problem, even when both F and X are stochastic, that comes with a 'natural' interpretation and leads, at least in the case we shall consider, to implementable algorithmic procedures. For each realization ξ of the random quantities, let $g(\xi, x)$ be a function that measures the compliance gap, i.e., a nonnegative function such that $g(\xi, x) = 0$ if and only if $-F(\xi, x) \in N_{X_\xi}(x)$. The values to assign to $g(\xi, x)$ could depend on the specific application but usually it would be a relative of the gap function and solving the problem would be to minimize $E[g(\xi, \cdot)]$ or some other risk measure associated with the random variable $g(\xi, \cdot)$. It is this latter approach that will be developed in this chapter for the particular class of stochastic VIs described in the following sections.

The main contribution of this chapter is to show that the ERM formulation,

$$\min_{x\in D} \varphi(x) = E[f(\xi, x)], \tag{2.1.1}$$

defined by the new residual function has various desirable properties.

In this chapter, we focus on a new residual function $f(\xi, x)$ defined by the gap function given in Section 2.2. We show that the function $f(\xi, x)$ is a residual function for the stochastic VI, and it is measurable in ξ for any fixed x and locally Lipschitz continuous in x. In Section 2.3, we show the objective function $\varphi(x)$ of the ERM formulation defined by the residual function $f(\xi, x)$ is Lipschitz continuous and semismooth. Moreover, we prove the existence of solutions of the ERM formulation. For the linear case when $F(\xi, x) = M_\xi x + q_\xi$, we show that $\varphi(x)$ is convex if $E[M_\xi]$ is positive semi-definite. The ERM formulation defined by the residual function for the stochastic CP, the special case of the stochastic VI, is also introduced in Section 2.4.

2.2 A new residual function

To allow for the dependence of the set X on $\xi \in \Xi$, one needs to extend Definition 1.1.1 of the residual function for the classical VI to the stochastic VI.

Definition 2.2.1 *Let $D \subseteq R^n$ be a closed and convex set. $f : \Xi \times D \to R_+$ is a residual function of the stochastic VI, if the following conditions hold,*

(i) For any $x \in D$, $\text{prob}\{ f(\xi, x) \geq 0\} = 1$.

(ii) $\exists\ u : \Xi \times D \to R^n$ such that for any $x \in D$ and almost every $\xi \in \Xi$, $f(\xi, x) = 0$ if and only if $u(\xi, x)$ solves the VI$(X_\xi, F(\xi, \cdot))$.

From Definition 1.1.1, we can see that Definition 2.2.1 is a natural extension of Definition 1.1.1. Moreover, the residual function can be used to provide error bounds on the distance from x to the solution set of VI$(X_\xi, F(\xi, \cdot))$. See [26].

The 'natural' residual function

$$\|x - \text{proj}_{X_\xi}(x - F(\xi, x))\|^2$$

is a residual function for the stochastic VI with $D = R^n$ and $u(\xi, x) = x$. Here proj_{X_ξ} is the orthogonal projection of R^n onto X_ξ and $\|\cdot\|$ is the ℓ_2 norm. When $X_\xi = R^n_+$, one has

$$x - \text{proj}_{X_\xi}(x - F(\xi, x)) = \min(x, F(\xi, x)).$$

Other possible residual functions may be defined via the KKT conditions in the primal-dual variable $(x, v) \in R^{n+m}$

$$0 \leq F(\xi, x) + A^T v \perp x \geq 0, \quad Ax - b_\xi = 0.$$

However, in the 'natural' residual function and the KKT condition, there are not recourse variables.

In this thesis, we rely on the gap function [26, Section 1.5] to define a new residual function. The gap function provides a measure for the deviations that will be needed to 'adjust' the solution of the VI as it is affected by the circumstances, i.e., the random components of the problem.

For given ξ, the gap function for the VI$(X_\xi, F(\xi, \cdot))$ is defined by

$$g(\xi, x) = \max\{\, (x - y)^T F(\xi, x) \mid y \in X_\xi\}.$$

It is easy to see that $g(\xi, x) \geq 0$ for $x \in X_\xi$ and it is known that the VI$(X_\xi, F(\xi, \cdot))$ is equivalent to the minimization problem [26, Section 1.5.3]

$$\min\nolimits_{x \in X_\xi} g(\xi, x). \tag{2.2.1}$$

This minimization problem (2.2.1) can be written as a two stage optimization problem

$$\begin{aligned} \min \quad & x^T F(\xi, x) + Q(\xi, x) \\ \text{s.t.} \quad & x \in X_\xi \\ & Q(\xi, x) = \max\{\, -y^T F(\xi, x) \mid y \in X_\xi \}; \end{aligned} \tag{2.2.2}$$

from linear programming duality it follows that Q can also be written,

$$Q(\xi, x) = \min\{\, z^T b_\xi \mid A^T z + F(\xi, x) \geq 0\}. \tag{2.2.3}$$

Suppose that for any $x \in R^n$ the recourse variable $u(\xi, x)$ is defined by the projection of x on the set X_ξ. For any fixed $\xi \in \Xi$, to get $u(\xi, x)$ we should solve the following optimization problem:

$$\begin{aligned} \min_u \quad & \frac{1}{2}\|u - x\|^2 \\ & Au = b_\xi \\ & u \geq 0. \end{aligned}$$

We find that it is not easy to get the explicit form of $u(\xi, x)$ and for almost every $\xi \in \Xi$, we need to solve the above optimization problem. To avoid the complicated computation, we consider the recourse variable $u(\xi, x)$ as the projection of $x \in R^n$ on the set $\{x | Ax = b_\xi\}$ and obtain the following optimization problem

$$\begin{aligned} \min_u \quad & \frac{1}{2}\|u - x\|^2 \\ & Au = b_\xi, \end{aligned}$$

and we obtain $u(\xi, x) = (I - A^\dagger A)x + A^\dagger b_\xi$, where $A^\dagger$ is a generalized inverse matrix of the matrix A. If the matrix A has full-row rank, we can get $A^\dagger = A^T(AA^T)^{-1}$. Since $u(\xi, x)$ is a feasible point, to guarantee the nonnegativity of $u(\xi, x)$ we suppose that the point x belongs to a constraint set D.

Let

$$D = \{\, x \,|\, (A^{\dagger}A - I)x \leq \underline{c}\},$$

where $\underline{c}_i = \min_{\xi\in\Xi}(A^{\dagger}b_{\xi})_i$, for $i = 1, \ldots, m$.

In this thesis, we suppose that the set D is nonempty, and the following two conditions can guarantee $D \neq \emptyset$:

i. For any $\xi \in \Xi$, $A^{\dagger}b_{\xi} \geq 0$ holds. (Application OK!)

ii. For any $\xi \in \Xi$, $X_{\xi} \neq \emptyset$ and $\mathrm{argmin}_{\xi\in\Xi}(\mathrm{A}^{\dagger}\mathrm{b}_{\xi})_{\mathrm{i}} \bigcap \mathrm{argmin}_{\xi\in\Xi}(\mathrm{A}^{\dagger}\mathrm{b}_{\xi})_{\mathrm{j}} \neq \emptyset$ for $i \neq j$.

It is not difficult to verify that $u(\xi, x)$ satisfies the KKT conditions

$$0 \leq u - x + A^T v \perp u \geq 0 \quad \text{and} \quad Au = b_{\xi},$$

with Lagrange multiplier $v = (AA^T)^{-1}(Ax - b_{\xi})$, of the following convex minimization problem

$$\min\{\frac{1}{2}\|u - x\|^2 \,|\, Au = b_{\xi}, \quad u \geq 0\}$$

for a fixed $x \in D$. Hence, for any $x \in D$ and almost every $\xi \in \Xi$,

$$u(\xi, x) = \mathrm{proj}_{X_{\xi}}(x). \tag{2.2.4}$$

In this thesis, we rely on the residual function defined by the above gap function as follows:

$$f(\xi, x) = u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, u(\xi, x)), \tag{2.2.5}$$

where $u(\xi, x)$ and $Q(\xi, u(\xi, x))$ are defined by the above formulations.

Assumption 2.2.1 *Assume that for all $x \in D$ and for almost every $\xi \in \Xi$,*

$$\exists\, y(\xi, x) \text{ such that } Q(\xi, u(\xi, x)) = -y(\xi, x)^T F(\xi, u(\xi, x)).$$

Rather than assuming that the second stage program is feasible for all $u \in X_\xi$, Assumption 2.2.1 only requires that it is feasible for a much more restricted class, namely, those $u = \mathrm{proj}_{X_\xi}(x)$ when $x \in D$. In Chapter 6, we show that Assumption 2.2.1 holds for a class of matrices A and vectors b_ξ that arise from traffic equilibrium problems.

Theorem 2.2.1 *When Assumption 2.2.1 is satisfied, $f : \Xi \times D \to R$, as defined earlier $f(\xi, x) = u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, u(\xi, x))$, is a residual function for our stochastic VI.*

Proof. Let $x \in D$. By the definition of $u(\xi, x)$, we have $Au(\xi, x) = b_\xi$ and

$$u(\xi, x) = (I - A^\dagger A)x + A^\dagger b_\xi \geq (I - A^\dagger A)x + \underline{c} \geq 0.$$

Hence $u(\xi, x) \in X_\xi$. By definition of $f(\xi, x)$ and Assumption 2.2.1, for almost every $\xi \in \Xi$, there is $y(\xi, x) \in R^n$ such that

$$\begin{aligned} f(\xi, x) &= u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, u(\xi, x)) \\ &= u(\xi, x)^T F(\xi, u(\xi, x)) - y(\xi, x)^T F(\xi, u(\xi, x)) \\ &= \max\{(u(\xi, x) - y)^T F(\xi, u(\xi, x)) \,|\, y \in X_\xi\} \\ &\geq 0, \end{aligned}$$

where the last inequality follows from $u(\xi, x) \in X_\xi$. Hence, we obtain $\mathrm{prob}\{f(\xi, x) \geq 0\} = 1$. Moreover, $f(\xi, x) = 0$ if and only if $u(\xi, x)$ solves the $\mathrm{VI}(X_\xi, F(\xi, \cdot))$ a.s. ■

Assumption 2.2.2 *(i) There are $\underline{b}$, $\bar{b} \in R^m$ such that $\underline{b} \leq b_\xi \leq \bar{b}$ for $\forall \xi \in \Xi$; $E[\|b_\xi\|] < \infty$,*

(ii) $\exists\, d : \Xi \to R_+$ such that $\|F(\xi, u)\| \leq d(\xi)$ for all $u \in U$ and $E[d(\xi)] < \infty$,

(iii) $\exists\, d_1 : \Xi \to R_+$, bounded, such that $\|\nabla F(\xi, u)\| \leq d_1(\xi)$ for all $u \in U$,

(iv) $\exists\, \gamma > 0$ such that $X_\xi \subset U_0 = \{u \in R^n \,|\, \|u\|_\infty \leq \gamma\}$ for any $\xi \in \Xi$.

Assumption 2.2.2(i)-(iii) are pretty standard and are in no way restrictive as far as applications are concerned. Assumption 2.2.2(iv) is not quite as common but, in particular, is satisfied by the class of problems considered in Chapter 6.

Since $u(\xi, x) = (I - A^{\dagger}A)x + A^{\dagger}b_{\xi}$ is a linear function of x and $u(\xi, x) \in U$ for any $x \in D$, for almost every $\xi \in \Xi$, we immediately obtain the following proposition.

Proposition 2.2.1 *$F(\xi, u(\xi, x))$ is measurable in ξ for every $x \in D$. Moreover, for any fixed $\xi \in \Xi$, the following hold.*

(i) $F(\xi, u(\xi, x))$ is continuously differentiable with respect to x.

(ii) If (ii) and (iii) of Assumption 2.2.2 hold, then for all $x \in D$,

$$\|F(\xi, u(\xi, x))\| \leq d(\xi) \quad \text{and} \quad \|\nabla_x F(\xi, u(\xi, x))\| \leq \|I - A^{\dagger}A\| d_1(\xi).$$

Theorem 2.2.2 *Assume that Assumption 2.2.1 holds. Then, the function f is measurable in ξ for any $x \in D$ and locally Lipschitz continuous in x a.s.; actually, under Assumption 2.2.2(iii), the functions $\{f(\xi, \cdot) : D \to R,\ \xi \in \Xi\}$ are then also equi-locally Lipschitz continuous a.s.*

Proof. Since $u(\xi, x)$ is linear in x, by Proposition 2.2.1, we only need to consider $F(\xi, u)$ for $u \in U$.

For any $u, v \in U$ and almost every $\xi \in \Xi$, there are $z(\xi, u), z(\xi, v) \in R^m$ such that $Q(\xi, u) = b_{\xi}^T z(\xi, u)$ and $Q(\xi, v) = b_{\xi}^T z(\xi, v)$. By perturbation error analysis for linear programs in [41], there is a constant $\nu_A > 0$, that only depends on the matrix A, such that

$$\|Q(\xi, u) - Q(\xi, v)\| \leq \|b_{\xi}\| \|z(\xi, u) - z(\xi, v)\| \leq \|b_{\xi}\| m \nu_A \|F(\xi, u) - F(\xi, v)\| \ a.s. \quad (2.2.6)$$

Since for any fixed $\xi \in \Xi$, $F(\xi, \cdot)$ is continuously differentiable in x, $Q(\xi, \cdot)$ is locally Lipschitz continuous in x a.s. with, in view of Assumption 2.2.2(iii), the (local) Lipschitz constant not depending on ξ. From this it follows that for any fixed $\xi \in \Xi$, the two terms in $f(\xi, \cdot)$ are locally Lipschitz continuous in x with Lipschitz constant not depending on ξ. Hence, the collection $\{f(\xi, \cdot),\ \xi \in \Xi\}$ is then equi-locally Lipschitz continuous in x, a.s. Recall that $F(\xi, x)$ is measurable in ξ for every $x \in R^n$ and b_ξ is measurable in ξ. We have that $Q(\xi, u)$ is measurable in ξ for any $u \in U$, cf. [57, Theorem 19, Chapter 1]. Hence the function $f(\xi, x)$ is measurable in ξ for any $x \in R^n$. ■

2.3 The ERM formulation for stochastic VI

By the residual function f, we get our ERM formulation (2.1.1) with the objective function:

$$\varphi(x) = E[f(\xi, x)] = E[u(\xi, x)^T F(u(\xi, x))] + E[Q(\xi, u(\xi, x))].$$

By Theorem 2.2.1, $\varphi(x) \geq 0$ for all $x \in D$ and if $\varphi(x) = 0$ then, $u(\xi, x)$ solves the VI($X_\xi, F(\xi, \cdot)$) for almost every $\xi \in \Xi$. Hence the "here and now" solution is

$$x_{\text{ERM}} = E[u(\xi, x^*)] = x^* + A^\dagger(E[b_\xi] - Ax^*),$$

where x^* is a solution of the ERM formulation (2.1.1). By definition of $u(\xi, x)$,

$$Ax_{\text{ERM}} = E[b_\xi] \quad \text{and} \quad x_{\text{ERM}} \geq 0. \tag{2.3.7}$$

Moreover, the following proposition shows that x_{ERM} is also a solution of our ERM formulation (2.1.1).

Proposition 2.3.1 *Under Assumption 2.2.1, if (2.1.1) has a solution x^*, then*

$$x_{ERM} \in \operatorname{argmin}_{x \in D} \varphi(x). \tag{2.3.8}$$

Proof. For $x \in D$, let $\bar{u} = E[u(\xi, x)] == (I - A^\dagger A)x + A^\dagger E[b_\xi]$, and we have

$$\begin{aligned}(A^\dagger A - I)\bar{u} &= (A^\dagger A - I)((I - A^\dagger A)x + A^\dagger E[b_\xi]) \\ &= (A^\dagger A - I)(I - A^\dagger A)x + (A^\dagger A - I)A^\dagger E[b_\xi] \\ &= (A^\dagger A - I)x + 0 \\ &\leq \underline{c},\end{aligned}$$

where the last inequality holds because $x \in D$.

Hence, $\bar{u} \in D$. Then, from (2.2.4)

$$u(\xi, \bar{u}) = \mathrm{proj}_{X_\xi}(\bar{u}) = \mathrm{proj}_{X_\xi}(E[\mathrm{proj}_{X_\xi}(x)]).$$

Moreover, we obtain

$$\begin{aligned}u(\xi, \bar{u}) - u(\xi, x) &= (I - A^\dagger A)\bar{u} + A^\dagger b_\xi - (I - A^\dagger A)x - A^\dagger b_\xi \\ &= (I - A^\dagger A)((I - A^\dagger A)x + A^\dagger E[b_\xi]) - (I - A^\dagger A)x \\ &= (I - A^\dagger A)A^\dagger E[b_\xi] = 0,\end{aligned}$$

where the last two equalities use $(I - A^\dagger A)(I - A^\dagger A) = I - A^\dagger A$ and $(I - A^\dagger A)A^\dagger = 0$.

Hence for any $x \in D$ and almost every $\xi \in \Xi$, we have

$$\mathrm{proj}_{X_\xi}(x) = \mathrm{proj}_{X_\xi}(E[\mathrm{proj}_{X_\xi}(x)]). \tag{2.3.9}$$

From (2.3.9), for every $\xi \in \Xi$,

$$u(\xi, x_{\mathrm{ERM}}) = \mathrm{proj}_{X_\xi}(x_{\mathrm{ERM}}) = \mathrm{proj}_{X_\xi}(x^*) = u(\xi, x^*),$$

which, together with $\varphi(x^*) = \min_{x \in D} \varphi(x)$, implies

$$\varphi(x_{\mathrm{ERM}}) = \min_{x \in D} \varphi(x),$$

which in turn yields (2.3.8). ■

It is interesting to note that $x_{\text{ERM}} = x^*$ if and only if $A^\dagger(E[b_\xi] - Ax^*) = 0$. From (2.3.8), if the ERM formulation (2.1.1) has a solution and $A^\dagger(E[b_\xi] - Ax^*) \neq 0$, then (2.1.1) has a multiplicity of solutions.

Again, with $\bar{c}_i \geq \max_{\xi\in\Xi}(A^\dagger b_\xi)_i,\ i = 1, \cdots, m$, let

$$U = \{\, u = \Lambda \underline{c} + (I - \Lambda)\bar{c} + (I - A^\dagger A)x \,|\, \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n), \lambda_i \in [0,1], x \in D \,\}$$

and observe that for any $x \in D$ and $\xi \in \Xi$: $u(\xi, x) \in U$.

Theorem 2.3.1 *Assume that Assumption 2.2.1 holds. Moreover, under Assumption 2.2.2 (i)-(ii) the following hold.*

(i) If each component $F_i(\xi, u)$ of $F(\xi, u)$ is concave in u, then $Q(\xi, u)$ is convex in u.

(ii) If $F(\xi, x) = M_\xi x + q_\xi$ and $E[M_\xi]$ is positive semi-definite, then the objective function φ is a finite valued convex function on D.

Proof. We prove the theorem in the two aspects. (i) For any $u, v \in U$, $\lambda \in [0, 1]$ and almost every $\xi \in \Xi$,

$$\min\{\, b_\xi^T z \,|\, A^T z + F(\xi, u) \geq 0\} \qquad \text{and} \qquad \min\{\, b_\xi^T z \,|\, A^T z + F(\xi, v) \geq 0\}$$

have solutions. Let $z(\xi, u)$ and $z(\xi, v)$ be solutions of these two problems, respectively. Since the functions $F_i(\xi, x)$ are concave in x a.s.,

$$\begin{aligned} 0 &\leq \lambda(A^T z(\xi, u) + F(\xi, u)) + (1 - \lambda)(A^T z(\xi, v) + F(\xi, v)) \\ &\leq A^T(\lambda z(\xi, u) + (1 - \lambda)z(\xi, v)) + F(\xi, \lambda u + (1 - \lambda)v) \end{aligned}$$

holds a.s. This implies that $\lambda z(\xi, u) + (1-\lambda)z(\xi, v) \in \{z | A^T z + F(\xi, \lambda u + (1-\lambda)v) \geq 0\}$

a.s. Hence, we obtain the convexity of $Q(\xi, x)$,

$$\begin{aligned} Q(\xi, \lambda u + (1-\lambda)v) &\leq b_\xi^T(\lambda z(\xi, u) + (1-\lambda)z(\xi, v)) \\ &= \lambda Q(\xi, u) + (1-\lambda)Q(\xi, v), \quad \text{a.s.} \end{aligned}$$

(ii) With $B = A^\dagger A - I$, one has

$$\begin{aligned} f(\xi, x) &= (-Bx + A^\dagger b_\xi)^T(M_\xi(-Bx + A^\dagger b_\xi) + q_\xi) + Q(\xi, -Bx + A^\dagger b_\xi) \\ &= x^T B^T M_\xi Bx - (A^\dagger b_\xi)^T(M_\xi + M^T(\xi))Bx - q_\xi^T Bx \\ &\quad +(A^\dagger b_\xi)^T(M_\xi A^\dagger b_\xi + q_\xi) + Q(\xi, -Bx + A^\dagger b_\xi). \end{aligned}$$

By conditions (i) and (ii) of Assumption 2.2.2, there exists $d_2(\xi)$ such that $0 \leq f(\xi, x) \leq d_2(\xi)$ for all $x \in D$ and $E[d_2(\xi)] < \infty$. Taking the expected value of f, we see that φ is finite valued and there are a vector $c \in R^n$ and a constant c_0 such that

$$\varphi(x) = x^T B^T E[M_\xi]Bx + c^T x + c_0 + E[Q(\xi, -Bx + A^\dagger b_\xi)].$$

Since $Q(\xi, u)$ is convex in u for almost every $\xi \in \Xi$, $Q(\xi, -Bx + A^\dagger b_\xi)$ is convex in x for almost every $\xi \in \Xi$. Hence, when $E[M_\xi]$ is positive semi-definite it implies that φ is convex. ■

Theorem 2.3.2 *Under Assumptions 2.2.1 and 2.2.2, φ is globally Lipschitz on D, i.e.,*

$$|\varphi(x) - \varphi(y)| \leq \kappa\|x - y\|, \quad x, y \in D \tag{2.3.10}$$

where

$$\kappa = (E[d(\xi)] + E[d_1(\xi)](E[\|b_\xi\|]\, m\nu_A + \gamma\sqrt{n}))\|I - A^\dagger A\|;$$

recall that A is an $m \times n$-matrix and for the constant ν_A refer to (2.2.6).

Proof. For the first term in φ, we have

$$\begin{aligned}|u^T F(\xi,u) - v^T F(\xi,v)| &\le |u^T(F(\xi,u) - F(\xi,v))| + |(u-v)^T F(\xi,v)| \\ &\le \|u\| d_1(\xi)\|u-v\| + d(\xi)\|u-v\| \\ &\le (\gamma\sqrt{n} d_1(\xi) + d(\xi))\|u-v\|.\end{aligned}$$

For the second term, from (2.2.6), we have

$$|Q(\xi,u) - Q(\xi,v)| \le \|b_\xi\|\, m\nu_A d_1(\xi)\|u-v\|.$$

Combining these two inequalities,

$$\begin{aligned}&|\varphi(x) - \varphi(y)| \le E[|f(\xi,x) - f(\xi,y)|] \\ &\le E[|u(\xi,x)^T F(\xi,u(\xi,x)) - u(\xi,y)^T F(\xi,u(\xi,y))|] + E[|Q(\xi,u(\xi,x)) - Q(\xi,u(y,\xi))|] \\ &\le (\gamma\sqrt{n}E[d_1(\xi)] + E[d(\xi))] + m\nu_A E[\|b_\xi\|]E[d_1(\xi)])\|I - A^\dagger A\|\|x-y\|,\end{aligned}$$

completes the proof. ∎

Definition 2.3.1 [44] *Suppose that $\phi : X \subseteq R^m \to R$ is a locally Lipschitz continuous function, then ϕ is* semismooth *at $x \in \mathrm{int}\, X$ if ϕ is directionally differentiable at x and for any $g \in \partial\phi(x+h)$,*

$$\phi(x+h) - \phi(x) - g^T h = o(\|h\|),$$

where int X *denotes the interior of X and $\partial\phi$ denotes the Clarke generalized gradient.*

Theorem 2.3.3 *Suppose Assumptions 2.2.1 and 2.2.2 hold. Then the function φ is semismooth on D.*

Proof. Following Proposition 1 and (3.1)-(3.2) in [52], we only need to show that the following three conditions hold:

(i) There exists an integrable function κ_1 such that

$$|f(\xi, x) - f(\xi, y)| \le \kappa_1(\xi)\|x - y\|, \quad \text{for all}\, x, y \in D, \quad a.s.$$

(ii) $f(\xi, \cdot)$ is semismooth at $x \in D$ a.s.

(iii) The directional derivative $f_\xi'(x; h)$ of $f(\xi, \cdot)$ at x in direction h satisfies

$$\frac{|f_\xi'(x + h; h) - f_\xi'(x; h)|}{\|h\|} \le \kappa_2(\xi),$$

where $E[\kappa_2(\xi)] < \infty$.

For (i), as follows from the proof of Theorem 2.3.2,

$$|f(\xi, x) - f(\xi, y)| \le (d(\xi) + d_1(\xi)\sqrt{n}\gamma + m\nu(A)d_1(\xi)\|b_\xi\|)\|I - A^\dagger A\|\|x - y\|$$

for all $x, y \in D$ and almost every $\xi \in \Xi$.

For (ii), since $F(\xi, \cdot)$ is continuously differentiable at x, it suffices to worry about $Q(\xi, \cdot)$ and by [12, Theorem 5.8, Section 3.1] this function is piecewise smooth. Since piecewise smooth implies semismooth and the addition of semismooth functions is also a semismooth function, $f(\xi, \cdot)$ is semismooth on D a.s.

For (iii), from Assumption 2.2.2, we find that the first term of $f_\xi'(x+h; h)$ is bounded by the integrable function $(d(\xi) + \sqrt{n}\gamma d_1(\xi))\|I - A^\dagger A\|\|h\|$. The second term of f is the directional derivative of $Q(\xi, x)$, by [53, Lemma 2.2] and the formula 2.3.7, this term can be bounded by $m\nu(A)d_1(\xi)\|b_\xi\|\|I - A^\dagger A\|\|h\|$. Thus, we set $\kappa_2(\xi) = 2(d(\xi) + \sqrt{n}\gamma d_1(\xi) + m\nu(A)d_1(\xi)\|b_\xi\|)\|I - A^\dagger A\|$ and this yields (iii). $\blacksquare$

Theorem 2.3.4 *Suppose Assumptions 2.2.1 and 2.2.2(i-ii, iv) hold. Then, (2.1.1) has a solution in the compact set*

$$D_1 = \{\, y \,|y = (I - A^\dagger A)x,\; x \in D\}.$$

Moreover,

$$D_1 \subseteq D \quad \text{and} \quad \text{argmin}_{y\in D_1} \varphi(y) \subseteq \text{argmin}_{x\in D} \varphi(x). \tag{2.3.11}$$

Proof. For any $x \in D$, $u(\xi, x) = (I - A^\dagger A)x + A^\dagger b_\xi \in X_\xi$ and $y = (I - A^\dagger A)x \in D_1$, we get $y = u(\xi, x) - A^\dagger b_\xi$. Under Assumption 2.2.2(i) and (iv), we know that $u(\xi, x)$ and $A^\dagger b_\xi$ are bounded, so we obtain the set D_1 is closed and bounded.

For any $x \in D$, we can gain $0 \le \varphi(x) < \infty$ from Theorem 2.2.1 and Theorem 2.2.2. From the definition of $u(\xi, x)$, we have that $u(\xi, x) \in X_\xi$ and there are two constants $\underline{b}$ and $\bar{b}$ such that $\underline{b} \le b_\xi \le \bar{b}$ for $\forall\ \xi \in \Xi$. Hence, the vector

$$(I - A^\dagger A)x = u(\xi, x) - A^\dagger b_\xi$$

is in the compact set D_1. From $(I - A^\dagger A)(I - A^\dagger A) = (I - A^\dagger A)$ and $D = \{x|(I - A^\dagger A)x + \underline{c} \ge 0\}$, we have $y = (I - A^\dagger A)x \in D$ which implies $D_1 \subseteq D$. Moreover, from

$$(I - A^\dagger A)(I - A^\dagger A)x + A^\dagger b_\xi = (I - A^\dagger A)x + A^\dagger b_\xi = u(\xi, x),$$

we obtain

$$\min_{x\in D} \varphi(x) = \min_{y\in D_1} \varphi(y). \tag{2.3.12}$$

Since D_1 is compact and φ is continuous, $\text{argmin}_{D_1} \varphi \ne \emptyset$ and any $y^* \in \text{argmin}_{D_1} \varphi$ also minimizes φ on D since $D_1 \subseteq D$. Finally, from (2.3.12) one obtains (2.3.11). ■

2.4 The ERM formulation for stochastic CP

The stochastic CP, as a special case of the stochastic VI, deals with finding a vector $x \in R^n$, such that

$$x \ge 0, \quad F(\xi, x) \ge 0, \quad x^T F(\xi, x) = 0 \tag{2.4.13}$$

holds for every $\xi \in \Xi$.

The ERM formulation for the stochastic CP has been studied in [16, 20, 27, 67, 69].

The residual function of these papers is defined as $f(\xi, x) = ||G(\xi, x)||^2$ with

$$G(\xi, x) = \begin{pmatrix} \phi(F_1(\xi, x), x_1) \\ \vdots \\ \phi(F_n(\xi, x), x_n) \end{pmatrix},$$

and $\phi : R^2 \to R$ is an NCP function, which satisfies

$$\phi(a, b) = 0 \iff a \geq 0,\ b \geq 0,\ ab = 0.$$

Among various NCP functions, the "min" function ϕ_1 and the Fischer-Burmeister(FB) function ϕ_2 [28] are popular, which are given as follows:

$$\phi_1(a, b) := \min(a, b)$$

and

$$\phi_2 := a + b - \sqrt{a^2 + b^2}.$$

Another NCP function defined based on the FB function is presented by Chen-Chen-Kanzow in [13]

$$\phi_3 := \lambda(a + b - \sqrt{a^2 + b^2}) + (1 - \lambda)a_+b_+, \qquad \lambda \in (0, 1),$$

which is called the penalized FB function.

The ERM formulation for the stochastic CP defined by the "min" function, FB function and the penalized FB function has different properties such as the smoothness and the boundedness. We can find the related results about these properties in [16, 20, 27, 67].

When the feasible set of the stochastic VI is defined by $X := \{x | x \geq 0\}$, the recourse variable $u(\xi, x)$ reduces to x and the residual function (2.2.5) is as follows:

$$f_1(\xi, x) = x^T F(\xi, x) + Q(\xi, x), \tag{2.4.14}$$

where $Q(\xi, x) = \max_{y \geq 0} -y^T F(\xi, x)$.

Moreover, the constraint set of the ERM formulation is given by $D := \{x | x \geq 0\}$.

The ERM formulation for the stochastic CP defined by the residual function (2.4.14) is

$$\min_{x \geq 0} \varphi(x) = E[f_1(\xi, x)]. \tag{2.4.15}$$

We can see that under Assumption 2.2.1, the function $f_1(\xi, x)$ and $\varphi(x)$ are well defined. Furthermore, if Assumption 2.2.1 holds for the stochastic CP, it means that for a fixed $x \geq 0$ and almost every $\xi \in \Xi$, $F(\xi, x)$ should be nonnegative and $Q(\xi, x) = 0$. The residual function (2.4.14) becomes to $f_1(\xi, x) = x^T F(\xi, x)$ and the ERM formulation (2.4.15) for the stochastic CP reduces to

$$\begin{aligned} \min \quad & \varphi(x) = x^T E[F(\xi, x)] \\ & x \geq 0. \end{aligned} \tag{2.4.16}$$

The EV formulation for the stochastic CP is to find a vector $x \in R^n$ such that

$$x \geq 0, \quad E[F(\xi, x)] \geq 0, \quad x^T E[F(\xi, x)] = 0. \tag{2.4.17}$$

Under Assumption 2.2.1, we can define the gap function of the EV formulation for the stochastic CP as

$$g_1(x) = \max_{y \geq 0} (x - y)^T E[F(\xi, x)] = x^T E[F(\xi, x)]. \tag{2.4.18}$$

Hence, we can see that the gap function of the EV formulation is the same as the objective function of the ERM formulation. It means that under Assumption 2.2.1, solutions of the ERM formulation for the stochastic CP is the same as that of the EV formulation.

The EV formulation (2.4.17) is a deterministic CP. If Assumption 2.2.1 holds, we get the feasibility of the CP (2.4.17) and under some suitable conditions, such as co-

coercivity of the function $E[F(\xi, x)]$ we get that the solution set of (2.4.17) is nonempty and compact [26]. In other words, under Assumption 2.2.1 and conditions of existence of solutions for the CP (2.4.17) we can guarantee the existence of the ERM formulation for the stochastic CP.

Remark 2.4.1 *To define a deterministic optimization formulation for finding a "here and now" solution for the stochastic VI, we need a deterministic feasible set and a deterministic objective function. The feasible set D defined in Section 2.2 after (2.2.3) can ensure that*

(i) $u(\xi, x) = proj_{X_\xi}(x) \geq 0$, *for any* $x \in D$*;*

(ii) existence of solutions and finding a solution on a bounded subset $D_1 \subseteq D$.

The new function $f(\xi, x)$ *in (2.2.5) is defined by the recourse variable* $u(\xi, x)$ *which is dependent on the first level variable* x *and random variable* ξ*. Hence the degree of inadequacy or "loss" of a given* x *for a given* ξ *can be measured by* $f(\xi, x)$*. In Chapter 6, we show that* $\max\{-y^T F(\xi, x) | y \in X_\xi\}$ *has a closed form and* $f(\xi, x)$ *can be written explicitly for Wardrop's equilibrium for traffic assignment.*

Chapter 3

Smoothing sample average approximations (SSAA) for the stochastic VI

3.1 Introduction

Let $\xi^1, \cdots, \xi^N$ be a sampling of ξ. The Sample Average Approximation (SAA) method has been used to find a solution of the EV formulation (1.1.2) over a deterministic feasible set X [31, 35, 63]. The SAA method for the EV formulation of the stochastic VI uses the sample average value

$$\hat{F}^N(x) = \frac{1}{N}\sum_{i=1}^{N} F(\xi^i, x)$$

to approximate the expected value $E[F(\xi, x)]$ and solves

$$(y - x)^T \hat{F}^N(x) \geq 0, \qquad \text{for all } \ y \in X.$$

The classical law of large numbers ensures that $\hat{F}^N(x)$ converges with probability 1 to $E[F(x, \xi)]$ when the sample is iid.

Similarly, one can apply the SAA method to the ERM formulation (1.1.3) and denote the sample average value by

$$\hat{\varphi}^N(x) = \frac{1}{N}\sum_{i=1}^{N} f(\xi^i, x).$$

By the assumption that F is continuously differentiable in x for every $\xi \in \Xi$, $E[F(\xi, x)]$ and $\hat{F}^N(x)$ are continuously differentiable. However, the assumption of continuous differentiability of F does not imply that our (objective) function φ and its sample average approximation $\hat{\varphi}^N(x)$ are differentiable. In what follows, we introduce a smoothing sample average approximation (SSAA)

$$\Phi_\mu^N(x) = \frac{1}{N}\sum_{i=1}^{N} \tilde{f}(\xi^i, x, \mu), \tag{3.1.1}$$

where $\tilde{f} : \Xi \times R^n \times R_+$ is a smoothing approximation of f.

Definition 3.1.1 *Let $g : R^n \to R$ be a locally Lipschitz continuous function. We call $\tilde{g} : R^n \times R_+ \to R$ a* smoothing function *of g, if $\tilde{g}$ is continuously differentiable on R^n for any $\mu \in R_{++}$ and for any $x \in R^n$,*

$$\lim_{z\to x,\ \mu\downarrow 0} \tilde{g}(z, \mu) = g(x). \tag{3.1.2}$$

For the stochastic linear VI, under the condition that $E[M_\xi]$ is positive semi-definite, we employe the Moreau-Yosida regularization which is a smooth version of the ERM formulation to solve the ERM formulation. We analyze the convergence of optimal solutions of the SAA problem for the Moreau-Yosida regularization to that of the ERM formulation in Chapter 4. However, the Moreau-Yosida regularization cannot be used when the function $F(\xi, x)$ is nonlinear in x or $E[M_\xi]$ is not positive semi-definite. The SSAA function is based on the smoothing function, and we will give smoothing functions for traffic equilibrium problems in Chapter 6.

In this chapter, we define the SSAA function and prove the existence of solutions to SSAA minimization problems. Moreover, we show that any sequence of solutions of SSAA minimization problems has a cluster point and any such cluster point is a solution of the ERM formulation (2.1.1) (a.s.). We also show that any cluster point of a sequence of stationary points of SSAA minimization problems is a stationary point of the ERM formulation (2.1.1) (a.s.).

3.2 Minimizers of SSAA problem

In this section, we consider the existence and the convergence of solutions of the following SAA problems

$$\min_{x\in D} \hat{\varphi}^N(x) \tag{3.2.3}$$

and SSAA problems

$$\min_{x\in D} \Phi_\mu^N(x). \tag{3.2.4}$$

Let $X \subseteq R^n$ be an open set and $\overline{R} = [-\infty, \infty]$.

Definition 3.2.1 [56] *A sequence of functions $\{g^N : X \to \overline{R}, N \in \mathbb{N}\}$ epi-converges to $g : X \to \overline{R}$, written $g^N \xrightarrow{e} g$, if for all $x \in X$,*

(i) $\liminf_{N\to\infty} g^N(x^N) \geq g(x)$ *for all $x^N \to x$; and*

(ii) $\limsup_{N\to\infty} g^N(x^N) \leq g(x)$ *for some $x^N \to x$.*

Definition 3.2.2 [37] *A function $g : \Xi \times X \to \overline{R}$ is a random lsc (lower semicontinuous) function if*

(i) g is jointly measurable in (ξ, x),

(ii) $g(\xi, \cdot)$ is lsc for every $\xi \in \Xi$.

Definition 3.2.3 [37] *A sequence of random lsc functions $\{g^N : \Xi \times X \to \overline{R},\ N \in \mathbb{N}\}$ epi-converges to $g : X \to \overline{R}$ a.s., written $g^N \xrightarrow{e} g$ a.s., if for almost every $\xi \in \Xi$, $\{g^N(\xi, \cdot) : X \to \overline{R}, N \in \mathbb{N}\}$ epi-converges to $g : X \to \overline{R}$.*

Let $\delta_D(x) = 0$ when $x \in D$ and $\delta_D(x) = \infty$ otherwise; δ_D is the *indicator function* of the set D. For a given $x \in R^n$ and a positive number r, we denote the closed ball with center x and radius r by

$$B(x, r) = \{ y \in R^n \mid \|y - x\| \leq r \}.$$

Let $\bar{\mu}$ be a positive number. Let

$$\varphi_\mu(x) = E[\tilde{f}(\xi, x, \mu)].$$

Lemma 3.2.1 *Let $\tilde{f}$ be a smoothing function of f. Then Φ_μ^N and φ_μ are smoothing functions of $\hat{\varphi}^N$ and φ, respectively. If the sample is iid then for any fixed $\mu \in [0, \bar{\mu}]$, we have*

$$\Phi_\mu^N \quad \xrightarrow{e} \quad \varphi_\mu, \quad \text{on } D, \quad \text{a.s.} \tag{3.2.5}$$

Proof. By Definition 3.1.1, it is easy to see that Φ_μ^N and φ_μ are smoothing functions of $\hat{\varphi}^N$ and φ, respectively.

The proof for (3.2.5) is based on the convergence of inf-projections. Let

$$c_{x,r} = \inf\nolimits_{B(x,r)} \varphi_\mu + \delta_D, \qquad c_{x,r}^N = \inf\nolimits_{B(x,r)} \Phi_\mu^N + \delta_D.$$

Let Q^n be the set of rational n-dimensional vectors and $Q_{++} = R_{++} \cap Q^1$. For any $x \in Q^n$, $r \in Q_{++}$, since the samples are iid, the random variables $\{c_{x,r}^N\}$ are iid [37]. From the Law of Large Number follows

$$c_{x,r}^N \quad \longrightarrow \quad c_{x,r} \quad \text{as } N \to \infty \quad \text{a.s.}$$

Since $\Phi_\mu^N + \delta_D$ and $\varphi_\mu + \delta_D$ are random lsc functions, both functions can be completely identified by a countable collection of their inf-projections [37, 56, Chapter 14]. Hence we obtain (3.2.5). ■

For any locally Lipschitz continuous function $g : R^n \to R$, we can construct a smoothing function $\tilde{g} : R^n \times R_+ \to R$ satisfying the gradient consistent property

$$\{\lim_{z\to x,\mu\downarrow 0} \nabla \tilde{g}(z,\mu)\} \subseteq \partial g(x) \tag{3.2.6}$$

by convolution [56, Theorem 9.67], where ∂g denotes the Clarke generalized gradient. Moreover, for many locally Lipschitz continuous functions, we can easily construct computable smoothing functions satisfying (3.2.6). See examples in Chapter 6 and (ii) of Lemma 6.1.2. In the remainder of this chapter, we assume that the smoothing functions Φ_μ^N and φ_μ satisfy the gradient consistent property (3.2.6).

Lemma 3.2.2 *Under Assumptions 2.2.1 and 2.2.2(iii), whatever be the sample* $\{\xi^1, \cdots, \xi^N\}$ *that defines the functions* $\hat{\varphi}^N$ *and* Φ_μ^N*, the collection of functions* $\{\hat{\varphi}^N,\ N \in \mathbb{N}\}$*, as well as the collection* $\{\Phi_\mu^N,\ \mu > 0,\ N \in \mathbb{N}\}$*, are equi-locally Lipschitz continuous on* D*.*

In particular, this implies that when, for all N*, the samples are iid, the functions* Φ_μ^N *not only epi-converge almost surely to* φ_μ *on* D*, but converge also pointwise almost surely.*

Proof. The statements about the collections being equi-locally Lipschitz follow directly from Theorem 2.2.2 and the gradient consistent property (3.2.6), since they imply that both the collections of functions $\{f(\xi,\cdot) : D \to R,\ \xi \in \Xi\}$ and $\{\tilde{f}(\xi,\cdot\mu) : D \to R,\ \mu > 0,\ \xi \in \Xi\}$, that define $\hat{\varphi}^N$ and Φ_μ^N, via finite sums, are equi-locally Lipschitz continuous.

The almost sure pointwise convergence then follows immediately from [56, Theorem 7.10] and Lemma 3.2.1 which imply that under equi-lower semicontinuity of the approximating functions, epi-converges implies pointwise convergence. ■

Lemma 3.2.3 *Under the assumptions of Theorem 2.3.4, for any* $\mu \in [0, \bar{\mu}]$ *and* $N \in \mathbb{N}$*, the SAA minimization problem (3.2.3) and the SSAA minimization problem (3.2.4) admit optimal solutions.*

Proof. Since for any $\xi \in \Xi$, $f(\xi, \cdot)$ is a continuous function on D and measurable in ξ for any $x \in D$, the SAA function $\hat{\varphi}^N$ and the SSAA function Φ_μ^N are continuous functions on D for any $\mu \in [0, \bar{\mu}]$ and $N \in \mathbb{N}$ and consequently are also random lsc functions [56, Example 14.15]. Moreover by the same arguments as in the proof of Theorem 2.3.4, one obtains

$$\min_{x \in D} \hat{\varphi}^N(x) = \min_{y \in D_1} \hat{\varphi}^N(y) \tag{3.2.7}$$

and

$$\min_{x \in D} \Phi_\mu^N(x) = \min_{y \in D_1} \Phi_\mu^N(y). \tag{3.2.8}$$

Since D_1 is compact, there are y^*, y^{**} such that

$$y^* \in \mathrm{argmin}_{y \in D_1} \hat{\varphi}^N(y) \quad \text{and} \quad y^{**} \in \operatorname*{argmin}_{y \in D_1} \Phi_\mu^N(y),$$

respectively. Moreover, from $D_1 \subseteq D$ and (3.2.7), (3.2.8), y^* and y^{**} are thus solutions of (3.2.3) and (3.2.4), respectively. ■

Let S^*, S^N and S_μ^N be the sets of solutions of (2.1.1), (3.2.3) and (3.2.4) in D_1. In the following, we analyze the convergence of S^N and S_μ^N to S^*. For two sets Y and Z, we denote the distance from $z \in R^n$ to Y and the *excess* of the set Y on the set Z by

$$\mathrm{dist}(z, Y) = \inf_{y \in Y} \|z - y\|, \quad \text{and} \quad \mathbb{e}(Y, Z) = \sup_{y \in Y} \mathrm{dist}(y, Z).$$

Since φ, $\hat{\varphi}^N$ and Φ_μ^N are continuous and D_1 is compact, we have

$$\min_{x \in R^n} h(x) + \delta_{D_1}(x) \Longleftrightarrow \min_{x \in D_1} h(x),$$

for $h = \varphi$, $h = \hat{\varphi}^N$ or $h = \Phi_\mu^N$.

Theorem 3.2.1 *Under Assumptions 2.2.1 and 2.2.2, if the sample is iid, then the following hold.*

(i) Any sequence $\{x_\mu^N \in S_\mu^N\}$ has a cluster point as $N \to \infty$ and $\mu \downarrow 0$ a.s.

(ii) Any cluster point of $\{x_\mu^N \in S_\mu^N\}$ is an optimal solution of (2.1.1) a.s.

(iii) $\mathbb{e}(S_\mu^N, S^*) \longrightarrow 0$ a.s., as $N \to \infty$ and $\mu \downarrow 0$.

Proof. By the definition of the smoothing functions of $\varphi(x)$, $\lim_{x\to\bar{x},\mu\downarrow 0}\varphi_\mu(x) = \varphi(\bar{x})$ for any $x, \bar{x} \in D_1$. Moreover, from Lemmas 3.2.1, 3.2.2 and

$$|\Phi_\mu^N(x) - \varphi(\bar{x})| \le |\Phi_\mu^N(x) - \varphi_\mu(x)| + |\varphi_\mu(x) - \varphi(\bar{x})|,$$

we obtain

$$\Phi_\mu^N(x) \quad \longrightarrow \quad \varphi(\bar{x}), \quad \text{as} \quad x \to \bar{x}, N \to \infty, \mu \downarrow 0, \qquad \text{a.s.}$$

which means Φ_μ^N epi-converges to φ as $N \to \infty$ and $\mu \downarrow 0$, a.s. Hence by [56, Theorem 7.11], one has

$$\Phi_\mu^N + \delta_{D_1} \qquad \xrightarrow{e} \qquad \varphi + \delta_{D_1}, \quad \text{a.s.}$$

Moreover, by the continuity and nonnegativity of φ on the compact set D_1 and Theorem 2.3.4, one also has

$$-\infty < \min_{x\in R^n} \varphi(x) + \delta_{D_1}(x) = \min_{x\in D_1}\varphi(x) < \infty.$$

Hence, from [56, Theorem 7.31], we obtain

$$\begin{aligned}
\limsup_{N\to\infty,\mu\downarrow 0} \operatorname{argmin}_{x\in D_1} \Phi_\mu^N(x) &= \limsup_{N\to\infty,\mu\downarrow 0} \operatorname{argmin}_{x\in D_1}(\Phi_\mu^N(x) + \delta_{D_1}(x)) \\
&\subset \operatorname{argmin}_{x\in D_1}(\varphi(x) + \delta_{D_1}(x)) \\
&= \operatorname{argmin}_{x\in D_1}\varphi(x), \quad \text{a.s.}
\end{aligned}$$

By the compactness of D_1, the sequence $\{x_\mu^N\}$ has a cluster point and any such cluster point lies in the solution set of $\min_{x\in D_1}\varphi(x)$ a.s. Using Theorem 2.3.4 again, any such cluster point is also in the solution set of (2.1.1). The statement (iii) follows from (i) and (ii) of this theorem and the compactness of D_1. ■

In some cases, the expectation can be defined by multi-dimensional integrals and we can apply efficient quasi-Monte Carlo methods [60] to find approximate values of the expectation at each point x over a compact set. By error analysis of quasi-Monte Carlo methods for numerical evaluation of continuous integrals, we have

$$\lim_{N\to\infty} \Phi_\mu^N(x) = \varphi_\mu(x), \quad x \in D_1, \quad \mu \in [0, \bar{\mu}], \tag{3.2.9}$$

in the sense that for any given $\epsilon > 0$, there is a $\bar{\nu} > 0$, such that for any $N \geq \bar{\nu}$, we have

$$|\Phi_\mu^N(x) - \varphi_\mu(x)| < \epsilon, \quad \text{for any} \quad x \in D_1, \quad \mu \in [0, \bar{\mu}].$$

Theorem 3.2.2 *Under Assumptions 2.2.1 and 2.2.2, if (3.2.9) holds, so do the following.*

(i) Any sequence $\{x_\mu^N\} \subseteq S_\mu^N$ has a cluster point as $N \to \infty$ and $\mu \downarrow 0$.

(ii) Any cluster point of $\{x_\mu^N\}$ is an optimal solution of (2.1.1).

(iii) $\mathbb{e}(S_\mu^N, S^*) \longrightarrow 0$, as $N \to \infty$ and $\mu \downarrow 0$.

Proof. By definition of the smoothing functions associated with $\varphi(x)$, $\lim_{x\to\bar{x},\mu\downarrow 0} \varphi_\mu(x) = \varphi(\bar{x})$ for any $\bar{x} \in D_1$. Moreover, from (3.2.9) and

$$|\Phi_\mu^N(x) - \varphi(\bar{x})| \leq |\Phi_\mu^N(x) - \varphi_\mu(x)| + |\varphi_\mu(x) - \varphi(\bar{x})|,$$

we find

$$\lim_{x\to\bar{x},N\to\infty,\mu\downarrow 0} \Phi_\mu^N(x) = \varphi(\bar{x}),$$

which means $\Phi_\mu^N + \delta_{D_1}$ continuously converges to φ as $N \to \infty$ and $\mu \downarrow 0$ and continuous convergence implies epi-convergence. The remaining part of the proof is then similar to the proof of Theorem 3.2.1. ■

3.3 Stationary points of SSAA problem

In this section, we analyze the convergence of stationary points, that so far has only received perfunctory attention in the approximation theory for variational problems.

Recall [56, Section 8.A] that the subderivative of a function $g : R^n \to \overline{R}$ at a point $\bar{x}$ at which $g(\bar{x})$ is finite, is the function $dg(\bar{x}; \cdot)$ defined by

$$dg(\bar{x}; h) \;=\; \liminf_{\substack{\tau \downarrow 0 \\ h' \to h}} \; \Delta_\tau g(x; h') \;\text{ or, equivalently, }\; dg(\bar{x}; \cdot) = \text{ epi-}\liminf_{\tau \downarrow 0} \; \Delta_\tau g(\bar{x}; \cdot)$$

where $\Delta_\tau g(x; w)$ is the difference quotient function:

$$\Delta_\tau g(x; h) := \frac{g(x + \tau h) - g(x)}{\tau} \;\text{ for } \tau > 0.$$

One refers to $\bar{x} \in X \subset R^n$ as a stationary point of g on a closed set X, if

$$dg(\bar{x}; h) \geq 0 \;\text{ for all } h \in T_X(\bar{x}), \tag{3.3.10}$$

where $T_X(\bar{x})$ is the tangent cone of X at $\bar{x} \in X$ [26]. When, X is convex, one can exploit the polarity between the tangent and the normal cones [56, Theorem 6.9] and reformulate this condition as

$$dg(\bar{x}; z - \bar{x}) \geq 0 \;\text{ for all } z \in X.$$

We work with this latter inequality since our X, the sets D and D_1, are convex. Moreover, the functions $f(\xi, \cdot)$, from Theorem 2.2.2, and, a fortiori, $\tilde{f}(\xi, x, \mu)$ that are used to build our sample average approximation are locally Lipschitz (a.s.). We are going to assume that they are also Clarke regular at the points of interest. Of course, this would be the case when $Q(\xi, \cdot)$ is regular since, by assumption, $F(\xi, \cdot)$ is continuously differentiable. This occurs in a variety of situations, for example, when $F(\xi, \cdot)$ is linear, when for $i = 1, \ldots, n$, the functions $F_i(\xi, \cdot)$ are concave and, in particular, when $Q(\xi, \cdot)$ can be expressed as a max-function as in our applications in Chapter 6.

In view of [56, Theorem 9.16], when g is locally Lipschitz and Clarke regular at $\bar{x}$, then the subderivative coincides with the directional derivative,

$$dg(\bar{x}; h) \;=\; \lim_{\tau \downarrow 0} \Delta_\tau g(x; h) = g'(x; h).$$

Moreover, $dg(\cdot, h)$ is usc (upper semicontinuous); in fact, [56, Theorem 9.16] asserts a bit more but that's not needed here.

In addition to these properties, the proof of the next theorem relies like Lemma 3.2.1 on the law of large numbers for random lsc functions, more precisely, random usc functions, and two inequalities: The first one, comes about from the interchange of subdifferentiation and taking expectation, the second one results from the choice of a smoothing function that will satisfy

$$\lim_{\mu \downarrow 0} d\tilde{f}_\mu(\xi, x; h) \leq df(\xi, x; h) \;\text{ for all } x, h. \tag{3.3.11}$$

In Chapter 6, we show that $Q(\xi, \cdot)$ is regular and the exponential smoothing function [19, 47] satisfies (3.3.11) for piecewise maxima functions.

Theorem 3.3.1 *Suppose Assumptions 2.2.1 and 2.2.2 hold and $Q(\xi, \cdot)$ is regular for any fixed $\xi \in \Xi$. Then for any $\mu \geq 0$ and $N \in \mathbb{N}$, the SAA problem (3.2.3) and the SSAA problem (3.2.4) have stationary points in the compact set D_1. Let $\{x_\mu^N\} \subset D_1$ be a sequence of stationary point of (3.2.4). If the sample is iid, then any cluster point of $\{x_\mu^N\}$ is a stationary point of (2.1.1), a.s.*

Proof. The existence of stationary points follows directly from the existence of minimizers of (3.2.3) and (3.2.4).

By the regularity of Q and continuous differentiability of F, we deduce that f, φ, $\hat{\varphi}^N$ are Clarke regular [21, Definition 2.3.4, Proposition 2.3.6] in D.

Since f is globally Lipschitz in D, there are constants $\bar{t} > 0$ and β such that $t^{-1}[f(\xi, x+h) - f(\xi, x)] \geq \beta$, a.s. for all h in a neighborhood of 0 and $0 < \bar{t} \leq t$. By

Proposition 2.9 in [62, Section 2], we obtain

$$E[df(\xi, x; y-x)] \le d\varphi(x; y-x), \quad \forall\, x, y \in D. \tag{3.3.12}$$

By the continuous differentiability of $\tilde{f}(\xi, x, \mu)$ for $\mu > 0$ and upper semicontinuity of $df(\xi, x; h)$ on x for each fixed h, we deduce that for any fixed $\mu \in [0, \bar{\mu}]$ and $h \in R^n$, $d\Phi_\mu^N(\cdot; h) = \frac{1}{N}\sum_{i=1}^N d\tilde{f}_\mu(\xi^i, \cdot; h)$ is upper semicontinuous. Hence, we can use the same technique as in the proof of Lemmas 3.2.1 and 3.2.2, to show that

$$d\Phi_\mu^N(\cdot; h) \quad \xrightarrow{e,p} \quad d\varphi_\mu(\cdot; h), \quad \text{in}\, D, \quad \text{a.s.}, \tag{3.3.13}$$

where e, p stands for epi- and pointwise convergence.

Let $\hat{x}$ be a cluster point of $\{x_\mu^N\}$. For a $y \in D$, let $h = y - \hat{x}$. One might have to restrict the argument to a subsequence but to simplify the notation, assume that $\{x_\mu^N\}$ converges to $\hat{x}$. Then, one has

$$\begin{aligned}
&0 \le d\Phi_\mu^N(x_\mu^N; y - x_\mu^N) \\
&\le \sigma\|\hat{x} - x_\mu^N\| + d\Phi_\mu^N(x_\mu^N; h) - d\varphi_\mu(\hat{x}; h) + d\varphi_\mu(\hat{x}; h) - d\varphi(\hat{x}; h) + d\varphi(\hat{x}; h),
\end{aligned}$$

where σ is a Lipschitz constant of Φ_μ^N near $\hat{x}$ for all $\mu \ge 0$ and $N \in \mathbb{N}$; the existence of such σ follows from the global Lipschitz continuity of Φ_μ^N and φ.

The third and second terms give $d\Phi_\mu^N(x_\mu^N; h) - d\varphi_\mu(\hat{x}; h) \to 0$ as $N \to \infty$ and $\mu \downarrow 0$, a.s. by using (3.3.13).

From (3.3.12) and (3.3.11), the fifth and fourth terms give

$$d\varphi_\mu(\hat{x}; h) - d\varphi(\hat{x}; h) \le E[df_\mu(\xi, \hat{x}; h) - df(\hat{x}; h)] \le 0, \quad \text{as} \quad \mu \downarrow 0.$$

Hence we obtain $d\varphi(\hat{x}; h) \ge 0$ as $N \to \infty$ and $\mu \downarrow 0$. ■

Remark 3.3.1 *From the properties of smoothing functions, we can define*

$$\tilde{f}(\xi, x, 0) = \lim_{\mu \downarrow 0} \tilde{f}(\xi, x, \mu)$$

at any $x \in D$ *and* $\xi \in \Xi$. *Hence, we can consider* $\hat{\varphi}^N(x) = \Phi_0^N(x) = \lim_{\mu \downarrow 0} \Phi_\mu^N(x)$ *at any* $x \in D$. *Since our convergence results include* $\mu \equiv 0$, *the same convergence results hold for SAA-solutions and SAA-stationary points as a special case.*

Remark 3.3.2 *The conclusions of Proposition 6* [57, Chapter 6] *are similar to that of Theorem 3.2.1 but require the a.s.-uniform convergence of the SAA-functions* $\hat{\varphi}^N$ *whereas essentially our only requirement is 'iid samples' and then, we followed the pattern already laid out in* [5].

Remark 3.3.3 *In* [64], *Xu and Zhang proposed a SSAA method for solving a general class of one stage nonsmooth stochastic problems and derived the exponential rate of convergence of the SSAA method. We believe that the exponential rate can be also derived for residual minimization SSAA method for stochastic variational inequalities. However, this is by no means straightforward and, as far as we can tell, it requires non-classical analysis. This certainly will require a separate treatment that we plan to deal with in a separate paper.*

Chapter 4

Moreau-Yosida regularization for stochastic linear VI

4.1 Introduction

In this chapter, we focus on a class of linear VIs in a stochastic environment, in which both the function and the feasible set have uncertainties. The stochastic linear VI is closely linked to the study of stochastic linear and quadratic programs, and has a wide range of applications, which we can find in pricing competition among several firms providing substitutable goods or services [14] and the transportation stochastic user equilibrium [35, 69]. Moreover, some problems in oligopolistic transit market can be reformulated as generalized Nash equilibrium with stochastic linear VI constraints [38, 70].

The function and the feasible set of the stochastic linear VIs are defined by:

$$F(\xi, x) = M_\xi x + q_\xi$$

and

$$X_\xi = \{x \mid Ax = b_\xi, \quad x \geq 0\}.$$

Here $A \in R^{m\times n}$, $M_\xi \in R^{n\times n}$, $q_\xi \in R^n$, and $b_\xi \in R^m$ for any fixed $\xi \in \Xi \subseteq R^\ell$, a set representing future state of knowledge.

In general, there is no $x \in X_\xi$ such that

$$(y-x)^T(M_\xi x + q_\xi) \geq 0, \quad \forall\, y \in X_\xi \tag{4.1.1}$$

holds for all random variables $\xi \in \Xi$. Such stochastic linear VI includes the stochastic linear complementarity problem as a special subclass [16, 20, 27].

Assume that M_ξ, q_ξ and b_ξ are measurable in ξ. We denote the expected values as:

$$\bar{M} = E[M_\xi], \quad \bar{q} = E[q_\xi], \quad \bar{b} = E[b_\xi].$$

Assumption 4.1.1 *Assume that the matrix A has full-row rank, and there is a positive constant γ, such that $X_\xi \subseteq U_0 = \{u \in R^n \,|\, \|u\|_\infty \leq \gamma\}$ holds for any $\xi \in \Xi$.*

Remark 4.1.1 *Assumption 4.1.1 holds for a class of matrices A and vectors b_ξ in traffic equilibrium problems* [42, 69]. *See Chapter 6.*

Under Assumption 4.1.1, the Moore-Penrose generalized inverse of A can be defined as $A^\dagger = A^T(AA^T)^{-1}$. The residual function defined by (2.2.5) has the following form for the linear case,

$$f(\xi, x) = u(\xi, x)^T(M_\xi u(\xi, x) + q_\xi) + Q(\xi, u(\xi, x)), \tag{4.1.2}$$

where $u(\xi, x) = x + A^\dagger(b_\xi - Ax)$ and

$$\begin{aligned} Q(\xi, u(\xi, x)) &= \max_y\{\, -y^T(M_\xi u(\xi, x) + q_\xi) \,|\, Ay = b_\xi,\ y \geq 0\,\} \\ &= \min_z\{\, z^T b_\xi \,|\, A^T z + M_\xi u(\xi, x) + q_\xi \geq 0\,\}. \end{aligned}$$

Under Assumption 4.1.1, X_ξ is bounded. Combining it with the continuity of $-y^T(M_\xi u(\xi, x) + q_\xi)$, we can ensure that there is $y(\xi, x)$, such that $Q(\xi, u(\xi, x)) = -y(\xi, x)^T(M_\xi u(\xi, x) +$

q_ξ). Moreover, for any $\xi \in \Xi$, there exist $\underline{b}$ and $\tilde{b}$ such that $\underline{b} \leq b_\xi \leq \tilde{b}$. Hence, there is a vector $\underline{c}$, such that $\underline{c}_i = \min_{\xi\in\Xi}(A^\dagger b_\xi)_i$, $i = 1, \cdots, m$. Let

$$D = \{\, x \mid A^\dagger Ax - x \leq \underline{c}\}.$$

Assumption 4.1.2 *Suppose that M_ξ and q_ξ are measurable in ξ with the following property*

$$E[||M_\xi||] < \infty \quad \text{and} \quad E[||q_\xi||] < \infty.$$

We consider the following ERM formulation with the residual function (4.1.2)

$$\min_{x\in D} \varphi(x) := E[f(\xi, x)]. \tag{4.1.3}$$

For the ERM formulation, we set the here-and-now solution as

$$x_{ERM} = x^* + A^\dagger(E[b_\xi] - Ax^*),$$

where $x^* \in \mathrm{argmin}_{x\in D}\varphi(x)$.

From Theorem 2.3.1 in Chapter 2, we know that the objective function φ is a finite convex function on the set D, if the matrix $\bar{M}$ is positive semi-definite, which means

$$x^T\bar{M}x \geq 0, \qquad \forall\, x \in R^n. \tag{4.1.4}$$

In [2], Agdeppa, et al. formulate the ERM formulation using the D-gap function [49] as follows:

$$\min_{x\in R^n} \Psi(x) := E[\theta_\tau(\xi, x)],$$

where $\theta_\tau(\xi, x) = g_\tau(\xi, x) - g_{\frac{1}{\tau}}(\xi, x)$ and $\tau > 1$. Here

$$g_\tau(\xi, x) = \max_{y\in X_\xi}\ (M_\xi x + q_\xi)^T(x - y) - \frac{1}{2\tau}\|x - y\|^2$$

is the regularized gap function [29] for a fixed $\xi \in \Xi$. The following condition

$$\inf_{\xi\in\Xi,\ \|x\|=1} x^T M_\xi x \geq \beta_0 > 0 \tag{4.1.5}$$

is used in [2] to prove the convexity of the function Ψ.

Note that condition (4.1.4) is weaker than condition (4.1.5). See Example 4.2.1 in Section 4.2.

To guarantee the convexity of the SAA problem of (4.1.3), we adopt the Tikhonov regularization to (4.1.3) as follows

$$\min_{x\in D} \varphi_\varepsilon(x) := E[f(\xi, x)] + \frac{\varepsilon}{2} x^T x, \tag{4.1.6}$$

where $\varepsilon > 0$. For any $\varepsilon > 0$, the function φ_ε is strongly convex under condition (4.1.4). Moreover, under Assumptions 4.1.1 and 4.1.2, the function φ_ε is globally Lipschitz continuous and semismooth on D. The convexity enables us to employ the Moreau-Yosida regularization to define the SAA problem of the following smooth convex minimization problem

$$\min_{x\in D} \hat{\varphi}_\varepsilon(x), \tag{4.1.7}$$

where

$$\hat{\varphi}_\varepsilon := \min\{\varphi_\varepsilon(y) + \frac{\mu}{2}\|x - y\|^2 | y \in D\}$$

is the Moreau-Yosida regularization of φ_ε.

For any $\mu > 0$, problem (4.1.6) is equivalent to problem (4.1.7).

4.2 Sample average approximation for the ERM formulation

Let $\Xi, \cdots, \xi_N$ be a sample of ξ. The SAA method for the ERM formulation (4.1.3) uses the sample average value

$$\Phi^N(x) = \frac{1}{N}\sum_{i=1}^{N} f(\xi_i, x)$$

to approximate the expected value $\varphi(x) = E[f(\xi, x)]$ and solves

$$\min_{x \in D} \Phi^N(x). \tag{4.2.8}$$

The classical law of large number for random functions ensures that for any fixed $x \in D$, $\Phi^N(x)$ converges with probability 1 to the expected value $\varphi(x) = E[f(\xi, x)]$, when the sample is independent and identically distributed (iid).

To ensure the convexity of the SAA problem, we adopt the Tikhonov regularization and consider (4.1.6). The SAA problem of (4.1.6) is denoted by

$$\min_{x \in D} \Phi_\varepsilon^N(x) = \frac{1}{N}\sum_{i=1}^{N} f(\xi_i, x) + \frac{\varepsilon}{2} x^T x. \tag{4.2.9}$$

Under Assumption 4.1.1 and condition (4.1.4), for any $\varepsilon > 0$, there is $N_\varepsilon > 0$ such that for any $N \geq N_\varepsilon$, the function Φ_ε^N is convex. For the Monte Carlo method, from Theorem 1.1 in [48] we get that the following error estimate

$$\Phi_\varepsilon^N(x) - \varphi_\varepsilon(x) = \Phi^N(x) - \varphi(x) = O(x, \frac{1}{\sqrt{N}}) \tag{4.2.10}$$

holds for a fixed $x \in D$. For a fixed $\xi \in \Xi$, we use $y(\xi, x)$ to denote an optimal solution of $\max_{y \in X_\xi} -y^T(M_\xi u(\xi, x) + q_\xi)$. From Theorem 2.4 in [41], $y(\xi, x)$ is Lipschitz continuous in b_ξ and the Lipschitz constant is only dependent on A. Suppose that the matrix M_ξ, vectors q_ξ and b_ξ are Lipschitz continuous in ξ and bounded on Ξ. Then, $y(\xi, x)$ is Lipschitz continuous in ξ. Moreover, from $u(\Xi, x) - u(\xi_2, x) = A^\dagger(b_\Xi - b_{\xi_2})$, we get the Lipschitz continuity of $u(\xi, x)$ in ξ. Under Assumption 4.1.1, we know $u(\xi, x)$ and $y(\xi, x)$ are bounded for any $\xi \in \Xi$. Hence

$$\begin{aligned}
& ||u(\Xi, x)^T(M_\Xi u(\Xi, x) + q_\Xi) - u(\xi_2, x)^T(M_{\xi_2} u(\xi_2, x) + q_{\xi_2})|| \\
\leq \;& (||u(\Xi, x)||||M_\Xi|| + ||u(\xi_2, x)||||M_{\xi_2}|| + ||q_\Xi||)||u(\Xi, x) - u(\xi_2, x)|| \\
& + ||u(\Xi, x)||||u(\xi_2, x)||||M_\Xi - M_{\xi_2}|| + ||u(\xi_2, x)||||q_\Xi - q_{\xi_2}|| \\
\leq \;& \gamma_1||\Xi - \xi_2||
\end{aligned}$$

holds, which means the Lipschitz continuity of $u(\xi, x)^T(M_\xi u(\xi, x) + q_\xi)$. Similarly, we can obtain $Q(\xi, x) = -y(\xi, x)^T(M_\xi u(\xi, x) + q_\xi)$ is Lipschitz continuous in ξ. Thus, the function $f(\xi, x) = u(\xi, x)^T(M_\xi u(\xi, x) + q_\xi) + Q(\xi, x)$ is Lipschitz continuous in ξ and the Lipschitz constant is independent on x. Hence, the error estimate $O(\frac{1}{\sqrt{N}})$ is independent on x.

Combining this with the strong convexity of φ_ε, we obtain

$$\begin{aligned}
\Phi_\varepsilon^N(z) &= \varphi_\varepsilon(z) + O(\frac{1}{\sqrt{N}}) \\
&\leq \lambda\varphi_\varepsilon(x) + (1-\lambda)\varphi_\varepsilon(y) - \frac{\varepsilon}{2}\lambda(1-\lambda)\|x - y\|^2 + O(\frac{1}{\sqrt{N}}) \\
&= \lambda\Phi_\varepsilon^N(x) + (1-\lambda)\Phi_\varepsilon^N(y) - \frac{\varepsilon}{2}\lambda(1-\lambda)\|x - y\|^2 + O(\frac{1}{\sqrt{N}}), \quad (4.2.11)
\end{aligned}$$

where $z = \lambda x + (1-\lambda)y$ for $0 \leq \lambda \leq 1$. Roughly speaking, if $N \geq \varepsilon^{-3}$, it is very likely that $-\frac{\varepsilon}{2}\lambda(1-\lambda)\|x - y\|^2 + O(\frac{1}{\sqrt{N}}) \leq 0$, which implies the convexity of Φ_ε^N. The quasi-Monte Carlo method [48] yields a better error bound $O(\frac{1}{N})$ for suitably chosen sets of nodes for some problems. Similarly, for this case, choosing $N \geq \varepsilon^{-2}$ might ensure the convexity of Φ_ε^N.

Note that under condition (4.1.5) in [2], problem (4.2.8) is a strongly convex problem. Moreover, if the matrix M_ξ is positive semi-definite for almost every $\xi \in \Xi$, problem (4.2.8) is also a convex problem for all N. For these two cases, for any $\varepsilon > 0$, we can ensure the strong convexity of (4.2.9) for any N.

The following example satisfies condition (4.1.4), but fails for condition (4.1.5) imposed in [2]. For this example, it is not difficult to find the value of ε and the sample size N such that Φ_ε^N is convex.

Example 4.2.1 *Consider the linear function $F(\xi, x) = M_\xi x + q_\xi$ with the random variable $\xi = (\hat{\xi}, \tilde{\xi})^T \in \Xi \subseteq R^2$, and*

$$M_\xi = \begin{pmatrix} -4 + (14 + \hat{\xi}) \max(0, \operatorname{sign}(\hat{\xi})) & 0 & 0 \\ 0 & -4 - (16 + \hat{\xi}) \min(0, \operatorname{sign}(\hat{\xi})) & 0 \\ 0 & 0 & \hat{\xi} \end{pmatrix},$$

$$q_\xi = \begin{pmatrix} 3 + \hat{\xi} \\ \hat{\xi} \\ -4\hat{\xi} \end{pmatrix},$$

where $\hat{\xi} \in \Xi := [-1, 1]$ and $\hat{\xi}$ is uniformly distributed in Ξ. The matrix A and the vector b_ξ are given as follows:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad b_\xi = \begin{pmatrix} \frac{2}{1+e^{-\tilde{\xi}^2}} \\ \frac{2}{1+\tilde{\xi}^2} \end{pmatrix},$$

where $\tilde{\xi}$ is normally distributed in R. For any fixed $\xi \in \Xi := \Xi \times R$, the constraint set is

$$X_\xi := \{x | x \geq 0,\ x_1 + x_2 = \frac{2}{1 + e^{-\tilde{\xi}^2}},\ x_3 = \frac{2}{1 + \tilde{\xi}^2}\}.$$

Moreover, the vector b_ξ is bounded as $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \leq b_\xi \leq \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ for any $\xi \in \Xi$.

Thus

$$A^\dagger = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}, \qquad \underline{c} = \min_{\xi \in \Xi} A^\dagger b_\xi = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix},$$

and we obtain the feasible set $D := \{x \mid |x_1 - x_2| \leq 1,\ x_3 \in R\}$*.*

It is easy to see that

$$M_\xi = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 12+\hat{\xi} & 0 \\ 0 & 0 & \hat{\xi} \end{pmatrix} \text{ for } \hat{\xi} < 0, \quad \mathrm{M}_\xi = \begin{pmatrix} 10+\hat{\xi} & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & \hat{\xi} \end{pmatrix} \text{ for } \hat{\xi} > 0,$$

$$M_\xi = \begin{pmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ for } \hat{\xi} = 0, \quad \bar{\mathrm{M}} = \mathrm{E}[\mathrm{M}_\xi] = \begin{pmatrix} 3.25 & 0 & 0 \\ 0 & 3.75 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, there is no $\xi \in \Xi$ *such that* M_ξ *is a positive semi-definite matrix, which implies that condition (4.1.5) does not hold. For this example, the objective function* $\Psi(x)$ *of the ERM formulation defined by the D-gap function in* [2] *is not convex.*

For every $\xi \in \Xi$*, the vector* b_ξ *is nonnegative, so we obtain*

$$Q(\xi, x) = \min\{b_\xi^T z | A^T z + M_\xi u(\xi, x) + q_\xi \geq 0\} = b_\xi^T z(\xi, x),$$

where $u(\xi, x) = (I - A^\dagger A)x + A^\dagger b_\xi$ *and* $z(\xi, x) = (\max\{-y_1(\xi, x), -y_2(\xi, x)\}, -y_3(\xi, x))^T$*. Here* $y_i(\xi, x) = (M_\xi u(\xi, x) + q_\xi)_i$ *means the i-th element of the vector* $M_\xi u(\xi, x) + q_\xi$*. Based on the above analysis, the residual function is*

$$\begin{aligned} & f(\xi, x) \\ = \ & u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, x) \\ = \ & x^T (I - A^\dagger A) M_\xi (I - A^\dagger A)x + (2M_\xi A^\dagger b_\xi + q_\xi)^T (I - A^\dagger A)x \\ & + (A^\dagger b_\xi)^T (M_\xi A^\dagger b_\xi + q_\xi) + b_\xi^T z(\xi, x) \\ = \ & x^T \tilde{M}_\xi x + (2M_\xi A^\dagger b_\xi + q_\xi)^T (I - A^\dagger A)x + b_\xi^T z(\xi, x) + (A^\dagger b_\xi)^T (M_\xi A^\dagger b_\xi + q_\xi), \end{aligned}$$

where

$$\begin{aligned}
\tilde{M}_\xi &= (I - A^\dagger A) M_\xi (I - A^\dagger A) \\
&= \begin{pmatrix} -2 + (\frac{7}{2} + \frac{\hat{\xi}}{4})a - (4 + \frac{\hat{\xi}}{4})b & 2 - (\frac{7}{2} + \frac{\hat{\xi}}{4})a + (4 + \frac{\hat{\xi}}{4})b & 0 \\ 2 - (\frac{7}{2} + \frac{\hat{\xi}}{4})a + (4 + \frac{\hat{\xi}}{4})b & -2 + (\frac{7}{2} + \frac{\hat{\xi}}{4})a - (4 + \frac{\hat{\xi}}{4})b & 0 \\ 0 & 0 & 0 \end{pmatrix},
\end{aligned}$$

$a = \max(0, \mathrm{sign}(\hat{\xi}))$ *and* $b = \min(0, \mathrm{sign}(\hat{\xi}))$. *The expected value of the matrix* $\tilde{M}_\xi$ *is as follows:*

$$E[\tilde{M}_\xi] = \begin{pmatrix} 1.75 & -1.75 & 0 \\ -1.75 & 1.75 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is positive semi-definite, so our ERM formulation (4.1.3) is a convex minimization problem. Suppose that the sample of $\hat{\xi}$ *is given as* $\hat{\xi}_i = -1 + (i-1)\frac{2}{N-1}$, *for* $i = 1, \cdots, N$, *where* $N > 1$ *is the sample size. The SAA problem of (4.1.3) is described as* $\Phi^N(x) = \frac{1}{N}\sum_{i=1}^N f(\xi_i, x)$ *and its quadratic item is* $\frac{1}{N}\sum_{i=1}^N x^T \tilde{M}_{\xi_i} x = x^T \frac{1}{N}\sum_{i=1}^N \tilde{M}_{\xi_i} x$. *Furthermore, we know*

$$\frac{1}{N}\sum_{i=1}^N \tilde{M}_{\xi_i} = \begin{pmatrix} 1.75 & -1.75 & 0 \\ -1.75 & 1.75 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The matrix $\frac{1}{N}\sum_{i=1}^N \tilde{M}_{\xi_i}$ *is positive semi-definite, which implies* Φ^N *is convex. So for any* $\varepsilon > 0$ *and sample size* $N > 1$, Φ_ε^N *is strongly convex.*

We denote $x_{EV} \in SOL(\bar{X}, \bar{M}x + \bar{q})$ *and* $x_{ERM} = (I - A^\dagger A)x^* + A^\dagger E[b_\xi]$, *where* x^* *is a solution of (4.1.6). In Table 4.1, we list the solutions* x_{ERM}, x_{EV}, *and the risk criteria for the ERM and EV solutions with different sample sizes* N *and different regularized parameters* ε. *We use the residual function* f *and CVaR* [54]

$$\alpha^*(x) \in \mathrm{argmin}_{\alpha \in R}\ \alpha + \frac{1}{1-\beta} E[(f(\xi, x) - \alpha)_+] \tag{4.2.12}$$

to compare the ERM and EV solutions, where $(t)_+ = \max(t, 0)$. *We find that the ERM* *solution performs better than the EV solution in the aspects of the risk criteria.*

Table 4.1: Solutions and criteria for $x = x_{ERM}$ and $x = x_{EV}$ with $\beta = 0.98$

	$N = 10^3$ $\varepsilon = 10^{-2}$		$N = 10^4$ $\varepsilon = 10^{-4}$		$N = 10^6$ $\varepsilon = 10^{-6}$	
	x_{ERM}	x_{EV}	x_{ERM}	x_{EV}	x_{ERM}	x_{EV}
x_1	0.9609	0.7458	0.9672	0.7502	0.9797	0.7651
x_2	1.2620	1.4771	1.2615	1.4786	1.2497	1.4643
x_3	0.6397	0.6397	0.6534	0.6534	0.6529	0.6529
$\alpha^*(x)$	34.6273	42.1975	34.3611	41.9291	34.0628	41.5353
$\mathrm{CVaR}(x, \alpha^*(x))$	35.4821	43.1837	35.2542	43.0134	34.8736	42.5204
$E[f(\xi, x)]$	18.8417	19.1656	18.9022	19.2308	18.9307	19.2525

4.3 Moreau-Yosida regularization

For simplicity, we assume that Φ_ε^N is convex in our discussion. In this section, we consider the Moreau-Yosida regularization of the SAA problem (4.2.9).

Let δ_D be the indicator function of the set D, that is,

$$\delta_D(x) = \begin{cases} 0, & \text{if} \quad x \in D, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then we have

$$\min_{x \in D} \varphi_\varepsilon(x) \Longleftrightarrow \min_{x \in R^n} \varphi_\varepsilon(x) + \delta_D(x) \tag{4.3.13}$$

and

$$\min_{x \in D} \Phi_\varepsilon^N(x) \Longleftrightarrow \min_{x \in R^n} \Phi_\varepsilon^N(x) + \delta_D(x). \tag{4.3.14}$$

Since Φ_ε^N is not necessarily differentiable, we employ the Moreau-Yosida regularization to define a smooth convex function. For $\mu > 0$, the Moreau-Yosida regularization of Φ_ε^N is defined by

$$\hat{\Phi}_\varepsilon^N(x) \;\; := \;\; \min\{\Phi_\varepsilon^N(y) + \frac{\mu}{2}\|x - y\|^2 | y \in D\}. \tag{4.3.15}$$

By the smoothing property of the Moreau-Yosida regularization, the functions $\hat{\varphi}_\varepsilon$ and $\hat{\Phi}_\varepsilon^N$ are convex and continuously differentiable on D. Moreover, for any $x \in D$, we have

$$\nabla\hat{\varphi}_\varepsilon(x) = \mu(x - p_\varepsilon(x)), \qquad \nabla\hat{\Phi}_\varepsilon^N(x) = \mu(x - q_\varepsilon^N(x)),$$

where $p_\varepsilon(x)$ and $q_\varepsilon^N(x)$ denote the unique optimal solutions of (4.1.8) and (4.3.15), respectively. From Exercise 12.23 in [56], we know p_ε and q_ε^N are globally Lipschitz continuous, which implies the gradients of $\hat{\varphi}_\varepsilon$ and $\hat{\Phi}_\varepsilon^N$ are globally Lipschitz continuous. Based on the above analysis and the convexity of $\varphi_\varepsilon(x)$ and $\Phi_\varepsilon^N(x)$, solving optimization problems (4.3.13) and (4.3.14) is equivalent to solving the following nonlinear equations on D

$$\nabla\hat{\varphi}_\varepsilon(x) = \mu(x - p_\varepsilon(x)) = 0, \tag{4.3.16a}$$

$$\nabla\hat{\Phi}_\varepsilon^N(x) = \mu(x - q_\varepsilon^N(x)) = 0, \tag{4.3.16b}$$

respectively. Suppose that

$$x_\varepsilon^* = \mathrm{argmin}_{x\in D}\varphi_\varepsilon(x), \qquad \hat{x}_\varepsilon = \mathrm{argmin}_{x\in D}\hat{\varphi}_\varepsilon(x),$$

$$x_\varepsilon^{N*} \in \mathrm{argmin}_{x\in D}\Phi_\varepsilon^N(x), \qquad \hat{x}_\varepsilon^N \in \mathrm{argmin}_{x\in D}\hat{\Phi}_\varepsilon^N(x).$$

It is known that minimizing φ_ε and minimizing $\hat{\varphi}_\varepsilon$ are equivalent, in the sense that

$$\hat{x}_\varepsilon = x_\varepsilon^*.$$

Similarly, we have

$$\mathrm{argmin}_{x\in D}\Phi_\varepsilon^N(x) = \mathrm{argmin}_{x\in D}\hat{\Phi}_\varepsilon^N(x).$$

See [33]. Now, we consider the following problem

$$\min_{x\in D}\hat{\Phi}_\varepsilon^N(x). \tag{4.3.17}$$

Let S^*, S_ε^* and $\hat{S}_\varepsilon^N$ be the sets of solutions of (4.1.3), (4.1.6) and (4.3.17). In the following, we analyze the convergence of $\hat{S}_\varepsilon^N$ to S^*. For two sets Y and Z, we denote the distance from $z \in R^n$ to Y and the *excess* of the set Y on the set Z by

$$\mathrm{dist}(z, Y) = \inf_{y \in Y} \|z - y\|$$

and

$$\mathbb{e}(Y, Z) = \sup_{y \in Y} \mathrm{dist}(y, Z).$$

Let

$$D_1 = \{y \mid y = (I - A^\dagger A)x,\ x \in D\}.$$

Under Assumption 4.1.1, for any $y \in D_1$, we have $y = u(\xi, x) - A^\dagger b_\xi$ and b_ξ is bounded. Furthermore, we know $u(\xi, x) \in X_\xi$, so we obtain the set D_1 is closed and bounded.

Lemma 4.3.1 *If Assumption 4.1.1 and condition (4.1.4) hold, we have*

$$\mathrm{argmin}_{y \in D_1} \varphi_\varepsilon(y) = \mathrm{argmin}_{x \in D} \varphi_\varepsilon(x).$$

Proof. From $(I - A^\dagger A)(I - A^\dagger A) = (I - A^\dagger A)$, we know that $D_1 \subseteq D$. Because of the continuity of the function $\varphi_\varepsilon(x)$ and the fact that the set D_1 is closed and bounded, we have

$$\min_{y \in D_1} \varphi_\varepsilon(y) \geq \min_{x \in D} \varphi_\varepsilon(x). \tag{4.3.18}$$

For $x_\varepsilon^* = \mathrm{argmin}_{x \in D} \varphi_\varepsilon(x)$, let $y_\varepsilon^* = (I - A^\dagger A)x_\varepsilon^*$. Then we have $y_\varepsilon^* \in D_1$ and

$$\begin{aligned} u(\xi, y_\varepsilon^*) &= (I - A^\dagger A)y_\varepsilon^* + A^\dagger b_\xi \\ &= (I - A^\dagger A)(I - A^\dagger A)x_\varepsilon^* + A^\dagger b_\xi \\ &= (I - A^\dagger A)x_\varepsilon^* + A^\dagger b_\xi \\ &= u(\xi, x_\varepsilon^*). \end{aligned}$$

Hence, we have

$$\begin{aligned}
\varphi_\varepsilon(x_\varepsilon^*) - \varphi_\varepsilon(y_\varepsilon^*) &= \frac{\varepsilon}{2}(x_\varepsilon^{*T} x_\varepsilon^* - y_\varepsilon^{*T} y_\varepsilon^*) \\
&= \frac{\varepsilon}{2}(x_\varepsilon^{*T} x_\varepsilon^* - x_\varepsilon^{*T}(I - A^\dagger A)^T (I - A^\dagger A) x_\varepsilon^*) \\
&= \frac{\varepsilon}{2}(x_\varepsilon^{*T} x_\varepsilon^* - x_\varepsilon^{*T}(I - A^\dagger A) x_\varepsilon^*) \\
&= \frac{\varepsilon}{2} x_\varepsilon^{*T} A^\dagger A x_\varepsilon^* \\
&\geq 0,
\end{aligned}$$

where the last inequality uses that the matrix A has full-row rank and the matrix $A^\dagger A$ is positive semi-definite.

Thus,

$$\min_{y \in D_1} \varphi_\varepsilon(y) \leq \varphi_\varepsilon(y_\varepsilon^*) \leq \varphi_\varepsilon(x_\varepsilon^*) = \min_{x \in D} \varphi_\varepsilon(x). \tag{4.3.19}$$

Combining (4.3.18) with (4.3.19), we obtain $\min_{y \in D_1} \varphi_\varepsilon(y) = \min_{x \in D} \varphi_\varepsilon(x)$. Moreover, from the strong convexity of $\varphi_\varepsilon(x)$ we get $\text{argmin}_{y \in D_1} \varphi_\varepsilon(y) = \text{argmin}_{x \in D} \varphi_\varepsilon(x)$. ■

Lemma 4.3.2 *If the sample is iid, for any fixed $\varepsilon > 0$ we have*

$$\Phi_\varepsilon^N \xrightarrow{e} \varphi_\varepsilon, \quad \text{in } D_1, \quad \text{a.s.} \tag{4.3.20}$$

Proof. The proof is based on the convergence of the inf-projections. Let

$$c_{x,r} = \inf_{B(x,r)} \varphi_\varepsilon + \delta_{D_1}, \qquad c_{x,r}^N = \inf_{B(x,r)} \Phi_\varepsilon^N + \delta_{D_1}.$$

Let Q^n be the set of rational n-dimensional vectors and $Q_{++} = R_{++} \cap Q^1$. For any $x \in Q^n$, $r \in Q_{++}$, since the samples are iid, the random variables $\{c_{x,r}^N\}$ are iid [37]. From the Law of Large Number follows

$$c_{x,r}^N \quad \longrightarrow \quad c_{x,r} \quad \text{as } N \to \infty \quad \text{a.s..}$$

Since $\Phi_\varepsilon^N + \delta_{D_1}$ and $\varphi_\varepsilon + \delta_{D_1}$ are random lsc functions, both functions can be completely identified by a countable collection of their inf-projections [37, 56, Chapter 14]. Hence, we obtain our desirable result. ■

Theorem 4.3.1 *Under Assumption 4.1.1 and condition (4.1.4), if the sample is iid, then for any fixed $\varepsilon > 0$ the sequence $\{\hat{x}_\varepsilon^N \in \hat{S}_\varepsilon^N\}$ converges to the optimal solution of $\min_{x\in D}\varphi_\varepsilon(x)$ a.s. as $N \to \infty$.*

Proof. First, from Lemma 4.3.2, we know that Φ_ε^N epi-converges to φ_ε as $N \to \infty$. Hence, one has

$$\Phi_\varepsilon^N + \delta_{D_1} \xrightarrow{e} \varphi_\varepsilon + \delta_{D_1}, \qquad a.s.$$

From Assumption 4.1.1 and condition (4.1.4), we have that φ_ε is a convex function on D_1 and $\varphi_\varepsilon + \delta_{D_1}$ is convex.

By the definition of the Moreau-Yosida regularization, we obtain

$$\min_{x\in D_1} \hat{\varphi}_\varepsilon(x) \Longleftrightarrow \min_{x\in D_1} \varphi_\varepsilon(x)$$

and

$$\min_{x\in D_1} \hat{\Phi}_\varepsilon^N(x) \Longleftrightarrow \min_{x\in D_1} \Phi_\varepsilon^N(x).$$

Furthermore, by the continuity of φ_ε on the compact set D_1 and the fact $\min_{y\in D_1}\varphi_\varepsilon(y) = \min_{x\in D}\varphi_\varepsilon(x)$, we have

$$-\infty < \min_{x\in R^n} \varphi_\varepsilon(x) + \delta_{D_1}(x) = \min_{x\in D_1} \varphi_\varepsilon(x) < +\infty.$$

Hence, from [56], we obtain

$$\begin{aligned}
\limsup_{N\to\infty}\text{argmin}_{x\in D_1}\hat{\Phi}_\varepsilon^N(x) &= \limsup_{N\to\infty}\text{argmin}_{x\in R^n}(\Phi_\varepsilon^N(x) + \delta_{D_1}(x)) \\
&\subset \text{argmin}_{x\in R^n}(\varphi_\varepsilon(x) + \delta_{D_1}(x)) \\
&= \text{argmin}_{x\in D_1}\varphi_\varepsilon(x), \qquad a.s.
\end{aligned}$$

Since D_1 is closed and bounded, the sequence $\{\hat{x}_\varepsilon^N\}$ has a cluster point and by Lemma 4.3.1 this cluster point is also in the solution set of $\min_{x\in D}\varphi_\varepsilon(x)$ a.s. Because of the strong convexity of $\varphi_\varepsilon(x)$, we know that the solution set S_ε^* is singleton. From the argument above, we have our desirable result. ∎

From Theorem 4.3.1, we deduce

$$\mathbb{e}(\hat{S}_\varepsilon^N, S_\varepsilon^*) \longrightarrow 0, \text{ a.s., as } N \to \infty.$$

Moreover, we also obtain that the optimal solution x_ε^* of (4.1.6) exists. Based on the above analysis, we get that the solution set of (4.1.3) has a *least l_2-norm solution*, which is the vector that belongs to the solution set and is the closest one to the origin in the l_2-norm.

Lemma 4.3.3 *x_ε^* converges to the least l_2-norm solution of (4.1.3) as $\varepsilon \to 0$.*

Proof. For any $\hat{x} \in S^*$, where S^* is the set of the optimal solutions of (4.1.3), we have the following inequalities

$$\varphi(\hat{x}) \leq \varphi(x_\varepsilon^*) \leq \varphi(x_\varepsilon^*) + \frac{\varepsilon}{2} x_\varepsilon^{*T} x_\varepsilon^* \leq \varphi(\hat{x}) + \frac{\varepsilon}{2} \hat{x}^T \hat{x}. \tag{4.3.21}$$

Since φ is a continuous function on D, any cluster point of $\{x_\varepsilon^*\}$ is a solution of (4.1.3), which means $\lim_{\varepsilon\to 0} \operatorname{dist}(x_\varepsilon^*, S^*) = 0$.

Let x^* be a limit of a subsequence $\{x_{\varepsilon_k}^*\}$ as $\varepsilon_k \to 0$. We assume that there exists $\tilde{x} \in S^*$ such that $\|\tilde{x}\| \leq \|x^*\|$ and $\varphi(\tilde{x}) = \varphi(x^*)$. We have

$$\varphi(x_{\varepsilon_k}^*) + \frac{\varepsilon_k}{2} x_{\varepsilon_k}^{*T} x_{\varepsilon_k}^* \leq \varphi(\tilde{x}) + \frac{\varepsilon_k}{2} \tilde{x}^T \tilde{x} \leq \varphi(x^*) + \frac{\varepsilon_k}{2} x^{*T} x^*.$$

Combining this with (4.3.21), we obtain

$$0 \;\leq\; \frac{2}{\varepsilon_k}(\varphi(x_{\varepsilon_k}^*) - \varphi(\tilde{x})) \leq \tilde{x}^T \tilde{x} - x_{\varepsilon_k}^{*T} x_{\varepsilon_k}^* \leq x^{*T} x^* - x_{\varepsilon_k}^{*T} x_{\varepsilon_k}^* \quad \to \quad 0 \quad \text{as} \quad \varepsilon_k \to 0.$$

This implies that $\|x_{\varepsilon_k}^*\|$ converges to $\|\tilde{x}\|$ as $\varepsilon_k \to 0$. Hence, we obtain $\|\tilde{x}\| = \|x^*\|$. Moreover, from the convexity of D and φ, the least l_2-norm solution of (4.1.3) is unique.

Therefore, we get that x^* is the least l_2-norm solution of (4.1.3). Because of the arbitrariness of the chosen subsequence, we can claim that x_ε^* converges to the least l_2-norm solution x^* as $\varepsilon \to 0$. ■

Theorem 4.3.2 *Suppose that Assumption 4.1.1 and condition (4.1.4) hold, and the sample is iid. $\hat{x}_\varepsilon^N$ converges to the least l_2-norm solution x^* of (4.1.3) a.s., as $N \to \infty$ and $\varepsilon \to 0$.*

Proof. The desirable result can be derived from Theorem 4.3.1 and Lemma 4.3.3. ■

4.4 Semismoothness of the gradient of the Moreau-Yosida regularization

In this section, we prove the semismoothness of the gradient of the Moreau-Yosida regularization $\hat{\Phi}_\varepsilon^N$. A remarkable property is that using the semismoothness we can obtain the superlinear or quadratic convergence of generalized Newton methods and quasi-Newton methods for solving nonsmooth equations (4.3.16b) [15, 45].

There are several different forms for the definition of piecewise smooth functions. We state one from [43] below.

Definition 4.4.1 [43] *A continuous function $\psi : R^n \to R^l$ is said to be a piecewise C^k (k times continuously differentiable) function on a set $Y \subseteq R^n$ if there exists a finite index set $J = \{1, \cdots, r\}$, closed sets $Y_1, \cdots, Y_r$, open sets $U_1, \cdots, U_r$ (or relatively open with respect to the affine hull of Y), and functions $\psi_1, \cdots, \psi_r$ such that*

(i) $Y \subseteq \cup_{j=1}^r Y_j$ and $Y_j \subseteq U_j$ for each $j \in J$,

(ii) $\psi_j \in C^k(U_j)$ for each $j \in J$,

(iii) $\psi(u) = \psi_j(u)$ for any $u \in Y \cap Y_j$ and each $j \in J$.

We refer to $\{(Y_j, U_j, \psi_j)\}_{j\in J}$ *as a representation of* ψ.

For any $u \in D$, we write

$$J(u) = \{j \in J : \ \Phi_\varepsilon^N(u) = \Phi_{\varepsilon j}^N(u)\}.$$

To study the semismoothness of the gradient of the Moreau-Yosida regularization $\hat{\Phi}_\varepsilon^N$, the following constraint qualification-Affine Independence Preserving Constraint Qualification (AIPCQ) [45] is important.

Definition 4.4.2 [45] *AIPCQ is said to hold for a piecewise smooth function* ψ *at* u *if for every subset* $K \subseteq J(u)$ *for which there exist a sequence* $\{u^k\}$ *with* $\{u^k\} \to u$, $K \subseteq J(u^k)$, *and the vectors*

$$\left\{ \begin{pmatrix} \nabla\psi_i(u^k) \\ 1 \end{pmatrix} : i \in K \right\} \tag{4.4.22}$$

being linearly independent, it follows that the vectors

$$\left\{ \begin{pmatrix} \nabla\psi_i(u) \\ 1 \end{pmatrix} : i \in K \right\} \tag{4.4.23}$$

are linearly independent.

Theorem 4.4.1 *Suppose that Assumption 4.1.1 and condition (4.1.4) hold. Then the gradient* $\nabla\hat{\Phi}_\varepsilon^N$ *of the Moreau-Yosida regularization* $\hat{\Phi}_\varepsilon^N$ *is semismooth on* D.

Proof. By the definition of $u(\xi, x)$,

$$\begin{aligned} Q(\xi, x) &= \max_y \{ -y^T(M_\xi u(\xi,x) + q_\xi) \,|\, Ay = b_\xi,\ y \geq 0 \} \\ &= \max_y \{ -y^T(M_\xi(I - A^\dagger A)x + M_\xi A^\dagger b_\xi + q_\xi) \,|\, Ay = b_\xi,\ y \geq 0 \}. \end{aligned}$$

Hence, $Q(\xi, x)$ is a piecewise linear function on x, and so is the sum of $Q(\xi_i, x)$ which means that $Q^N(x) = \sum_{i=1}^N Q(\xi_i, x)$ is a piecewise linear function. Moreover, the other part of the function Φ_ε^N is a quadratic function on x and continuously differentiable. Therefore, we get that Φ_ε^N is a piecewise C^2 function.

Since $Q^N(x)$ is piecewise linear and it can be represented by $\{Q_j^N\}_{j\in J}$ using Definition 4.4.1, where $Q_j^N(x) = (\alpha_j^N)^T x - \beta_j^N$ for some $\alpha_j^N \in R^n$. For any $w \in D$ and any index set $K \subseteq J(w)$, we have

$$\left\{ \begin{pmatrix} (\nabla \Phi_\varepsilon^N(w))_j \\ 1 \end{pmatrix} : j \in K \right\} = \left\{ \begin{pmatrix} \dfrac{Pw + h + \alpha_j^N}{N} \\ 1 \end{pmatrix} : j \in K \right\}, \tag{4.4.24}$$

where

$$P = N\varepsilon I + \sum_{i=1}^N [(I - A^\dagger A)(M_{\xi_i} + M_{\xi_i}^T)(I - A^\dagger A)]$$

and

$$h = (I - A^\dagger A) \sum_{i=1}^N [q_{\xi_i} + (M_{\xi_i} + M_{\xi_i}^T) A^\dagger b_{\xi_i}].$$

It is easy to find that vectors in (4.4.24) are linearly independent if and only if vectors $\left\{ \begin{pmatrix} \dfrac{\alpha_j^N}{N} \\ 1 \end{pmatrix} : j \in K \right\}$ are linearly independent, from

$$\begin{pmatrix} I & -\dfrac{(Pw+h)}{N} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \dfrac{Pw + h + \alpha_{j_1}^N}{N} & \cdots & \dfrac{Pw + h + \alpha_{j_{|K|}}^N}{N} \\ 1 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \dfrac{\alpha_{j_1}^N}{N} & \cdots & \dfrac{\alpha_{j_{|K|}}^N}{N} \\ 1 & \cdots & 1 \end{pmatrix}.$$

Suppose that there exists a sequence $\{w^k\} \to w$, and vectors in (4.4.24) with w^k are linearly independent. Then we can get that vectors $\left\{ \begin{pmatrix} \dfrac{\alpha_j^N}{N} \\ 1 \end{pmatrix} : j \in K \right\}$ are linearly independent. Based on the above argument, we obtain that vectors in (4.4.24) with w are linearly independent. So AIPCQ holds at any $w \in D$.

From Theorem 1 in [45], we can get our desirable result, which is the gradient of the Moreau-Yosida regularization of $\hat{\Phi}_{\varepsilon}^{N}$ is semismooth. ■

Example 4.4.1 *Suppose that there is $\Gamma > 0$, such that for any $\xi \in \Xi$, $b_\xi \geq 0$, $E[\|b_\xi\|_\infty] \leq \Gamma$, and A can be split into two submatrices A_1 and A_2, where A_1 is an $m \times m$ M-matrix and A_2 is an $m \times (n-m)$ nonnegative matrix whose each column has only one positive entry. Without loss of generality we assume that $A = [A_1,\ A_2]$. Such matrix can be found in traffic equilibrium* [42, 69].

Let

$$\begin{aligned} c(\xi, x) &= M_\xi u(\xi, x) + q_\xi \\ &= M_\xi (I - A^\dagger A)x + M_\xi A^\dagger b_\xi + q_\xi \end{aligned}$$

and

$$J_k = \{j|\ a_{k,m+j} \neq 0,\ 1 \leq j \leq n-m\}.$$

Then

$$\begin{aligned} Q(\xi, x) = \min_z \{b_\xi^T z | A^T z + c(\xi, x) \geq 0\} \\ = \sum_{k=1}^{m} (b_\xi)_k \max\{(-A_1^{-T}\tilde{c}(\xi, x))_k, \max_{j \in J_k}\{-\frac{c_{m+j}(\xi, x)}{a_{k,m+j}}\}\}, \end{aligned}$$

where $\tilde{c}(\xi, x) = (c_1(\xi, x), \cdots, c_m(\xi, x))^T$. Hence, the function $Q^N(x)$ can be written as

$$Q^N(x) = \sum_{i=1}^{N} \sum_{k=1}^{m} (b_{\xi_i})_k \max\{(-A_1^{-T}\tilde{c}(\xi_i, x))_k, \max_{j \in J_k}\{-\frac{c_{m+j}(\xi_i, x)}{a_{k,m+j}}\}\}.$$

For any fixed $\xi \in \Xi$ and any $u \in X_\xi$, we have $[A_1, A_2]u = b_\xi$ and $u \geq 0$. Since A_1 is an M-matrix, we can obtain $[I_m, A_1^{-1}A_2]u = A_1^{-1}b_\xi \geq 0$. Moreover, because $u \geq 0$, $E[\|b_\xi\|_\infty] \leq \Gamma$ and $[I_m, A_1^{-1}A_2] \geq 0$, we get $\|u\|_\infty \leq \gamma$ for some $\gamma > 0$.

Hence, for this example, Assumption 4.1.1 holds. Moreover, AIPCQ holds at any $x \in D$ and from Theorem 4.4.1, $\nabla \hat{\Phi}_{\varepsilon}^{N}$ is semismooth on D.

Remark 4.4.1 *When the stochastic vectors b_ξ and $M_\xi x + q_\xi$ are independent for any $x \in R^n$, we can get that the solution set of the EV formulation is the same as the solution set of the ERM formulation (4.1.3).*

For the EV formulation, it is equivalent to solve a determine VI, and its residual function is:

$$f(x) = \min_{x \in \bar{X}} x^T(\bar{M}x + \bar{q}) + \bar{b}^T y(x), \tag{4.4.25}$$

where $y(x) = \mathrm{argmin}\{\bar{b}^T y | A^T y + \bar{M}x + \bar{q} \geq 0\}$, and $\bar{X} := \{x | x \geq 0, Ax = \bar{b}\}$.

For the ERM formulation, we know the solution u should satisfy $u \geq 0$ and $Au = \bar{b}$. Moreover, the expected residual function is:

$$\varphi(x) = E[f(\xi, u)] = E[u(\xi, x)^T(M_\xi u(\xi, x) + q_\xi)] + E[b_\xi^T y(\xi, x)], \tag{4.4.26}$$

where $y(\xi, x) = \mathrm{argmin}\{b_\xi^T y | A^T y + M_\xi u(\xi, x) + q_\xi \geq 0\}$. Since $b_\xi \geq 0$, b_ξ and $M_\xi x + q_\xi$ are independent for any $x \in R^n$, from the form of $y(\xi, x)$ given in Example 4.4.1, we can get that

$$\begin{aligned} \varphi(x) &= \bar{u}^T(\bar{M}\bar{u} + \bar{q}) + \bar{b}^T E[y(\xi, x)] \\ &= \bar{u}^T(\bar{M}\bar{u} + \bar{q}) + \bar{b}^T \bar{y}(\bar{u}), \end{aligned}$$

where $\bar{y}(\bar{u}) = \mathrm{argmin}\{\bar{b}^T y | A^T y + \bar{M}\bar{u} + \bar{q} \geq 0\}$.

From the above analysis, the EV formulation and the ERM formulation have the same objective functions and feasible sets if $M_\xi x + q_\xi$ and b_ξ are independent random variables for any $x \in R^n$. However, such case rarely happens in the real world. For example, in stochastic traffic equilibrium problems, the uncertainty of the vector $M_\xi x + q_\xi$ is due to the capacity on the links and the uncertainty of the vector b_ξ arising from the demand. In general, both the capacity and the demand are influenced by weather, accidents, etc., so the vectors $M_\xi x + q_\xi$ and b_ξ are dependent in general.

Chapter 5

Distributionally robust CVaR formulation for the stochastic linear VI

5.1 Introduction

Value-at-Risk (VaR) is a widely used measure of risk in finance and economics. For a specific probability level β, the β-VaR of a decision is the smallest number α, such that the loss will not exceed α with the probability β. Let $f(\xi, x)$ be the loss associated with the decision $x \in D \subseteq R^n$ and the random variable $\xi \in \Xi$. The probability of $f(\xi, x)$ not exceeding α is given as follows:

$$\Psi(x, \alpha) = \text{prob}\{f(\xi, x) \leq \alpha\}.$$

From the property of the probability function, we know $\Psi(x, \alpha)$ is right continuous with respect to α. The β-VaR associated with the decision x and any probability level $\beta \in (0, 1)$ will be denoted by $\alpha_\beta(x)$, and it is given by

$$\alpha_\beta(x) = \min\{\alpha \in R : \ \Psi(x, \alpha) \geq \beta\}. \tag{5.1.1}$$

CVaR [54, 55] is one of risk measures which is coherent. A coherent risk measure is a function that satisfies properties of monotonicity, sub-additivity, homogeneity, and translational invariance. For a given β-quantile, β-CVaR measures the expected loss of a decision given that a loss is occurring at or below the β-quantile. The β-CVaR denoted by $\phi_\beta(x)$ is given by [54]

$$\phi_\beta(x) = (1-\beta)^{-1} \int_{f(\xi,x)\geq \alpha_\beta(x)} f(\xi,x) d\mathcal{F}(\xi), \tag{5.1.2}$$

where $\mathcal{F}(\xi)$ is the distribution function of the random variable ξ. From the definitions of VaR and CVaR, we find that the β-VaR is never more than the β-CVaR, which means low CVaR must have low VaR.

Rockafellar and Uryasev in [54] define the function G_β on $D \times R$

$$G_\beta(x,\alpha) = \alpha + (1-\beta)^{-1} E[(f(\xi,x)-\alpha)_+], \tag{5.1.3}$$

which characterizes some properties of $\alpha_\beta(x)$ and $\phi_\beta(x)$. Theorem 1 in [54] and Theorem 10 in [55] show that $\phi_\beta(x) = \min_{\alpha\in R} G_\beta(x,\alpha)$ and $\alpha_\beta(x) = \text{left point of } \operatorname*{argmin}_{\alpha\in R} G_\beta(x,\alpha)$.

In this chapter, We use the residual function (2.2.5) as a loss function. We define the CVaR formulation by the ERM formulation. In Section 5.2, we discuss the relationship between the ERM formulation and the CVaR formulation, and properties of the CVaR formulations for the stochastic linear VI, such as Lipschitz continuity and gradient consistency. In Section 5.3, we employ the sublinear expectation to discuss the robust CVaR formulation and prove the existence of optimal solutions of the robust CVaR formulation.

5.2 CVaR and ERM formulations

In this section, we focus on the CVaR formulation of the stochastic linear VI which is to find

$$x \in X_\xi := \{x | Ax = b_\xi, \quad x \geq 0\}$$

such that

$$(y - x)^T (M_\xi x + q_\xi) \geq 0, \quad \forall\, y \in X_\xi. \tag{5.2.4}$$

Assumption 5.2.1 *Assume that there exists a positive number γ, such that $X_\xi \subseteq U_0 = \{u \in R^n | \|u\|_\infty \leq \gamma\}$ holds for any $\xi \in \Xi$.*

If the matrix A has full-row rank, we can define the Moore-Penrose generalized inverse of A by $A^\dagger = A^T (AA^T)^{-1}$.

The residual function (2.2.5) has the following form

$$f(\xi, x) = u(\xi, x)^T (M_\xi u(\xi, x) + q_\xi) + Q(\xi, u(\xi, x)), \tag{5.2.5}$$

where $u(\xi, x) = x + A^\dagger (b_\xi - Ax)$ and $Q(\xi, x) = \min\{b_\xi^T z | A^T z + M_\xi u(\xi, x) + q_\xi \geq 0\}$. Under Assumption 5.2.1, for any fixed $\xi \in \Xi$, b_ξ is bounded, so we can ensure there exists $z(\xi, x)$ such that $Q(\xi, x) = b_\xi^T z(\xi, x)$ and there is $\underline{c}$ such that $\underline{c}_i = \min_{\xi \in \Xi} (A^\dagger b_\xi)_i$, $i = 1, \cdots, m$. We define the feasible set by

$$D = \{\, x \,|\, A^\dagger A x - x \leq \underline{c}\}.$$

The ERM formulation defined by the residual function (5.2.5) is as follows:

$$\min_{x \in D} E[f(\xi, x)]. \tag{5.2.6}$$

We define the here and now solution of the ERM formulation by $x_e = x^* + A^\dagger (E[b_\xi] - Ax^*)$, where $x^* \in \mathrm{argmin}_{x \in D} E[f(\xi, x)]$. From Proposition 2.3.1 in Chapter 2, we know that x_e is an optimal solution of the ERM formulation.

Then, the CVaR formulation defined by the ERM function (5.2.6) is given as the following one:

$$\min_{x \in D,\ \alpha \in R} G_\beta(x, \alpha) \quad = \quad \alpha + (1-\beta)^{-1} E[(f(\xi, x) - \alpha)_+]. \tag{5.2.7}$$

From Theorem 1 in [54], if the probability $\Psi(x, \alpha)$ is continuous in α, we know for any fixed $x \in D$, $G_\beta(x, \alpha)$ is convex and continuously differentiable on α. For any fixed x_e, we denote $\alpha_e \in \operatorname{argmin} G_\beta(x_e, \alpha)$, and $(x_\beta^*, \alpha_\beta^*) \in \operatorname{argmin}_{x \in D, \alpha \in R} G_\beta(x, \alpha)$.

Proposition 5.2.1 *Suppose that $\Psi(x, \alpha)$ is continuous with respect to α, and (x_0^*, α_0^*) is an optimal solution of (5.2.7) as $\beta = 0$. Then x_0^* is an optimal solution of the ERM formulation and $(x_e, \alpha_e) \in \operatorname{argmin}_{x \in D, \alpha \in R} G_\beta(x, \alpha)$.*

Proof. For $\beta = 0$, we have $G_0(x, \alpha) = \alpha + E[(f(\xi, x) - \alpha)_+]$. For any fixed $x \in D$, $G_0(x, \alpha)$ is continuous with respect to α and the derivative is

$$\frac{\partial G_0(x, \alpha)}{\partial \alpha} = 1 + (\Psi(x, \alpha) - 1) = \Psi(x, \alpha).$$

Since (x_0^*, α_0^*) is an optimal solution of (5.2.7) as $\beta = 0$, we have

$$\operatorname{prob}\{\xi | f(\xi, x_0^*) \le \alpha_0^*\} = 0,$$

and $G_0(x_0^*, \alpha_0^*) = E[f(\xi, x_0^*)]$.

Similarly, we have $\operatorname{prob}\{\xi | f(\xi, x_e) \le \alpha_e\} = 0$, and

$$G_\beta(x_e, \alpha_e) = E[f(\xi, x_e)].$$

Moreover, $G_\beta(x_e, \alpha_e) \ge G_\beta(x_0^*, \alpha_0^*)$ holds, which implies

$$E[f(\xi, x_0^*)] \le E[f(\xi, x_e)].$$

Since x_e is an optimal solution of $\min_{x\in D} E[f(\xi,x)]$, we obtain $E[f(\xi,x_e)] = E[f(\xi,x_0^*)]$, which implies our desirable result. ■

Generally speaking, the values of α_e and α_β^* are both not unique, because optimal solutions of $\min_{\alpha\in R} G_\beta(x_e,\alpha)$ and (5.2.7) may have multiple optimal solutions. In the following analysis, we assume that our solutions α_e and α_β^* are the left points of the sets of optimal solutions, respectively, which means that they are the β-VaR of the decisions x_e and x_β^*.

In the following analysis, we define the indicator function by

$$I_{\hat{\Xi}} = \begin{cases} 1, & \text{if } \xi \in \hat{\Xi}, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, we get that the expectation of the random variable $h(\xi)$ on the set $\hat{\Xi}$ is given as $E[h(\xi)I_{\hat{\Xi}}]$ and $E[I_{\hat{\Xi}}] = \text{prob}\{\xi|\xi\in\hat{\Xi}\}$.

For any $x \in D$ and $\alpha \in R$, we denote the sets $\Xi(x,\alpha)$ and $\Xi^C(x,\alpha)$ by

$$\Xi(x,\alpha) := \{\xi | f(\xi,x) \le \alpha\}$$

and

$$\Xi^C(x,\alpha) := \{\xi | f(\xi,x) > \alpha\}.$$

We define the difference of the two sets $\Xi(x,\alpha_1)$ and $\Xi(x,\alpha_2)$ for $\alpha_2 \le \alpha_1$ by

$$\Xi(x,\alpha_1)\backslash\Xi(x,\alpha_2) = \{\xi|\alpha_2 < f(\xi,x) \le \alpha_1\}.$$

Theorem 5.2.1 *Suppose that $\Psi(x,\alpha)$ is continuous with respect to α. For any fixed $\beta\in(0,1)$, x_β^*, x_e are optimal solutions of the ERM formulation and (5.2.7) respectively, if and only if $\alpha_\beta^* \le \alpha_e$.*

Proof. From Theorem 1 in [54], we have

$$\text{prob}\{f(\xi, x_\beta^*) \leq \alpha_\beta^*\} = \text{prob}\{f(\xi, x_e) \leq \alpha_e\} = \beta$$

and

$$\begin{aligned}
G_\beta(x_\beta^*, \alpha_\beta^*) &= \alpha_\beta^* + (1-\beta)^{-1} E[(f(\xi, x_\beta^*) - \alpha_\beta^*)_+] \\
&= \alpha_\beta^* + (1-\beta)^{-1} E[(f(\xi, x_\beta^*) - \alpha_\beta^*) I_{\Xi^C(x_\beta^*, \alpha_\beta^*)}] \\
&= \alpha_\beta^* + (1-\beta)^{-1} E[f(\xi, x_\beta^*) I_{\Xi^C(x_\beta^*, \alpha_\beta^*)}] - (1-\beta)^{-1} \alpha_\beta^* E[I_{\Xi^C(x_\beta^*, \alpha_\beta^*)}] \\
&= (1-\beta)^{-1} E[f(\xi, x_\beta^*) I_{\Xi^C(x_\beta^*, \alpha_\beta^*)}].
\end{aligned}$$

Similarly, we get $G_\beta(x_e, \alpha_e) = (1-\beta)^{-1} E[f(\xi, x_e) I_{\Xi^C(x_e, \alpha_e)}]$. Hence, we have

$$E[f(\xi, x_\beta^*) I_{\Xi^C(x_\beta^*, \alpha_\beta^*)}] \leq E[f(\xi, x_e) I_{\Xi^C(x_e, \alpha_e)}].$$

In the following, we discuss our result in two aspects, and the first one is $\alpha_\beta^* \leq \alpha_e$.

We have

$$\begin{aligned}
& E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*)}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_e)}] \\
= \; & E[(f(\xi, x_\beta^*) - \alpha_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*)}] + E[(\alpha_\beta^* - f(\xi, x_e)) I_{\Xi(x_e, \alpha_\beta^*)}] \\
& + E[(\alpha_\beta^* - f(\xi, x_e)) I_{\Xi(x_e, \alpha_e) \setminus \Xi(x_e, \alpha_\beta^*)}] \\
= \; & E[(f(\xi, x_\beta^*) - \alpha_\beta^*) I_{\Xi(x_\beta^*, \alpha_1)}] + E[(f(\xi, x_\beta^*) - \alpha_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*) \setminus \Xi(x_\beta^*, \alpha_1)}] \\
& + E[(\alpha_\beta^* - f(\xi, x_e)) I_{\Xi(x_e, \alpha_\beta^*)}] + E[(\alpha_\beta^* - f(\xi, x_e)) I_{\Xi(x_e, \alpha_e) \setminus \Xi(x_e, \alpha_\beta^*)}] \\
= \; & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_1)}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_\beta^*)}] \\
& + E[(f(\xi, x_\beta^*) - \alpha_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*) \setminus \Xi(x_\beta^*, \alpha_1)}] + E[(\alpha_\beta^* - f(\xi, x_e)) I_{\Xi(x_e, \alpha_e) \setminus \Xi(x_e, \alpha_\beta^*)}],
\end{aligned}$$

where $\text{prob}\{\xi | \xi \in \Xi(x_\beta^*, \alpha_1)\} = \text{prob}\{\xi | \xi \in \Xi(x_e, \alpha_\beta^*)\} = \beta_1$, $\beta_1 \leq \beta$ and $\alpha_1 \leq \alpha_\beta^* \leq \alpha_e$.

We consider a sequence $\{\alpha_i\}_{i=0}^{\infty}$, which satisfies the following properties

$$\operatorname{prob}\{\xi|\xi \in \Xi(x_\beta^*, \alpha_{i+1})\} = \operatorname{prob}\{\xi|\xi \in \Xi(x_e, \alpha_i)\} = \beta_{i+1}. \tag{5.2.8}$$

We use the sequence $\{\alpha_i\}_{i=0}^{\infty}$ to decompose the support set Ξ and choose $\alpha_0 = \alpha_\beta^*$, $\beta_{i+1} \leq \beta_i$ and $\lim_{i\to\infty} \beta_{i+1} = 0$. We get

$$\begin{aligned}
& E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_{i+1})}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_i)}] \\
= \; & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta, \alpha_{i+2})}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_{i+1})}] \\
& + E[(f(\xi, x_\beta^*) - \alpha_{i+1}) I_{\Xi(x_\beta^*, \alpha_{i+1}) \setminus \Xi(x_\beta^*, \alpha_{i+2})}] + E[(\alpha_{i+1} - f(\xi, x_e)) I_{\Xi(x_e, \alpha_i) \setminus \Xi(x_e, \alpha_{i+1})}].
\end{aligned}$$

It is not difficult to see that

$$E[(f(\xi, x_\beta^*) - \alpha_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*) \setminus \Xi(x_\beta^*, \alpha_1)}] + E[(\alpha_\beta^* - f(\xi, x_e)) I_{\Xi(x_e, \alpha_e) \setminus \Xi(x_e, \alpha_\beta^*)}] \leq 0$$

and

$$E[(f(\xi, x_\beta^*) - \alpha_{i+1}) I_{\Xi(x_\beta^*, \alpha_{i+1}) \setminus \Xi(x_\beta^*, \alpha_{i+2})}] + E[(\alpha_{i+1} - f(\xi, x_e)) I_{\Xi(x_e, \alpha_i) \setminus \Xi(x_e, \alpha_{i+1})}] \leq 0,$$

which means

$$\begin{aligned}
& E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_{i+1})}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_i)}] \\
\leq \; & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_{i+2})}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_{i+1})}], \quad \text{for } i = 0, 1, 2, \cdots.
\end{aligned}$$

From Theorem 2.2.1, $f(\xi, x)$ is a residual function, which means $f(\xi, x) \geq 0$ for any $x \in D$ and almost every $\xi \in \Xi$. Hence, there exist $\underline{\alpha}_\beta^*$ and $\underline{\alpha}_e$ such that

$$\operatorname{prob}\{\xi|\xi \in \Xi(x_\beta^*, \underline{\alpha}_\beta^*)\} = \operatorname{prob}\{\xi|\xi \in \Xi(x_e, \underline{\alpha}_e)\} = 0.$$

Thus, we get

$$
\begin{aligned}
& E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_\beta^*)}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_e)}] \\
\leq\ & E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\underline{\alpha}_\beta^*)}]-E[f(\xi,x_e)I_{\Xi(x_e,\underline{\alpha}_e)}] \\
=\ & 0.
\end{aligned}
$$

Combining the fact of $E[f(\xi,x_\beta^*)I_{\Xi^C(x_\beta^*,\alpha_\beta^*)}] \leq E[f(\xi,x_e)I_{\Xi^C(x_e,\alpha_e)}]$, we get $E[f(\xi,x_\beta^*)] \leq E[f(\xi,x_e)]$. Moreover, we know x_e is an optimal solution of $\min_{x\in D} E[f(\xi,x)]$, so we have $E[f(\xi,x_\beta^*)] = E[f(\xi,x_e)]$ and $E[f(\xi,x_\beta^*)I_{\Xi^C(x_\beta^*,\alpha_\beta^*)}] = E[f(\xi,x_e)I_{\Xi^C(x_e,\alpha_e)}]$. Hence, we get x_β^* is an optimal solution of $\min_{x\in D} E[f(\xi,x)]$ and x_e is an optimal solution of (5.2.7).

The other aspect is that $\alpha_e < \alpha_\beta^*$. We get

$$
\begin{aligned}
& E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_\beta^*)}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_e)}] \\
=\ & E[(f(\xi,x_\beta^*)-\alpha_e)I_{\Xi(x_\beta^*,\alpha_e)}]+E[(\alpha_e-f(\xi,x_e))I_{\Xi(x_e,\alpha_e)}] \\
& +E[(f(\xi,x_\beta^*)-\alpha_e)I_{\Xi(x_\beta^*,\alpha_\beta^*)\setminus\Xi(x_\beta^*,\alpha_e)}] \\
=\ & E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_e)}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_1)}] \\
& +E[(f(\xi,x_\beta^*)-\alpha_e)I_{\Xi(x_\beta^*,\alpha_\beta^*)\setminus\Xi(x_\beta^*,\alpha_e)}]+E[(\alpha_e-f(\xi,x_e))I_{\Xi(x_e,\alpha_e)\setminus\Xi(x_e,\alpha_1)}].
\end{aligned}
$$

Similarly, we consider the sequence $\{\alpha_i\}_{i=1}^\infty$ defined by (5.2.8) with $\alpha_0=\alpha_e$, $\beta_{i+1}\leq\beta_i$ and $\lim_{i\to\infty}\beta_{i+1}=0$. Moreover, since α_e and α_β^* are both left points of optimal sets, we get

$$
\text{prob}\{\xi|\xi\in\Xi(x_\beta^*,\alpha_\beta^*)\}-\text{prob}\{\xi|\xi\in\Xi(x_\beta^*,\alpha_e)\}>0
$$

and

$$
E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_\beta^*)}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_e)}]>E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_e)}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_1)}].
$$

Furthermore, we also get

$$
E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_i)}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_{i+1})}]\geq E[f(\xi,x_\beta^*)I_{\Xi(x_\beta^*,\alpha_{i+1})}]-E[f(\xi,x_e)I_{\Xi(x_e,\alpha_{i+2})}],
$$

for $i = 0, 1, 2, \cdots$.

From the above analysis, we get

$$\begin{aligned} & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*)}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_e)}] \\ > \ & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \underline{\alpha}_\beta^*)}] - E[f(\xi, x_e) I_{\Xi(x_e, \underline{\alpha}_e)}] \\ = \ & 0. \end{aligned}$$

On the other hand, we employ the sequence $\{\alpha_i\}_{i=1}^{\infty}$ with $\alpha_0 = \alpha_\beta^*$, $\beta_i \le \beta_{i+1}$ and $\lim_{i\to\infty} \beta_{i+1} = 1$. Similarly, we get

$$\begin{aligned} 0 \ < \ & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_\beta^*)}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_e)}] \\ \le \ & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_1)}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_\beta^*)}] \end{aligned}$$

and

$$\begin{aligned} & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_{i+1})}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_i)}] \\ \le \ & E[f(\xi, x_\beta^*) I_{\Xi(x_\beta^*, \alpha_{i+2})}] - E[f(\xi, x_e) I_{\Xi(x_e, \alpha_{i+1})}] \\ \le \ & E[f(\xi, x_\beta^*)] - E[f(\xi, x_e)], \end{aligned}$$

for $i = 0, 1, 2, \cdots$.

Hence, we get

$$E[f(\xi, x_\beta^*)] > E[f(\xi, x_e)].$$

Thus, x_β^* is not an optimal solution of the ERM formulation. Moreover, we have

$$\text{prob}\{\xi | f(\xi, x_e) \le \alpha_\beta^*\} > \text{prob}\{\xi | f(\xi, x_e) \le \alpha_e\} = \beta,$$

which implies x_e is not an optimal solution of (5.2.7).

So we obtain our desirable result. ■

Suppose that for any $x \in D$ and $\alpha \in R$ the expectation of the random matrix M_ξ over the set $\Xi(x, \alpha)$ is positive semi-definite, which means

$$E[M_\xi I_{\Xi^C(x,\alpha)}] \succeq 0. \tag{5.2.9}$$

It is easy to see that if the matrix M_ξ is positive semi-definite for almost every $\xi \in \Xi$, condition (5.2.9) holds.

Proposition 5.2.2 *(i) For any fixed $\alpha \in R$, the function $G_\beta(x, \alpha)$ is locally Lipschitz continuous and for any fixed $x \in D$ it is also locally Lipschitz continuous.*

(ii) Suppose that condition (5.2.9) holds. Then $G_\beta(x, \alpha)$ is convex with respect to (x, α).

Proof. (i) For any fixed $\alpha \in R$ and $x, y \in D$, we have

$$\begin{aligned}
|G_\beta(x,\alpha) - G_\beta(y,\alpha)| &= (1-\beta)^{-1}|E[(f(\xi,x)-\alpha)_+] - E[(f(\xi,y)-\alpha)_+]| \\
&= (1-\beta)^{-1}|E[(f(\xi,x)-\alpha)_+ - (f(\xi,y)-\alpha)_+]| \\
&\le (1-\beta)^{-1}E[|(f(\xi,x)-\alpha)_+ - (f(\xi,y)-\alpha)_+|] \\
&\le (1-\beta)^{-1}E[|f(\xi,x) - f(\xi,y)|].
\end{aligned}$$

Moreover, from Theorem 2.2.2, we know the residual function f is locally Lipschitz continuous with respect to x. Hence, we obtain the function $G_\beta(x, \alpha)$ is locally Lipschitz continuous with respect to x.

For any fixed $x \in D$ and $\alpha_1, \alpha_2 \in R$, in the similar way, we have

$$\begin{aligned}
& |G_\beta(x,\alpha_1) - G_\beta(x,\alpha_2)| \\
\le\ & |\alpha_1 - \alpha_2| + (1-\beta)^{-1}|E[(f(\xi,x)-\alpha_1)_+] - E[(f(\xi,x)-\alpha_2)_+]| \\
\le\ & |\alpha_1 - \alpha_2| + (1-\beta)^{-1}E[|\alpha_1 - \alpha_2|] \\
=\ & (1 + (1-\beta)^{-1})|\alpha_1 - \alpha_2|.
\end{aligned}$$

So we obtain our desirable result.

(ii) Under condition (5.2.9), we know $E[(f(\xi,x)-\alpha)_+] = E[(f(\xi,x)-\alpha)I_{\Xi^C(x,\alpha)}]$ is convex in x and α, so $G_\beta(x,\alpha)$ is convex with respect to (x,α). ■

From Theorem 2.2.2, for any fixed $\xi \in \Xi$, $f(\xi,x)$ is Lipschitz continuous but not differentiable, so we get $G_\beta(x,\alpha)$ is Lipschitz continuous but not differentiable with respect to x. If the probability function $\Psi(x,\alpha)$ is continuous with respect to α, $G_\beta(x,\alpha)$ is continuously differentiable in α. Otherwise, for any fixed $x \in D$, $G_\beta(x,\alpha)$ is Lipschitz continuous but not differentiable. For any fixed $\xi \in \Xi$, we assume that $\tilde{f}(\xi,x,\mu)$ is a smoothing function of $f(\xi,x)$ and $\tilde{s}(t,\alpha,\mu)$ is a smoothing function of $(t-\alpha)_+$.

Then, we define

$$\tilde{G}_\beta(x,\alpha,\mu) = \alpha + (1-\beta)^{-1}E[\tilde{s}(\tilde{f}(\xi,x,\mu),\alpha,\mu)]. \tag{5.2.10}$$

The Clark subdifferential of G_β at $(x,\alpha) \in D \times R$ is denoted by

$$\partial G_\beta(x,\alpha) = \mathrm{con}\partial_B G_\beta(x,\alpha),$$

where

$$\partial_B G_\beta(x,\alpha) = \{\begin{pmatrix} w \\ v \end{pmatrix} \mid \begin{pmatrix} \nabla_x G_\beta(z,\nu) \\ \nabla_\alpha G_\beta(z,\nu) \end{pmatrix} \to \begin{pmatrix} w \\ v \end{pmatrix}\}.$$

Here G_β is differentiable at $(z,\nu), z \to x, \nu \to \alpha$.

Moreover, we denote the subdifferential associated with a smoothing function by

$$G_{\tilde{G}_\beta}(x,\alpha) = \mathrm{con}\{\lim_{x^k\to x,\alpha^k\to\alpha,\mu_k\downarrow 0} \begin{pmatrix} w^k \\ v^k \end{pmatrix}\},$$

where $\begin{pmatrix} w^k \\ v^k \end{pmatrix} = \begin{pmatrix} \nabla_x \tilde{G}_\beta(x^k,\alpha^k,\mu_k) \\ \nabla_\alpha \tilde{G}_\beta(x^k,\alpha^k,\mu_k) \end{pmatrix}$.

Theorem 5.2.2 *The function $\tilde{G}_\beta(x,\alpha,\mu)$ defined by (5.2.10) is a smoothing function of*

$G_\beta(x,\alpha)$. *Moreover, for any* $(x,\alpha) \in D \times R$, $\{\lim_{x^k\to x,\alpha^k\to\alpha,\mu_k\downarrow 0} \begin{pmatrix} w^k \\ v^k \end{pmatrix}\}$ *is nonempty and bounded, and the gradient consistency* $\partial G_\beta(x,\alpha) = G_{\tilde{G}_\beta}(x,\alpha)$ *holds.*

Proof. First, we get

$$\begin{aligned}
& |\tilde{G}_\beta(z,\nu,\mu) - G_\beta(x,\alpha)| \\
\le\ & |\nu-\alpha| + (1-\beta)^{-1}|E[\tilde{s}(\tilde{f}(\xi,z,\mu),\nu,\mu) - (f(\xi,x)-\alpha)_+]| \\
\le\ & |\nu-\alpha| + (1-\beta)^{-1}E[|\tilde{s}(\tilde{f}(\xi,z,\mu),\nu,\mu) - (f(\xi,x)-\alpha)_+|] \\
\le\ & |\nu-\alpha| + (1-\beta)^{-1}E[|\tilde{s}(\tilde{f}(\xi,z,\mu),\nu,\mu) - (\tilde{f}(\xi,z,\mu)-\nu)_+| \\
& +|(\tilde{f}(\xi,z,\mu)-\nu)_+ - (f(\xi,x)-\alpha)_+|] \\
\le\ & |\nu-\alpha| + (1-\beta)^{-1}(E[|\tilde{s}(\tilde{f}(\xi,z,\mu),\nu,\mu) - (\tilde{f}(\xi,z,\mu)-\nu)_+|] \\
& +E[|\tilde{f}(\xi,z,\mu) - f(\xi,x)|] + |\nu-\alpha|).
\end{aligned}$$

Since $\tilde{s}(t,\alpha,\mu)$ is a smoothing function of $(t-\alpha)_+$, we get $\lim_{\mu\to 0}|\tilde{s}(\tilde{f}(\xi,z,\mu),\nu,\mu) - (\tilde{f}(\xi,z,\mu)-\nu)_+| = 0$. Similarly, we get $\lim_{z\to x,\mu\to 0}|\tilde{f}(\xi,z,\mu) - f(\xi,x)| = 0$. Moreover, we have $\lim_{\nu\to\alpha}|\nu-\alpha|$=0, so $\lim_{z\to x,\nu\to\alpha,\mu\to 0}|\tilde{G}_\beta(z,\nu,\mu) - G_\beta(x,\alpha)| = 0$, which implies $\tilde{G}_\beta$ is a smoothing function of G_β. Furthermore, we also get the gradient of the smoothing function $\tilde{G}_\beta$ as follows:

$$\begin{pmatrix} w \\ v \end{pmatrix} = \begin{pmatrix} (1-\beta)^{-1}E[\nabla_t\tilde{s}(\xi,t,\alpha)|_{t=\tilde{f}(\xi,x,\mu)}\nabla_x\tilde{f}(\xi,x,\mu)] \\ 1+(1-\beta)^{-1}E[\nabla_\alpha\tilde{s}(\xi,t,\alpha)|_{t=\tilde{f}(\xi,x,\mu)}] \end{pmatrix}.$$

Hence, we know

$$\{\lim_{x^k\to x,\alpha^k\to\alpha,\mu_k\downarrow 0}\begin{pmatrix} w^k \\ v^k \end{pmatrix}\} \subseteq \begin{pmatrix} (1-\beta)^{-1}E[\partial_t(t-\alpha)_+|_{t=f(\xi,x)}\partial_x f(\xi,x)] \\ 1+(1-\beta)^{-1}E[-\partial_\alpha(t-\alpha)_+|_{t=f(\xi,x)}] \end{pmatrix}.$$

For the stochastic linear VI and any fixed $\xi \in \Xi$, we have

$$f(\xi,x) = u(\xi,x)^T(M_\xi u(\xi,x) + q_\xi) + Q(\xi,x),$$

where $u(\xi, x) = x + A^{\dagger}(b_{\xi} - Ax)$ and $Q(\xi, x) = \min\{b_{\xi}^T z | A^T z + M_{\xi} u(\xi, x) + q_{\xi} \geq 0\}$ is linear in x. Since the first part of $f(\xi, x)$ is strictly differentiable and $Q(\xi, x)$ is convex in x, by Proposition 2.3.6 in [21], we get $f(\xi, x)$ is regular at x. Moreover, we know $(t - \alpha)_+$ is regular at t and $\forall p \in \partial(\cdot)_+$ is nonnegative, so by the chain rule Theorem 2.3.9 in [21] we get

$$(1 - \beta)^{-1} E[\partial_t (t - \alpha)_+ |_{t=f(\xi,x)} \partial_x f(\xi, x)] = \partial_x G_{\beta}(x, \alpha).$$

Similarly, we obtain

$$1 + (1 - \beta)^{-1} E[-\partial_{\alpha}(t - \alpha)_+ |_{t=f(\xi,x)}] = \partial_{\alpha} G_{\beta}(x, \alpha).$$

Hence, we obtain

$$\text{con}\{\lim_{x^k \to x, \alpha^k \to \alpha, \mu_k \to 0} \begin{pmatrix} w^k \\ v^k \end{pmatrix}\} = G_{\tilde{G}_{\beta}}(x, \alpha) \subseteq \partial G_{\beta}(x, \alpha).$$

On the other hand, from Proposition 5.2.2 we know G_{β} is locally Lipschitiz continuous and $\partial G_{\beta}(x, \alpha)$ is nonempty. According to Theorem 9.61 and Corollary 8.47 in [56], we get $\partial G_{\beta}(x, \alpha) \subseteq G_{\tilde{G}_{\beta}}(x, \alpha)$. Then, we get the gradient consistency $\partial G_{\beta}(x, \alpha) = G_{\tilde{G}_{\beta}}(x, \alpha)$. ■

Corollary 5.2.1 *Suppose that $\Psi(x, \alpha)$ is continuous with respect to α and $\tilde{G}_{\beta}(x, \alpha, \mu) = \alpha + (1 - \beta)^{-1} E[(\tilde{f}(\xi, x, \mu) - \alpha)_+]$ is a smoothing function of $G_{\beta}(x, \alpha)$. For any $(x, \alpha) \in D \times R$, $\{\lim_{x^k \to x, \mu_k \downarrow 0} \begin{pmatrix} w^k \\ v \end{pmatrix}\}$ is nonempty and bounded, and the gradient consistency $\partial G_{\beta}(x, \alpha) = G_{\tilde{G}_{\beta}}(x, \alpha)$ holds.*

Proof. This is a special case of Theorem 5.2.2 and the proof is similar with Theorem 5.2.2, so we omit it here. ■

For some applications, such as traffic equilibrium problem, because of the special structure of the matrix A, the recourse problem $Q(\xi, u(\xi, x))$ has explicit form. For the transportation problem, the elements of the matrix A is either 1 or 0 and every column of the matrix A only has a positive number, so we define the index set I_i as follows:

$$I_i = \{j : \ A(i,j) \neq 0, \ \forall \ j = 1, \cdots, n\}.$$

The explicit form of the recourse problems is $Q(\xi, u(\xi, x)) = \sum_{i=1}^{m} b_{\xi i} \max_{j \in I_i}\{-F_j(\xi, u(\xi, x))\}$, where $F_j(\xi, u(\xi, x))$ denotes the j-th element of the vector $F(\xi, u(\xi, x))$. Then we get the explicit form of the residual function as

$$f(\xi, x) = g(\xi, x) + \sum_{i=1}^{m} b_{\xi i} \max_{j \in I_i}\{-F_j(\xi, u(\xi, x))\}. \tag{5.2.11}$$

Then, $G_\beta(x, \alpha)$ defined by the residual function (5.2.11) is given as the following one:

$$\begin{aligned} & G_\beta(x, \alpha) \\ = \quad & \alpha + (1-\beta)^{-1} E[(g(\xi, x) + \sum_{i=1}^{m} b_{\xi i} \max_{j \in I_i}\{-F_j(\xi, u(\xi, x))\} - \alpha)_+] \end{aligned} \tag{5.2.12}$$

We know that one of the smoothing functions of $p(y) = \max_{1 \leq i \leq k}\{y_i\}$ is defined by $\tilde{p}(y, \mu) = \mu \ln \sum_{i=1}^{k} e^{\frac{y_i}{\mu}}$ in [47].

We define the smoothing function of the CVaR formulation (5.2.12) as follows:

$$\begin{aligned} & \tilde{G}(x, \alpha, \mu) \\ = \quad & \alpha + \mu(1-\beta)^{-1} E[\ln(1 + e^{\frac{g(\xi,x)-\alpha}{\mu}} (\prod_{i=1}^{m} \sum_{j \in I_i} e^{\frac{-b_{\xi i} F_j(\xi, u(\xi, x))}{\mu}})]. \end{aligned} \tag{5.2.13}$$

Remark 5.2.1 *If we do not get the explicit form of the recourse problem, we suppose that $\tilde{f}(\xi, x, \mu)$ is a smoothing function of $f(\xi, x)$. Hence, $\tilde{G}(x, \alpha, \mu) = \alpha + \mu(1 - \beta)^{-1} E[ln(1 + e^{\frac{\tilde{f}(\xi,x,\mu)-\alpha}{\mu}})]$ is a smoothing function of $G_\beta(x, \alpha)$.*

5.3 Distributionally robust stochastic linear VI

In this section, we consider the distributionally robust CVaR formulation of stochastic linear VI using the sublinear expectation.

By the sublinear expectation, the β-CVaR formulation for the robust case is denoted by the following one:

$$\min_{x \in D, \alpha \in R} G_\beta^R(x, \alpha) = \alpha + (1-\beta)^{-1}\mathbb{E}[(f(\xi, x) - \alpha)_+], \tag{5.3.14}$$

where $\mathbb{E}[\cdot]$ denotes the sublinear expectation.

We generalize Definition 1.2.1 as follows:

Definition 5.3.1 *For any fixed x and $h_i(\xi, x) : \Xi \times R^n \to R$ where $i = 1, 2$, a sublinear expectation $\mathbb{E}$ is a functional $\mathbb{E} : \mathcal{H} \to R$ satisfying*

(i) *Monotonicity: $\mathbb{E}[h_1(\xi, x)] \geq \mathbb{E}[h_2(\xi, x)]$ if $h_1(\xi, x) \geq h_2(\xi, x)$.*

(ii) *Constant preserving: $\mathbb{E}[c] = c$ for $c \in R$.*

(iii) *Sub-additivity: For each $h_1(\cdot, x), h_2(\cdot, x) \in \mathcal{H}$, $\mathbb{E}[h_1(\xi, x) + h_2(\xi, x)] \leq \mathbb{E}[h_1(\xi, x)] + \mathbb{E}[h_2(\xi, x)]$.*

(iv) *Positive homogeneity: $\mathbb{E}[\lambda h_1(\xi, x)] = \lambda \mathbb{E}[h_1(\xi, x)]$ for $\lambda \geq 0$.*

Proposition 5.3.1 *For any fixed x, $c \in R$ and $h(\xi, x) : \Xi \times R^n \to R$, we have $\mathbb{E}[h(\xi, x) + c] = \mathbb{E}[h(\xi, x)] + c$.*

Proof. From Definition 5.3.1, we get

$$\begin{aligned}
2\mathbb{E}[h(\xi, x)] &= \mathbb{E}[2h(\xi, x)] = \mathbb{E}[h(\xi, x) + c + h(\xi, x) - c] \\
&\leq \mathbb{E}[h(\xi, x) + c] + \mathbb{E}[h(\xi, x) - c] \\
&\leq \mathbb{E}[h(\xi, x)] + \mathbb{E}[c] + \mathbb{E}[h(\xi, x)] + \mathbb{E}[-c] \\
&= 2\mathbb{E}[h(\xi, x)].
\end{aligned}$$

From the above formulation, we get all of equalities should hold. Moreover, from Definition 5.3.1 we know $\mathbb{E}[h(\xi, x)+c] \leq \mathbb{E}[h(\xi, x)]+c$ and $\mathbb{E}[h(\xi, x)-c] \leq \mathbb{E}[h(\xi, x)]+\mathbb{E}[-c]$. If $\mathbb{E}[h(\xi, x)+c] \neq \mathbb{E}[h(\xi, x)]+c$, we have $\mathbb{E}[h(\xi, x)+c]+\mathbb{E}[h(\xi, x)-c] < \mathbb{E}[h(\xi, x)]+c+\mathbb{E}[h(\xi, x)]+\mathbb{E}[-c]$ which conflicts with that all of equalities hold. Based on the above analysis, we get $\mathbb{E}[h(\xi, x)+c] = \mathbb{E}[h(\xi, x)]+c$ holds. ■

Lemma 5.3.1 [51] *Let $\mathbb{E}$ be a functional defined on a linear space $\mathcal{H}$ satisfying sub-additivity and positive homogeneity. Then there exists a family of linear functionals $\{E_\theta : \theta \in \Theta\}$ defined on $\mathcal{H}$ such that*

$$\mathbb{E}[X] = \sup_{\theta \in \Theta} E_\theta[X], \qquad \text{for } X \in \mathcal{H}$$

and, for each $X \in \mathcal{H}$, there exists $\theta_X \in \Theta$ such that $\mathbb{E}[X] = E_{\theta_X}[X]$.

Furthermore, if $\mathbb{E}$ is a sublinear expectation, then the corresponding E_θ is a linear expectation.

Theorem 5.3.1 *For any fixed $x \in D$, $G_\beta^R(x, \alpha)$ is finite and convex with respect to α, and the robust β-CVaR of the loss associated with $x \in D$ can be given from*

$$\phi_\beta^R(x) = \min_{\alpha \in R} G_\beta^R(x, \alpha),$$

and the optimal set is a nonempty, closed and bounded interval.

Proof. For any fixed $x \in D$, since the linear expectation $E_\theta[f(\xi, x)] < \infty$ holds for any distributions $\theta \in \Theta$, we have $\mathbb{E}[f(\xi, x)]$ is also finite, which implies the finiteness of the function $G_\beta^R(x, \alpha)$. Moreover, by Lemma 5.3.1 we get the explicit form of the sublinear expectation $\mathbb{E}[(f(\xi, x)-\alpha)_+] = \sup_{\theta \in \Theta} E_\theta[(f(\xi, x)-\alpha)_+]$, and we know for any fixed $x \in D$ the function $E_\theta[(f(\xi, x)-\alpha)_+]$ is convex in α. Hence, we obtain the convexity.

From Definition 5.3.1 and Proposition 5.3.1, we get

$$\begin{aligned} G_\beta^R(x,\alpha) &= \alpha + (1-\beta)^{-1}\mathbb{E}[(f(\xi,x)-\alpha)_+] \\ &= (1-\beta^{-1})\mathbb{E}[(1-\beta)\alpha + (f(\xi,x)-\alpha)_+]. \end{aligned}$$

Moreover, from Lemma 5.3.1 and Theorem 10 in [55], for any fixed $x \in D$, we get the existence of optimal solutions of $\min_\alpha G_\beta^R(x,\alpha)$.

We get our desirable results. ■

Theorem 5.3.2 *Suppose that condition (5.2.9) holds. Then the robust β-CVaR formulation defined by (5.3.14) is convex over all $(x,\alpha) \in D \times R$, and $\min_{x\in D} \phi_\beta^R(x) = \min_{x\in D,\alpha\in R} G_\beta^R(x,\alpha)$ holds.*

Proof. From Proposition 5.2.2 and Lemma 5.3.1, if condition (5.2.9) holds, we get $\mathbb{E}[(f(\xi,x)-\alpha)_+]$ is convex with respect to (x,α), which implies the convexity of $G_\beta^R(x,\alpha)$ over all $(x,\alpha) \in D \times R$. From Theorem 5.3.1, optimal solutions of $\min_{\alpha_R} G_\beta^R(x,\alpha)$ exist and can be obtained, so we get the equivalence of the two optimization problems.

Hence, we get the desirable result. ■

Proposition 5.3.2 *The robust β-CVaR defined by the sublinear expectation is a coherent risk measure when the function $f(\xi,x)$ is linear in x, and if $f(\xi,x) = c$, then $\phi_\beta^R(x) = c$. Moreover, if $f(\xi,x_1) \le f(\xi,x_2)$, then $\phi_\beta^R(x_1) \le \phi_\beta^R(x_2)$.*

Generally speaking, the distributions of random variable $\xi \in \Xi$ has the following four parameters:

$$\bar{\mu} := \mathbb{E}[\xi], \qquad \underline{\mu} := -\mathbb{E}[-\xi],$$

and

$$\bar{\sigma}^2 := \mathbb{E}[\xi^2], \qquad \underline{\sigma}^2 := -\mathbb{E}[-\xi^2].$$

Suppose that the random variable ξ follows the normal distribution such that the mean

$\mu = 0$ and the standard deviation $\sigma^2 \in [\underline{\sigma}^2, \bar{\sigma}^2]$, which is also called the G-normal distribution [50].

Proposition 5.3.3 [50] *Suppose that the random variable ξ follows the G-normal distribution. If the function $(f(\xi, x) - \alpha)_+$ is convex in ξ, then for any fixed $x \in D$ and $\alpha \in R$,*

$$\begin{aligned} G_\beta(x, \alpha) &= \alpha + (1 - \beta)^{-1}\mathbb{E}[(f(\xi, x) - \alpha)_+] \\ &= \alpha + (1 - \beta)^{-1}E_{\theta^*}[(f(\xi, x) - \alpha)_+], \end{aligned}$$

where E_{θ^*} denotes the linear expectation, the density function of which is the normal distribution $\mathcal{N}(0, \bar{\sigma}^2)$; but if the function $(f(\xi, x) - \alpha)_+$ is concave in ξ, the above $\bar{\sigma}^2$ should be replaced by $\underline{\sigma}^2$.

In reality, lots of random variables follow the normal distribution. Since the parameters of the normal distribution may have uncertainty, we can use the G-normal distribution to reduce the risk of a decision. Moreover, when $(f(\xi, x) - \alpha)_+$ is convex or concave in ξ, the explicit form of the sublinear expectation can be given and we can use the SAA method to obtain the value of $\mathbb{E}[(f(\xi, x) - \alpha)_+]$.

Chapter 6

Applications and numerical experiments

In this chapter, we employ the traffic equilibrium to illustrate our applications. Numerical results for the linear or nonlinear cases show that our ERM formulation has some good properties, such as robustness and high probability.

For the numerical experiments, we use the semi-smooth Newton method [53] to get a solution x_{EV} of the EV formulation. We use examples coming from traffic equilibrium assignment to illustrate the ERM formulation (2.1.1) and the SSAA-method. We derive an explicit expression for $Q(\xi, x)$ and its smoothing approximation for a class of stochastic VI and show that all conditions used in the above chapters are satisfied.

Moreover, we present numerical results to compare the solution of ERM formulation defined by our new residual function with that of the EV formulation.

It is remarkable that for all the applications being considered the only requirement is that the sampling should be independent and identically distributed, (abbreviated iid) whereas related convergence results require strong conditions, for example, uniform convergence of the approximating functions.

6.1 Application

6.1.1 Stochastic VI for traffic equilibrium

The model of traffic or transportation obtained by the transportation networks is usually used to forecast the future traffic flows to avoid the congestions. For the transportation assignment, the Wardrop's user equilibrium is widely used to define an equilibrium point. The Wardrop's user equilibrium principle states that the user-optimized traffic pattern with the equilibrium property that, once established, no user may decrease his travel cost by making a unilateral decision to change his route. We can easily find that the model is based on the demand and the cost functions. Because of the influence of the weather, accidents and so on, we should consider the uncertainty of the traffic networks and obtain a stochastic user equilibrium. It can be represented as the following stochastic VI:

$$\langle F(\xi, x), y - x \rangle \geq 0 \quad \forall\, y \in \{x \mid Ax = b_\xi,\ x \geq 0\}, \tag{6.1.1}$$

where $F : \Xi \times R^n \to R^n$ is the cost function, b_ξ denotes the demand of the traffic network and $A \in R^{m\times n}$ is the OD-path incidence matrix whose entries are given by

$$A_{ir} = \begin{cases} 1, & \text{if the path } r \text{ connects the OD pair } i, \\ 0, & \text{otherwise,} \end{cases}$$

x denotes the travel flow, and the link flow vector $v = \Delta x$, where Δ is the link-path incidence matrix whose components are given by

$$\Delta_{ak} = \begin{cases} 1, & \text{if the link } a \text{ on the path } k, \\ 0, & \text{otherwise.} \end{cases}$$

Traffic equilibrium models are built based on travel demand between every OD-pair and travel capacity on each link. The demand and capacity depend heavily on various uncertain parameters, such as weather, accidents, etc. Let $\Xi \subseteq R^L$ denote the set of uncertain factors. Let $(b_\xi)_i > 0$ denote the stochastic travel demand on the ith OD pair and $(c_\xi)_k$ denote the stochastic capacity of link k.

The link travel time function $T(\xi, v)$ is a stochastic vector and each of its entries $T_a(\xi, v)$ is assumed to follow a generalized Bureau of Public Roads (GBPR) function,

$$T_a(\xi, x) = t_a^0(1 + b_a(\frac{v_a}{(c_\xi)_a})^{n_a}), \tag{6.1.2}$$

where t_a^0, b_a and n_a are given parameters, and $(c_\xi)_a$ denotes the stochastic capacity on the link a for uncertainty $\xi \in \Xi$. The path travel cost function is defined by

$$F(\xi, x) = \eta_1 \Delta^T T(\xi, \Delta x), \tag{6.1.3}$$

where $\eta_1 > 0$ is the time-based operating costs factor.

A traffic network consists of a set of nodes and a set of links. We denote by W the origin-destination (OD) pairs and K the set of all paths between OD-pairs. The network in Figure 6.1 from [65] has 5 nodes, 7 links, 2 OD-pairs ($1 \to 4$, $1 \to 5$) and 6 paths $p_1 = \{3, 7, 6\}, p_2 = \{3, 1\}, p_3 = \{4, 6\}, p_4 = \{3, 7, 2\}, p_5 = \{3, 5\}, p_6 = \{4, 2\}$.

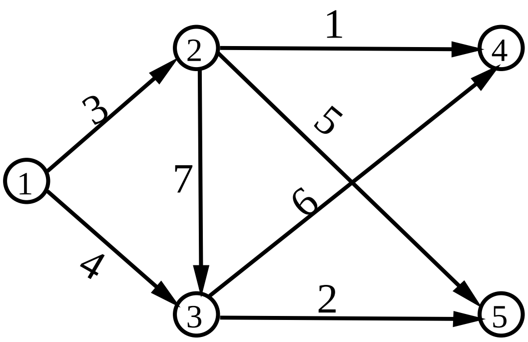

Figure 6.1: The 7-links, 6-paths network

For a realization of random vectors $b_\xi \in R^2$ and $c_\xi \in R^7$, $\xi \in \Xi$, an assignment of flows to all paths is denoted by the vector $x \in R^6$, whose component x_j denotes the flow on path j, while an assignment of flows to all links is represented by the vector v

whose component v_k denotes the stochastic flow on link k. The incidence matrices for the network in Figure 6.1 are given respectively as follows.

$$\Delta = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

If $n_k = 1$, then $F(\xi, x) = M_\xi x + q$, where

$$M_\xi = 0.15\eta_1 \Delta^T \text{diag}\left(\frac{t_k^0}{(c_\xi)_k}\right)\Delta \quad \text{and} \quad q = \eta_1 t_1^0 \Delta^T e.$$

Moreover, for Figure 6.1, we note that rank(Δ)=5 for any $\xi \in \Xi$. $M_\xi \in R^{6\times 6}$ is a positive semi-definite matrix with rank(M_ξ) = 5. Obviously, $E[M_\xi]$ is positive semi-definite, but condition (4.1.5) used in [2] does not hold.

In a stochastic environment, ξ belongs to a set Ξ representing future states of knowledge. In general, we cannot find a vector $\bar{x}$ such that $f(\xi, \bar{x}) = 0$ for all $\xi \in \Xi$. The ERM formulation is to find a vector x^* which minimizes the expected value of $f(\xi, \bar{x})$ over Ξ. The main role of traffic model is to provide a forecast for future traffic states. The solution of the ERM formulation is a "here and now" solution which provides a robust forecast and has advantages over other models for long term planning.

6.1.2 Efficiency of the ERM formulation

In this section, we give sufficient conditions on A and b_ξ that guarantee that Assumptions 2.2.1, 2.2.2, 4.1.1, 4.1.2 and 5.2.1 hold. Such conditions hold for the OD-path incidence matrix and random demand vector.

Definition 6.1.1 [22] *A set $S \subseteq R^m$ is a meet semi-sublattice under the componentwise ordering of R^m if*

$$u, v \in S \quad \Rightarrow \quad w = \min(u, v) \in S.$$

The vector w is called the meet of u and v.

Lemma 6.1.1 [22] *If S is a nonempty meet semi-sublattice that is closed and bounded below, then S has a least element.*

Theorem 6.1.1 *Suppose* $\text{prob}\{b_\xi > 0, \|b_\xi\|_\infty \leq \beta\} = 1$ *for some $\beta > 0$ and A can be split into two submatrices A_K and A_J, where A_K is an $m \times m$ M-matrix and A_J is an $m \times (n-m)$ nonnegative matrix whose columns have only one positive entry. Let*

$$\gamma_0 = \min_{i,j}\{(A_K^{-1}A_J)_{ij} \,|\, (A_K^{-1}A_J)_{ij} > 0,\, j \in J,\, 1 \leq i \leq m\}, \quad \gamma = \max(1, \gamma_0^{-1})\beta\|A_K^{-1}\|_\infty.$$

Then,

$$X_\xi \subseteq \{x \,|\, 0 \leq x \leq \gamma e\} =: U_0. \tag{6.1.4}$$

Further, if for some $\kappa > 0$ and any $u \in U_0$, $\text{prob}\{\|F(\xi, u)\|_\infty \leq \kappa\} = 1$, *then Assumption 2.2.1 holds with $Q(\xi, u(\xi, x)) = b_\xi^T z(\xi, u(\xi, x))$ and*

$$\|z(\xi, u(\xi, x))\|_\infty \leq \theta = \kappa \max(1, \gamma_0^{-1})\|A_K^{-T}\|_\infty \tag{6.1.5}$$

for any $x \in D$ and almost every $\xi \in \Xi$.

Proof. Let P be $n \times n$ permutation matrix such that $AP = [A_K, A_J]$. For fixed $\xi \in \Xi$, consider a vector $x \in X_\xi$ with $x_{j_0} = \max_j x_j = \|x\|_\infty$. By definition,

$$A_K^{-1}b_\xi = A_K^{-1}APPx = A_K^{-1}[A_K, A_J]Px = [I, A_K^{-1}A_J]Px. \tag{6.1.6}$$

Since $[I, A_K^{-1}A_J]$ is a nonnegative matrix and its each column has at least one positive element, $[I, A_K^{-1}A_J]Px \geq 0$. Hence, there is a positive element $(I, A_K^{-1}A_J)_{i,j_0} = B_{i,j_0} \geq$

$\min(1, \gamma_0)$, such that

$$\min(1, \gamma_0)\|x\|_\infty \le B_{i,j_0} x_{j_0} \le \|[I, A_K^{-1} A_J] P x\|_\infty \le \|A_K^{-1} b_\xi\|_\infty \le \|A_K^{-1}\|_\infty \beta \text{ a.s.}$$

This implies $X_\xi \subseteq U_0$ a.s.

Let $S_{\xi,u} = \{z \,|\, A^T z + F(\xi, u) \ge 0\}$ denote the feasible set. For $w, v \in S_{\xi,u}$, let $s = \min(w, v)$ be their meet. We consider an arbitrary index $i \in \{1, \cdots, n\}$. By the assumptions of this theorem, there is at most one positive element $a_{ki} > 0$. Without loss of generality, we assume $s_k = v_k$. Then,

$$\begin{aligned}
(A^T s + F(\xi, u))_i &= F_i(\xi, u) + \sum_{j \ne k}^{m} a_{ji} s_j + a_{ki} s_k \\
&\ge F_i(\xi, u) + \sum_{j \ne k}^{m} a_{ji} v_j + a_{ki} v_k \\
&\ge 0.
\end{aligned}$$

This establishes the feasibility of the vector s and the meet semi-sublattice property of $S_{\xi,u}$.

Let $e \in R^m$ and $\tilde{e} \in R^n$ be vectors with all of their elements 1. Let $t = \kappa \max(1, \gamma_0^{-1}) A_K^{-T} e$. Note that $A_J^T A_K^{-T}$ is a nonnegative matrix. Then

$$PA^T t = \kappa \max(1, \gamma_0^{-1}) \begin{pmatrix} e \\ A_J^T A_K^{-T} e \end{pmatrix} \ge \kappa \tilde{e} \ge -PF(\xi, u) \text{ a.s.}$$

Hence $t \in S_{\xi,u}$ and thus $S_{\xi,u}$ is nonempty, a.s.

Let $C = [A_K^{-T}, 0] \in R^{m \times n}$. For any $z \in S_{\xi,u}$,

$$CP(A^T z + F(\xi, u)) = z + CPF(\xi, u) \ge 0,$$

which implies

$$z \ge -CPF(\xi, u) \ge -LA_K^{-T} e \ge -\max(1, \gamma_0^{-1}) \kappa A_K^{-T} e. \tag{6.1.7}$$

Hence $S_{\xi,u}$ is closed and bounded below. By Lemma 6.1.1, $S_{\xi,u}$ has a unique least element $z(\xi,u)$, a.s. Moreover, by the assumption $b_\xi > 0$ a.s., $z(\xi,u)$ is the unique solution of (2.2.3) a.s.

Furthermore, using $z(\xi,u) \le t$ and (6.1.7),

$$\|z(u,\xi)\|_\infty \le \kappa \max(1,\gamma_0^{-1})\|A_K^{-T}\|_\infty = \theta \text{ a.s.} \tag{6.1.8}$$

which completes the proof. ■

In traffic flow problem [1, 65, 69], we often have the following constraints

$$X_\xi = \{x \,|\, \sum_{j\in I_i} x_j = (b_\xi)_i, \quad i = 1,\cdots,m\} \tag{6.1.9}$$

with

$$\bigcup_{i=1}^{m} I_i = \{1,2,\cdots,n\}, \qquad I_i \cap I_j = \emptyset,\, i \ne j,$$

where b_ξ is a demand vector which comes with uncertainties due to weather, accidents, etc., $x_j, j \in I_i$ are traffic flows on the j path connecting the i-th original-destination (OD) pair. The constraints (6.1.9), can be written as $Ax = b_\xi$, where A is called the OD-path incidence matrix. Each column of A has only one nonzero element 1 and the ith row has $|I_i|$ elements. Such matrix satisfies the assumption on A in Theorem 6.1.1. Moreover, if $b_\xi > 0$, then from $A^T z + F(\xi,u) \ge 0$, the solution $z(\xi,u)$ of (4.1.3) has a closed form

$$z_i(\xi,u) = \max\{-F_j(\xi,u), j \in I_i\}, \quad i = 1,\cdots,m. \tag{6.1.10}$$

Moreover, If $F(\xi,x) = M_\xi x + q_\xi$, then φ is a convex function.

Now, we define a smoothing function of

$$f(\xi,x) = u(\xi,x)^T F(\xi,u(\xi,x)) + \sum_{i=1}^{m} b_i(\xi) \max_{j\in I_i}\{-F_j(\xi,u(\xi,x))\}. \tag{6.1.11}$$

Consider the following nonsmooth function for a vector $y \in R^k$

$$p(y) = \max_{1 \le i \le k} \{y_i\}.$$

We define a smoothing function of p as follows [47]: for $\mu > 0$,

$$\tilde{p}(y,\mu) = \mu \ln(\sum_{i=1}^{k} e^{y_i/\mu}).$$

Lemma 6.1.2 [19] *$\tilde{p}$ is continuously differentiable with respect to x for any fixed $\mu > 0$. Moreover, the following hold.*

(i)

$$0 \le \tilde{p}(y,\mu) - p(y) = \mu \ln\big(\sum_{i=1}^{k} e^{\frac{y_i - p(y)}{\mu}}\big) \le \mu \ln k.$$

(ii) $\{\lim_{z \to x,\ \mu \downarrow 0} \nabla_x \tilde{p}(z,\mu)\}$ is nonempty and bounded. Moreover, $\tilde{p}$ satisfies the gradient consistent property, that is,

$$\{\lim_{y \to \bar{y}, \mu \downarrow 0,} \nabla_y \tilde{p}(y,\mu)\} \subset \partial p(\bar{y}),$$

where ∂p denotes the Clarke generalized gradient.

Lemma 6.1.3 *The directional derivative $\tilde{p}'_\mu(y;h)$ of $\tilde{p}$ satisfies*

$$\lim_{\mu \downarrow 0} \tilde{p}'_\mu(y;h) \le p'(y;h), \quad \forall\ y, h \in R^k. \tag{6.1.12}$$

Proof. For any given $y, h \in R^k$, let $K = \{i \,|\, y_i = p(y)\,\}$ and $h_0 = \max_{i \in K} h_i$. The directional derivative $p'(y;h) = h_0$. For $\mu > 0$, $\tilde{p}$ is continuously differentiable and

$$\lim_{\mu \downarrow 0} \tilde{p}_\mu(y;h) = \lim_{\mu \downarrow 0} \nabla \tilde{p}_\mu(y)^T h = \sum_{i=1}^{k} h_i \sum_{j=1}^{k} \frac{1}{e^{(y_j - y_i)/\mu}} \le \frac{1}{|K|} \sum_{i \in K} h_i \le h_0 = p'(y;h).$$

This completes the proof. ■

Let

$$\tilde{f}(\xi, x, \mu) = u(\xi, x)^T F(\xi, u(\xi, x)) + \mu \sum_{i=1}^{m} (b_\xi)_i \ln \sum_{j \in I_i} e^{-F_j(\xi, u(\xi, x))/\mu}. \qquad (6.1.13)$$

Theorem 6.1.2 *When X_ξ is defined by (6.1.9) and $\tilde{f}$ is defined by (6.1.13), the assumptions of Theorem 6.1.1 hold and φ_μ and Φ_μ^N are smoothing functions of φ and $\hat{\varphi}^N$, respectively. Moreover, $Q(\xi, u(\xi, x))$ is regular in x for any fixed $\xi \in \Xi$ and $\tilde{f}$ satisfies (3.3.11).*

Proof. The matrix A can be split into two submatrices A_K and A_J, where $A_K = I \in R^{m\times m}$ whose i-th column is the first column of A_{I_i} and A_J is an $m\times(n-m)$ nonnegative matrix whose columns have only one positive element.

From Lemma 6.1.2, it is easy to verify that $\tilde{f}$ is a smoothing function of f defined in (6.1.11). By definitions, φ_μ and Φ_μ^N are smoothing functions of φ and $\hat{\varphi}^N$.

The regularity of $Q(\xi, u(\xi, x)) = \sum_{i=1}^m b_i(\xi) \max_{j\in I_i}\{-F_j(\xi, u(\xi, x))\}$ follows directly from the Chain Rule [21, Theorem 2.3.9] since $b_\xi > 0$, p is convex and F is continuously differentiable.

Next, we show (3.3.11) holds. Note that by the regularity of f, $df(\xi, x; h) = f'(\xi, x; h)$. Since the first term of f is continuously differentiable, we only need to consider the second term. Without loss of generality, we assume $I_1 = K = \{1, \cdots k\}$ and thus $z_1(\xi, u) = \max\{-F_j(\xi, u), j \in K\}$. For a fixed ξ, let $g(u) = (-F_1(\xi, u), \cdots, -F_k(\xi, u))^T$ and $q(u) = p(g(u)) = \max(g_1(u), \cdots, g_k(u))$. Since $b_i(\xi) > 0$, for $i = 1, \cdots, m$, it is sufficient to show that

$$\lim_{\mu\downarrow 0} \tilde{q}'_\mu(u; h) \le q'(u; h), \quad \forall\ u, h \in R^k. \qquad (6.1.14)$$

By continuously differentiability of g, the directional derivative of q satisfies

$$q'(u, h) = \lim_{t\downarrow 0} \frac{p(g(u + th)) - p(g(u))}{t}$$

$$= \lim_{t \downarrow 0} \frac{p(g(u) + tg'(u)h + o(t)) - p(g(u))}{t} = p(g(u); g'(u)h).$$

For $\mu > 0$,

$$\lim_{\mu \downarrow 0} \tilde{q}'_\mu(u; h) = \lim_{\mu \downarrow 0} \nabla \tilde{p}_\mu(g(u))^T g'(u)h \leq p(g(u); g'(u)h) = q(u; h)$$

that follows from Lemma 6.1.3. ■

6.2 Numerical experiments

In this section, X_ξ is defined by (6.1.9) and $\tilde{f}$ is defined by (6.1.13). The EV formulation for the examples is to find an $x \in X = \{x \,|\, Ax = E[b_\xi]\,\}$ such that

$$(y - x)^T E[F(\xi, x)] \geq 0, \qquad y \in X. \tag{6.2.15}$$

We solve the following minimization problem

$$\min_{x \in X} g(x) := \max\{(x - y)^T E[F(\xi, x)] \,|\, y \in X\} \tag{6.2.16}$$

and set a minimizer to be x_{EV}.

For the ERM formulation, we solve the ERM problem (2.1.1) and set $x_{\text{ERM}} = (I - A^\dagger A)x^* + A^\dagger E[b_\xi]$, where x^* is a solution of (2.1.1).

We use the residual function f and conditional value-at-risk(CVaR) to compare the two formulations; for fixed x,

$$\alpha^*(x) \in \operatorname*{argmin}_{\alpha \in R} \text{CVaR}(x, \alpha) := \alpha + \frac{1}{1 - \beta} E\{[f(\xi, x) - \alpha]_+\}.$$

6.2.1 SSAA methods

Example 6.2.1 *This example is the 7-link, 6-paths problem in Figure 6.1. The free travel time* t_k^0 *and the mean of the capacity* $E[c_{k(\xi)}]$ *of the network are the same as those used in* [65]*, which are listed in Table 6.1.*

Table 6.1: Link cost parameters in Figure 6.1

Link number	k	1	2	3	4	5	6	7
Free-flow time	T_k	6	4	3	5	6	4	1
Mean	C_k	15	15	30	30	15	15	15

For the travel demand vector, we set $E[b_\xi] = [200\ 220]^T$*, where the components follow the order of the OD-pairs* $1 \longrightarrow 4$ *and* $1 \longrightarrow 5$*. The link capacity and the demand vector both have a beta distribution. For the demand vector* b_ξ*, the lower bound is* $\underline{b} = [150\ 180]^T$ *and the parameters for the beta distribution are* $\alpha = 5$*,* $\beta = 1$*. For the link capacity* c_ξ*, the lower bound is* $\underline{c} = [10\ 10\ 20\ 20\ 10\ 10\ 10]^T$ *and the parameters for the beta distribution are* $\alpha = 2$*,* $\beta = 2$*.*

Results in Table 6.2 and Table 6.3 were obtained by using the same sampling with size $N = 1000$*. Table 6.2 gives EV and ERM solutions for different values of* n_a*. Table 6.3 lists robustness and risk criteria for the EV and ERM solutions in Table 6.2;* x_ξ^* *means solution of the variational inequalities for each fixed* $\xi \in \Xi$*.*

Table 6.2: Solutions for sampling size N=1000

	$n_a = 2$			$n_a = 4$		
	x_{EV}	x_{ERM}	x^*	x_{EV}	x_{ERM}	x^*
x_1	18.85	27.28	16.61	2.89	14.87	23.60
x_2	90.32	88.11	77.44	95.09	92.38	101.12
x_3	90.83	84.61	73.95	102.03	92.75	101.49
x_4	26.61	28.29	10.95	20.37	19.64	17.31
x_5	99.65	97.53	80.20	104.87	102.73	100.40
x_6	93.74	94.18	76.85	94.76	97.63	95.30

Table 6.3: Criteria for $x = x_{EV}, x = x_{ERM}, x = x^*$ with $N = 1000$, $\beta = 0.9$

	$n_a = 2$ $\varepsilon = 4.5E3$			$n_a = 4$ $\varepsilon = 5E5$		
	$x_{\rm EV}$	$x_{\rm ERM}$	x^*	$x_{\rm EV}$	$x_{\rm ERM}$	x^*
$\text{prob}\{f(\xi, x) \le \varepsilon\}$	0.58	0.61	0.61	0.56	0.59	0.59
$E[\lVert x - x_\xi^* \rVert]$	46.94	39.63	54.00	35.28	33.01	36.65
$E[\lVert u(\xi, x) - x_\xi^* \rVert]$	47.03	39.72	39.72	35.41	33.15	33.15
$E[f(\xi, x)]$	4.316E3	4.198E3	4.198E3	5.064E5	4.907E5	4.907E5
$\alpha^*(x)$	7.395E3	7.132E3	7.132E3	1.071E6	1.037E6	1.037E6
$\text{CVaR}(x, \alpha^*(x))$	8.691E3	8.515E3	8.515E3	1.254E6	1.229E6	1.229E6

In Figure 6.2, we graph $\text{prob}\{f(\xi, x) \le \varepsilon\}$ with different values of ε. We can see the ERM formulation has higher probability than the EV formulation for each ε.

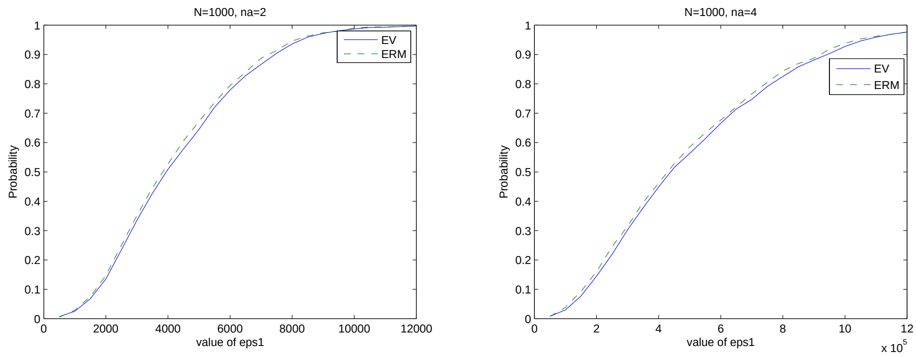

Figure 6.2: $\text{prob}\{f(\xi, x) \le \varepsilon\}$ with different values of ε for $x_{\rm EV}$ and $x_{\rm ERM}$.

Example 6.2.2 *This example uses the Nguyen and Dupuis network given in Figure 6.3, which contains 13 nodes, 19 links, 25 paths and 4 OD movements.*

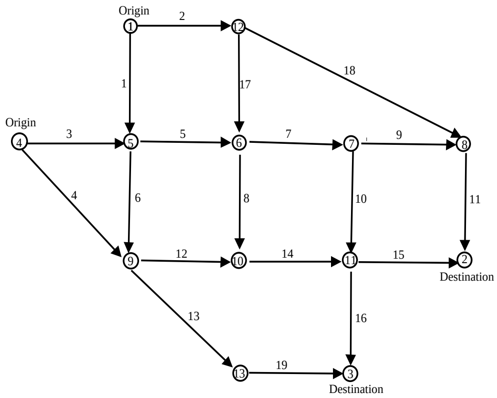

Figure 6.3: Nguyen and Dupuis Network

We use the free-flow travel time t_a^0, the mean $E[c_\xi]$, the coefficient of variation $CV[(c_\xi)_a]$ and the free-flow travel time t_a^0 are given in Table 6.4.

Table 6.4: Link cost parameters

Link	Free-flow travel time t_a^0	Link capacity, c_a Mean	Link capacity, c_a Coefficient of Variation
1	7.0	800	4
2	9.0	400	2
3	9.0	200	2
4	12.0	800	4
5	3.0	350	2
6	9.0	400	2
7	5.0	800	2
8	13.0	250	2
9	5.0	250	2
10	9.0	300	2
11	9.0	550	4
12	10.0	550	4
13	9.0	600	2
14	6.0	700	4
15	9.0	500	4
16	8.0	300	2
17	7.0	200	2
18	14.0	400	2
19	11.0	600	2

The mean of the demand vector $E[b_\xi]$ of the network is $E[b_\xi] = [400, 800, 600, 450]^T$. The link capacity has three possible scenarios which denotes different conditions of the network such as weather, accidents and so on, and we give the three scenarios with probability $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$ and $p_3 = \frac{1}{4}$.

$c_a^1 = 100 \times [8,\ 3.2,\ 3.2,\ 8,\ 4,\ 3.2,\ 8,\ 2,\ 2,\ 2,\ 4,\ 4,\ 8,\ 6,\ 4,\ 4,\ 1.6,\ 3.2,\ 8]^T;$

$c_a^2 = 100 \times [10,\ 4.4,\ 1.4,\ 10,\ 3,\ 4.4,\ 10,\ 2,\ 2,\ 4,\ 7,\ 7,\ 7,\ 7,\ 4,\ 3.5,\ 2.2,\ 4.4,\ 7]^T;$

$c_a^3 = 100 \times [4,\ 4,\ 2,\ 4,\ 4,\ 4,\ 4,\ 4,\ 4,\ 2,\ 4,\ 4,\ 2,\ 8,\ 8,\ 1,\ 2,\ 4,\ 2]^T.$

The demand vector follows the beta distribution $b_{\xi^i} \sim \underline{b} + \hat{b} * \mathrm{beta}(\alpha, \beta)$ *with the lower bound* $\underline{b} = [300, 700, 500, 350]^T$ *and parameters* $\alpha = 50$, $\beta = 10$ *and* $\hat{b} = [120, 120, 120, 120]^T$. *We rely on the Monte-Carlo method to randomly generate* N *samples of* $(b_{\xi^i}, (c_{\xi^i})_a)$ *for* $i = 1, 2, \cdots, N$, *where* $(c_{\xi^i})_a$ *is sampled from the three possibilities with given probability and* b_{ξ^i} *is sampled from the beta distribution.*

We list the probability $prob\{f(\xi, x) \leq 3.3 * 10^3\}$, *the total residual* $E[f(\xi, x)]$ *and the risk criteria in Table 6.5. Form the table, we can find that solutions of ERM solution has better properties than that of EV formulation.*

Table 6.5: Criteria for $\beta = 0.9$, $n_a = 2$, $\varepsilon = 3.3E3$

		x_{EV}	x_{ERM}
$N = 10^3$ $\mu = 10^{-4}$	$\mathrm{prob}\{f(\xi, x) \leq \varepsilon\}$	0.508	0.952
	$E[f(\xi, x)]$	3.498E3	2.938E3
	α^*	7.935E3	3.226E3
	$\mathrm{CVaR}(x, \alpha^*)$	8.154E3	3.333E3
$N = 5 * 10^3$ $\mu = 10^{-5}$	$\mathrm{prob}\{f(\xi, x) \leq \varepsilon\}$	0.510	0.908
	$E[f(\xi, x)]$	3.498E3	2.983E3
	α^*	7.918E3	3.286E3
	$\mathrm{CVaR}(x, \alpha^*)$	8.121E3	3.403E3
$N = 10^4$ $\mu = 10^{-6}$	$\mathrm{prob}\{f(\xi, x) \leq \varepsilon\}$	0.509	0.927
	$E[f(\xi, x)]$	3.505E3	2.976E3
	α^*	7.978E3	3.253E3
	$\mathrm{CVaR}(x, \alpha^*)$	8.168E3	3.359E3

Example 6.2.3 *We consider the Sioux Falls network as shown in Figure 6.4 (left), which consists of 24 nodes, 76 links, 528 OD movements. The total of 1179 paths are pre-generated as possible travel routes between different OD pairs. The parameters of the GBPR function are the same as that in* [69] *except* n_a=4. *We consider the stochastic settings for the OD demands and the capacity of the links. Each* $(b_\xi)_i$ *is supposed to follow a log-norm distribution, and the coefficients of variation for each* $(b_\xi)_i$ *are 5. For the capacity, we use the beta distribution to generate the samples. The link flow patterns*

obtained by the ERM (2.1.1) are displayed in Figure 6.4 (right). Here the link flow is displayed on each link with the unit $1.0 * 10^3$*, and the width of each link is proportional to the link flow. By the property of* x_{ERM}*, we know that the ERM flow patterns satisfy the average of travel demand as* $Ax_{ERM} = E[b_\xi]$*. Moreover, the ERM flow patterns satisfy the stochastic travel demand on all OD pairs with high probability:*

$$0.848 \geq \text{prob}\{(Ax_{ERM} - b_\xi)_i \geq 0\} \geq 0.780, \quad i = 1, \cdots, 528.$$

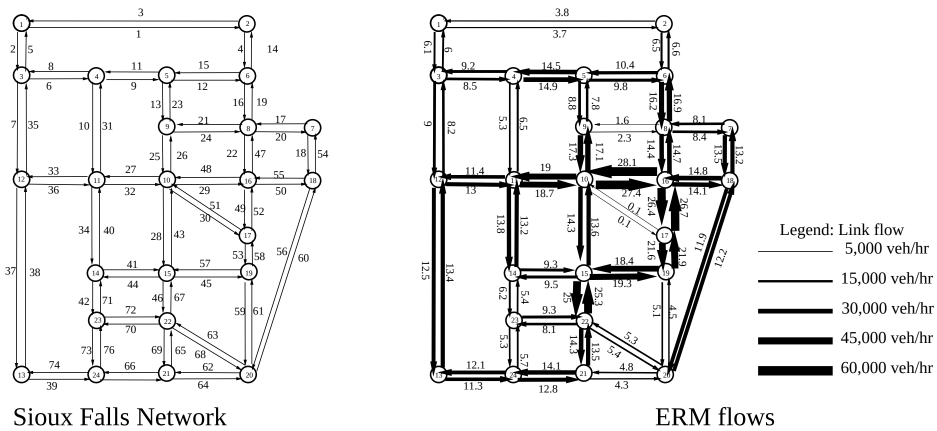

Figure 6.4: Sioux Falls Network

6.2.2 Moreau-Yosida regularization for stochastic linear VI

To demonstrate the properties of the ERM formulation, we employ the proximal quasi-Newton method [15] to get a solution x_{ERM} of the ERM formulation (4.1.3) for the Moreau-Yosida regularization.

The Nguyen and Dupuis network, which contains 13 nodes, 19 directed links, and 4 OD movements $1 \longrightarrow 2$, $1 \longrightarrow 3$, $4 \longrightarrow 2$, and $4 \longrightarrow 3$, is used in our numerical experiments. See Figure 6.3.

We use the free-flow travel time t_a^0 given in [66]. The demand vector b_ξ and the link capacity c_ξ both follow log-norm distributions. The mean $E[c_\xi]$, the coefficient of variation $CV[(c_\xi)_a]$ and the free-flow travel time t_a^0 are the same as that in [66], and we list them in Table 6.4.

The mean of the demand vector b_ξ is given as $E[b_\xi] = [400, 800, 600, 450]^T$. We set $b_{\xi,1} = \frac{1}{2}c_1(\xi)$, $b_{\xi,2} = c_1(\xi)$, $b_{\xi,3} = \frac{3}{4}c_1(\xi)$ and $b_{\xi,4} = \frac{9}{16}c_1(\xi)$. From this, we know that the stochastic vectors b_ξ and c_ξ are dependent which means the co-variance matrix of b_ξ and c_ξ are not 0. We employ the sample size $N = 10,000$ and the Tikhonov regularization parameter $\varepsilon = 0.01$ in our numerical experiments. We show the link flow patterns of the ERM formulation and the EV formulation in Figure 6.5, and we can see the difference between the two patterns.

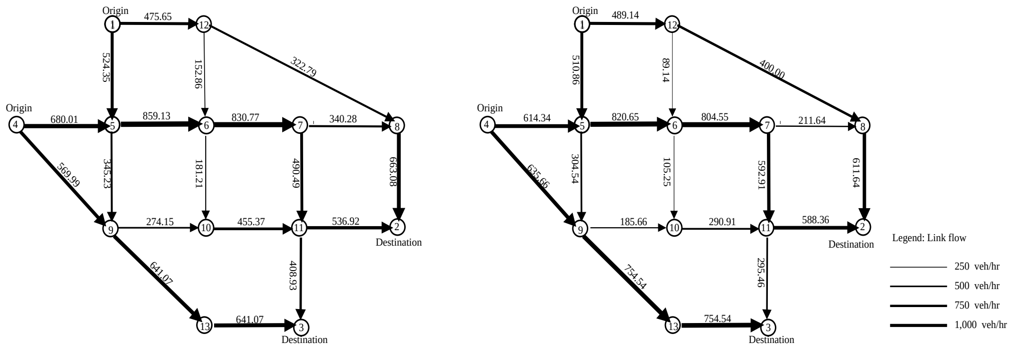

Figure 6.5: Link flows for the ERM formulation (left) and the EV formulation (right)

In Figure 6.6, we graph prob$\{f(\xi, x) \le \delta\}$ with different values of δ. We can see the ERM formulation has higher probability than the EV formulation for each δ.

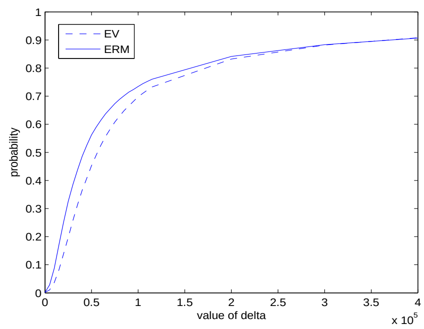

Figure 6.6: prob$\{f(\xi, x) \le \delta\}$ with different values of δ for x_{EV} and x_{ERM}.

In Table 6.6, we list the values $E[f(\xi, x)]$, $E[||x - x_\xi||]$ and $E[||V - V_\xi||]$, which denote the distance of a traffic assignment pattern under uncertainty to the Wardrop's user equilibrium for each scenario. From Table 6.6, we can see that the expected distance between x_{ERM} and Wardrop's user equilibrium assignment for all possible realizations is smaller than that of x_{EV}.

Table 6.6: Robust criteria of various traffic assignment patterns

Various criteria	x_{ERM}	x_{EV}
$E[f(\xi, x)]$	5.855E+05	5.887E+05
$E[\|\|x - x_\xi\|\|]$	1.197E+03	1.248E+03
$E[\|\|V - V_\xi\|\|]$	2.672E+03	2.697E+03

Remark 6.2.1 *The three examples are often used in transportation research. They satisfy all our assumptions of the theoretical analysis for the ERM formulation in the above chapters. Moreover, our preliminary numerical results show that the ERM solution performs better than the EV solution both as far as robustness and risk analysis are concerned.*

Bibliography

[1] R. P. Agdeppa, N. Yamashita, and M. Fukushima, *The traffic equilibrium problem with nonadditive costs and its monotone mixed complementarity problem formulation*, Transp. Res. B, 41 (2007), pp. 862–874.

[2] ——, *Convex expected residual models for stochastic affine variational inequality problems and its application to the traffic equilibrium problem*, Pac. J. Optim., 6 (2010), pp. 3–19.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, *Coherent measures of risk*, Math. Financ., 9 (1999), pp. 203–228.

[4] A. Atamtürk, *Strong formulations of robust mixed 0-1 programming*, Math. Program., 108 (2006), pp. 235–250.

[5] H. Attouch and R. J.-B. Wets, *Epigraphical processes: laws of large numbers for random lsc functions*, sém. Anal. Convexe, 13 (1991), pp. 13.1–13.29.

[6] A. Bel-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769–805.

[7] ——, *Robust solutions to uncertain programs*, Oper. Res. Lett., 25 (1999), pp. 1–13.

[8] ——, *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Program., 88 (2000), pp. 411–424.

[9] D. Bertsimas and M. Sim, *Robust discrete optimization and network flows*, Math. Program., 98 (2003), pp. 49–71.

[10] ——, *Price of robustness*, Oper. Res., 52 (2004), pp. 35–53.

[11] ——, *Tractable approximations to robust conic optimization problems*, Math. Program., 107 (2006), pp. 5–36.

[12] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*, Springer, New York, 1997.

[13] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer-Burmeister NCP function*, Math. Program., 88 (2000), pp. 211–216.

[14] F. Y. Chen, H. Yan, and L. Yao, *A newsvendor pricing game*, IEEE Trans. Syst., Man, Cybern.: Part A, 34 (2004), pp. 450–456.

[15] X. Chen and M. Fukushima, *Proximal quasi-Newton methods for nondifferentiable convex optimization*, Math. Program., 85 (1999), pp. 313–334.

[16] ——, *Expected residual minimization method for stochastic linear complementarity problems*, Math. Oper. Res., 30 (2005), pp. 1022–1038.

[17] X. Chen and G. Lin, *CVaR-based formulation and approximation method for stochastic variational inequalities*, Numer. Algebra. Contr. Optim., 1 (2011), pp. 35–48.

[18] X. Chen, M. Sim, and P. Sun, *A robust optimization perspective on stochastic programming*, Oper. Res., 55 (2007), pp. 1058–1071.

[19] X. Chen, R. S. Womersley, and J. Ye, *Minimizing the condition number of a Gram matrix*, SIAM J. Optim., 21 (2011), pp. 127–148.

[20] X. Chen, C. Zhang, and M. Fukushima, *Robust solution of monotone stochastic linear complementarity problems*, Math. Program., 117 (2009), pp. 51–80.

[21] F. H. Clarke, *Optimization and nonsmooth analysis*, John Wiley, New York, 1983.

[22] R. W. Cottle, J. S. Pang, and R. E. Stone, *The linear complementarity problem*, Academic Press, New York, 1992.

[23] E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res., 58 (2010), pp. 595–612.

[24] L. El-Ghaoui and H. Lebret, *Robust solutions to least-square problems to uncertain data matrices*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 1035–1064.

[25] L. El-Ghaoui, F. Oustry, and H. Lebret, *Robust solutions to uncertain semidefinite programs*, SIAM J. Optim., 9 (1998), pp. 33–52.

[26] F. Facchinei and J. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.

[27] H. Fang, X. Chen, and M. Fukushima, *Stochastic R_0 matrix linear complementarity problems*, SIAM J. Optim., 18 (2007), pp. 482–506.

[28] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

[29] M. Fukushima, *Merit functions for variational inequality and complementarity problems*, in Nonlinear optimization and Applications, G. Pillo and F. Giannessi, eds., Plenum Press, New York, 1996, pp. 155–170.

[30] G. GÜrkan, A. Y. Özge, and S. M. Robinson, *Sample-path solution of stochastic variational inequalities with applications to option pricing*, in Proceedings of the Winter Simulation Conference, J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, eds., IEEE Computer Society, 1996, pp. 337–344.

[31] ——, *Sample-path solution of stochastic variational inequalities*, Math. Program., 84 (1999), pp. 313–333.

[32] D. Goldfarb and G. Iyengar, *Robust convex quadratically constrained programs*, Math. Program. Ser. B, 97 (2003), pp. 495–515.

[33] J. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin Heidelberg, 1993.

[34] P. Huber, *Robust Statistics*, John Wiley and Sons, 1981.

[35] H. Jiang and H. Xu, *Stochastic approximation approaches to the stochastic variational inequality problem*, IEEE. Trans. Autom. Control., 53 (2008), pp. 1462–1475.

[36] P. Kall and S. Wallace, *Stochastic Programming*, John Wiley and Sons, Chichester, 1994.

[37] L. A. Korf and R. J.-B. Wets, *Random lsc functions: an ergodic theorem*, Math. Oper. Res., 26 (2001), pp. 421–445.

[38] Z. Li, W. H. K. Lam, and A. Sumalee, *Modeling impact of transit operator fleet size under various market regimes with uncertainty in network*, Transp. Res. Rec., 2063 (2008), pp. 18–27.

[39] C. Ling, L. Qi, G. Zhou, and L. Caccetta, *The SC^1 property of an expected residual function arising from stochastic complementarity problems*, Oper. Res. Lett., 36 (2008), pp. 456–460.

[40] M. J. Luo and G. H. Lin, *Expected residual minimization method for stochastic variational inequality problems*, J. Optim. Theory Appl., 140 (2009), pp. 103–116.

[41] O. L. Mangasarian and T. H. Shiau, *Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems*, SIAM J. Control Optim., 25 (1987), pp. 583–595.

[42] P. Marcotte, *Adavantages and drawbacks of variational inequalities formulations*, in Variational Inequalities and Network Equilibrium Problems, F. Giannessi and A. Maugeri, eds., Plenum Press, 1995, pp. 179–194.

[43] F. Meng, G. Z. M. Goh, and R. D. Souza, *Lagrangian-dual functions and Moreau-Yosida regularization*, SIAM J. Optim., 19 (2008), pp. 39–61.

[44] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 957–972.

[45] R. Mifflin, L. Qi, and D. Sun, *Properties of the Morear-Yosida regularization of a piecewise C^2 convex function*, Math. Program., 84 (1999), pp. 269–281.

[46] K. Natarajan, D. Pachamanova, and M. Sim, *Incorporating asymmetric distributional information in robust value-at-risk optimization*, Manage. Sci., 54 (2008), pp. 573–585.

[47] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.

[48] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 1992.

[49] J. Peng, *Equivalence of variational inequality problems to unconstrained optimization*, Math. Program., 78 (1997), pp. 347–355.

[50] S. Peng, *G-expectation, G-Brownian motion and related stochastic calculus of Itô type*, Stoch. Anal. Appl., 2 (2007), pp. 541–567.

[51] ——, *Nonlinear expectations and stochastic calculus under uncertainty*, Preprint, in arXiv:1002.4546v1, 2010.

[52] L. Qi, A. Shapiro, and C. Ling, *Differentiability and semismoothness properties of integrable functions and their applications*, Math. Program., 102 (2005), pp. 223–248.

[53] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[54] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk, 2 (2000), pp. 493–517.

[55] ——, *Conditional value-at-risk for general loss distributions*, J. Bank. Financ., 26 (2002), pp. 1443–1471.

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

[57] A. Ruszczyński and A. Shapiro, *Stochastic Program., Handbooks in Operations Research and Management Science*, Elsevier, 2003.

[58] H. Scarf, *A min-max solution of an inventory problem*, in Studies in the Mathematical Theory of Inventory and Production, K. Arrow, S. Karlin, and H. Scarf, eds., Stanford University Press, 1958, pp. 201–209.

[59] A. Shapiro and T. Homem-De-Mell, *On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, SIAM J. Optim., 11 (2000), pp. 70–86.

[60] I. H. Sloan and H. Wozniakowskic, *When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?*, J. Complex., 14 (1998), pp. 1–33.

[61] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, 1991.

[62] R. J. B. Wets, *Stochastic programming*, in Handbooks in Operations Research and Management Sciences, G. Nemhauser, A. R. Kan, and M. Todd, eds., vol. 1, Elsevier, 1989, pp. 573–629.

[63] H. Xu, *Sample average approximation methods for a class of stochastic variational inequality problems*, Asia Pac. J. Oper. Res., 27 (2010), pp. 103–119.

[64] H. Xu and D. Zhang, *Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications*, Math. Program., 119 (2009), pp. 371–401.

[65] H. Yang, *Sensitivity analysis for the elastic-demand network equilibrium problem with applications*, Transp. Res. B, 31 (1997), pp. 55–70.

[66] Y. F. Yin, S. M. Madanat, and X. Y. Lu, *Robust inprovement schemes for road networks under demand uncertainty*, Eur. J. Oper. Res., 198 (2009), pp. 470–479.

[67] C. Zhang and X. Chen, *Stochastic nonlinear complementarity problem and applications to traffic equilibrium under uncertainty*, J. Optim. Theory Appl., 137 (2008), pp. 277–295.

[68] ——, *Smoothing projected gradient method and its application to stochastic linear complementarity problems*, SIAM J. Optim., 20 (2009), pp. 627–649.

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

[70] J. Zhou, W. H. K. Lam, and B. G. Heydecker, *The generalized Nash equilibrium model for oligopolistic transit market with elastic demand*, Trans. Res. B, 39 (2005), pp. 519–544.