STOCHASTIC VARIATIONAL INEQUALITIES: RESIDUAL MINIMIZATION
SMOOTHING/SAMPLE AVERAGE APPROXIMATIONS

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

20 February 2011, 3 April 2012

Abstract. The stochastic variational inequality (VI) 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 the linear stochastic VI 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 denoted by $\text{VI}(X, F)$, also written as,

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

here $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuous function, $X \subseteq \mathbb{R}^n$ a (nonempty) closed, convex set and $N_X(x)$ is the normal cone to $X$ at $x$. The $\text{VI}(X, F)$ is often solved via a deterministic optimization problem by using a residual function for the $\text{VI}(X, F)$.

Definition 1.1. [10] A residual function for the $\text{VI}(X, F)$ on a (closed) set $D \supseteq X$ is a nonnegative function $f : D \rightarrow \mathbb{R}_+$ such that $f(x) = 0$ if and only if $x \in D$ solves the $\text{VI}(X, F)$.

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.

*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 supported by ARO-grant W911NS1010246.
‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.(09900332p@polyu.edu.hk).
Even, when just $F$ involves stochastic parameters, say $\xi$, one might be led to consider a variety of formulations: find $x \in X$ such that

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

or still

$$E[-F(\xi, x)] \in N_X(x),$$

(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 article. The first formulation with $\alpha = 1$ could be converted to a large variational inequality, involving an infinite number of inequalities when $\xi$ 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 $\alpha$. When, also the set $X$ depends on $\xi$, 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 $X_\xi$ by its expectation or try to compute $E[N_X(x) + F(\xi, x)]$.

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 $\xi$ 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 $- F(\xi, x) \in N_X(x)$. 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 $\xi$ 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 $b_\xi$ 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
  \begin{equation}
  (y - x)^T E[F(\xi, x)] \geq 0, \quad y \in X.
  \end{equation}

- **Expected Residual Minimization** ERM-formulation [1, 5, 7, 11, 15, 16, 33, 34]:
  \begin{equation}
  \min_{x \in X} E[f(\xi, x)],
  \end{equation}

  $f(\xi, \cdot) : X \to R_+$ 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 occurrences 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 Definition 1.1 of the residual function for the classical VI to stochastic VI.

**Definition 1.2.** 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, we can see that Definition 1.2 is a natural extension of Definition 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 [10]. In this paper, we will not study the theoretical error bounds, but we will provide numerical results for $E[\|u(\xi, x) - x^*\|]$ and $E[f(\xi, x^*)]$.

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$_{X\xi}$ is the canonical projection of $R^n$ onto $X_\xi$ and $\| \cdot \|$ is the $\ell_2$ 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.

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 article, we rely on the gap function [10, Section 1.5] to define a new residual function

\begin{equation}
 f(\xi, x) = u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, u(\xi, x))
\end{equation}

where

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

is a recourse variable 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 \( A \). The gap function provides a measure for the deviations that will be needed to ‘adjust’ the solution of the variational inequality as it is affected by the circumstances, i.e., the random components of the problem.

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,

\begin{equation}
 x^T E[M_\xi] x \geq 0, \quad \forall x \in \mathbb{R}^n.
\end{equation}

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

\begin{equation}
 \inf_{\xi \in \Xi, \|x\|=1} x^T M_\xi x \geq \beta_0
\end{equation}

for some positive constant \( \beta_0 \).

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 \( M_\xi \) that satisfies condition (1.5), but there is no \( \xi \in \Xi \) for which \( M_\xi \) is positive semidefinite. Hence, condition (1.5) is much weaker than (1.6). Moreover, the new residual function (1.4) can be used when \( X_\xi \) is a random set.

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

\begin{equation}
 \min_{x \in D} \varphi(x) = E[f(\xi, x)],
\end{equation}

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 \( \varphi \) 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 $\varphi$ is convex if $E[M_\xi]$ is positive semi-definite.

In Section 3, 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 (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 VIs 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 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.

Throughout the paper, $\|\cdot\|$ represents the $\ell_2$ 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 $\text{proj}_C$ denote the orthogonal projection from $R^n$ onto $C$, that is, $\text{proj}_C(x) = \arg\min_{y \in C} \|y - x\|$.

2. A new residual function. For given $\xi$, 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]

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

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

$$\min \quad x^T F(\xi, x) + Q(\xi, x)$$

s.t. $\quad x \in X_\xi$

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

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

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

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

$$D = \{ x \mid (A^T A - I)x \leq \xi \}$$

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

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

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

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

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

**Assumption 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 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 $b_\xi$ that arise from traffic equilibrium problems.

**Theorem 2.1.** When Assumption 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

$$u(\xi, x) = (I - A^TA)x + A^Tb_\xi \geq (I - A^TA)x + \xi \geq 0.$$ 

Hence $u(\xi, x) \in X_\xi$. By definition of $f(\xi, x)$ and Assumption 1, for almost every $\xi \in \Xi$, there is $y(\xi, x) \in R^n$ such that

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

$$= u(\xi, x)^T F(\xi, u(\xi, x)) - y(\xi, x)^T F(\xi, u(\xi, x))$$

$$= \max\{(u(\xi, x) - y)^T F(\xi, u(\xi, x)) \mid y \in X_\xi\} \geq 0,$$

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 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:

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

By Theorem 2.1, $\varphi(x) \geq 0$ for all $x \in D$ and if $\varphi(x) = 0$ then, $u(\xi, x)$ solves the VI($X_\xi, F(\xi, \cdot)$) for almost every $\xi \in \Xi$. Hence the “here and now” solution is

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

where $x^*$ is a solution of the ERM-formulation (1.7). By definition of $u(\xi, x)$,

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

Moreover, the following proposition shows that $x_{\text{ERM}}$ is also a solution of our ERM-formulation (1.7).
Proposition 2.2. Under Assumption 1, if (1.7) has a solution \( x^* \), then

\[
x_{\text{ERM}} \in \arg\min_{x \in D} \varphi(x).
\]

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

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

Moreover, we obtain

\[
u(\xi, \bar{u}) - \nu(\xi, x) = (I - A^\dagger A)\bar{u} + A^\dagger b_{\xi} - (I - A^\dagger A)x - A^\dagger b_{\xi}
\]

\[
= (I - A^\dagger A)((I - A^\dagger A)x + A^\dagger E[b_{\xi}]) - (I - A^\dagger A)x
\]

\[
= (I - A^\dagger A)A^\dagger E[b_{\xi}] = 0,
\]

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

\[
\operatorname{proj}_{X_\xi}(x) = \operatorname{proj}_{X_\xi}(E[\operatorname{proj}_{X_\xi}(x)]).
\]

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

\[
u(\xi, x_{\text{ERM}}) = \operatorname{proj}_{X_\xi}(x_{\text{ERM}}) = \operatorname{proj}_{X_\xi}(x^*) = \nu(\xi, x^*),
\]

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 \( A^\dagger (E[b_{\xi}] - A x^*) = 0 \). From (2.6), if the ERM-formulation (1.7) has a solution and \( A^\dagger (E[b_{\xi}] - A x^*) \neq 0 \), then (1.7) has a multiplicity of solutions.

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

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

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

Assumption 2.

(i) There are \( b, \bar{b} \in \mathbb{R}^m \) such that \( b \leq b_{\xi} \leq \bar{b} \) for \( \forall \xi \in \Xi \),

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

(iii) \( \exists d_1 : \Xi \to \mathbb{R}^+, \) bounded, such that \( \|\nabla F(\xi, u)\| \leq d_1(\xi) \) for all \( u \in U \),

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

Assumption 2(i)-(iii) are pretty standard and are in no way restrictive as far as applications are concerned. Assumption 2(iv) is not quite as common 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 \( \xi \in \Xi \), we immediately obtain the following proposition.

Proposition 2.3. \( F(\xi, u(\xi, x)) \) is measurable in \( \xi \) 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 hold, then for all $x \in D$, 
\[
\|F(\xi, u(\xi, x))\| \leq d(\xi) \quad \text{and} \quad \|\nabla_x F(\xi, u(\xi, x))\| \leq \|I - A^T A\| d(\xi).
\]

**Theorem 2.4.** Assume that Assumption 1 holds. Then, the function $f$ is measurable in $\xi$ for any $x \in D$ and locally Lipschitz continuous in $x$ a.s.; actually, under Assumption 2(iii), the functions \( \{ f(\xi, \cdot) : D \to R, \xi \in \Xi \} \) are then also equi-locally Lipschitz continuous a.s. Moreover, under Assumption 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 $\varphi$ 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 $\nu_\lambda > 0$, that only depends on the matrix $A$, such that
\[
\|Q(\xi, u) - Q(\xi, v)\| \leq \|b_\xi\| \|z(\xi, u) - z(\xi, v)\| \leq \|b_\xi\| \nu_\lambda \|F(\xi, u) - F(\xi, v)\| \quad \text{a.s.}
\]

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(ii), the (local) Lipschitz constant not depending on $\xi$. 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 $\xi$. 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 $\xi$ for every $x \in R^n$ and $b_\xi$ is measurable in $\xi$. We have that $Q(\xi, u)$ is measurable in $\xi$ for any $u \in U$, cf. [25, Theorem 19, Chapter 1]. Hence the function $f(\xi, x)$ is measurable in $\xi$ 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 $\xi \in \Xi$,
\[
\min\{b_\xi^T z | A^T z + F(\xi, u) \geq 0\} \quad \text{and} \quad \min\{b_\xi^T z | A^T z + F(\xi, v) \geq 0\}
\]
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.,
\[
0 \leq \lambda(A^T z(\xi, u) + F(\xi, u)) + (1 - \lambda)(A^T z(\xi, v) + F(\xi, v))
\]
\[
\leq A^T (\lambda z(\xi, u) + (1 - \lambda)z(\xi, v)) + F(\xi, \lambda u + (1 - \lambda)v)
\]
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 $Q(\xi, x)$,
\[
Q(\xi, \lambda u + (1 - \lambda)v) \leq b_\xi^T (\lambda z(\xi, u) + (1 - \lambda)z(\xi, v))
\]
\[
= \lambda Q(\xi, u) + (1 - \lambda)Q(\xi, v), \quad \text{a.s.}
\]

(ii) With $B = A^T A - I$, one has 
\[
f(\xi, x) = (-Bx + A^T b_\xi)^T (M_\xi (-Bx + A^T b_\xi) + q_\xi) + Q(\xi, -Bx + A^T b_\xi)
\]
\[
= x^T B M_\xi Bx - (A^T b_\xi)^T (M_\xi + M^T(\xi))Bx - q_\xi^T Bx
\]
\[
+ (A^T b_\xi)^T (M_\xi A^T b_\xi + q_\xi) + Q(\xi, -Bx + A^T b_\xi).
\]
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 $\varphi$ is finite valued and there are a vector $c \in \mathbb{R}^n$ and a constant $c_0$ such that

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

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

**Theorem 2.5.** Under Assumptions 1 and 2, $\varphi$ is globally Lipschitz on $D$, i.e.,

$$
|\varphi(x) - \varphi(y)| \leq \kappa \|x - y\|, \quad x, y \in D
$$

where

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

recall that $A$ is an $m \times n$-matrix and for the constant $\nu_A$ refer to (2.8).

Proof. For the first term in $\varphi$, we have

$$
|u^T F(\xi, u) - v^T F(\xi, v)| \leq |u^T (F(\xi, u) - F(\xi, v))| + |(u - v)^T F(\xi, v)|
$$

$$
\leq \|u\| d_1(\xi) \|u - v\| + d(\xi) \|u - v\|
$$

$$
\leq (\gamma \sqrt{n} d_1(\xi) + d(\xi)) \|u - v\|.
$$

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

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

Combining these two inequalities,

$$
|\varphi(x) - \varphi(y)| \leq E[|f(\xi, x) - f(\xi, y)|]
$$

$$
\leq E[|u(\xi, x)^T F(\xi, u(\xi, x)) - u(\xi, y)^T F(\xi, u(\xi, y))|] + E[|Q(\xi, u(\xi, x)) - Q(\xi, u(\xi, y))|]
$$

$$
\leq (\gamma \sqrt{n} E[d_1(\xi)] + E[d(\xi)]) + \nu_A E[\|b_\xi\| E[d_1(\xi)]] \|I - A^T A\| \|x - y\|,
$$

completes the proof. □

**Definition 2.6.** [18] Suppose that $\phi : X \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ is a locally Lipschitz continuous function, then $\phi$ is semismooth at $x \in \text{int } X$ if $\phi$ 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 $\text{int } X$ denotes the interior of $X$ and $\partial \phi$ denotes the Clarke generalized gradient.

**Theorem 2.7.** Suppose Assumptions 1 and 2 hold. Then the function $\varphi$ 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 $\kappa_1$ such that

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

\[
\frac{|f'_\xi(x + h; h) - f'_\xi(x; h)|}{\|h\|} \leq \kappa_2(\xi),
\]

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

For (i), as follows from the proof of Theorem 2.5,

\[
|f(\xi, x) - f(\xi, y)| \leq (d(\xi) + d_1(\xi)\sqrt{n\gamma} + m\nu(\xi))\|I - A^1A\|\|x - y\|
\]

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 \( (d(\xi) + \sqrt{n\gamma}d_1(\xi))\|I - A^1A\|\|h\| \). 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 \( m\nu(\xi)d_1(\xi)\|b_\xi\|\|I - A^1A\|\|h\| \). Thus, we set \( \kappa_2(\xi) = 2(d(\xi) + \sqrt{n\gamma}d_1(\xi) + m\nu(\xi)d_1(\xi)\|b_\xi\|\|I - A^1A\| \) 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

\[
D_1 = \{ y | y = (I - A^1A)x, x \in D \}.
\]

Moreover,

\[
D_1 \subseteq D \quad \text{and} \quad \text{argmin}_{y \in D_1} \varphi(y) \subseteq \text{argmin}_{x \in D} \varphi(x).
\]

**Proof.** From Theorem 2.1 and Theorem 2.4 follows \( 0 \leq \varphi(x) < \infty \) 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 \( \overline{b} \) such that \( \underline{b} \leq b_\xi \leq \overline{b} \) for \( \forall \xi \in \Xi \). Hence, the vector

\[
(I - A^1A)x = u(\xi, x) - A^1b_\xi
\]

is in the compact set \( D_1 \). From \( (I - A^1A)(I - A^1A) = (I - A^1A) \) and \( D = \{ x | (I - A^1A)x + \xi \geq 0 \} \), we have \( y = (I - A^1A)x \in D \) which implies \( D_1 \subseteq D \). Moreover, from

\[
(I - A^1A)(I - A^1A)x + A^1b_\xi = (I - A^1A)x + A^1b_\xi = u(\xi, x),
\]

we obtain

\[
\text{min}_{x \in D} \varphi(x) = \text{min}_{y \in D_1} \varphi(y).
\]

Since \( D_1 \) is compact and \( \varphi \) is continuous, \( \text{argmin}_{D_1} \varphi \neq \emptyset \) and any \( y^* \in \text{argmin}_{D_1} \varphi \) also minimizes \( \varphi \) on \( D \) since \( D_1 \subseteq D \). Finally, from (2.11) one obtains (2.10).

**Remark 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 this section after (2.3) can ensure that
(i) \( u(\xi, x) = \text{proj}_{X_1}(x) \geq 0 \), 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 (1.4) is defined by the recourse variable \( u(\xi, x) \) which is dependent on the first level variable \( x \) and random variable \( \xi \). Hence the degree of inadequacy or “loss” of a given \( x \) for a given \( \xi \) can be measured by \( f(\xi, x) \). In section 4, we show that \( \max\{-y^T F(\xi, x)|y \in X_1\} \) has a closed form and \( f(\xi, x) \) can be written explicitly for Wardrop’s equilibrium for traffic assignment.

3. Smoothing sample average approximation (SSAA). Let \( \xi^1, \ldots, \xi^N \) be a sampling of \( \xi \). 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

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

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

\[
\hat{\phi}^N(x) = \frac{1}{N} \sum_{i=1}^{N} f(\xi^i, x).
\]

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 \( \phi \) and its sample average approximation \( \hat{\phi}^N(x) \) are differentiable. In what follows, we introduce a smoothing sample average approximation (SSAA)

\[
\Phi^N_\mu(x) = \frac{1}{N} \sum_{i=1}^{N} \tilde{f}(\xi^i, x, \mu), \tag{3.1}
\]

where \( \tilde{f} : \Xi \times \mathbb{R}^n \times \mathbb{R}^+ \rightarrow \mathbb{R} \) is a smoothing approximation of \( f \).

**Definition 3.1.** Let \( g : \mathbb{R}^n \rightarrow \mathbb{R} \) be a locally Lipschitz continuous function. We call \( \tilde{g} : \mathbb{R}^n \times \mathbb{R}^+ \rightarrow \mathbb{R} \) a smoothing function of \( g \), if \( \tilde{g} \) is continuously differentiable on \( \mathbb{R}^n \) for any \( \mu \in \mathbb{R}^+ \) and for any \( x \in \mathbb{R}^n \),

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

If \( f(\xi, \cdot) \) is convex for almost every \( \xi \in \Xi \), then we can use the Moreau-Yoshida regularization to define a smoothing function for \( \Phi^N_\mu \). However, the Moreau-Yoshida
regularization cannot be used when \( f(\xi, \cdot) \) is not convex for almost every \( \xi \in \Xi \). We will give smoothing functions for traffic equilibrium problems in section 4.

We consider the existence and the convergence of solutions of the following SAA problems
\[
(3.3) \quad \min_{x \in D} \varphi^N(x)
\]
and SSAA problems
\[
(3.4) \quad \min_{x \in D} \Phi^N_\mu(x).
\]

Let \( X \subseteq \mathbb{R}^n \) be an open set and \( \mathcal{R} = [-\infty, \infty] \).

**Definition 3.2.** [23] A sequence of functions \( \{g^N: X \to \mathcal{R}, N \in \mathbb{N}\} \) epi-converges to \( g: X \to \mathcal{R} \), written \( g^N \srtr{e} g \), 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 \( g: \Xi \times X \to \mathbb{R} \) is a random lsc (lower semicontinuous) function if

(i) \( g \) is jointly measurable in \( (\xi, x) \),
(ii) \( g(\xi, \cdot) \) is lsc for every \( \xi \in \Xi \).

**Definition 3.4.** [14] A sequence of random lsc functions \( \{g^N: \Xi \times X \to \mathbb{R}, N \in \mathbb{N}\} \) epi-converges to \( g: X \to \mathcal{R} \) a.s., written \( g^N \srtr{e} g \) a.s. if for almost every \( \xi \in \Xi \),
\[
\{g^N(\xi, \cdot): X \to \mathcal{R}, N \in \mathbb{N}\} \text{ epi-converges to } g: X \to \mathcal{R}.
\]

Let \( \delta_D(x) = 0 \) when \( x \in D \) and \( \delta_D(x) = \infty \) otherwise; \( \delta_D \) is the indicator function of the set \( D \). For a given \( x \in \mathbb{R}^n \) and a positive number \( r \), we denote the closed ball with center \( x \) and radius \( r \) by \( B(x, r) = \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\} \). Let \( \mu \) be a positive number. Let
\[
\varphi_\mu(x) = E[\tilde{f}(\xi, x, \mu)].
\]

**Lemma 3.5.** Let \( \tilde{f} \) be a smoothing function of \( f \). Then \( \Phi^N_\mu \) and \( \varphi_\mu \) are smoothing functions of \( \hat{\varphi}^N \) and \( \varphi \), respectively. If the sample is iid then for any fixed \( \mu \in [0, \bar{\mu}] \), we have
\[
(3.5) \quad \Phi^N_\mu \srtr{e} \varphi_\mu, \text{ on } D, \text{ a.s.}
\]

**Proof.** By Definition 3.1, it is easy to see that \( \Phi^N_\mu \) and \( \varphi_\mu \) are smoothing functions of \( \hat{\varphi}^N \) and \( \varphi \), respectively.

The proof for (3.5) is based on the convergence of inf-projections. Let
\[
\varphi_\mu(x) = \inf_{B(x, r)} \varphi_\mu + \delta_D, \quad \varphi^N_\mu(x) = \inf_{B(x, r)} \varphi^N + \delta_D.
\]
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 \( \{c^N_{x, r}\} \) are iid [14]. From the Law of Large Number follows
\[
c^N_{x, r} \srtr{e} c_{x, r} \text{ as } N \to \infty \text{ a.s.}
\]
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). \( \square \)

For any locally Lipschitz continuous function \( g : \mathbb{R}^n \to \mathbb{R} \), we can construct a smoothing function \( \tilde{g} : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R} \) satisfying the gradient consistent property (3.6)

\[
\lim_{z \to x; \mu \downarrow 0} \nabla \tilde{g}(z, \mu) \subseteq \partial g(x)
\]

by convolution [23, Theorem 9.67], where \( \partial g \) denotes the Clarke generalized gradient. Moreover, for many locally Lipschitz continuous functions, we can easily construct computable smoothing functions satisfying (3.6). See examples in section 4 and (ii) of Lemma 4.4. In the remainder of this paper, we assume that the smoothing functions \( \Phi^N_{\mu} \) and \( \varphi_{\mu} \) satisfy the gradient consistent property (3.6).

**Lemma 3.6.** Under Assumptions 1 and 2(iii), whatever be the sample \( \{\xi^1, \ldots, \xi^N\} \) that defines the functions \( \hat{\varphi}^N \) and \( \Phi^N_{\mu} \), the collection of functions \( \{\hat{\varphi}^N, N \in \mathbb{N}\} \), as well as the collection \( \{\Phi^N_{\mu}, \mu > 0, N \in \mathbb{N}\} \), are equi-locally Lipschitz continuous on \( D \).

In particular, this implies that when, for all \( N \), the samples are iid, the functions \( \Phi^N_{\mu} \) not only epi-converge almost surely to \( \varphi_{\mu} \) on \( D \), but converge also pointwise almost surely.

**Proof.** The statements about the collections being equi-locally Lipschitz follow directly from Theorem 2.4 and the gradient consistent property (3.6), since they imply that both the collections of functions \( \{f(\xi, \cdot) : D \to \mathbb{R}, \xi \in \Xi\} \) and \( \{\tilde{f}(\xi, \cdot, \mu) : D \to \mathbb{R}, \mu > 0, \xi \in \Xi\} \), that define \( \hat{\varphi}^N \) and \( \Phi^N_{\mu} \), via finite sums, are equi-locally Lipschitz continuous.

The almost sure pointwise convergence then follows immediately from [23, Theorem 7.10] and Lemma 3.5 which imply that under equi-lower semicontinuity of the approximating functions, epi-converges implies pointwise convergence. \( \square \)

**Lemma 3.7.** Under the 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 optimal solutions.

**Proof.** Since for any \( \xi \in \Xi \), \( f(\xi, \cdot) \) is a continuous function on \( D \) and measurable in \( \xi \) for any \( x \in D \), the SAA function \( \hat{\varphi}^N \) and the SSAA function \( \Phi^N_{\mu} \) 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

\[
\min_{x \in D} \hat{\varphi}^N(x) = \min_{y \in D_1} \hat{\varphi}^N(y)
\]

and

\[
\min_{x \in D} \Phi^N_{\mu}(x) = \min_{y \in D_1} \Phi^N_{\mu}(y).
\]

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

\[
y^* \in \arg\min_{y \in D_1} \hat{\varphi}^N(y) \quad \text{and} \quad y^{**} \in \arg\min_{y \in D_1} \Phi^N_{\mu}(y),
\]
respectively. Moreover, from $D_1 \subseteq D$ and (3.7), (3.8), $y^*$ and $y^{**}$ are thus solutions of (3.3) and (3.4), respectively. □

Let $S^*$, $S^N$ and $S^*_\mu$ 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 $S^*_\mu$ 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

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

Since $\varphi$, $\varphi^N$ and $\Phi^N_\mu$ are continuous and $D_1$ is compact, we have

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

for $h = \varphi$, $h = \varphi^N$ or $h = \Phi^N_\mu$.

**Theorem 3.8.** Under Assumptions 1 and 2, if the sample is iid, then the following hold.

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

(ii) Any cluster point of $\{x^N_\mu \in S^N\}$ is an optimal solution of (1.7) a.s.

(iii) $\epsilon(S^N, S^*) \to 0$ a.s., as $N \to \infty$ and $\mu \downarrow 0$.

**Proof.** By the definition of the smoothing functions of $\varphi(x)$, $\lim_{x \to x_\mu} \varphi_\mu(x) = \varphi(\bar{x})$ for any $x, \bar{x} \in D_1$. Moreover, from Lemmas 3.5, 3.6 and

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

we obtain

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

which means $\Phi^N_\mu$ epi-converges to $\varphi$ as $N \to \infty$ and $\mu \downarrow 0$, a.s. Hence by [23, Theorem 7.11], one has

$$\Phi^N_\mu + \delta_{D_1} \sr{e}{\to} \varphi + \delta_{D_1}, \quad \text{a.s.}$$

Moreover, by the continuity and nonnegativity of $\varphi$ 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

$$\limsup_{N \to \infty, \mu \downarrow 0} \arg\min_{x \in D_1} \Phi^N_\mu(x) = \limsup_{N \to \infty, \mu \downarrow 0} \arg\min_{x \in D_1} (\Phi^N_\mu(x) + \delta_{D_1}(x))$$

$$\subset \arg\min_{x \in D_1} (\varphi(x) + \delta_{D_1}(x)) = \arg\min_{x \in D_1} \varphi(x), \quad \text{a.s.}$$

By the compactness of $D_1$, the sequence $\{x^N_\mu\}$ 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

\[
\lim_{N \to \infty} \Phi^N_\mu(x) = \varphi_\mu(x), \quad x \in D_1, \quad \mu \in [0, \bar{\mu}],
\]

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

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

**Theorem 3.9.** Under Assumptions 1 and 2, if (3.9) holds, so do the following.

(i) Any sequence \( \{x^N_\mu\} \subseteq S^N_\mu \) has a cluster point as \( N \to \infty \) and \( \mu \downarrow 0 \).

(ii) Any cluster point of \( \{x^N_\mu\} \) is an optimal solution of (1.7).

(iii) \( e(S^N_\mu, S^*) \to 0 \), as \( N \to \infty \) and \( \mu \downarrow 0 \).

**Proof.** By definition of the smoothing functions associated with \( \varphi(x) \),

\[
\lim_{x \to x; N \to \infty; \mu \downarrow 0} \Phi^N_\mu(x) = \varphi(x),
\]

which means \( \Phi^N_\mu + \delta_{D_1} \) continuously converges to \( \varphi \) 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 3.8. \( \Box \)

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.

Recall [23, Section 8.A] that the subderivative of a function \( g : \mathbb{R}^n \to \mathbb{R} \) at a point \( \bar{x} \) at which \( g(\bar{x}) \) is finite, is the function \( dg(\bar{x}; \cdot) \) defined by

\[
dg(\bar{x}; h) = \lim_{\tau \downarrow 0} \inf_{h' \to h} \Delta_\tau g(x; h') \quad \text{or, equivalently,} \quad dg(\bar{x}; \cdot) = \text{epi-} \lim_{\tau \downarrow 0} \Delta_\tau g(\bar{x}; \cdot)
\]

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

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

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

\[
dg(\bar{x}; h) \geq 0 \quad \text{for all } h \in T_X(\bar{x}),
\]

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

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

We work with this latter inequality since our \( X \), the sets \( D \) and \( D_1 \), are convex. Moreover, the functions \( f(\xi, \cdot) \), cf. Theorem 2.4, and, a fortiori, \( f(\xi, x, \mu) \) 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,

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

Moreover, \( \text{dg}(\cdot, h) \) is usc (upper semicontinuous); in fact, [23, 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 3.5 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 (3.11) for piecewise maxima functions.

In Section 4, we show that \( Q(\xi, \cdot) \) is regular and the exponential smoothing function \([6, 19]\) satisfies (3.11) for piecewise maxima functions.

**Theorem 3.10.** Suppose Assumptions 1 and 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 (3.3) and the SSAA problem (3.4) have stationary points in the compact set \( D_1 \). Let \( \{x^N_\mu\} \subset D_1 \) be a sequence of stationary points of (3.4). If the sample is iid, then any cluster point of \( \{x^N_\mu\} \) is a stationary point of (1.7), a.s.

**Proof.** The existence of stationary points follows 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 \( f, \varphi, \hat{\varphi}^N \) are Clarke regular \([8, \text{Definition 2.3.4, Proposition 2.3.6}] \) in \( D \).

Since \( f \) is globally Lipschitz in \( D \), there are constants \( \bar{t} > 0 \) and \( \beta \) 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 \([28, \text{Section 2}] \), we obtain

\[
E[|df(\xi, x; y - x)|] \leq d\varphi(x; y - x), \quad \forall x, y \in D.
\]

By the continuous differentiability of \( \hat{f}(\xi, x, \mu) \) 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 \( h \in \mathbb{R}^n \), \( d\Phi^N(\cdot; h) = \frac{1}{N} \sum_{i=1}^N df_i(\xi_i, \cdot; h) \) is upper semicontinuous. Hence, we can use the same technique as in the proof of Lemmas 3.5 and 3.6, to show that

\[
d\Phi^N(\cdot; h) \xrightarrow{e.p.} d\varphi(\cdot; h), \quad \text{in } D, \quad \text{a.s.,}
\]

where \( e.p \) stands for epi- and pointwise convergence.
Let \( \hat{x} \) be a cluster point of \( \{x^N_\mu\} \). 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 \( \{x^N_\mu\} \) converges to \( \hat{x} \). Then, one has

\[
0 \leq d\Phi^N_\mu(x^N_\mu; y - x^N_\mu) \\
\leq \sigma \|\hat{x} - x^N_\mu\| + d\Phi^N_\mu(x^N_\mu; h) - d\varphi_\mu(\hat{x}; h) + d\varphi_\mu(\hat{x}; h) - d\varphi(\hat{x}; h) + d\varphi(\hat{x}; h),
\]

where \( \sigma \) is a Lipschitz constant of \( \Phi^N_\mu \) near \( \hat{x} \) for all \( \mu \geq 0 \) and \( N \in \mathbb{N} \); the existence of such \( \sigma \) follows from the global Lipschitz continuity of \( \Phi^N_\mu \) and \( \varphi \).

The third and second terms give

\[
d\varphi_\mu(\hat{x}; h) - d\varphi(\hat{x}; h) \to 0 \text{ as } N \to \infty \text{ and } \mu \downarrow 0, \text{ a.s. by using (3.13)}.
\]

From (3.12) and (3.11), the fifth and fourth terms give

\[
d\varphi_\mu(\hat{x}; h) - d\varphi(\hat{x}; h) \leq E[d\mu(\xi_0; \hat{x}; h) - d\hat{f}(\hat{x}; h)] \leq 0, \quad \text{as } \mu \downarrow 0.
\]

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

**Remark 2.** From the properties of smoothing functions, we can define

\[
\hat{f}(\xi, x, 0) = \lim_{\mu \downarrow 0} \hat{f}(\xi, x, \mu)
\]

at any \( x \in D \) and \( \xi \in \Xi \). Hence, we can consider \( \hat{\varphi}^N(x) = \Phi^N_0(x) = \lim_{\mu \downarrow 0} \Phi^N_\mu(x) \) 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.** 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].

**Remark 4.** In [30], 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 that would certainly lead to substantially exceeding page limitations. This certainly will require a separate treatment that we plan to deal with in a separate paper.

4. **Application and numerical experiments.** In this section, we use three 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 \( p_1 = \{3, 7, 6\}, p_2 = \{3, 1\}, p_3 = \{4, 6\}, p_4 = \{3, 7, 2\}, p_5 = \{3, 5\}, p_6 = \{4, 2\} \).
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 $\Xi \subseteq \mathbb{R}^L$ denote the set of uncertain factors. Let $(b_i)_i > 0$ denote the stochastic travel demand on the $i$th OD pair and $(c_k)_k$ denote the stochastic capacity of link $k$.

For a realization of random vectors $b_\xi \in \mathbb{R}^2$ and $c_\xi \in \mathbb{R}^7$, $\xi \in \Xi$, an assignment of flows to all paths is denoted by the vector $x \in \mathbb{R}^{|K|}$, 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 $i$th OD and $a_{i,j} = 0$ otherwise. The incidence matrices for the network in Figure 4.1 are given respectively as follows.

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

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,

$$(4.1) \quad T_k(\xi, v) = t_{k}^{0} \left(1.0 + 0.15 \left(\frac{v_k}{(c_\xi)_k}\right)^{n_k}\right), \quad k = 1, \ldots, 7$$

where $t_{k}^{0}$ and $n_k$ are given parameters. The path travel cost function is defined by

$$(4.2) \quad F(\xi, x) = \eta_1 \Delta^T T(\xi, \Delta x),$$

where $\eta_1 > 0$ is the time-based operating costs factor. If $n_k = 1, k = 1, \ldots, 7$, then $F(\xi, x) = M_\xi x + q$, where

$M_\xi = 0.15\eta_1 \Delta^T \text{diag} \left(\frac{v_{k}^{0}}{(c_\xi)_k}\right) \Delta$ \quad and \quad $q = \eta_1 t_{1}^{0} \Delta^T e$.  

Fig. 4.1. The 7-links, 6-paths network
Note that \( \text{rank}(\Delta) = 5 \) for any \( \xi \in \Xi \). \( M_{\xi} \in \mathbb{R}^{6 \times 6} \) is a positive semi-definite matrix with \( \text{rank}(M_{\xi}) = 5 \). 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 \( \text{VI}(X_{\xi}, F(\xi, \cdot)) \), seeks an equilibrium path flow \( x_{\xi} \in X_{\xi} \) such that

\[
(y - x_{\xi})^T F(\xi, x_{\xi}) \geq 0, \quad \text{for all } y \in X_{\xi} = \{ x \mid Ax = b_{\xi}, \quad x \geq 0 \},
\]

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, \( \xi \) belongs to a set \( \Xi \) representing future states of knowledge. In general, we cannot find a vector \( \bar{x} \) such that \( f(\xi, x) = 0 \) for all \( \xi \in \Xi \). The ERM-formulation is to find a vector \( \bar{x} \) which minimizes the expected value of \( f(\xi, x) \) over \( \Xi \). 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 \( b \) 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 \( S \subseteq \mathbb{R}^m \) is a meet semi-sublattice under the componentwise ordering of \( \mathbb{R}^m \) if

\[
u, v \in S \implies w = \min(u, v) \in S.
\]

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 \( \text{prob}\{b_{\xi} > 0, \|b_{\xi}\|_\infty \leq \beta\} = 1 \) 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

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

Then,

\[
X_{\xi} \subseteq \{ x \mid 0 \leq x \leq \gamma e \} =: U_0.
\]

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

\[
\|z(\xi, u(\xi, x))\|_\infty \leq \theta = \kappa \max(1, \gamma_0^{-1}) \| A^{-1}_K \|_\infty
\]

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

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

\[
A^{-1}_K b_{\xi} = A^{-1}_K AP x = A^{-1}_K [A_K, A_J] P x = [I, A^{-1}_K A_J] P x.
\]
Since \( [I, A_K^{-1}A_J] \) is a nonnegative matrix and its each column has at least one positive element, \([I, A_K^{-1}A_J]P x \geq 0\). 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
\[
\min(1, \gamma_0)\|x\|_{\infty} \leq B_{i,j_0} x_{j_0} \leq \|([I, A_K^{-1}A_J]P x\|_{\infty} \leq \|A_K^{-1}b_x\|_{\infty} \leq \|A_K^{-1}\|_{\infty}\beta \text{ a.s.}
\]

This implies \(X_\xi \subseteq U_0\) a.s.

Let \(S_{\xi,u} = \{z \mid A^Tz + F(\xi, u) \geq 0\} \) denote the feasible set. For \(w, v \in S_{\xi,u}\), let \(s = \min(w, v)\) be their meet. We consider an arbitrary index \(i \in \{1, \ldots, n\}\). 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,
\[
(A^T s + F(\xi, u))_i = F_i(\xi, u) + \sum_{j \neq k} a_{ji}s_j + a_{ki}s_k
\]
\[
\geq F_i(\xi, u) + \sum_{j \neq k} a_{ji}v_j + a_{ki}v_k
\]
\[
\geq 0.
\]

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

Let \(e \in R^m\) and \(\check{e} \in R^n\) 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
\[
PA^T t = \kappa \max(1, \gamma_0^{-1}) \left( e_{A_J^T, A_K^{-T}e} \right) \geq \kappa \check{e} \geq -PF(\xi, u) \text{ a.s.}
\]

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
\[
\begin{align*}
(4.7) \quad z & \geq -CPF(\xi, u) \geq -LA_K^{-T}e \geq -\max(1, \gamma_0^{-1})\kappa A_K^{-T}e.
\end{align*}
\]

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

which completes the proof. \(\Box\)

In traffic flow problem \([2, 31, 34]\), we often have the following constraints
\[
(4.9) \quad X_\xi = \{x \mid \sum_{j \in \ell_i} x_j = (b_\ell)_i, \quad i = 1, \ldots, m\}
\]
with
\[ \bigcup_{i=1}^{m} I_i = \{1, 2, \ldots, n\}, \quad I_i \cap I_j = \emptyset, \ i \neq j, \]

where \( b_\xi \) 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 (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 \( i \)th 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
\[ (4.10) \quad z_i(\xi, u) = \max \{-F_j(\xi, u), j \in I_i\}, \quad i = 1, \ldots, m. \]
Moreover, if \( F(\xi, x) = M_\xi x + q_\xi \), then \( \varphi \) is a convex function.

Now, we define a smoothing function of (4.11)
\[ f(\xi, x) = u(\xi, x)^T F(\xi, u(\xi, x)) + \sum_{i=1}^{m} b_i(\xi) \max_{j \in I_i} \{-F_j(\xi, u(\xi, x))\}. \]

Consider the following nonsmooth function for a vector \( y \in \mathbb{R}^k \)
\[ p(y) = \max_{1 \leq i \leq k} \{y_i\}. \]
We define a smoothing function of \( p \) as follows [19]: for \( \mu > 0 \),
\[ \tilde{p}(y, \mu) = \mu \ln \left( \sum_{i=1}^{k} e^{y_i/\mu} \right). \]

**Lemma 4.4.** [6] \( \tilde{p} \) is continuously differentiable with respect to \( x \) for any fixed \( \mu > 0 \). Moreover, the following hold.
(i)
\[ 0 \leq \tilde{p}(y, \mu) - p(y) = \mu \ln \left( \sum_{i=1}^{k} e^{y_i/\mu} \right) \leq \mu \ln k. \]
(ii) \( \{ \lim_{z \to y, \mu \downarrow 0} \nabla z \tilde{p}(z, \mu) \} \) is nonempty and bounded. Moreover, \( \tilde{p} \) satisfies the gradient consistent property, that is,
\[ \{ \lim_{y \to \bar{y}, \mu \downarrow 0} \nabla y \tilde{p}(y, \mu) \} \subseteq \partial p(\bar{y}), \]
where \( \partial p \) denotes the Clarke generalized gradient.

**Lemma 4.5.** The directional derivative \( \tilde{p}'_\mu(y; h) \) of \( \tilde{p} \) satisfies
\[ (4.12) \quad \lim_{\mu \downarrow 0} \tilde{p}'_\mu(y; h) \leq p'(y; h), \quad \forall \ y, h \in \mathbb{R}^k. \]
Proof. For any given \( y, h \in \mathbb{R}^k \), let \( K = \{ i \mid y_i = p(y) \} \) 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

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

This completes the proof. \( \square \)

Let

\[(4.13) \quad \tilde{f}(\xi, x, \mu) = u(\xi, x)^T F(\xi, u(\xi, x)) + \mu \sum_{i=1}^m (b_i), \ln \sum_{j \in I_i} e^{-F_j(\xi, u(\xi, x))/\mu}.
\]

**Theorem 4.6.** When \( X_\xi \) is defined by (4.9) and \( \tilde{f} \) is defined by (4.13), the assumptions of Theorem 4.3 hold and \( \varphi_\mu \) and \( \Phi_\mu^N \) are smoothing functions of \( \varphi \) 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 (3.11).

*Proof.* The matrix \( A \) can be split into two submatrices \( A_K \) and \( A_J \), where \( A_K = I \in \mathbb{R}^{m \times m} \) whose \( i \)th column is the first column of \( A_I \) 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, \( \varphi_\mu \) and \( \Phi_\mu^N \) are smoothing functions of \( \varphi \) 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 (3.11) 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 \( I_1 = K = \{ 1, \ldots, k \} \) and thus \( z_1(\xi, u) = \max\{ -F_j(\xi, u), j \in K \} \). For a fixed \( \xi \), let \( g(u) = (-F_1(\xi, u), \ldots, -F_k(\xi, u))^T \) and \( g(u) = p(g(u)) = \max\{ g_1(u), \ldots, g_k(u) \} \). Since \( b_i(\xi) > 0 \), for \( i = 1, \ldots, m \), it is sufficient to show that

\[(4.14) \quad \lim_{\mu \downarrow 0} \tilde{q}_\mu(u; h) \leq q'(u; h), \quad \forall \ u, h \in \mathbb{R}^k.
\]

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 \( \mu > 0 \),

\[
\lim_{\mu \downarrow 0} \tilde{q}_\mu(u; h) = \lim_{\mu \downarrow 0} \nabla \tilde{p}_\mu(g(u)) \cdot g'(u)h = p(g(u); g'(u)h) = q(u; h)
\]

that follows from Lemma 4.5. \( \square \)
4.2. Numerical experiment. In Examples 4.1-4.2, \( X \) is defined by (4.9) and \( \tilde{f} \) is defined by (4.13). The EV-formulation for the two examples is to find an \( x \in X = \{ x \mid Ax = E[b] \} \) such that

\[(4.15) \quad (y - x)^T E[F(\xi, x)] \geq 0, \quad y \in X.\]

We solve the following minimization problem

\[(4.16) \quad \min_{x \in X} g(x) := \max\{(x - y)^T E[F(\xi, x)] \mid y \in X\}\]

and set a minimizer to be \( x_{EV} \).

For the ERM-formulation, we solve the ERM problem (1.7) and set \( x_{ERM} = (I - A^\dagger A)x^* + A^\dagger E[b] \), 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 \( x \),

\[\alpha^*(x) \in \arg\min_{\alpha \in \mathbb{R}} \text{CVaR}(x, \alpha) := \alpha + \frac{1}{1 - \beta} E\{[f(\xi, x) - \alpha]^+ \}.\]

For the GBPR function, we set \( n_a = n_k, k = 1, \ldots, k_v \), where \( k_v \) is the number of links.

**Example 4.1.** This example is the 7-links, 6-paths problem in Figure 4.1. The free travel time \( t_{0k} \) and the mean of the capacity \( E[(c_k)] \) of the network are the same as those used in [31], which are listed in Table 4.1.

**Table 4.1**

<table>
<thead>
<tr>
<th>Link number</th>
<th>( k )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-flow time ( t_{0k} )</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mean ( E[(c_k)] )</td>
<td>15</td>
<td>15</td>
<td>30</td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

For the travel demand vector, we set \( E[b] = [200 \ 220]^T \), where the components follow the order of the OD-pairs 1 \( \rightarrow \) 4 and 1 \( \rightarrow \) 5. The link capacity and the demand vector both have a beta distribution. For the demand vector \( b_\xi \), the lower bound is \( \bar{b} = [150 \ 180]^T \) and the parameters for the beta distribution are \( \alpha = 5, \beta = 1 \). For the link capacity \( c_\xi \), the lower bound is \( \bar{c} = [10 \ 10 \ 20 \ 20 \ 10 \ 10 \ 10]^T \) and the parameters for the beta distribution are \( \alpha = 2, \beta = 2 \).

**Table 4.2**

| Solutions for sampling size \( N = 1000 \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | \( n_a = 2 \)   |                | \( n_a = 4 \)   |                |                |                |
| \( x_1 \)      | 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           |
The ERM-formulation has higher probability than the EV-formulation for each $p$ of the network such as weather, accidents and so on, and we give the three scenarios $x$. Table 4.3 lists robustness and risk criteria for the EV and ERM solutions in Table 4.1 with size $N=1000$, $n_a=2$.

<table>
<thead>
<tr>
<th>$n_a = 2$</th>
<th>$n_a = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon = 4.5E3$</td>
<td>$\varepsilon = 5E5$</td>
</tr>
<tr>
<td>$x_{EV}$</td>
<td>$x_{ERM}$</td>
</tr>
<tr>
<td>prob{f($\xi$, $x$) $\leq \varepsilon$}</td>
<td>0.58</td>
</tr>
<tr>
<td>$E[|x - x^*_t|]$</td>
<td>46.94</td>
</tr>
<tr>
<td>$E[|u($,$x$, $x^*_t|)]$</td>
<td>47.03</td>
</tr>
<tr>
<td>$E[f($,$x$, $x^*_t|)]$</td>
<td>4.316E3</td>
</tr>
<tr>
<td>$\alpha^*$($x$)</td>
<td>7.395E3</td>
</tr>
<tr>
<td>CVaR($x$, $\alpha^*$($x$))</td>
<td>8.691E3</td>
</tr>
</tbody>
</table>

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 $n_a$. Table 4.3 lists robustness and risk criteria for the EV and ERM solutions in Table 4.1; $x^*_t$ means a solution of the variational inequalities for each fixed $\xi \in \Xi$.

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

**Example 4.2.** This example uses the Nguyen and Dupuis network, which contains 13 nodes, 19 links, 25 paths and 4 OD-pairs. See Figure 4.3.

We use the free-flow travel time $t^0_k$ as that used by Yin [32], and the mean of the demand vector $E[b_k]$ of the network is $E[b_k] = [400, 800, 600, 450]^T$.

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

- $c_\xi = 100 \times [8, 3.2, 3.2, 8, 4, 3.2, 8, 2, 2, 2, 4, 4, 8, 6, 4, 4, 1.6, 3.2, 8]^T$;
- $c_\xi = 100 \times [10, 4.4, 1.4, 10, 3, 4.4, 10, 2, 2, 4, 7, 7, 7, 4, 3.5, 2.2, 4.4, 7]^T$;
- $c_\xi = 100 \times [4, 4, 2, 4, 4, 4, 4, 4, 4, 4, 4, 2, 4, 2, 8, 8, 1, 2, 4, 2]^T$

corresponding to probabilities $p_1 = \frac{1}{3}$, $p_2 = \frac{1}{4}$ and $p_3 = \frac{1}{4}$, respectively.

The demand vector follows the beta distribution $b_\xi \sim \beta + \tilde{b} \ast \text{beta}(\alpha, \beta)$ with
the lower bound $\bar{b} = [300, 700, 500, 350]^T$ 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 $(b_\xi, c_\xi)$ for $i = 1, 2, \cdots, N$, where $c_\xi$ is sampled from the three possibilities with given probability and $b_\xi$ is sampled from the beta distribution.

**Example 4.2.** Criteria for $\beta = 0.9$ in CVaR, $n_a = 2$, $\varepsilon = 3.3E3$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\mu$</th>
<th>$\text{prob}(f(\xi, x) \leq \varepsilon)$</th>
<th>$E[f(\xi, x)]$</th>
<th>$\alpha^*(x)$</th>
<th>$\text{CVaR}(x, \alpha^*(x))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^3$</td>
<td>$10^{-4}$</td>
<td>0.508</td>
<td>3.498E3</td>
<td>7.935E3</td>
<td>8.154E3</td>
</tr>
<tr>
<td>$5 \times 10^3$</td>
<td>$10^{-5}$</td>
<td>0.510</td>
<td>3.498E3</td>
<td>7.918E3</td>
<td>8.121E3</td>
</tr>
<tr>
<td>$10^4$</td>
<td>$10^{-6}$</td>
<td>0.509</td>
<td>3.505E3</td>
<td>7.978E3</td>
<td>8.168E3</td>
</tr>
</tbody>
</table>

**Example 4.3.** We consider the Sioux Falls network as shown in Figure 4.4 (left), which consists of 24 nodes, 76 links, 528 OD-pairs. 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 [34] 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 (1.7) are displayed in Figure 4.4 (right). Here the link flow is displayed on each link with the unit 1.0e3, and the width of each link is proportional to the link flow. By the property of $x_{\text{ERM}}$, we known that the ERM flow patterns satisfy the average of travel demand as $Ax_{\text{ERM}} = E[b_\xi]$. Moreover, the ERM flow patterns satisfy the stochastic travel demand on all OD pairs with high
probability:

\[ 0.848 \geq \text{prob}\{ (Ax_{ERM} - b_{xi})_i \geq 0 \} \geq 0.780, \ i = 1, \ldots , 528. \]

Remark 5. The three 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.

Acknowledgement We would like to thank the associate editor and anonymous referees for valuable comments.

REFERENCES