STOCHASTIC VARIATIONAL INEQUALITIES: RESIDUAL MINIMIZATION SMOOTHING/SAMPLE AVERAGE APPROXIMATIONS

XIAOJUN CHEN*, ROGER J-B WETS†, AND YANFANG ZHANG‡

20 February 2011

Abstract. The stochastic variational inequality (SVI) has been used widely, in engineering and economics, as an effective mathematical model for a number of equilibrium problems involving uncertain data. This paper presents a new expected residual minimization (ERM) formulation for a class of SVI. The objective of the ERM-formulation is Lipschitz continuous and semismooth which helps us guarantee the existence of a solution and convergence of approximation methods. We propose, a globally convergent (a.s.) smoothing sample average approximation (SSAA) method to minimize the residual function; this minimization problem is convex for linear SVI if the expected matrix is positive semi-definite. We show that the ERM problem and its SSAA problems have minimizers in a compact set and any cluster point of minimizers and stationary points of the SSAA problems is a minimizer and a stationary point of the ERM problem (a.s.). Our examples come from applications involving traffic flow problems. We show that the conditions we impose are satisfied and that the solutions, efficiently generated by the SSAA-procedure, have desirable properties.

Key words. Stochastic variational inequalities, epi-convergence, lower/upper semicontinuous, semismooth, smoothing sample average approximation, expected residual minimization, stationary point.

AMS subject classifications. 90C33, 90C15.

1. Introduction. In a deterministic environment, one refers to the problem of finding $x \in X$ that satisfies the inclusion $-F(x) \in N_X(x)$ as a *variational inequality*, 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. A good formulation of a variational inequality, 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

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

$$\text{or still} \quad E[-F(\xi,x)] \in N_X(x), \tag{1.1}$$

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.(maxjchen@polyu.edu.hk). The author's work was supported in part by Hong Kong Research Grant Council grant PolyU5003/09p.

†Department of Mathematics, University of California, Davis, CA 95616.(rjbwets@ucdavis.edu). Partially suported by ARO-grant W911NS1010246.

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.(09900332r@polyu.edu.hk).

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 article. The first formulation with $\alpha = 1$ could be converted to a large variational inequality, 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 $E[N_{X_\xi}(x) + F(\xi, x)]$.

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* [10, Section 1.5.3] 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 paper for the particular class of variational inequalities described below.

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

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

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 nonlinear complementarity problem:

$$x \geq 0, \quad F(\xi, x) \geq 0, \quad x^T F(\xi, x) = 0.$$

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. Using mean values or some other estimates for the uncertain parameters in the model may lead to seriously misleading decisions.

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

- *Expected Value* EV-formulation [12, 13, 25, 29]: find $x \in X$ such that

$$(y - x)^T E[F(\xi, x)] \geq 0, \qquad y \in X. \tag{1.2}$$

- *Expected Residual Minimization* ERM-formulation [1, 5, 7, 11, 15, 16, 33, 34]:

$$\min_{x \in X} E[f(\xi, x)], \tag{1.3}$$

$f(\xi, \cdot) : X \to R_+$ is a *residual function* for the VI$(X, F(\xi, \cdot))$ for fixed $\xi \in \Xi$ [10, Section 6.1].

As already pointed out earlier, the EV-formulation can be viewed as a deterministic VI$(X, \bar{F})$ with the expectational function $\bar{F}(x) = E[F(\xi, x)]$. The ERM-formulation minimizes the expected values of the 'loss' for all possible occurences 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 [7].

To allow for the dependence of the set X on $\xi \in \Xi$, one needs to extend the definition of the residual function.

DEFINITION 1.1. *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) $\exists\ u : D \times \Xi \to R_+$ 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$(X_\xi, F(\xi, \cdot))$.

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 canonical projection of R^n onto X_ξ and $\| \cdot \|$ is the Euclidean norm. When $X_\xi = R^n_+$, one has

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

The ERM-formulation with this 'natural' residual function would be a nonsmooth, nonconvex minimization problem.

In this paper, we rely on the gap function [10, Section 1.5] to define a *new residual function*

$$f(\xi, x) = u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, u(\xi, x)) \tag{1.4}$$

where

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

and

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

$A^\dagger = A^T(AA^T)^{-1}$ is the Moore-Penrose generalized inverse of A.

In Section 2, we show that f is a residual function for the stochastic VI. Moreover, in the affine case where $F(\xi, x) = M_\xi x + q_\xi$, we show that $E[f(\xi, x)]$ is convex if the expectation matrix $E[M_\xi]$ is a positive semi-definite matrix, that is,

$$x^T E[M_\xi] x \geq 0, \qquad \forall\ x \in R^n. \tag{1.5}$$

Luo and Lin [16] dealt with an ERM-formulation for the stochastic VI, with X deterministic, by using the regularized gap function as a residual function. Agdeppa, Yamashita and Fukushima [1] showed that the ERM-formulation using the regularized gap function is convex when $F(\xi, x) = M_\xi x + q_\xi$ and

$$\inf_{\xi \in \Xi, \|x\|=1} x^T M_\xi x \geq \beta_0 \tag{1.6}$$

for some positive constant β_0.

Obviously, in the affine case, (1.6) implies (1.5). However, the converse is not true. It is worth noting that (1.5) does not imply that the probability

$$\text{prob}\{ M_\xi \text{ positive semidefinite}\} > 0.$$

Example 1.1 in [7] exhibits a stochastic matrix M_ξ that satisfies condition (1.5), but there is no $\xi \in \Xi$ for which M_ξ is positive semidefinite. Hence, condition (1.5) is much weaker than (1.6). Moreover, the new residual function (1.4) can be used when X_ξ is a random set.

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

$$\min_{x \in D} \varphi(x) = E[f(\xi, x)], \tag{1.7}$$

defined by the new residual function (1.4), has various desirable properties and to prove the convergence of smoothing sample average approximation SSAA-methods to solve (1.7) by relying on an epi-convergence argument and the properties of inf-projections [23]. Moreover, we provide efficient methods to solve a class of stochastic variational inequalities with applications to traffic flow problems. In particular, we give explicit forms of $Q(\xi, u(\xi, x))$ and smoothing approximations of $f(\xi, x)$.

In Section 2, we show that the function f is a residual function for the stochastic VI and the objective function φ is Lipschitz continuous and semismooth. Moreover, we prove the existence of solutions of (1.7). For the case where $F(\xi, x) = M_\xi x + q_\xi$, we show that φ is convex if $E[M_\xi]$ is positive semi-definite.

In Section 3, we define the SSAA-function and prove the existence of solution 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 (1.7) (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 (1.7) (a.s.).

In Section 4, we use examples coming from traffic equilibrium assignment to illustrate the ERM-formulation (1.7) 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 Sections 2 and 3 are satisfied. Moreover, we present numerical results to compare the solution of (1.7) with 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.

Throughout the paper, $\|\cdot\|$ represents the Euclidean norm, $R^n_+ = \{x \in R^n \mid x \geq 0\}$, e denotes the vector whose elements are all 1, I denotes the identity matrix. For a given matrix $A = [a_{ij}] \in R^{m \times n}$, let $A_K \in R^{m \times |K|}$ be the submatrix of A with column-index in the index set $K \subseteq \{1, \ldots, n\}$ of cardinality $|K|$. Let proj_C denote the orthogonal projection from R^n onto C, that is, $\text{proj}_C(x) = \text{argmin}_{y \in C} \|y - x\|$.

2. A new residual function. 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 [10, Section 1.5.3]

$$\min_{x\in X_\xi} g(\xi,x). \tag{2.1}$$

This minimization problem (2.1) can be written as a two stage optimization problem

$$\begin{array}{ll} \min & x^T F(\xi,x) + Q(\xi,x) \\ \text{s.t.} & x \in X_\xi \\ & Q(\xi,x) = \max\{\, -y^T F(\xi,x) \mid y \in X_\xi\}; \end{array} \tag{2.2}$$

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

$$Q(\xi,x) = \min\{\, z^T b_\xi \mid A^T z + F(\xi,x) \geq 0\}. \tag{2.3}$$

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

$$D = \{\, x \,|\, (A^\dagger A - I)x \leq \underline{c}\} \text{ where for } i = 1,\ldots,m,\ \ \underline{c}_i \leq \min_{\xi\in\Xi}(A^\dagger b_\xi)_i.$$

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 $\xi \in \Xi$,

$$u(\xi,x) = \text{proj}_{X_\xi}(x). \tag{2.4}$$

ASSUMPTION 1. *Assume that for all $x \in D$ and for all $\xi \in_{\text{a.s.}} \Xi$,*

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

Rather than assuming that the second stage program is feasible for all $u \in X_\xi$, Assumption 1 only requires that it is feasible for a much more restricted class, namely, those $u = \text{proj}_{X_\xi}(x)$ when $x \in D$. In Section 4, we show that Assumption 1 holds for a class of matrices A and vectors b_ξ that arise from traffic equilibrium problems.

THEOREM 2.1. *When Assumption 1 is satisfied, $f : D \to R$, 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 1, for almost every $\xi \in \Xi$, there is $y(x, \xi) \in R^n$ such that

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

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

It is this residual function f that gets used in our ERM-formulation (1.7) with the objective function:

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

By Theorem 2.1, $\varphi(x) \geq 0$ for all $x \in D$ and if $\varphi(x) = 0$ then, $u(\xi, x)$ solves the $\text{VI}(X_\xi, F(\cdot, \xi))$ 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 (1.7). By definition of $u(\xi, x)$,

$$Ax_{\text{ERM}} = E[b_\xi] \quad \text{and} \quad x_{\text{ERM}} \geq 0. \tag{2.5}$$

Moreover, the following proposition shows that x_{ERM} is also a solution of our ERM-formulation (1.7).

PROPOSITION 2.2. *Under Assumption 1, if (1.7) has a solution x^*, then*

$$x_{ERM} \in \text{argmin}_{x \in D}\, \varphi(x). \tag{2.6}$$

Proof. For $x \in D$, let $\bar{u} = E[u(\xi, x)]$. Then, from (2.4)

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

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

$$\text{proj}_{X_\xi}(x) = \text{proj}_{X_\xi}(E[\text{proj}_{X_\xi}(x)]). \tag{2.7}$$

From (2.7), for every $\xi \in \Xi$,

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

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

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

which in turn yields (2.6). □

It is interesting to note that $x_{\text{ERM}} = x^*$ if and only if $A^\dagger(E[b_\xi] - Ax^*) = 0$. From (2.6), if the ERM-formulation (1.7) has a solution and $A^\dagger(E[b_\xi] - Ax^*) \neq 0$, then (1.7) has a multiplicity of solutions.

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

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

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

Assumption 2.

(i) the range of b_ξ is bounded: $E[\|b_\xi\|] < \infty$,
(ii) $\exists\, d(\xi)$ such that $\|F(\xi,u)\| \leq d(\xi)$ for all $u \in U$ and $E[d(\xi)] < \infty$,
(iii) $\exists\, d_1(\xi)$ such that $\|F'(\xi,u)\| \leq d_1(\xi)$ for all $u \in U$ and $E[d_1(\xi)] < \infty$,
(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(i)-(iii) are pretty standard and are not really restrictive as far as applications are concerned. Assumption 2(iv) is not quite as general but, in particular, is satisfied by the class of problems considered in Section 4.

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 $\xi \in \Xi$, it immediately follows,

Proposition 2.3. *$F(\xi, u(\xi,x))$ 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 differential with respect to x.
(ii) If (ii) and (iii) of Assumption 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 2.4. *Assume that Assumption 1 holds. Then, the function f is measurable in ξ for any $x \in D$ and locally Lipschitz continuous in x a.s. Moreover, under Assumption 2 (i)-(ii) the following hold.*

(i) If $F_i(\xi,u),\ i = 1,\dots,n$ are 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. Since $u(\xi,x)$ is linear in x, by Proposition 2.3, 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 [17], there is a constant $\nu_A > 0$, that only depends on the matrix A, such that

$$\|Q(\xi,u) - Q(\xi,v)\| \leq \|b_\xi\|\|z(\xi,u) - z(\xi,v)\| \leq \|b_\xi\| m\nu_A \|F(\xi,u) - F(\xi,v)\| \ a.s. \tag{2.8}$$

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. Moreover, it is easy to see that the two terms in

$f(\xi,\cdot)$ are locally Lipschitz in x for any fixed $\xi \in \Xi$. Hence $f(\xi,\cdot)$ is 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. [25, Theorem 19, Chapter 1]. Hence the function $f(\xi,x)$ is measurable in ξ for any $x \in R^n$.

Now we prove the second part of this theorem. (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, 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. $\square$

THEOREM 2.5. *Under Assumptions 1 and 2, φ is globally Lipschitz on D, i.e.,*

$$|\varphi(x) - \varphi(y)| \leq \kappa\|x - y\|, \quad x, y \in D \tag{2.9}$$

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 (2.8).

Proof. For the first term in φ, we have

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

For the second term, from (2.8), we have

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

Combining these two inequalities,

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

completes the proof. □

Definition 2.6. [18] *Suppose that $\phi: X\subseteq R^m\to R$ is a locally Lipschitz continuous function, then ϕ is* semismooth *at $x\in \text{int } X$ if ϕ is directionally differentiable at x and for any $g\in\partial\phi(x+h)$,*

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

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

Theorem 2.7. *Assume that Assumptions 1 and 2 hold. Then the function φ is semismooth on D.*

Proof. Following Proposition 1 and (3.1)-(3.2) in [20], 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 2.5,

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

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 [4, 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, 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 [21, Lemma 2.2] and the formula (2.5), 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). □

THEOREM 2.8. *Suppose Assumptions 1 and 2(i-ii, iv) hold. Then, (1.7) has a solution in the compact set*

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

Moreover,

$$D_1 \subseteq D \quad \text{and} \quad \operatorname{argmin}_{y\in D_1} \varphi(y) \subseteq \operatorname{argmin}_{x\in D} \varphi(x). \tag{2.10}$$

Proof. From Theorem 2.1 and Theorem 2.4 follows $0 \le \varphi(x) < \infty$ for any $x \in D$. 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

$$\min_{x\in D} \varphi(x) = \min_{y\in D_1} \varphi(y). \tag{2.11}$$

Since D_1 is compact and φ is continuous, $\operatorname{argmin}_{D_1} \varphi \ne \emptyset$ and any $y^* \in \operatorname{argmin}_{D_1} \varphi$ also minimizes φ on D since $D_1 \subset D$. Finally, from (2.11) one obtains (2.10). □

3. Smoothing sample average approximation (SSAA). Let $\xi^1, \dots, \xi^N$ be a sampling of ξ. The Sample Average Approximation (SAA) method has been used to find a solution of the EV-formulation (1.2) over a deterministic feasible set X [12, 13, 29]. 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

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

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 (1.3) 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)

$$\Phi_\mu^N(x) = \frac{1}{N}\sum_{i=1}^{N} \tilde{f}(\xi^i, x, \mu), \tag{3.1}$$

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

Definition 3.1. *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 $x \in R^n$,*

$$\lim_{z\to x,\ \mu\downarrow 0} \tilde{g}(z, \mu) = g(x). \tag{3.2}$$

We consider the existence and the convergence of solutions of the following SAA problems

$$\min_{x\in D} \hat{\varphi}^N(x) \tag{3.3}$$

and SSAA problems

$$\min_{x\in D} \Phi_\mu^N(x). \tag{3.4}$$

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

Definition 3.2. [23] *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 3.3. [14] *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 3.4. [14] *A sequence of random lsc functions $\{g^N : X \times \Xi \to \overline{R},\ N \in \mathbb{N}\}$ 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 $g : X \to \bar{R}$.*

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.

Lemma 3.5. *If the sample is iid then for any fixed $\mu \in [0, \bar{\mu}]$, we have*

$$\Phi_\mu^N \quad \xrightarrow{e} \quad \varphi_\mu, \quad \text{in}\, D, \quad \text{a.s.}. \tag{3.5}$$

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

$$c_{x,r} = \inf_{B(x,r)} \varphi + \delta_D, \qquad c^N_{x,r} = \inf_{B(x,r)} \Phi^N_\mu + \delta_D.$$

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 $\{c^N_{x,r}\}$ are iid [14]. From the Law of Large Number follows

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

Since $\Phi^N_\mu + \delta_D$ and $\varphi_\mu + \delta_D$ are random lsc functions, both functions can be completely identified by a countable collection of their inf-projections [14, 23, Chapter 14]. Hence we obtain (3.5). □

LEMMA 3.6. *Under assumptions of Theorem 2.8, for any $\mu \in [0, \bar{\mu}]$ and $N \in \mathbb{N}$, the SAA minimization problem (3.3) and the SSAA minimization problem (3.4) admit solutions.*

Proof. Since for any $\xi \in \Xi$, $f(\cdot, \xi)$ 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^N_μ are continuous functions on D for any $\mu \in [0, \bar{\mu}]$ and $N \in \mathbb{N}$ and consequently are also random lsc functions [23, Example 14.15]. Moreover by the same arguments as in the proof of Theorem 2.8, one obtains

$$\min_{x\in D} \hat{\varphi}^N(x) = \min_{y\in D_1} \hat{\varphi}^N(y) \tag{3.6}$$

and

$$\min_{x\in D} \Phi^N_\mu(x) = \min_{y\in D_1} \Phi^N_\mu(y). \tag{3.7}$$

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

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

respectively. Moreover, from $D_1 \subseteq D$ and (3.6), (3.7), y^* and y^{**} are thus solutions of (3.3) and (3.4), respectively. □

Let S^*, S^N and S^N_μ be the sets of solutions of (1.7), (3.3) and (3.4) in D_1. In the following, we analyze the convergence of S^N and S^N_μ 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

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

Since φ, $\hat{\varphi}^N$ and Φ^N_μ 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 $h = \Phi^N_\mu$.

THEOREM 3.7. *Under assumptions of Theorem 2.8, if the sample is iid, then the following hold.*

(i) Any sequence $\{x^N_\mu \in S^N_\mu\}$ has a cluster point as $N \to \infty$ and $\mu \downarrow 0$ a.s.
(ii) Any cluster point of $\{x^N_\mu \in S^N_\mu\}$ is an optimal solution of (1.7) 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 Lemma 3.5 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 [23, 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 2.8, 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 [23, 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 2.8 again, any such cluster point is also in the solution set of (1.7). 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 [26] 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

$$\lim_{N\to\infty} \Phi_\mu^N(x) = \varphi_\mu(x), \quad x \in D_1, \quad \mu \in [0, \bar{\mu}], \tag{3.8}$$

in the sense that for any given $\epsilon > 0$, there is a $\bar{\nu} > 0$, such that for any $N \ge \bar{\nu}$, 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 3.8. *Under assumptions of Theorem 2.8, if (3.8) 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 (1.7).*
(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 (3.8) and

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

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 3.7. □

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

For $g : R^n \to \overline{R}$ and $\bar{x}$ at which $g(\bar{x})$ is finite, recall [23, Section 8.A] that the *subderivative* of a function $g : R^m \to \overline{R}$ at a point $\bar{x}$ at which it 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

$$dg(\bar{x}; h) \geq 0 \;\text{ for all } h \in T_X(\bar{x}), \tag{3.9}$$

where $T_X(\bar{x})$ is the tangent cone of X at $\bar{x} \in X$ [10]. When, X is convex, one can exploit the polarity between the tangent and the normal cones [23, 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's, the sets D and D_1, are convex. Moreover, the functions $f(\xi, \cdot)$, cf. Theorem 2.4, and, a fortiori, $\tilde{f}(\xi, x, \mu)$ that are used to build our sample average approximations 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 Section 4.

In view of [23, 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, [23, Theorem 9.16] asserts a bit more but it would not be needed here.

In addition to these properties, the proof of the next theorem relies like Lemma 3.5 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

$$\lim_{\mu\downarrow 0} d\tilde{f}_\mu(\xi, x; h) \le df(\xi, x; h) \ \text{ for all } x, h. \tag{3.10}$$

In Section 4, we show that $Q(\xi,\cdot)$ is regular and the exponential smoothing function [6, 19] satisfies (3.10) for piecewise maxima functions.

THEOREM 3.9. *Suppose Assumptions 1 and 2 hold and $Q(\xi,\cdot)$ is regular for any fixed $\xi \in \Xi$. Then for any $\mu \ge 0$ and $N \in \mathbb{N}$, the SAA problem (3.3) and the SSAA problem (3.4) have stationary points in the compact set D_1. Let $\{x_\mu^N\} \subset D_1$ be a sequence of stationary point of (3.4). If the sample is iid, then any cluster point of $\{x_\mu^N\}$ is a stationary point of (1.7), a.s.*

Proof. The existence of stationary points is directly from the existence of minimizers of (3.3) and (3.4).

By the regularity of Q and continuous differentiability of F, we deduce that f, φ, $\hat{\phi}^N$ are Clarke regular [8, 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 $t^{-1}[f(\xi, x+h) - f(\xi, x)] \ge \beta$, a.s. for all h in a neighborhood of 0 and $0 < \bar{t} \le t$. By Proposition 2.9 in [28, Section 2], we obtain

$$E[df(\xi, x; y-x)] \le d\varphi(x; y-x), \quad \forall\, x, y \in D. \tag{3.11}$$

By the continuously 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 $h \in R^n$, $d\Phi_\mu^N(\cdot; h) = \frac{1}{N}\sum_{i=1}^N d\tilde{f}_\mu(\xi, \cdot; h)$ is upper semicontinuous. Hence we can use the same technique as in the proof of Lemma 3.5, to show

$$d\Phi_\mu^N(\cdot; h) \quad \xrightarrow{e} \quad d\varphi_\mu(\cdot; h), \quad \text{in } D, \quad \text{a.s.} \tag{3.12}$$

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, we have

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

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 (3.12).

From (3.11) and (3.10), 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 1. *From the properties of 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 2. *The conclusions of Proposition 6* [25, Chapter 6] *are similar to that of Theorem 3.7 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* [3].

4. Application and numerical experiments. In this section, we use two examples in traffic network analysis to illustrate the new ERM-formulation (1.7) and the theoretical results derived in the preceding sections. We first use an example with 7 links and 6 variables to explain the theory and its application in detail. Next we present numerical results for this example and one more example with 19 links and 25 variables to show the efficiency of the SSAA approach.

4.1. Application. 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 4.1 from [31] has 5 nodes, 7 links, 2 OD-pairs ($1 \to 4$, $1 \to 5$) and 6 paths $p_1 = \{3, 7, 6\}, p_2 = \{3, 1\}, p_3 = \{4, 6\}, p_4 = \{3, 7, 2\}, p_5 = \{3, 5\}, p_6 = \{4, 2\}$.

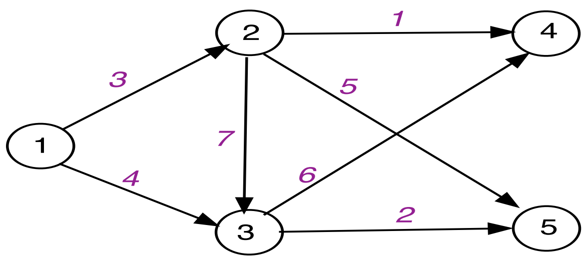

FIG. 4.1. *The 7-links network*

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.

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 $x \in R^{|K|}$, 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 relation between x and v is given by

$$v = \Delta x,$$

where $\Delta = (\delta_{k,j})$ is the link-path incidence matrix with entries $\delta_{k,j} = 1$ if link k is on path j and $\delta_{k,j} = 0$ otherwise. Let $A = (a_{i,j})$ denote the OD-path incidence matrix with entries $a_{i,j} = 1$ if path j connects the ith OD and $a_{i,j} = 0$ otherwise. The incidence matrices for the network in Figure 4.1 are given respectively as follows.

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

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

$$T_k(\xi, v) = t_k^0\Big(1.0 + 0.15\Big(\frac{v_k}{(c_\xi)_k}\Big)^{n_k}\Big), \quad k = 1, \dots, 7 \tag{4.1}$$

where t_k^0 and n_k are given parameters. The path travel cost function is defined by

$$F(\xi, x) = \eta_1 \Delta^T T(\xi, \Delta x), \tag{4.2}$$

where $\eta_1 > 0$ is the time-based operating costs factor. If $n_k = 1$, then $F(\xi, x) = M_\xi x + q$, where

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

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$(M_\xi) = 5$. Obviously, $E[M_\xi]$ is positive semi-definite, but condition (1.6) used in [1] does not hold.

For a fixed $\xi \in \Xi$, the VI formulation for Wardrop's user equilibrium, denoted by VI($X_\xi, F(\cdot, \xi)$), seeks an equilibrium path flow $x_\xi \in R^n$ such that

$$(y - x_\xi)^T F(\xi, x_\xi) \geq 0, \qquad \text{for all } y \in X_\xi = \{x \,|\, Ax = b_\xi, \quad x \geq 0\}, \tag{4.3}$$

which is equivalent to find a solution such that the residual function $f(\xi, x) = 0$. The residual function is nonnegative and regarded as a cost function.

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.

Now we give sufficient conditions on A and b_ξ that guarantee that Assumption 1 and Assumption 2 hold. Such conditions hold for the OD-path incidence matrix and random demand vector.

DEFINITION 4.1. [9] *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 4.2. [9] *If S is a nonempty meet semi-sublattice that is closed and bounded below, then S has a least element.*

THEOREM 4.3. *Suppose* $\text{prob}\{b_\xi > 0, \|b_\xi\|_\infty \le \beta\} = 1$ *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*

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

Then,

$$X_\xi \subseteq \{x \,|\, 0 \le x \le \gamma e\} =: U_0. \tag{4.4}$$

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

$$\|z(\xi, u(\xi, x))\|_\infty \le \theta = \kappa \max(1, \gamma_0^{-1})\|A_K^{-T}\|_\infty \tag{4.5}$$

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,

$$A_K^{-1}b_\xi = A_K^{-1}APPx = A_K^{-1}[A_K, A_J]Px = [I, A_K^{-1}A_J]Px. \tag{4.6}$$

Since $[I, A_K^{-1}A_J]$ is a nonnegative matrix and its each column has at least one positive element, $[I, A_K^{-1}A_J]Px \ge 0$. Hence, there is a positive element $(I, A_K^{-1}A_J)_{i,j_0} = B_{i,j_0} \ge \min(1, \gamma_0)$, such that

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

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

Let $S_{\xi,u} = \{z \,|\, A^Tz + F(\xi, u) \ge 0\}$ denote the feasible set. For $w, v \in S_{\xi,u}$, let $s = \min(w, v)$ be their meet. We consider an arbitrary index $i \in \{1, \dots, n\}$. 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,

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

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 $A_J^TA_K^{-T}$ is a nonnegative matrix. Then

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

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 $z \in S_{\xi,u}$,

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

which implies

$$z \geq -CPF(\xi,u) \geq -LA_K^{-T} e \geq -\max(1,\gamma_0^{-1})\kappa A_K^{-T} e. \tag{4.7}$$

Hence $S_{\xi,u}$ is closed and bounded below. By Lemma 4.2, $S_{\xi,u}$ 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 (2.3) a.s.

Furthermore, using $z(\xi,u) \leq t$ and (4.7),

$$\|z(u,\xi)\|_\infty \leq \kappa \max(1,\gamma_0^{-1})\|A_K^{-T}\|_\infty = \theta \ \text{ a.s.} \tag{4.8}$$

which completes the proof. □

In traffic flow problem [2, 31, 34], we often have the following constraints

$$X_\xi = \{x \mid \sum_{j\in I_i} x_j = (b_\xi)_i, \quad i = 1,\dots,m\} \tag{4.9}$$

with

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

where b 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 ith original-destination (OD) pair. The constraints (4.9), 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 4.3. Moreover, if $b_\xi > 0$, then from $A^T z + F(\xi,u) \geq 0$, the solution $z(\xi,u)$ of (2.3) has a closed form

$$z_i(\xi,u) = \max\{-F_j(\xi,u), j \in I_i\}, \quad i = 1,\dots,m. \tag{4.10}$$

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

Now, we define a smoothing function of

$$f(\xi,x) = u(\xi,x)^T F(\xi,u(\xi,x)) + \sum_{i=1}^m b_i(\xi) \max_{j\in I_i}\{-F_j(\xi,u(\xi,x))\}. \tag{4.11}$$

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

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

We define a smoothing function of p as follows [19]: for $\mu > 0$,

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

LEMMA 4.4. [6] *$\tilde{p}$ is continuously differentiable with respect to x for any fix $\mu > 0$. 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 4.5. *The directional derivative* $\tilde{p}'_\mu(y;h)$ *of* $\tilde{p}$ *satisfies*

$$\lim_{\mu\downarrow 0} \tilde{p}'_\mu(y;h) \le p'(y;h), \quad \forall\ y,h \in R^k. \tag{4.12}$$

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

$$\tilde{f}(\xi,x,\mu) = u(\xi,x)^T F(\xi,u(\xi,x)) + \mu \sum_{i=1}^{m} (b_\xi)_i \ln \sum_{j\in I_i} e^{-F_j(\xi,u(\xi,x))/\mu}. \tag{4.13}$$

THEOREM 4.6. *When* X_ξ *is defined by (4.9) and* $\tilde{f}$ *is defined by (4.13), the assumptions of Theorem 4.3 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 (3.10).*

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 ith 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 4.4, it is easy to verify that $\tilde{f}$ is a smoothing function of f defined in (4.11). 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 [8, Theorem 2.3.9] since $b_\xi > 0$, p is convex and F is continuously differentiable.

Next, we show (3.10) 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 $I_1 = K = \{1,\dots k\}$ and thus $z_1(\xi,u) = \max\{-F_j(\xi,u), j \in K\}$. For a fixed ξ, let $g(u) =$

$(-F_1(\xi,u),\dots,-F_k(\xi,u))^T$ and $q(u)=p(g(u))=\max(g_1(u),\dots,g_k(u))$. Since $b_i>0$, for $i=1,\dots,m$, it is sufficient to show that

$$\lim_{\mu\downarrow 0}\tilde{q}'_\mu(u;h)\le q'(u;h),\quad \forall\, u,h\in R^k. \tag{4.14}$$

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 $\mu>0$,

$$\lim_{\mu\downarrow 0}\tilde{q}'_\mu(u;h)=\lim_{\mu\downarrow 0}\nabla\tilde{p}_\mu(g(u))^Tg'(u)h\le p(g(u);g'(u)h)=q(u;h)$$

that follows from Lemma 4.5. □

4.2. Numerical experiment. In the following two examples, X_ξ is defined by (4.9) and $\tilde{f}$ is defined by (4.13). The EV-formulation for the two examples is to find an $x\in X=\{x\,|\,Ax=E[b_\xi]\,\}$ such that

$$(y-x)^TE[F(\xi,x)]\ge 0,\qquad y\in X. \tag{4.15}$$

We solve the following minimization problem

$$\min_{x\in X} g(x):=\max\{(x-y)^TE[F(\xi,x)]\,|\,y\in X\} \tag{4.16}$$

and set a minimizer to be $x_{\rm EV}$.

For the ERM-formulation, we solve the ERM problem (1.7) and set $x_{\rm ERM}=(I-A^\dagger A)x^*+A^\dagger E[b_\xi]$, where x^* is a solution of (1.7).

We use the residual function f and conditional value-at-risk(CVaR) to compare the two formulations; for fixed x,

$$\alpha^*\in\operatorname*{argmin}_{\alpha\in R}\text{CVaR}(x,\alpha):=\alpha+\frac{1}{1-\beta}E\{[f(\xi,x)-\alpha]_+\}.$$

Example 4.1. This example is the 7-link problem in Figure 4.1. 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 [31], which are listed in Table 4.1.

Table 4.1
Link cost parameters in Fig.4.1

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

TABLE 4.2
Solutions for sampling size N=1000

	$n_a = 1$		$n_a = 2$		$n_a = 4$	
	x_{EV}	x_{ERM}	x_{EV}	x_{ERM}	x_{EV}	x_{ERM}
x_1	44.24	45.34	18.85	27.28	2.89	14.87
x_2	79.82	79.42	90.32	88.11	95.09	92.38
x_3	75.94	75.25	90.83	84.61	102.03	92.75
x_4	50.31	51.37	26.61	28.29	20.37	19.64
x_5	88.20	87.88	99.65	97.53	104.87	102.73
x_6	81.49	80.75	93.74	94.18	94.76	97.63

TABLE 4.3
Criteria for N=1000, $\beta = 0.9$

	$n_a = 1$ $\varepsilon = 5.5E2$		$n_a = 2$ $\varepsilon = 4.5E3$		$n_a = 4$ $\varepsilon = 5E5$	
	x_{EV}	x_{ERM}	x_{EV}	x_{ERM}	x_{EV}	x_{ERM}
$\text{prob}\{f(\xi, x) \leq \varepsilon\}$	0.96	0.96	0.58	0.61	0.56	0.59
$E[\|x - u_\xi^*\|]$	23.61	23.69	46.94	39.63	35.28	33.01
$E[\|u(\xi, x) - u_\xi^*\|]$	23.79	23.87	47.03	39.72	35.41	33.15
$E[f(\xi, x)]$	2.901E2	2.897E2	4.316E3	4.198E3	5.064E5	4.907E5
α^*	4.795E2	4.790E2	7.395E3	7.132E3	1.071E6	1.037E6
$\text{CVaR}(x, \alpha^*)$	5.564E2	5.570E2	8.691E3	8.515E3	1.254E6	1.229E6

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 4.2 and Table 4.3 were obtained by using the same sampling with size $N = 1000$. Table 4.2 gives EV and ERM solutions for different values of n_a. Tbale 4.3 lists robustness and risk criteria for the EV and ERM solutions in Table 4.3; u_ξ^* means solution of the variational inequalities for each fixed $\xi \in \Xi$.

In Figure 4.2, we graph $\text{prob}\{f(\xi, x) \leq \varepsilon\}$ with different values of ε. We can see the ERM-formulation has higher probability than the EV-formulation for each ε.

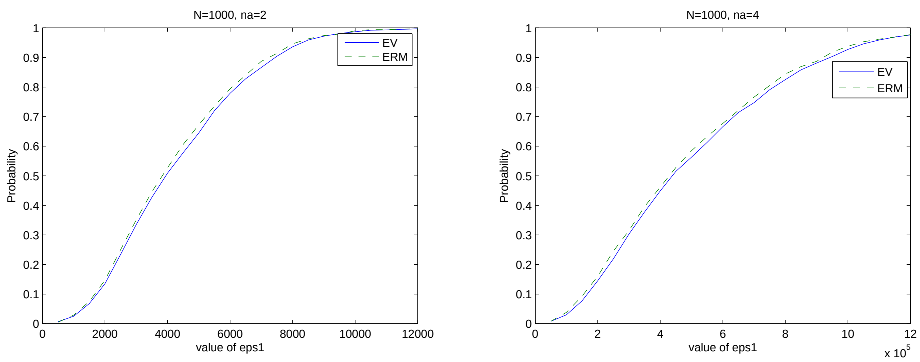

FIG. 4.2. $\text{prob}\{f(\xi, x) \leq \varepsilon\}$ *with different values of* ϵ *for* x_{EV} *and* x_{ERM}.

Example 4.2. This example uses the Nguyen and Dupuis network, which contains 13 nodes, 19 directed links, and 4 OD movements. See Figure 4.3.

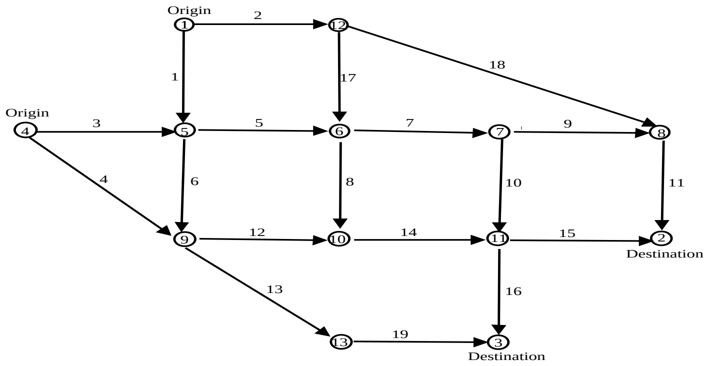

FIG. 4.3. *Nguyen and Dupuis Network*

We use the free-flow travel time t_a^0 as that used by Yin [32], and the mean of the demand vector $E[b(\omega)]$ of the network is $E[b(\omega)] = [400, 800, 600, 450]^T$.

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 $b(\xi) \sim \underline{b} + \hat{b} * \text{beta}(\alpha, \beta)$ 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 $(b(\xi^i), Ca(\xi^i))$ for $i = 1, 2, \cdots, N$, where $Ca(\xi^i)$ is sampled from the three possibilities with given probability and $b(\xi^i)$ is sampled from the beta distribution.

TABLE 4.4
Example 4.2. Criteria for $\beta = 0.9$, $n_a = 2$, $\varepsilon = 3.3E3$

		x_{EV}	x_{ERM}
$N = 10^3$, $\mu = 10^{-4}$	$\text{prob}\{f(x,\xi) \le \varepsilon\}$	0.508	0.952
	$E[f(x,\xi)]$	3.498E3	2.938E3
	α^*	7.935E3	3.226E3
	$\text{CVaR}(x, \alpha^*)$	8.154E3	3.333E3
$N = 5 * 10^3$, $\mu = 10^{-5}$	$\text{prob}\{f(x,\xi) \le \varepsilon\}$	0.510	0.908
	$E[f(x,\xi)]$	3.498E3	2.983E3
	α^*	7.918E3	3.286E3
	$\text{CVaR}(x, \alpha^*)$	8.121E3	3.403E3
$N = 10^4$, $\mu = 10^{-6}$	$\text{prob}\{f(x,\xi) \le \varepsilon\}$	0.509	0.927
	$E[f(x,\xi)]$	3.505E3	2.976E3
	α^*	7.978E3	3.253E3
	$\text{CVaR}(x, \alpha^*)$	8.168E3	3.359E3

Remark 3. *The two examples are often used in transportation research. They satisfy all our assumptions of the theoretical analysis for the ERM-formulation in Sections 2 and 3. 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.*

REFERENCES

[1] R.P. Agdeppa, N. Yamashita and M. Fukushima, *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.

[2] 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.

[3] H. Attouch and R. J-B Wets, *Epigraphical processes: laws of large numbers for random lsc functions*, Séminaire d'Analyse Convexe, 13(1991), pp. 13.1–13.29.

[4] J.R. Birge and F. Louveaux, Introduction to Stochastic Programming, Springer, New York, 1997.

[5] X. Chen and M. Fukushima, *Expected residual minimization method for stochastic linear complementarity problems*, Math. Oper. Res., 30(2005), pp. 1022–1038.

[6] X. Chen, R.S. Womersley and J. Ye, *Minimizing the condition number of a Gram matrix*, SIAM J. Optim., 21(2011), pp. 127–148.

[7] X. Chen, C. Zhang and M. Fukushima, *Robust solution of monotone stochastic linear complementarity problems*, Math. Program., 117(2009), pp. 51–80.

[8] F.H. Clarke, Optimization and Nonsmooth Analysis, John Wiley, New York, 1983.

[9] R.W. Cottle, J.S. Pang and R.E. Stone, The Linear Complementarity Problem, Academic Press, New York, 1992.

[10] F. Facchinei and J.S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer-Verlag, New York, 2003.

[11] H. Fang, X. Chen and M. Fukushima, *Stochastic R_0 matrix linear complementarity problems*, SIAM J. Optim., 18(2007), pp. 482–506.

[12] G. Gürkan, A.Y. Özge and S.M. Robinson, *Sample-path solution of stochastic variational inequalities*, Math. Program., 84(1999), pp. 313–333.

[13] H. Jiang and H. Xu, *Stochastic approximation approaches to the stochastic variational inequality problem*, IEEE. Trans. Autom. Control, 53(2008), pp. 1462–1475.

[14] L.A. Korf and R. J-B. Wets, *Random lsc functions: an ergodic theorem*, Math. Oper. Res., 26(2001), pp. 421–445.

[15] 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.

[16] 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.

[17] 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.

[18] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15(1977), pp. 957-972.

[19] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103(2005), pp. 127-152.

[20] L. Qi, A. Shapiro and C. Ling, *Differentiablility and semismoothness properties of integrable functions and their applications*, Math. Program., 102(2005), pp. 223–248.

[21] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58(1993), pp. 353–367.

[22] R.T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk, 2(2000), pp. 493–517.

[23] R.T. Rockafellar and R.J-B Wets, Variational Analysis, Springer, Berlin, 1998.

[24] R.Y.Rubinstein and A.Shapiro, *Discrete event sustems: sensitivity analysis ans stochastic optimization by the score function method*, John Wiley and Sons, New York, 1993.

[25] A. Ruszczynski and A. Shapiro, Stochastic Programming, Handbooks in Operations Research and Management Science, Elsevier, 2003.

[26] I.H. Sloan and H. Wozniakowskic, *When are quasi-Monte Carlo algorithms efficient for*

high dimensional integrals ? J. Complex., 14(1998), pp. 1–33.

[27] C. Suwansirikul, T.L. Friesz and R.L. Tobin, *Equilibrium decomposed optimization: A heuristic for the continuous equilibrium network design problems*, Transp. Sci., 21(4)(1987),pp. 254–263.

[28] R.J-B Wets, *Stochastic programming*, G.L. Nemhauser et al., eds., Handbooks in Operations Research and Management Sciences, Vol 1., Elsevier, 1989. pp. 573–629.

[29] H. Xu, *Sample average approximation methods for a class of stochastic variational inequality problems*, Asia Pac. J. Oper. Res., 27(2010), pp. 103 –119.

[30] 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.

[31] H. Yang, *Sensitivity analysis for the elastic-demand network equilibrium problem with applications*, Transp. Res. B, 31(1997), pp. 55–70.

[32] Y.F. Yin, S.M. Madanat and X.-Y. Lu, *Robust improvement schemes for road networks under demand uncertainty*, Eur. J. Oper. Res., 198(2)(2009), pp. 470–479.

[33] C. Zhang and X. Chen, *Smoothing projected gradient method and its application to stochastic linear complementarity problems*, SIAM J. Optim., 20(2009), pp. 627–649.

[34] C. Zhang, X. Chen and A. Sumalee, *Robust Wardrops user equilibrium assignment under stochastic demand and supply: expected residual minimization approach*, Transp. Res. B, 45(2011), pp. 534-552.