Confidence regions for stochastic variational inequalities

Shu Lu
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 355 Hanes Building, CB#3260, Chapel Hill, NC 27599-3260
email: shulu@email.unc.edu http://www.unc.edu/~shulu/

Amarjit Budhiraja
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 357 Hanes Building, CB#3260, Chapel Hill, NC 27599-3260
email: budhiraj@email.unc.edu http://www.unc.edu/~budhiraj/

The sample average approximation (SAA) method is a basic approach for solving stochastic variational inequalities (SVI). It is well known that under appropriate conditions the SAA solutions provide asymptotically consistent point estimators for the true solution to an SVI. It is of fundamental interest to use such point estimators along with suitable central limit results to develop confidence regions of prescribed level of significance for the true solution. However, standard procedures are not applicable since the central limit theorem that governs the asymptotic behavior of SAA solutions involves a discontinuous function evaluated at the true solution of the SVI. This paper overcomes such a difficulty by exploiting the precise geometric structure of the variational inequalities and by appealing to certain large deviations probability estimates, and proposes a method to build asymptotically exact confidence regions for the true solution that are computable from the SAA solutions. We justify this method theoretically by establishing a precise limit theorem, apply it to complementarity problems, and test it with a linear complementarity problem.

Key words: variational inequalities; stochastic variational inequalities; central limit theorems in Banach spaces; large deviations; confidence regions; statistical inference

MSC2000 Subject Classification: Primary: 90C33; Secondary: 90C15

OR/MS subject classification: Primary: Programming: complementarity, Programming: stochastic; Secondary: Programming: nonlinear; theory

1. Introduction. This paper studies the problem of constructing confidence regions for the solution to a stochastic variational inequality defined over a polyhedral convex set. Let \((\Omega, \mathcal{F}, P)\) be a probability space, and \(\xi\) be a random vector that is defined on \(\Omega\) and supported on a closed subset \(\Xi\) of \(\mathbb{R}^d\). Let \(O\) be an open subset of \(\mathbb{R}^n\), and \(F\) be a measurable function from \(O \times \Xi\) to \(\mathbb{R}^n\), such that for each \(x \in O\) the expectation \(E[\|F(x, \xi)\|] < \infty\). Let \(S\) be a polyhedral convex set in \(\mathbb{R}^n\) defined by

\[
S = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \{x \in \mathbb{R}^n \mid \langle a_i, x \rangle \leq b_i, i = 1, \cdots, m\},
\]

where \(A\) is an \(m \times n\) matrix whose rows are given by \(a_1^T, \cdots, a_m^T\), and \(b = (b_1, \cdots, b_m)\) is a column vector in \(\mathbb{R}^m\). Suppose that \(S \cap O \neq \emptyset\), and let \(f_0 : O \to \mathbb{R}^n\) be defined as

\[
f_0(x) = E[F(x, \xi)].
\]

The stochastic variational inequality (SVI) problem is to find:

\[
x \in S \cap O, \text{ such that } 0 \in f_0(x) + N_S(x),
\]

where \(N_S(x) \subset \mathbb{R}^n\) denotes the normal cone to \(S\) at \(x\):

\[
N_S(x) = \{v \in \mathbb{R}^n \mid \langle v, s - x \rangle \leq 0 \text{ for each } s \in S\}.
\]

The normal cone \(N_S(x)\) is the polar cone of the tangent cone to \(S\) at \(x\), which is given by

\[
T_S(x) = \{u \in \mathbb{R}^n \mid \langle a_i, u \rangle \leq 0, i \in I(x)\},
\]

where \(I(x)\) denotes the active index set at \(x\), i.e., the set of all indices \(i \in \{1, \cdots, m\}\) such that \(\langle a_i, x \rangle = b_i\).

Indeed, the function \(f_0\) defined in (2) is deterministic, and the problem (3) is essentially a deterministic variational inequality. However, evaluating \(f_0(x)\) for each given value of \(x\) requires finding the expected value of a random vector, which in most problems of interest does not have a closed form expression and in general requires a numerical approximation. Such approximations are usually provided by sampling. Depending on how sampling is incorporated with the algorithm, solution methods for SVIs can be classified into two basic categories. The first category consists of the stochastic approximation (SA) methods, which perform sampling in an 'interior' manner, by applying an algorithm for deterministic variational
inaccuracies and resorting to sampling whenever the algorithm requires values or gradients of \( f_0 \) at given points. The second category corresponds to the sample average approximation (SAA) methods, which sample in an ‘exterior’ manner. These methods replace \( f_0 \) in (3) by a sample average function to obtain the SAA problem, and then use a solution to the SAA problem as an estimate of a solution to the true problem. In this paper, we will focus on solutions obtained by SAA methods, and will use these solutions to develop asymptotically exact confidence regions for the true solutions.

To formally define the SAA problem, let \( \xi_1, \cdots, \xi_N \) be independent and identically distributed (i.i.d.) random variables with distribution same as that of \( \xi \). Define the sample average function \( f_N : O \times \Omega \rightarrow \mathbb{R}^n \) by

\[
f_N(x, \omega) = N^{-1} \sum_{i=1}^{N} F(x, \xi^i(\omega)).
\]

The SAA problem is then to find \( x \in S \cap O \) such that

\[
0 \in f_N(x, \omega) + N_S(x).
\]

For brevity, we will write \( f_N(x, \omega) \) as \( f_N(x) \) when clear from the context. Also, we refer to a solution to (5) as an SAA solution, and a solution to (3) as a true solution.

SAA methods are known to be consistent. That is, under certain regularity conditions, SAA solutions will almost surely converge to a true solution as the sample size \( N \) goes to \( \infty \), see Gürkan, Özge and Robinson [4], King and Rockafellar [5], and Shapiro, Dentcheva and Ruszczynski [18, Section 5.2.1]. Moreover, [5, Theorem 2.7] and [18, Section 5.2.2] obtained the asymptotic distribution of SAA solutions. These papers showed that the difference between the SAA solution and the true solution, normalized by \( N^{-1/2} \), weakly converges to a random vector, which is the image of a normally distributed random vector under a certain function. From a different viewpoint, Xu [20] showed that the probability for the distance between the SAA solution and the set of true solutions to exceed any given number \( \epsilon \) is no more than \( ce^{-N\beta} \), where \( c \) and \( \beta \) are parameters depending on \( \epsilon \). Namely, SAA solutions converge to the true solution in probability at an exponential rate.

Above results say that SAA solutions provide ‘good’ point estimators for the true solutions, for large \( N \). It is of fundamental interest to use such point estimators along with the associated central limit theory to develop confidence regions of prescribed level of significance for the true solution. One can obtain an expression for confidence regions of the true solution, based on the asymptotic distribution of SAA solutions (see Demir [2]). However, such an expression is not directly usable for specifying confidence regions, because it contains a function that depends discontinuously on the true solution. Due to such discontinuity, it is problematic to replace the true solution in that expression by the SAA solution, as such a replacement may result in a region with probabilities quite different from the desired confidence level.

To overcome this difficulty, we design a sequence of functions that depends on the \( N \)-sample SAA solution. Using certain large deviations probability estimates (see Theorem 4.2 and equation (24)), these functions evaluated at the SAA solutions are shown to converge, as \( N \rightarrow \infty \), to the above discontinuous function evaluated at the true solution (see Corollary 5.1). This enables us to build confidence regions of the true solution that are computable from the SAA solutions. Development of this method is based on a close examination of the geometric structure of variational inequalities and the limiting behavior of SAA solutions. We justify this method theoretically by establishing a precise limit theorem (Theorem 3.1), apply it to complementarity problems, and test it with a linear complementarity problem. The paper is organized as follows. Section 2 below introduces some background about variational inequalities. Next, Section 3 presents the main result of this paper. Following that, Section 4 summarizes probability results used in this paper, and Section 5 develops and justifies the main method. Section 6 then specializes the method to complementarity problems and implements it in a numerical example. The appendix contains proofs of two theorems.

The method developed in this paper deals with situations in which the true problem (3) has a locally unique solution. If the true solution is not locally unique, then the current method will not work and new ideas will be needed.

Throughout this paper, we use \( \text{ri} \ C \) to denote the relative interior of a convex set \( C \). For a convex and closed set \( C \subset \mathbb{R}^n \) and a point \( z \in \mathbb{R}^n \), \( \Pi_C(z) \) denotes the Euclidean projection of \( z \) onto \( C \), namely the
point in \( C \) nearest to \( z \) in Euclidean norm. We use \( \| \cdot \| \) to denote the norm of an element in a normed space; unless explicitly stated otherwise, it can be any norm, as long as the same norm is used in all related contexts. We use \( N(0, \Sigma) \) to denote a Normal random vector with covariance matrix \( \Sigma \). Weak convergence of random variables \( Y_n \) to \( Y \) will be denoted as \( Y_n \Rightarrow Y \).

2. Preliminaries. This section contains preliminary results that form the foundation for subsequent developments.

2.1 The normal map and normal manifold. Throughout this paper, we will use the normal map formulation of a variational inequality. This subsection introduces this formulation and related concepts.

Given the function \( f_0 \) and the set \( S \) as defined in the introduction, the normal map induced by \( f_0 \) and \( S \) is a function \((f_0)_S : \Pi_S^{-1}(O) \rightarrow \mathbb{R}^n\), defined as

\[
(f_0)_S(z) = f_0(\Pi_S(z)) + (z - \Pi_S(z))
\]

for each \( z \in \Pi_S^{-1}(O) \), where we recall that \( \Pi_S(z) \) denotes the Euclidean projection of a point \( z \) on \( S \), and \( \Pi_S^{-1}(O) \) is the set of points \( z \in \mathbb{R}^n \) such that \( \Pi_S(z) \in O \).

Suppose for now that \( x \) is a solution of (3), and define \( z = x - f_0(x) \). It follows that \( \Pi_S(z) = x \) and that

\[
z \in \Pi_S^{-1}(O), \quad (f_0)_S(z) = 0.
\]

Conversely, if \( z \) satisfies (7), then \( x = \Pi_S(z) \) satisfies \( x - f_0(x) = z \) and solves (3). Thus, equation (7) is an equivalent formulation for (3).

In the following, we introduce the concept of the normal manifold. For detailed discussion on this, see [8, 12, 16].

The set \( S \) is polyhedral convex by hypothesis, so it has finitely many faces. Let \( \mathcal{F} \) be the collection of all of its nonempty faces. On the relative interior of each nonempty face \( F \), the normal cone to \( S \) is a constant cone, which we denote by \( N_S(\text{ri }F) \). For each \( F \in \mathcal{F} \), define

\[
C_S(F) = F + N_S(\text{ri }F).
\]

The sets \( C_S(F) \) satisfy the following properties.

(i) For each \( F \in \mathcal{F} \), the set \( C_S(F) \) is a polyhedral convex set of full dimension (i.e., of dimension \( n \)).

(ii) For each \( F_1 \in \mathcal{F} \) and \( F_2 \in \mathcal{F} \), the sets \( C_S(F_1) \) and \( C_S(F_2) \) intersect in a common face (possibly empty).

(iii) \( \bigcup_{F \in \mathcal{F}} C_S(F) = \mathbb{R}^n \).

The collection of all of these sets \( C_S(F) \) is called the normal manifold of \( S \), and each \( C_S(F) \) is called an \( n \)-cell in this normal manifold [12]. (The symbol \( n \) here refers to the dimension of these sets.) A \( k \)-dimensional face of an \( n \)-cell is called a \( k \)-cell in the normal manifold. Any \( k \)-cell, \( k = 0, \cdots, n \), is called a cell in the normal manifold.

On each \( n \)-cell \( C_S(F) \), the Euclidean projector \( \Pi_S \) coincides with the Euclidean projector onto the affine hull of \( F \) (the affine hull of \( F \) is the smallest affine set containing \( F \) ). The latter projector is an affine function. Consequently, \( \Pi_S \) is a piecewise affine function on \( \mathbb{R}^n \).

The following example illustrates above concepts with a set in \( \mathbb{R}^2 \).

Example 2.1 Let \( n = 2 \), \( m = 3 \), and \( S = \{ x \in \mathbb{R}^2 \mid x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0 \} \). The set \( S \) has 7 nonempty faces, including 3 vertices, 3 edges and itself. Its normal manifold contains seven 2-cells. The 2-cell corresponding to the edge joining \((0, 0)\) and \((0, 1)\) is the set \( \{ x \in \mathbb{R}^2 \mid x_1 \leq 0, 0 \leq x_2 \leq 1 \} \). On this cell, \( \Pi_S \) coincides with the Euclidean projector onto the \( x_2 \) axis.

2.2 Sensitivity of solutions to variational inequalities. Sensitivity analysis techniques for variational inequalities play a key role in understanding the behavior of SAA solutions. Theorem 2.1 below deals with sensitivity of solutions of a parametric variational inequality. Before stating this theorem, we
provide definitions of some related concepts. For more details and background we refer the reader to [10, 11].

Let $X$, $Y$ and $Z$ be normed linear spaces, and let $U$ and $V$ be open subsets of $X$ and $Y$ respectively. The injectivity modulus of a function $g : U \to Z$ on $U$ is defined to be

$$\inf \left\{ \frac{\|g(x_1) - g(x_2)\|}{\|x_1 - x_2\|} : x_1 \neq x_2, x_1, x_2 \in U \right\}.$$ 

We say that $g$ is B-differentiable at a point $x_0 \in U$ if there is a positively homogeneous function $dg(x_0) : X \to Z$, such that

$$g(x_0 + v) = g(x_0) + dg(x_0)v + o(v).$$

Recall that a function $dg(x_0)$ is positively homogeneous, if $dg(x_0)(\lambda v) = \lambda dg(x_0)(v)$ for each nonnegative real number $\lambda$ and each $v \in X$. If $dg(x_0)$ is a bounded linear function, then $g$ is Fréchet differentiable at $x_0$ and $dg(x_0)$ is the Fréchet derivative of $g$ at $x_0$. A function $h : U \times V \to Z$ is partially B-differentiable in $x$ at $(x_0,y_0) \in U \times V$, if the function $h(\cdot,y_0)$ is B-differentiable at $x_0$. The partial B-derivative is denoted by $d_x h(x_0,y_0)$. We say the partial B-derivative $d_x h(x_0,y_0)$ is strong, if for each $\epsilon > 0$ there exist neighborhoods $U'$ of $x_0$ in $X$ and $V'$ of $y_0$ in $Y$ such that

$$\|h(x,y) - h(x',y) - d_x h(x_0,y_0)(x-x')\| \leq \epsilon \|x-x'\|,$$

whenever $x$ and $x'$ belong to $U'$ and $y$ belongs to $V'$. A partial Fréchet derivative is strong if it satisfies the same condition as above.

Theorem 2.1 below is adapted from Theorems 3 and 4 of [13]. Sets $O$ and $S$ in this theorem are as defined at the beginning of this paper, and the normal map $L_K$ induced by a function $L$ and a set $K$ is defined in the same way as $(f_0)_S$ is in (6), with $L$ and $K$ in place of $f_0$ and $S$ respectively. The neighborhoods $X_0$, $Y$ and $Z$ constructed in this theorem may depend on $\lambda$, with larger values of $\lambda$ allowing for larger sizes of these neighborhoods.

**Theorem 2.1** Let $\Theta$ be an open subset of a normed linear space $P$, and $h$ be a function from $O \times \Theta$ to $\mathbb{R}^n$. Let $y_0 \in \Theta$ and $z_0 \in \mathbb{R}^n$, and define $x_0 = \Pi_S(z_0)$. Suppose that $x_0 \in O$ and that $h(\cdot,y_0)_S(z_0) = 0$, where $h(\cdot,y)_S$ denotes the normal map induced by $h(\cdot,y_0)$ and $S$. Assume that:

(i) For some positive number $\theta$ and each $x \in O$, $h(x, \cdot)$ is Lipschitz on $\Theta$ with modulus $\theta$.

(ii) $h$ has a strong partial Fréchet derivative in $x$ at $(x_0,y_0)$, denoted by $L = d_x h(x_0,y_0)$.

(iii) The normal map $L_K$, induced by $L$ and the critical cone $K = T_S(x_0) \cap \{z_0 - x_0\}^\perp$, is a homeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Then the normal map $L_K$ has a positive injectivity modulus $\delta$ on $\mathbb{R}^n$. Moreover, for each $\lambda > \delta^{-1} \theta$ there exist neighborhoods $X_0$ of $x_0$ in $O$, $Y$ of $y_0$ in $\Theta$ and $Z$ of $z_0$ in $\mathbb{R}^n$, and a function $z : Y \to \mathbb{R}^n$, such that:

(i) $z(y_0) = z_0$.

(ii) For each $y \in Y$, $z(y)$ is the unique point in $Z$ satisfying $h(\cdot,y)_S(z(y)) = 0$, and $x(y) = \Pi_S(z(y))$ is the unique point in $S \cap X_0$ satisfying $0 \in h(x(y),y) + N_S(x(y))$.

(iii) $z$ is Lipschitz on $Y$ with modulus $\lambda$.

Moreover, if $h$ has a partial B-derivative $d_y h(x_0,y_0)$ in $y$ at $(x_0,y_0)$, then the functions $z(y)$ and $x(y)$ are B-differentiable at $y_0$ with

$$dz(y_0) = (L_K)^{-1} \circ [-d_y h(x_0,y_0)]$$

and

$$dx(y_0) = \Pi_K \circ dz(y_0).$$

**3. Main result.** In this section we present the main result of this work. We begin by introducing the assumptions we make. In the rest of this paper, let $X$ be a nonempty compact subset of $O$. We use $C^1(X, \mathbb{R}^n)$ to denote the Banach space of continuously differentiable mappings $f : X \to \mathbb{R}^n$, equipped with the norm

$$\|f\|_{1,X} = \sup_{x \in X} \|f(x)\| + \sup_{x \in X} \|df(x)\|.$$  

(9)
Our first set of assumptions is Assumption 3.1 below. It is used to guarantee that (see Theorem 4.1) the function \( x \mapsto f_0(x) = E[F(x, \xi)] \) belongs to \( C^1(X, \mathbb{R}^n) \), that the function \( x \mapsto f_N(x, \omega) \) defined in (4) belongs to \( C^1(X, \mathbb{R}^n) \) for almost every \( \omega \), and that \( f_N \to f_0 \) almost surely in \( C^1(X, \mathbb{R}^n) \).

**Assumption 3.1**

(a) \( E \|F(x, \xi)\|^2 < \infty \) for all \( x \in O \).
(b) The map \( x \mapsto F(x, \xi(\omega)) \) is continuously differentiable on \( O \) for a.e. \( \omega \in \Omega \), and \( E \|dF_x(x, \xi)\|^2 < \infty \) for all \( x \in O \).
(c) There exists a square integrable random variable \( C \) such that

\[
\|F(x, \xi(\omega)) - F(x', \xi(\omega))\| + \|dF(x, \xi(\omega)) - dF(x', \xi(\omega))\| \leq C(\omega)\|x - x'\|
\]

for all \( x, x' \in O \) and a.e. \( \omega \in \Omega \).

The next assumption is crucial for establishing the solution uniqueness and stability needed for our development. There is a precise characterization, called the coherent orientation, for the normal map \( L_K \) under consideration to be a global homeomorphism; see [9, 12, 16]. A sufficient condition for this to hold is the restriction of \( L \) on the linear span of \( K \) being positive definite. For example, if \( f_0 \) is strongly monotone then \( L_K \) is always a global homeomorphism.

**Assumption 3.2** Suppose that \( x_0 \) solves the variational inequality (3) and that \( x_0 \) belongs to the interior of \( X \). Let \( z_0 = x_0 - f_0(x_0) \), \( L = df_0(x_0) \), \( K = T_S(x_0) \cap \{z_0 - x_0\}^\perp \), and assume that the normal map \( L_K \) induced by \( L \) and \( K \) is a homeomorphism from \( \mathbb{R}^n \) to \( \mathbb{R}^n \).

Lemma 3.1 below is an important consequence of Assumptions 3.1 and 3.2. It shows that the variational inequality (3) has a locally unique solution that is stable with respect to perturbations on \( f_0 \). Its proof is based on Theorem 2.1 and is given in Appendix A.

**Lemma 3.1** Under Assumptions 3.1 and 3.2 the normal map \( L_K \) has a positive injectivity modulus \( \delta \) on \( \mathbb{R}^n \). Moreover, for each \( \lambda > \delta^{-1} \) there exist neighborhoods \( X_0 \) of \( x_0 \) in \( X \), \( Z \) of \( z_0 \) in \( \mathbb{R}^n \) and \( \Gamma \) of \( f_0 \) in \( C^1(X, \mathbb{R}^n) \), and a function \( z : \Gamma \to \mathbb{R}^n \), such that:

(i) \( z(f_0) = z_0 \).
(ii) For each \( f \in \Gamma \), \( z(f) \) is the unique point in \( Z \) satisfying \( f(z(f)) = 0 \), and \( x(f) = \Pi_S(z(f)) \),\]

(iii) \( z \) is Lipschitz on \( \Gamma \) with modulus \( \lambda \).

Finally, the functions \( z \) and \( x \) are \( B \)-differentiable at \( f_0 \) with

\[
dz(f_0)(g) = (L_K)^{-1} \circ [-g(x_0)]
\]

and

\[
dx(f_0) = \Pi_K \circ dz(f_0).
\]

We will make the following non-degeneracy assumption.

**Assumption 3.3** Let \( \Sigma_0 \) denote the covariance matrix of \( F(x_0, \xi) \). Suppose that the determinant of \( \Sigma_0 \) is strictly positive.

Under above assumptions, \( f_N \) converges to \( f_0 \) almost surely in \( C^1(X, \mathbb{R}^n) \) as \( N \to \infty \), so \( f_N \) belongs to \( \Gamma \) for \( N \) large enough almost surely. It follows that \( z_N = z(f_N) \) is almost surely well defined for sufficiently large \( N \). Using Assumptions 3.1–3.3 it can be shown (see Theorem 5.1) that

\[
\sqrt{N} \Sigma_0^{-1/2} L_K(z_N - z_0) \to N(0, I_n).
\]

It then follows that, for large \( N \),

\[
N [L_K(z_N - z_0)]^T \Sigma_0^{-1} [L_K(z_N - z_0)] \tag{11}
\]

approximately follows the \( \chi^2 \) distribution with \( n \) degrees of freedom, and consequently the set

\[
\{ z \in \mathbb{R}^n \mid N [L_K(z_N - z)]^T \Sigma_0^{-1} [L_K(z_N - z)] \leq \chi^2_n(\alpha) \} \tag{12}
\]
defines an approximate $(1 - \alpha)100\%$ confidence region for $z_0$, where $\chi^2_n(\alpha)$ is defined to be the number that satisfies $P(U > \chi^2_n(\alpha)) = \alpha$ for a $\chi^2$ random variable $U$ with $n$ degrees of freedom.

However, the expression in (12) is not directly computable. In order for it to be useful, we need to find a good approximation for the matrix $\Sigma_0$, and a good approximation for $L_K$. It is natural to approximate $\Sigma_0$ by the sample covariance matrix of $\{F(x_N, \xi)\}^N_{i=1}$. Also, it is well known that the map $L_K$ is the same as the $B$-derivative of the normal map $(f_0)_S$, denoted as $d(f_0)_S$, evaluated at $z_0$. It is thus tempting to use $d(f_0)_S(z_N)$ as an estimate for $L_K$. However this is problematic since the function $d(f_0)_S(\cdot)$ is not continuous. The main objective of this paper is to demonstrate how one can overcome such a difficulty and develop a method for computing asymptotically true confidence regions for $z_0$. Towards that we now introduce an additional assumption.

**ASSUMPTION 3.4** (a) For each $t \in \mathbb{R}^n$ and $x \in X$, let

$$M_x(t) = E\left[\exp\left\{\langle t, F(x, \xi) - f_0(x) \rangle \right\}\right]$$

be the moment generating function of the random variable $F(x, \xi) - f_0(x)$. Assume

(i) There exists $\zeta > 0$ such that $M_x(t) \leq \exp\{\zeta^2\|t\|^2/2\}$ for every $x \in X$ and every $t \in \mathbb{R}^n$.

(ii) There exists a nonnegative random variable $\kappa$ such that

$$\|F(x, \xi(\omega)) - F(x', \xi(\omega))\| \leq \kappa(\omega)\|x - x'\|$$

for all $x, x' \in O$ and almost every $\omega \in \Omega$.

(iii) The moment generating function of $\kappa$ is finite valued in a neighborhood of zero.

(b) For each $T \in \mathbb{R}^{n \times n}$ and $x \in X$, let

$$M_x(T) = E\left[\exp\left\{\langle T, d_x F(x, \xi) - df_0(x) \rangle \right\}\right]$$

be the moment generating function of the random variable $d_x F(x, \xi) - df_0(x)$. Assume

(i) There exists $\zeta > 0$ such that $M_x(T) \leq \exp\{\zeta^2\|T\|^2/2\}$ for every $x \in X$ and every $T \in \mathbb{R}^{n \times n}$.

(ii) There exists a nonnegative random variable $\nu$ such that

$$\|d_x F(x, \xi(\omega)) - d_x F(x', \xi(\omega))\| \leq \nu(\omega)\|x - x'\|$$

for all $x, x' \in O$ and almost every $\omega \in \Omega$.

(iii) The moment generating function of $\nu$ is finite valued in a neighborhood of zero.

Assumption 3.4(a)(i) is satisfied, for example, if one of the following conditions hold.

(i) $\sup_{x \in X, \xi \in \Xi} \|F(x, \xi)\| < \infty$. That is, $F(x, \xi)$ is a bounded random variable uniformly in $x$.

(ii) $E[F_i(x, \xi) - (f_0(x))]/k = 0$ for each odd integer $k$ and each $i = 1, \ldots, n$, and there exists a normal random variable $Y$ with mean zero such that $\sup_{x \in X} \|F(x, \xi(\omega)) - f_0(x)\|^D \leq |Y|$, where we say $U \leq V$ if $V$ stochastically dominates $U$.

For a more general discussion on situations under which Assumption 3.4(a)(i) is satisfied, see [6, Chapter 5].

The following is the main result of this work. The proof is given in Section 5.3. Let $\Sigma_N$ be the sample covariance matrix for $\{F(x_N, \xi)\}^N_{i=1}$.

**THEOREM 3.1** Suppose that Assumptions 3.1 - 3.4 hold. Then, for every $N \in \mathbb{N}$, there is a function $\Phi_N : \mathbb{R}^n \rightarrow \mathbb{R}^n$, such that, writing $\Phi_N(z_N(\omega), \cdot, \omega)$ as $\Phi_N(z_N(\cdot))$, $\sqrt{N}\sum_{n}^{-1/2}\Phi_N(z_N) - z_0 \Rightarrow N(0, I_n)$.

The function $\Phi_N(z_N)(\cdot)$ can be explicitly evaluated given $z_N(\omega)$ and $df_N(\omega)$, and its precise definition is given in (44). As a result of this theorem, for each $\alpha \in (0, 1)$ the set

$$\{z \in \mathbb{R}^n | N[\Phi_N(z_N)(z_N - z)]^T \sum_{n}^{-1}[\Phi_N(z_N)(z_N - z)] \leq \chi^2_n(\alpha)\}$$

(14) defines an asymptotically exact $(1 - \alpha)100\%$ confidence region for $z_0$. 

\[ \text{Page dimensions: 612.0x792.0} \]
4. Probability background. This section summarizes some basic probability results that will be used in this paper. The first result is quite standard and we refer the reader to Theorems 7.44, 7.48 and 7.52 of [18] for its proof.

**Theorem 4.1** Let \( G : O \times \Xi \to \mathbb{R}^k \) be a measurable map such that \( G(x, \xi) \) is integrable for every \( x \in O \). Define \( g(x) = E[G(x, \xi)] \) for \( x \in O \).

(a) Let \( V \) be a nonempty compact set of \( O \). Suppose that \( x \to G(x, \xi(\omega)) \) is continuous on \( V \) for a.e. \( \omega \) and \( \sup_{x \in V} \|G(x, \xi(\omega))\| \leq \Psi_1(\omega) \), where \( \Psi_1 \) is an integrable random variable. Then \( g \) is continuous on \( V \). Let

\[
g_N(x, \omega) = \frac{1}{N} \sum_{i=1}^{N} G(x, \xi_i(\omega)), \quad x \in V, \ \omega \in \Omega.
\]

Then for a.e. \( \omega \), \( g_N(\cdot, \omega) \) converges to \( g(\cdot) \) uniformly on \( V \).

(b) Suppose that, for some integrable random variable \( \Psi_2 \),

\[
\|G(x, \xi(\omega)) - G(x', \xi(\omega))\| \leq \Psi_2(\omega)\|x - x'\|, \quad \text{for all } x, x' \in O, \text{ and a.e. } \omega \in \Omega.
\]

Then \( g \) is Lipschitz on \( O \) with modulus bounded by \( E\Psi_2 \).

(c) Suppose that \( G(\cdot, \xi(\omega)) \), in addition to satisfying the property in (b), is continuously differentiable on \( O \) for almost every \( \omega \in \Omega \). Then \( g \) is continuously differentiable on \( O \), with

\[
dg(x) = E[d_xG(x, \xi)].
\]

Furthermore, for every nonempty compact subset \( V \) of \( O \), \( g_N \) converges to \( g \) a.s. in \( C^1(V, \mathbb{R}^k) \).

We will apply parts of the above theorem with \( G \) replaced with \( F + d_xF \), where \( F \) is defined at the beginning of this paper. In particular the above theorem provides conditions that ensure almost sure uniform convergence of \( f_N \). The theorem below is about the rate of such convergence.

**Theorem 4.2** Suppose that \( F \) satisfies Assumption 3.1.

(a) Suppose that \( F \) also satisfies Assumption 3.4 (a). Then there exist positive real numbers \( \beta, \mu, M \) and \( \sigma \), such that the following holds for each \( \epsilon > 0 \) and each \( N \):

\[
\begin{align*}
\text{Prob}\left\{ \sup_{x \in X} \|f_N(x) - f_0(x)\| \geq \epsilon \right\} \leq \beta \exp\left\{-N\mu\right\} + \frac{M}{\epsilon^n} \exp\left\{-\frac{N\epsilon^2}{\sigma}\right\}.
\end{align*}
\]

(b) Suppose additionally that \( F \) satisfies Assumption 3.4 (b). Then there exist positive real numbers \( \beta_1, \mu_1, M_1 \) and \( \sigma_1 \), such that the following holds for each \( \epsilon > 0 \) and each \( N \):

\[
\begin{align*}
\text{Prob}\left\{ \|f_N - f_0\|_{1, X} \geq \epsilon \right\} \leq \beta_1 \exp\left\{-N\mu_1\right\} + \frac{M_1}{\epsilon^n} \exp\left\{-\frac{N\epsilon^2}{\sigma_1}\right\}.
\end{align*}
\]

**Proof.** (a) Applying [18, Theorem 7.67] (see also [19, Theorem 5.1]) to \( F_i(x, \xi) \) for each \( i = 1, \ldots, n \), we can obtain constants \( \mu > 0 \) and \( M > 0 \), such that the following holds for each \( i = 1, \ldots, n \), each \( \epsilon > 0 \) and each \( N \):

\[
\begin{align*}
\text{Prob}\left\{ \sup_{x \in X} |(f_N(x) - f_0(x))_i| \geq \epsilon \right\} \leq \exp\left\{-N\mu\right\} + \frac{M}{\epsilon^n} \exp\left\{-\frac{N\epsilon^2}{32\sigma_1^2}\right\},
\end{align*}
\]

where \( (f_N(x) - f_0(x))_i \) denotes the \( i \)th component of \( f_N(x) - f_0(x) \).

Now for each \( \epsilon > 0 \) and each \( N \), we have

\[
\begin{align*}
\text{Prob}\left\{ \sup_{x \in X} \|f_N(x) - f_0(x)\|_1 \geq \epsilon \right\} \leq \sum_{i=1}^{n} \text{Prob}\left\{ \sup_{x \in X} |(f_N(x) - f_0(x))_i| \geq \frac{\epsilon}{n} \right\} \\
\leq n \exp\left\{-N\mu\right\} + \frac{Mn^{n+1}}{\epsilon^n} \exp\left\{-\frac{N\epsilon^2}{32n^2\sigma^2}\right\},
\end{align*}
\]
where \( \|f_N(x) - f_0(x)\| \) denotes the 1-norm of \( f_N(x) - f_0(x) \). For any other norm on \( \mathbb{R}^n \), the inequality (17) still holds with proper choices of \( \sigma, \alpha \) and \( \beta \), since all norms are equivalent in \( \mathbb{R}^n \).

(b) By part (a), there exist positive real numbers \( \beta, \mu, M \) and \( \sigma \) such that (17) holds for each \( \epsilon > 0 \) and each \( N \). By Theorem 4.1, \( f_0 \) is continuously differentiable on \( O \) with \( df_0(x) = E[d\xi F(x, \xi)] \) for each \( x \in O \), and \( f_N \) converges to \( f_0 \) w.p. 1 as an element of \( C^1(X, \mathbb{R}^k) \). Now using Assumption 3.4(b) and arguments as for part (a) of the theorem to \( d\xi F(x, \xi) \), we can obtain positive real numbers \( \beta_2, \mu_2, M_2 \) and \( \sigma_2 \) such that

\[
\text{Prob}\left\{ \sup_{x \in X} \|df_N(x) - df_0(x)\| \geq \epsilon \right\} \leq \beta_2 \exp\left\{ -N\mu_2 \right\} + \frac{M_2}{\epsilon^n} \exp\left\{ -\frac{N\epsilon^2}{\sigma_2} \right\},
\]

for each \( \epsilon > 0 \) and each \( N \). The result follows on noting that

\[
\text{Prob}\{ \|f_N - f_0\|_{1, X} \geq \epsilon \} \leq \text{Prob}\left\{ \sup_{x \in X} \|f_N(x) - f_0(x)\| \geq \epsilon/2 \right\} + \text{Prob}\left\{ \sup_{x \in X} \|df_N(x) - df_0(x)\| \geq \epsilon/2 \right\}.
\]

The next key result is the following functional central limit theorem in the space \( C^1(X, \mathbb{R}^n) \). The proof is very similar to Corollary 7.17 of [1] which is for a setting where \( C^1(X, \mathbb{R}^n) \) is replaced by \( C(X, \mathbb{R}^n) \) (the space of continuous functions from \( X \) to \( \mathbb{R}^n \) with the uniform metric). For completeness we sketch the proof in Appendix B.

**THEOREM 4.3 (FUNCTIONAL CENTRAL LIMIT THEOREM)** Let \( F \) satisfy Assumption 3.1. Then there exists a \( C^1(X, \mathbb{R}^n) \) valued random variable \( Y \) such that for each finite subset \( \{X_1, X_2, \ldots, X_m\} \) of \( X \), the random vector \( (Y(X_1), \ldots, Y(X_m)) \) has a multivariate normal distribution with zero mean and the same covariance matrix as that of

\( \langle F(x_1, \xi), \ldots, F(x_m, \xi) \rangle \)

and, as \( N \to \infty \), \( \sqrt{N}(f_N - f_0) \) converges in distribution, in \( C^1(X, \mathbb{R}^n) \), to \( Y \).

The following theorem is taken from [18, Theorem 7.59]; see a more general version in [15]. We say a map \( G \) from a Banach space \( B_1 \) to \( B_2 \) is Hadamard directionally differentiable at a point \( \mu \in B_1 \), if for all directions \( v \in B_1 \) the directional derivative \( dG(\mu)(v) \) exists, and satisfies the following equality,

\[
dG(\mu)(v) = \lim_{t \downarrow 0, t \to v} \frac{G(\mu + tv) - G(\mu)}{t}.
\]

When \( G \) is Lipschitz continuous in a neighborhood of \( \mu \), Hadamard directionally differentiability is equivalent to \( B \)-differentiability [17].

**THEOREM 4.4 (FUNCTIONAL DELTA THEOREM)** Let \( B_1 \) and \( B_2 \) be Banach spaces. Let \( \{Y_N\} \) be a sequence of random elements in \( B_1 \), \( \{\tau_N\} \) be a sequence of positive numbers converging to \( +\infty \) as \( N \to \infty \), and \( G \) be a map from \( B_1 \) to \( B_2 \). Suppose that the space \( B_1 \) is separable, the map \( G \) is Hadamard directionally differentiable at a point \( \mu \in B_1 \), and the sequence \( \tau_N(Y_N - \mu) \) converges in distribution to a random element \( Y \) of \( B_1 \). Then \( \tau_N[G(Y_N) - G(\mu)] \) converges in distribution to \( dG(\mu)(Y) \), where \( dG(\mu) \) denotes the directional derivative of \( G \) at \( \mu \), and

\[
\tau_N[G(Y_N) - G(\mu)] = dG(\mu)(\tau_N[Y_N - \mu]) + o_p(1).
\]

In the above result \( o_p(1) \) denotes a \( B_2 \) valued random variable \( Z_N \) such that \( \|Z_N\| \) converges to zero in probability as \( N \to \infty \).

5. Methodology. This section discusses the asymptotic distribution of SAA solutions, and develops a method to build confidence regions for the true solution to an SVI. Instead of dealing with the SAA problem in its original form (5), we work with its normal map formulation

\[
(f_N)_S(z) = 0.
\]

Recall that, under Assumption 3.2, \( x_0 \) solves the variational inequality (3), with \( z_0 = x_0 - f_0(x_0) \). It follows that \( (f_0)_S(z_0) = 0 \).
5.1 Asymptotic distribution of SAA solutions. The following theorem describes the probabilistic behavior of SAA solutions.

**Theorem 5.1** Suppose that Assumptions 3.1 and 3.2 hold. Let \( \delta > 0 \) be as in the statement of Lemma 3.1. Choose \( \lambda > \delta^{-1} \), and let neighborhoods \( X_0, Z \) and \( \Gamma \) and the function \( z : \Gamma \to \mathbb{R}^n \) be as given in Lemma 3.1. For each \( \omega \in \Omega \) and integer \( N \) satisfying \( f_N \in \Gamma \), let \( z_N = z(f_N) \) and \( x_N = \Pi_S(z_N) \), so that \( z_N \) is the unique solution for (21) in \( Z \), \( x_N \) is the unique solution for the variational inequality \( 0 \in f_N(\cdot) + N_S(\cdot) \) in \( X_0 \). For almost every \( \omega \in \Omega \), there exists an integer \( N_\omega \), such that \( z_N \) and \( x_N \) are well defined for each \( N \geq N_\omega \). Moreover, \( \lim_{N \to \infty} z_N = z_0 \) almost surely,

\[
\sqrt{N}(z_N - z_0) \Rightarrow (L_K)^{-1} \circ [-Y(x_0)],
\]

where \( Y \) is as in the statement of Theorem 4.3, and

\[
\sqrt{N} d(f_0)_S(z_0)(z_N - z_0) \Rightarrow Y(x_0).
\]

Choose a positive real number \( \epsilon_0 \), such that each \( f \in C^1(X, \mathbb{R}^n) \) satisfying \( \| f - f_0 \|_{1,X} < \frac{\epsilon_0}{\lambda} \) belongs to \( \Gamma \). Suppose in addition that Assumption 3.4 holds. Then there exist positive real numbers \( \beta_0, \mu_0, M_0 \) and \( \sigma_0 \), such that the following holds for each \( \epsilon \in (0, \epsilon_0] \) and each \( N \):

\[
\text{Prob}\{\| x_N - x_0 \| < \epsilon \} \geq \text{Prob}\{\| z_N - z_0 \| < \epsilon \}
\geq 1 - \beta_0 \exp\{-N\mu_0\} - \frac{M_0}{\epsilon^n} \exp\left\{ -\frac{N\epsilon^2}{\sigma_0} \right\}.
\]

**Proof.** According to Lemma 3.1, for each \( f \) belonging to the neighborhood \( \Gamma \), \( z(f) \) is the unique point in \( Z \) satisfying \( f(\cdot)_S(z(f)) = 0 \), and \( x(f) = \Pi_S(z(f)) \) is the unique point in \( X_0 \) satisfying \( 0 \in f(\cdot) + N_S(\cdot) \). Moreover, the function \( z \) is B-differentiable at \( f_0 \) with its B-derivative given by (10), and is Lipschitz on \( \Gamma \) with modulus \( \lambda \). According to Theorem 4.1, \( f_N \) converges to \( f_0 \) w.p. 1 as an element of \( C^1(X, \mathbb{R}^n) \). Consequently, for almost every \( \omega \in \Omega \), there exists an integer \( N_\omega \), such that for each \( N \geq N_\omega \) the function \( f_N \) belongs to \( \Gamma \). In addition, the Lipschitz continuity of \( z \) ensures that \( \lim_{N \to \infty} z_N = z_0 \) w.p. 1.

From Theorem 4.3 \( \sqrt{N}(f_N - f_0) \) converges in distribution in \( C^1(X, \mathbb{R}^n) \) to \( Y \). Let \( \Gamma_1 \) be a closed subset of \( \Gamma \) such that \( f_0 \) belongs to the interior of \( \Gamma_1 \). Then \( z : \Gamma_1 \to \mathbb{R}^n \) is a Lipschitz map which is B-differentiable at \( z_0 \). From the remark above Theorem 4.4, \( z \) is Hadamard directionally differentiable at \( z_0 \). Define

\[
\tilde{f}_N(\omega) = f_N(\omega)1_{f_N(\omega) \in \Gamma_1} + f_01_{f_N(\omega) \in \Gamma_1}, \quad \omega \in \Omega.
\]

Then \( \sqrt{N}(\tilde{f}_N - f_0) \) converges in distribution in \( C^1(X, \mathbb{R}^n) \) to \( Y \). By Theorem 4.4,

\[
\sqrt{N}(z(\tilde{f}_N) - z(f_0)) \Rightarrow dz(f_0)(Y) = (L_K)^{-1} \circ [-Y(x_0)].
\]

Now since \( z(\tilde{f}_N) = z_N \) for all \( N \) large enough, a.s., we have (22). It was shown in [13] that \( L_K \) is exactly the B-derivative of the normal map \( f(\cdot)_S \) at \( z_0 \), denoted by \( d(f)_S(z_0) \). Applying the operator \( d(f)_S(z_0) \) on both sides of (22) yields (23).

Finally, from Theorem 4.2, there exist positive constants \( \beta_1, \mu_1, M_1 \) and \( \sigma_1 \) such that (18) holds for each \( \epsilon > 0 \) and each \( N \). The way we chose \( \epsilon_0 \) implies that, for each \( f \in C^1(X, \mathbb{R}^n) \) satisfying \( \| f - f_0 \|_{1,X} < \frac{\epsilon_0}{\lambda} \), \( z(f) \) is well defined and satisfies

\[
\| \Pi_S(z(f)) - x_0 \| \leq \| z(f) - z_0 \| \leq \lambda \| f - f_0 \|_{1,X},
\]

where the first inequality follows from the fact that \( \Pi_S \) is a non-expansive mapping, and the second follows from the Lipschitz continuity of \( z \). Accordingly, for each real number \( \epsilon \in (0, \epsilon_0] \) and each integer \( N \), we have

\[
\text{Prob}\{\| x_N - x_0 \| < \epsilon \} \geq \text{Prob}\{\| z_N - z_0 \| < \epsilon \}
\geq \text{Prob}\{\| f_N - f_0 \|_{1,X} < \epsilon/\lambda \}
\geq 1 - \beta_1 \exp\{-N\mu_1\} - \frac{M_1\lambda^2}{\epsilon^n} \exp\left\{ -\frac{N\epsilon^2}{\sigma_1\lambda^2} \right\},
\]

where the first and second inequalities follow from (25), and the third follows from (18). The inequality (24) follows by letting \( \beta_0 = \beta_1, \mu_0 = \mu_1, M_0 = M_1\lambda^n, \sigma_0 = \sigma_1\lambda^2 \). \( \square \)
Note that the definition of the function \( z \) in Lemma 3.1 ensures that \( \Pi_S(z(f)) \in X_0 \subset X \), whenever \( z(f) \) is defined. Thus, \( x_N \) belongs to \( X \) whenever it is well defined. For a given \( N \), \( z_N \) and \( x_N \) may not be well defined at some \( \omega \in \Omega \). The expression \( \text{Prob}\{\|x_N - x_0\| < \epsilon\} \) in (24) is a short version of
\[
\text{Prob}\{\|x_N - x_0\| < \epsilon \text{ and } x_N \text{ is well defined}\},
\]
and similarly for \( \text{Prob}\{\|z_N - z_0\| < \epsilon\} \). We use similar conventions in the rest of this paper for brevity.

5.2 B-derivatives of the Euclidean projector and normal maps. As noted in Section 3, in order to obtain computable confidence regions for \( z_0 \) we need to suitably approximate \( d(f_0)_S(z_0) \). To understand properties of B-derivatives of the normal maps \( (f_0)_S \) and \( (f_N)_S \), we first need to investigate properties of B-derivatives of the Euclidean projector \( \Pi_S \). As discussed in Section 2.1, \( \Pi_S \) is a piecewise affine function, so it is B-differentiable on \( \mathbb{R}^n \) [16, Chapter 2.2]. Moreover, if we define the critical cone to \( S \) at a point \( z \in \mathbb{R}^n \) as
\[
K(z) = T_S(\Pi_S(z)) \cap \{z - \Pi_S(z)\}^\perp,
\]
then we have
\[
d\Pi_S(z)(h) = \Pi_{K(z)}(h)
\]
for each \( h \in \mathbb{R}^n \); see [11, Corollary 4.5]. Theorem 5.2 below will show that \( K(z) \) is the same cone, for all points \( z \) belonging to the relative interior of a given face of a given \( n \)-cell in the normal manifold of \( S \). To develop this theorem we need the following lemma, which provides a characterization for faces of \( n \)-cells.

**Lemma 5.1** Let \( F \) be a nonempty face of \( S \), and let \( N = N_S(\text{ri} F) \) and \( C = C_S(F) \). Then, a nonempty set \( B \in \mathbb{R}^n \) is a face of \( C \), if and only if there is a nonempty face \( E \) of \( F \), and a nonempty face \( M \) of \( N \), such that \( B = E + M \).

**Proof.** It was shown in [12, Proposition 2.1] that
\[
\text{par} \, F = (\text{aff} \, N)^\perp,
\]
where \( \text{aff} \, N \) is the affine hull of \( N \), and \( \text{par} \, F \) is the subspace parallel to \( \text{aff} \, F \). Since \( C = F + N \), for any point \( z \) in \( C \) there is a unique decomposition \( z = z^F + z^N \) with \( z^F \in F \) and \( z^N \in N \). Points \( z^F \) and \( z^N \) are affine functions of \( z \).

Now, let \( E \) be a nonempty face of \( F \) and \( M \) be a nonempty face of \( N \). We need to prove that \( E + M \) is a face of \( C \). To this end, let \( c \in E + M \), and suppose that \( c_1, c_2 \in C \) satisfy \( c_1 + c_2 = 2c \). It follows that \( c_1^F + c_2^F = 2c^F \) and \( c_1^N + c_2^N = 2c^N \). The facts that \( c^F \in E \) and that \( E \) is a face of \( F \) imply \( c_1^F, c_2^F \in E \). Similarly, \( c_1^N, c_2^N \in M \). Consequently, \( c_1, c_2 \in E + M \). This proves that \( E + M \) is a face of \( C \).

Conversely, let \( B \) be a face of \( C \), and define
\[
E = \{z^F \mid z \in B\} \text{ and } M = \{z^N \mid z \in B\}.
\]
We now prove that \( E \) is a face of \( F \). Let \( e \in E \), and choose \( z \in B \) such that \( e = z^F \). Suppose \( f_1, f_2 \in F \) satisfy \( f_1 + f_2 = 2e \). Define \( c_1 = f_1 + z^N \) and \( c_2 = f_2 + z^N \). We have \( c_1, c_2 \in C \) and \( c_1 + c_2 = 2z \). The fact that \( B \) is a face of \( C \) implies \( c_1, c_2 \in B \). But \( f_1 = c_1^F \) and \( f_2 = c_2^F \), so \( f_1, f_2 \in E \). This proves that \( E \) is a face of \( F \). A similar argument proves that \( M \) is a face of \( N \). It remains to prove \( B = E + M \). It is clear that \( B \subset E + M \). For the other direction, let \( z_1, z_2 \in B \) and define \( \hat{z} = z_1^N + z_2^N \). It suffices to show \( \hat{z} \in B \). Write \( \hat{z} = z_1^F + z_2^F \), and note that \( z_1 + z_2 = \hat{z} + \hat{z} \). Because \( z_1, z_2 \in B \), \( \hat{z}, \hat{z} \in C \), and \( B \) is a face of \( C \), we have \( \hat{z} \in B \). This proves \( B = E + M \). \( \square \)

Theorem 5.2 below gives an explicit expression for \( K(z) \) for points \( z \) belonging to the relative interior of a face \( B \) of some \( n \)-cell \( C \). Recall that \( F \) denotes the collection of all nonempty faces of \( S \). Let \( \mathcal{I} = \{I(x) \mid x \in S\} \) be the family of all active index sets. There is a well known one-to-one correspondence between \( \mathcal{F} \) and \( \mathcal{I} \) given as follows. First, for each \( I \in \mathcal{I} \) consider the set
\[
F(I) = \{x \in \mathbb{R}^n \mid \langle a_i, x \rangle = b_i, \ i \in I, \ \langle a_i, x \rangle \leq b_i, \ i \in \{1, \ldots, m\} \setminus I, \}
\]
which is a nonempty face of \( S \). The active index set for any point belonging to \( \text{ri} F(I) \) is exactly \( I \). Conversely, for each nonempty face \( F \) of \( S \) there is a unique \( I \in \mathcal{I} \) such that \( F = F(I) \). For detailed discussion on this, see [7, Section 1] or [16, Chapter 2].
Let $I \in \mathfrak{I}$. It was shown in [16, Lemma 2.4.2] (see also [7, Proposition 1]) that each nonempty face of $F(I)$ is of the form $F(J)$, with $J \supset I$ and $J \in \mathfrak{I}$. The normal cone to $S$ on $ri F(I)$ is given by

$$N_S(ri F(I)) = \text{pos}\{a_i, i \in I\} = \{0\} \cup \left\{ \sum_{i \in I} \tau_i a_i : \tau_i \geq 0, i \in I \right\},$$

and each nonempty face of $N_S(ri F(I))$ is of the form $N_S(ri F(J))$, with $J \in \mathfrak{I}$ and $J \supset I$.

**Theorem 5.2** Let $I \in \mathfrak{I}$, $F = F(I)$, $N = N_S(ri F)$ and $C = C_S(F)$. Let $E$ be a nonempty face of $F$, $M$ be a nonempty face of $N$, and $B = E + M$. Let $I^+, I^- \in \mathfrak{I}$ be such that $E = F(I^+)$ and $M = N_S(ri F(I^-))$. For each $z \in ri B$,

$$K(z) = \{u \in \mathbb{R}^n | \langle a_i, u \rangle = 0, i \in I^-, \langle a_i, u \rangle \leq 0, i \in I^+ \setminus I^- \}. \tag{29}$$

**Proof.** By [14, Corollary 6.6.2], we have $ri B = ri E + ri M$. Let $x \in ri E$ and $y \in ri M$ satisfy $z = x + y$. Then $y \in N_S(x)$, $x = \Pi_{S}(z)$ and $K(z) = T_S(x) \cap \{y\}$. We have

$$ri E = \{x \in \mathbb{R}^n | \langle a_i, x \rangle = b_i, i \in I^+, \langle a_i, x \rangle < b_i, i \in \{1, \ldots, m\} \setminus I^+ \},$$

and

$$ri M = ri \text{pos}\{a_i, i \in I^-\} = \begin{cases} \{\sum_{i \in I^-} \tau_i a_i : \tau_i > 0, i \in I^-\} & \text{if } I^- \neq \emptyset, \\ \{0\} & \text{if } I^- = \emptyset. \end{cases}$$

The fact $x \in ri E$ implies $I(x) = I^+$, so

$$T_S(x) = \{u \in \mathbb{R}^n | \langle a_i, u \rangle \leq 0, i \in I^+ \}.$$  

Because $y \in ri M$, there exists $\tau_i > 0$ for each $i \in I^-$ such that $y = \sum_{i \in I^-} \tau_i a_i$. Note that $I^- \subset I \subset I^+$, so that for each $i \in I^-$ the vector $a_i$ belongs to the polar cone of $T_S(x)$. Thus, a vector $u \in T_S(x)$ satisfies $\langle y, u \rangle = 0$ if and only if $\langle a_i, u \rangle = 0$ for each $i \in I^-$. Equation (29) follows. \hfill \square

The right hand side of (29) depends on $I^+$ and $I^-$, but not $z$. Thus, $K(z)$ is the same for all $z \in ri B$. It then follows from (27) that $d\Pi_S(z)(\cdot)$ is the same function for all such $z$. Because $K(z)$ is a polyhedral convex cone, $d\Pi_S(z)(\cdot)$ is a piecewise linear function. The way $B$ is defined in Theorem 5.2 implies that it can be any $k$-cell of the normal manifold of $S$, for $k = 0, \ldots, n$.

For each $k = 0, \ldots, n$, we write the $k$-cells of the normal manifold of $S$ as $C_i^k$, $i = 1, \ldots, j(k)$, where $j(k)$ denotes the number of $k$-cells. For all $z$ belonging to the relative interior of a given $C_i^k$, $d\Pi_S(z)$ is the same function, which we denote by $\Psi_i^k$. Thus, for each $k = 0, \ldots, n$ and $i = 1, \ldots, j(k)$, $\Psi_i^k$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^n$, with

$$\Psi_i^k(\cdot) = d\Pi_S(z)(\cdot) \text{ if } z \text{ belongs to the relative interior of } C_i^k. \tag{30}$$

The next proposition shows that, for each point $z \in \mathbb{R}^n$, there is exactly one cell in the normal manifold of $S$ that contains $z$ in its relative interior. This cell is also the smallest among all cells containing $z$.

**Proposition 5.1** Let $z \in \mathbb{R}^n$. There is a unique cell, $C^q_j$, in the normal manifold of $S$, such that $z \in ri C^q_j$. Moreover, for any other cell $C^q_i$ that contains $z$, one has $C^q_j \subset C^q_i$ and $q < k$.

**Proof.** According to item (iii) below (8), the union of all $n$-cells in the normal manifold of $S$ is $\mathbb{R}^n$. Thus, $z$ belongs to at least one $n$-cell. Without loss of generality, suppose that $z$ belongs to $C_1^n, \ldots, C_m^n$. The intersection of these $n$-cells, $\cap_{i=1}^m C_i^n$, is a common face of theirs. Let $F$ be the unique face of $\cap_{i=1}^m C_i^n$ such that $z \in ri F$. The set $F$ is also a common face of $C_1^n, \ldots, C_m^n$, and is therefore the unique face of each $C_i^n$, $i = 1, \ldots, m$, that contains $z$ in its relative interior. Let $C_j^q = F$. This proves the existence and uniqueness of a cell $C_j^q$ such that $z \in ri C_j^q$.

Let $C_r^k$ be another cell containing $z$. Then $C_r^k$ is a face of one of the $n$-cells $C_1^n, \ldots, C_m^n$. Suppose it is a face of $C_i^n$. Because $C_j^q$ is a face of $C_i^n$, with $z \in ri C_j^q$, we have $C_j^q \subset C_r^k$ and $q < k$. \hfill \square

**Example 2.1 continued.** The set $S$ in Example 2.1 has nine 1-cells (edges of the 2-cells), and three 0-cells (vertices). So $j(2) = 7$, $j(1) = 9$, $j(0) = 3$. When $z$ belongs to the interior of any of the 2-cells, $d\Pi_S(z)(\cdot)$ is a linear function. For example, when $z$ belongs to the interior of the 2-cell \{ $x \in \mathbb{R}^2 | x_1 \leq 0, 0 \leq x_2 \leq 1$ \}, $d\Pi_S(h) = (0, h_2)$ for any $h \in \mathbb{R}^2$. When $z$ belongs to the relative
interior of any of the 1-cells, $d\Pi_S(z)(\cdot)$ is a piecewise linear function with two pieces. For example, when $h$ lies in the relative interior of the edge connecting $(0,0)$ and $(0,1)$, we have

$$d\Pi_S(z)(h) = \begin{cases} h, & \text{if } h_1 \geq 0, \\ (0,h_2), & \text{if } h_1 \leq 0. \end{cases}$$

When $z$ is any of the three vertices, $d\Pi_S(z)(\cdot)$ is a piecewise linear function with four pieces.

As is shown by the example, $d\Pi_S(z)(h)$ is continuous with respect to $h$ for a fixed $z$. However, on the boundaries of the $n$-cells, $d\Pi_S(z)$ is discontinuous with respect to $z$, due to the dramatic change of the structure of $K(z)$.

Now, suppose that Assumption 3.1 holds, so that $f_0$ and almost every $f_N$ are differentiable on $X$. By the definition of the normal map $(f_0)_S$ in (6) and the chain rule of B-differentiability (see, e.g., [16, Theorem 3.1.1]), the function $(f_0)_S(\cdot)$ is B-differentiable on $\Pi_S^{-1}(O)$, with

$$d(f_0)_S(z)(h) = d\Pi_S(z)(h) + d\Pi_S(z)(h)$$

for each $z \in \Pi_S^{-1}(O)$ and each $h \in \mathbb{R}^n$. Similarly, for almost every $\omega \in \Omega$, the function $(f_N)_S(\cdot)$ is B-differentiable on $\Pi_S^{-1}(O)$, with

$$d(f_N)_S(z)(h) = d\Pi_S(z)(h) + d\Pi_S(z)(h)$$

for each $z \in \Pi_S^{-1}(O)$ and each $h \in \mathbb{R}^n$. It follows that, the functions $d(f_0)_S(z)(h)$ and $d(f_N)_S(z)(h)$ are both continuous with respect to $h$ for fixed $z$, but they are discontinuous with respect to $z$ on the boundaries of the $n$-cells.

5.3 Approximations of the B-derivatives. This subsection constructs functions that are computable from the SAA solutions, and shows how to use these functions to approximate $d\Pi_S(z_0)$ and $d(f_0)_S(z_0)$. First, for each $k = 0, \ldots, n$ and $i = 1, \ldots, j(k)$, define a function $d^k_i : \mathbb{R}^n \to \mathbb{R}$ by

$$d^k_i(z) = d(z, C^k_i) = \min_{x \in C^k_i} \|z - x\|.$$
Lemma 5.2. For each \( N \in \mathbb{N} \), \( \Lambda_N \) is well defined and continuous on \( \mathbb{R}^n \times \mathbb{R}^n \).

Proof. Each \( z \in \mathbb{R}^n \) belongs to at least one cell \( C_i^k \) in the normal manifold, and for such \( k \) and \( i \), \( d_i^k(z) = 0 \). Thus, the denominator in (34) is positive for each \( N, z, h \), and \( \Lambda_N \) is well defined everywhere. For each \( k = 0, \ldots, n \) and each \( i = 1, \ldots, j(k) \), the functions \( h \mapsto \Psi_i^k(h) \) and \( z \mapsto d_i^k(z) \) are continuous on \( \mathbb{R}^n \), so \( \Lambda_N \) is continuous on \( \mathbb{R}^n \times \mathbb{R}^n \). \( \square \)

Theorem 5.3. Suppose that Assumptions 3.1 and 3.2 hold. For each \( N \in \mathbb{N} \), let \( \Lambda_N \) be as defined in (34). Let \( \gamma_0 > 0 \) be the minimum of \( d_i^k(z_0) \) among all of the cells \( C_i^k \) such that \( z_0 \notin C_i^k \). Then, there exists a positive real number \( \kappa \) and an integer \( N_0 \), such that for each \( N \geq N_0 \),

\[
\begin{align}
\text{Prob} \left\{ \sup_{h \in \mathbb{R}^n} \frac{\|\Lambda_N(z_N)(h) - d\Pi_S(z_0)(h)\|}{\|h\|} < \frac{\kappa}{g(N)} \right\} & \geq \text{Prob} \left\{ \|z_N - z_0\| < \gamma_0/2 \right\} + \text{Prob} \left\{ \|z_N - z_0\| < 1/(2g(N)) \right\} - 1.
\end{align}
\]

If Assumption 3.4 holds additionally, then

\[
\lim_{N \to \infty} \text{Prob} \left\{ \sup_{h \in \mathbb{R}^n} \frac{\|\Lambda_N(z_N)(h) - d\Pi_S(z_0)(h)\|}{\|h\|} < \frac{\kappa}{g(N)} \right\} = 1.
\]

Proof. Suppose that \( z_0 \) belongs to cells \( C_{i_1}^{k(1)}, \ldots, C_{i_q}^{k(q)} \) in the normal manifold of \( S \), and that \( z_0 \in \mathring{C}_{i_q}^{k(q)} \). It follows from Proposition 5.1 that \( C_{i_q}^{k(q)} \subset C_{i(j)}^{k(j)} \) and \( k(q) < k(j) \) for each \( j = 1, \ldots, q - 1 \), and from the definition of \( \Psi_i^k \) in (30) that

\[
d(\Pi_S)(z_0)(h) = \Psi_i^{k(q)}(h) \quad \text{for each} \quad h \in \mathbb{R}^n.
\]

The definition of \( g \) ensures that \( \lim_{N \to \infty} g(N) = \infty \), so there exists an integer \( N_0 \), such that \( g(N) \geq \max(2/\gamma_0, 1) \) for each \( N \geq N_0 \).

For each cell \( C_i^k \) with \( z_0 \notin C_i^k \), we have

\[
d_i^k(z_N) \geq d_i^k(z_0) - \|z_N - z_0\| \geq \gamma_0 - \|z_N - z_0\|.
\]

Consequently, for each \( N \geq N_0 \),

\[
\begin{align}
\text{Prob} \left\{ 1/g(N) - \min(d_i^k(z_N), 1/g(N)) = 0 \text{ for all } C_i^k \text{ not containing } z_0 \right\} & \geq \text{Prob} \left\{ \|z_N - z_0\| < \gamma_0/2 \right\}.
\end{align}
\]

Next, consider a cell that contains \( z_0 \), that is, one of the cells \( C_{i(j)}^{k(j)} \) for \( j = 1, \ldots, q \). The following inequality holds for each \( N \),

\[
0 \leq d_{i(j)}^{k(j)}(z_N) = d(z_N, C_{i(j)}^{k(j)}) \leq \|z_N - z_0\|,
\]

which implies, for each \( N \geq N_0 \),

\[
\begin{align}
\text{Prob} \left\{ 1/g(N) - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) > 1/(2g(N)) \right\} & \geq \text{Prob} \left\{ 1/g(N) - \min(\|z_N - z_0\|, 1/g(N)) > 1/(2g(N)) \right\} = \text{Prob} \left\{ \|z_N - z_0\| < 1/(2g(N)) \right\}.
\end{align}
\]

We then have

\[
\begin{align}
\text{Prob} \left\{ \sum_{j=1}^{q} \left[ 1/g(N) - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} > [1/(2g(N))]^{k(q)} \right\} & \geq \text{Prob} \left\{ \left[ 1/g(N) - \min(d_{i(q)}^{k(q)}(z_N), 1/g(N)) \right]^{k(q)} > [1/(2g(N))]^{k(q)} \right\} = \text{Prob} \left\{ 1/g(N) - \min(d_{i(q)}^{k(q)}(z_N), 1/g(N)) > 1/(2g(N)) \right\} \geq \text{Prob} \left\{ \|z_N - z_0\| < 1/(2g(N)) \right\}.
\end{align}
\]
It is clear that \( k(q) \leq n \). Proposition 5.1 implies that \( k(q) - k(j) \leq -1 \) for each \( j = 1, \ldots, q - 1 \). Hence, for each \( j = 1, \ldots, q - 1 \) and each \( N \geq N_0 \), we have

\[
[1/g(N)]^{k(j)} / [1/(2g(N))]^{k(q)} \leq 2^n g(N)^{-1}.
\]

Consequently, for each \( N \geq N_0 \),

\[
\text{Prob} \left\{ \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} \sum_{p=1}^{q} \left[ \frac{1}{g(N)} - \min(d_{i(p)}^{k(p)}(z_N), 1/g(N)) \right]^{k(p)} < 2^n g(N)^{-1} \right\}
\]

\[
\geq \text{Prob} \left\{ \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} > [1/(2g(N))]^{k(q)} \right\}
\]

\[
\geq \text{Prob} \left\{ \|z_N - z_0\| < 1/(2g(N)) \right\},
\]

where the second inequality follows from (39).

We are now ready to prove (35). For each \( k = 0, \ldots, n \) and \( i = 1, \ldots, j(k) \), \( \Psi^k_i \) defined in (30) is a piecewise linear function, so its norm defined as

\[
\|\Psi^k_i\| = \sup_{h \in \mathbb{R}^n} \frac{\|\Psi^k_i(h)\|}{\|h\|}
\]

is a finite number. Let \( \eta \) be the maximum of \( \|\Psi^k_i\| \) among all \( k \) and \( i \). For each \( h \in \mathbb{R}^n \), we have

\[
\|\Lambda_N(z_N(h)) - dI_S(z_0(h))\| = \|\Lambda_N(z_N(h)) - \Psi_i^{k(q)}(h)\|
\]

\[
= \left\| \sum_{k=0}^{n} \sum_{i=1}^{j(k)} \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} \Psi_{i(j)}^{k(j)}(h) - \Psi_i^{k(q)}(h) \right\|
\]

For each \( N \geq N_0 \), it follows from (37) that

\[
\text{Prob} \left\{ \|\Lambda_N(z_N(h)) - dI_S(z_0(h))\| \right\}
\]

\[
= \left\| \sum_{j=1}^{q} \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} \Psi_{i(j)}^{k(j)}(h) - \Psi_i^{k(q)}(h) \right\| \text{ for all } h \in \mathbb{R}^n \}
\]

\[
\geq \text{Prob} \left\{ \|z_N - z_0\| < \gamma_0/2 \right\}.
\]

For each \( h \in \mathbb{R}^n \), term (a) above is bounded from above by

\[
\left\| \sum_{j=1}^{q} \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} \Psi_{i(j)}^{k(j)}(h) \right\|
\]

\[
+ \left\| \sum_{j=1}^{q} \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(q)}(z_N), 1/g(N)) \right]^{k(q)} - 1 \right\| \Psi_i^{k(q)}(h),
\]

which is in turn bounded from above by

\[
2\eta \|h\| \left\| \sum_{j=1}^{q} \left[ \frac{1}{g(N)} - \min(d_{i(j)}^{k(j)}(z_N), 1/g(N)) \right]^{k(j)} \right\|
\]

From (40) we have, for each \( N \geq N_0 \),

\[
\text{Prob} \left\{ \text{term (b)} < 2^{n+1} \eta(q-1)g(N)^{-1}\|h\| \right\}
\]

\[
\geq \text{Prob} \left\{ \|z_N - z_0\| < 1/(2g(N)) \right\}.
\]
The two inequalities (42) and (43) and the fact that term (a) is always less than or equal to term (b) imply
\[
\begin{align*}
\Pr\{ \|A(N)(z)(h) - d\Pi S(z)(h)\| < 2^{n+1}\eta(q - 1)g(N)^{-1}\|h\| \text{ for all } h \in \mathbb{R}^n \} \\
\geq \Pr\{ \|z_N - z_0\| < \gamma_0/2 \} + \Pr\{ \|z_N - z_0\| < 1/(2g(N)) \} - 1
\end{align*}
\]
for each \(N \geq N_0\). Letting \(\kappa = 2^{n+1}\eta(q - 1)\) proves (35) for each \(N \geq N_0\).

If Assumption 3.4 holds additionally, then by Theorem 5.1 there exist positive constants \(c_0, \beta_0, \mu_0, M_0\) and \(\sigma_0\), such that (24) holds for each \(\epsilon \in (0, \epsilon_0)\) and each \(N\). For \(N\) large enough to satisfy \(z\gamma N^2 \leq \epsilon_0\), we have
\[
\Pr\{ \|z_N - z_0\| < 1/(2g(N)) \} \geq 1 - \beta_0 \exp\{-N\mu_0\} - 2^n M_0 g(N)^n \exp\left\{-\frac{N}{4\sigma_0 g(N)^2}\right\}.
\]
The right hand side of the above inequality converges to 1 as \(N \to \infty\) by the definition of \(g(N)\). Clearly, \(\Pr\{ \|z_N - z_0\| < \gamma_0/2 \} \) also converges to 1 as \(N \to \infty\). This proves (36) from (35).

Theorem 5.3 above proves that the probability for the norm of the piecewise linear function \(A_N(z)(z) - d\Pi S(z)(z)\) to be below \(\kappa/g(N)\) converges to 1 as \(N\) goes to infinity. Since \(g(N) \to \infty\) as \(N \to \infty\), this in particular says that the norm of \(A_N(z)(z) - d\Pi S(z)(z)\) converges to 0 in probability, consequently, \(A_N(z)(z)\) is a good approximation for \(d\Pi S(z)(z)\) for \(N\) large. Next, we define for each \(N \in \mathbb{N}\) a function \(\Phi_N : \Pi^{-1}(0) \times \mathbb{R}^n \times \Omega \to \mathbb{R}^n\) by
\[
\Phi_N(z, h, \omega) = df_N(\Pi S(z))(A_N(z)(h)) + h - A_N(z)(h).
\]
For convenience we will write \(\Phi_N(z_N(\omega), h, \omega)\) as \(\Phi_N(z_N(\omega))\). The corollary below shows that \(\Phi_N(z_N)\) provides a good approximation for \(df(0)S(z)(z)\).

**Corollary 5.1** Suppose that Assumptions 3.1, 3.2 and 3.4 hold. For each \(N \in \mathbb{N}\), let \(\Phi_N\) be as defined in (44). Then, there exists a positive real number \(\phi\), such that
\[
\lim_{N \to \infty} \Pr\left[ \sup_{h \in \mathbb{R}^n} \left\| \Phi_N(z_N(h), h - df(0)S(z)(h)) \right\| \leq \frac{\phi}{g(N)} \right] = 1.
\]

**Proof.** From Theorem 4.1, \(f_0\) belongs to \(C^1(X, \mathbb{R}^n)\) and \(df_0\) is Lipschitz with modulus bounded by \(E[C]\). Also, for almost every \(\omega\) the function \(f_N\) belongs to \(C^1(X, \mathbb{R}^n)\) for each \(N \in \mathbb{N}\). Next, by Theorem 4.2, there exist positive real numbers \(\beta_1, \mu_1, M_1\) and \(\sigma_1\) such that (18) holds for each \(\epsilon > 0\) and each \(N\). Now let positive constants \(\lambda, \epsilon_0, \beta_0, \mu_0, M_0\) and \(\sigma_0\), neighborhoods \(X_0\), \(Z\) and \(\Gamma\), and the function \(z\) be as defined in Theorem 5.1; recall that (24) holds for each \(\epsilon \in (0, \epsilon_0)\) and each \(N\). Note that from Theorem 5.3 there exists \(\kappa > 0\) such that (36) holds.

For each \(h \in \mathbb{R}^n\), it follows from (31) and (44) that
\[
\left\| \Phi_N(z_N(h)) - df(0)S(z)(h) \right\| = \left\| \left[ df_N(\Pi S(z_N))(A_N(z_N)(h)) - \Pi S(z)(h) \right] - df(0)S(z)(h) \right\|
\]
\[
\leq \left\| df_N(z_N)(L_N(z_N)(h)) - df(0)S(z)(h) \right\| + \left\| \Pi S(z)(h) - df(0)S(z)(h) \right\|.
\]

Recall the notation \(x_N = \Pi S(z)(z)\) and \(x_0 = \Pi S(z)(0)\). Then
\[
\left\| \Phi_N(z_N(h)) - df(0)S(z)(h) \right\|
\]
\[
\leq \left\| df_N(x_N)(L_N(z_N)(h)) - df(0)x_0(d\Pi S(z)(h)) \right\| + \left\| \Pi S(z)(h) - df(0)S(z)(h) \right\|.
\]

Term (a) above is bounded from above by
\[
\left\| df_N(x_N)(L_N(z_N)(h)) - df(0)S(z)(h) \right\|
\]
\[
+ \left\| df_N(x_N)(d\Pi S(z)(h)) - df(0)x_0(d\Pi S(z)(h)) \right\|.
\]

First, we examine term (c). For each \(h \in \mathbb{R}^n\), we have
\[
\text{term (c)} \leq \left\| df_N(x_N) \right\| \left\| L_N(z_N)(h) - df(0)S(z)(h) \right\|.
\]
\[\text{(47)}\]
Define \( \epsilon_1 = \min (2E|C|\epsilon_0, 2\epsilon_0/\lambda) \). For each \( \epsilon \in (0, \epsilon_1] \), we have
\[
\Pr \{ \|d f_N(x_N) - d f_0(x_0)\| < \epsilon \} 
\geq \Pr \{ \|d f_N(x_N) - d f_0(x_N)\| < \epsilon/2 \} + \Pr \{ \|d f_0(x_N) - d f_0(x_0)\| < \epsilon/2 \} - 1
\geq \Pr \{ \|f_N - f_0\|_{1, X} < \epsilon/2 \} + \Pr \left\{ \|x_N - x_0\| < \frac{\epsilon}{2E|C|} \right\} - 1.
\]
Replacing \( \epsilon \) with \( 1/g(N) \) in the above inequality, applying (18) and (24), and taking limits, we have
\[
\lim_{N \to \infty} \Pr \{ \|d f_N(x_N) - d f_0(x_0)\| < 1/g(N) \} = 1.
\]
It follows from (48) and (36) that
\[
\lim_{N \to \infty} \Pr \left\{ \sup_{h \in \mathbb{R}^n} \frac{\text{term (c)}}{\|h\|} < \left( \|d f_0(x_0)\| + \frac{\kappa}{g(N)} \right) \right\} = 1.
\] (49)
for each \( \epsilon > 0 \).

Next, we examine term (d). For each \( h \in \mathbb{R}^n \), we have \( \text{term (d)} \leq \|d f_N(x_N) - d f_0(x_0)\| \|\Pi_S(z_0)(h)\| \leq \|d f_N(x_N) - d f_0(x_0)\| \|\Pi_S(z_0)\| \|h\| \), where \( \|\Pi_S(z_0)\| \) is defined in the same way as \( \|\Psi_f\| \) is in (41). The inequality above and (48) imply
\[
\lim_{N \to \infty} \Pr \left\{ \sup_{h \in \mathbb{R}^n} \frac{\text{term (d)}}{\|h\|} < \frac{\|\Pi_S(z_0)\|}{g(N)} \right\} = 1.
\] (50)

Putting (36), (49) and (50) together proves (45).

**Corollary 5.2** Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Then
\[
\sqrt{N}\Phi_N(z_N) - z_0 \Rightarrow \mathcal{N}(0, \Sigma_0).
\] (51)
If Assumption 3.3 holds additionally, then
\[
\sqrt{N}\Sigma_0^{-1/2}\Phi_N(z_N) - z_0 \Rightarrow \mathcal{N}(0, I_n).
\] (52)

**Proof.** Since (52) follows from (51) easily when Assumption 3.3 holds, it suffices to prove (51) under Assumptions 3.1, 3.2 and 3.4. Under these assumptions, we have by (23),
\[
\sqrt{N}d(f_0)_S(z_0)(z_N - z_0) \Rightarrow \mathcal{N}(0, \Sigma_0).
\]
To prove (51) it suffices to prove
\[
\lim_{N \to \infty} \Pr \{ \sqrt{N}\|d(f_0)_S(z_0)(z_N - z_0) - \Phi_N(z_N)(z_N - z_0)\| > \epsilon \} = 0
\] (53)
for each \( \epsilon > 0 \). For brevity, we write
\[
V_N = \sup_{h \in \mathbb{R}^n} \frac{\|\Phi_N(z_N)(h) - d(f_0)_S(z_0)(h)\|}{\|h\|}.
\]
By Corollary 5.1, there exists a positive real number \( \phi \) such that
\[
\lim_{N \to \infty} \Pr \{ V_N > \frac{\phi}{g(N)} \} = 0.
\] (54)
We have
\[
\Pr \{ \sqrt{N}\|d(f_0)_S(z_0)(z_N - z_0) - \Phi_N(z_N)(z_N - z_0)\| > \epsilon \}
\leq \Pr \{ \sqrt{N}V_N\|z_N - z_0\| > \epsilon \}
\leq \Pr \{ V_N > \frac{\phi}{g(N)} \} + \Pr \{ \|z_N - z_0\| > \frac{\epsilon g(N)}{\phi \sqrt{N}} \}.
\] (55)
It follows from (24) that there exist positive real numbers \( \epsilon_0, \beta_0, \mu_0, M_0 \) and \( \sigma_0 \) such that
\[
\Pr \{ \|z_N - z_0\| > \frac{\epsilon g(N)}{\phi \sqrt{N}} \} \leq \beta_0 \exp\{ -N \mu_0 \} + \frac{M_0 \phi^n N^n/2}{\epsilon^n g(N)^n} \exp\left\{ -\frac{\epsilon^2 g(N)^2}{\sigma_0 \phi^n} \right\}.
\] (56)
as long as

\[ \frac{\varepsilon g(N)}{\phi \sqrt{N}} \leq \varepsilon_0. \]

(57)

From property (iii) of \( g \) given in Section 5.3 it follows that the inequality (57) holds for \( N \) sufficiently large, and from property (v) the right hand side of (56) converges to zero as \( N \) goes to \( \infty \). Thus we have

\[ \lim_{N \to \infty} \text{Prob}\{ \|z_N - z_0\| > \frac{\varepsilon g(N)}{\phi \sqrt{N}} \} = 0. \]

The above equality, combined with (54) and (55), implies (53).

We can now complete the proof of our main result, Theorem 3.1.

**Proof of Theorem 3.1.** In view of Corollary 5.2, it suffices to prove

\[ \lim_{N \to \infty} \text{Prob}\{ \|\Sigma_N - \Sigma_0\| > \varepsilon \} = 0 \text{ for each } \varepsilon > 0. \]

(58)

Define a function \( \Theta : O \times \Xi \to \mathbb{R}^{n \times n} \) by

\[ \Theta(x, z) = F(x, z)F(x, z)^T, \]

let \( \theta_0(x) = E[\Theta(x, \xi)] \), and for each \( N \in \mathbb{N} \) define the sample average function as

\[ \theta_N(x, \omega) = \frac{1}{N} \sum_{i=1}^{N} F(x, \xi_i(\omega))F(x, \xi_i(\omega))^T = \frac{1}{N} \sum_{i=1}^{N} \Theta(x, \xi_i(\omega)). \]

Let \( x_N = \Pi_S(z_N) \). The definitions of \( \Sigma_N \) and \( \Sigma_0 \) imply

\[ \Sigma_N = \frac{1}{N-1} \sum_{i=1}^{N} (F(x_N, \xi_i) - f_N(x_N))(F(x_N, \xi_i) - f_N(x_N))^T \]

\[ = \frac{1}{N-1} \sum_{i=1}^{N} F(x_N, \xi_i)^T F(x_N, \xi_i) - \frac{N}{N-1} f_N(x_N)f_N(x_N)^T \]

\[ = \frac{N}{N-1} \theta_N(x_N) - f_N(x_N)f_N(x_N)^T \]

and

\[ \Sigma_0 = E \left[ (F(x_0, \xi) - f_0(x_0))(F(x_0, \xi) - f_0(x_0))^T \right] \]

\[ = E[F(x_0, \xi)^T f_0(x_0) - f_0(x_0)^T f_0(x_0) - f_0(x_0) f_0(x_0)^T] = \theta_0(x_0) - f_0(x_0) f_0(x_0)^T. \]

From (24), \( x_N \) converges to \( x_0 \) in probability. Applying Theorem 4.1(a) with \( G \) replaced with \( FF^T \), we see that \( \theta_N \) converges uniformly in \( C(X, \mathbb{R}^{n \times n}) \) to \( \theta_0 \). Consequently, \( \theta_N(x_N) \) converges in probability to \( \theta_0(x_0) \). Similarly applying Theorem 4.1(a) to \( F \) we see that, \( N \to \infty \), \( f_N(x_N) \to f_0(x_0) \) in probability. Combining the above results we have (58) and the result follows.

**6. Application to complementarity problems.** This section applies techniques developed in Section 5 to build confidence regions for stochastic complementarity problems, which are an important class of stochastic variational inequalities in which the set \( S \) in (3) equals \( \mathbb{R}_+^p \times \mathbb{R}^{n-p} \). The general formulation of a stochastic complementarity problem is

\[ 0 \in f_0(x) + N_{\mathbb{R}_+^p \times \mathbb{R}^{n-p}}(x), \]

(59)

where \( f_0 \) is as defined at the beginning of this paper. Section 6.1 below utilizes the special structure of \( \mathbb{R}_+^p \times \mathbb{R}^{n-p} \) to simplify the formulas for computing \( \Lambda_N \) and \( \Phi_N \). Following that, Section 6.2 presents a numerical example to demonstrate the methodology and illustrate the results.

**6.1 Specialization of general formulas.** Given the special structure of \( \mathbb{R}_+^p \times \mathbb{R}^{n-p} \), its normal manifold consists of a total of \( 3^p \) cells. Each cell is characterized by a unique partition of the index set \( \{1, \ldots, p\} \) as the union of three disjoint subsets \( I_0, I_+, I_- \). More specifically, we define the following family

\[ \mathcal{F} = \left\{ (I_0, I_+, I_-) \mid I_0 \cup I_+ \cup I_- = \{1, \ldots, p\}, \ I_0 \cap I_+ = I_0 \cap I_- = I_+ \cap I_- = \emptyset \right\}. \]
Then, each triple \((I_0, I_+, I_-)\) that belongs to \(\Psi\) defines a cell in the normal manifold of \(\mathbb{R}^p_+ \times \mathbb{R}^{n-p}\), given by

\[
C(I_0, I_+, I_-) = \{ x \in \mathbb{R}^n \mid x_i = 0, i \in I_0; \ x_i \geq 0, i \in I_+; \ x_i \leq 0, i \in I_- \}.
\] (60)

Conversely, each cell in the normal manifold of \(\mathbb{R}^p_+ \times \mathbb{R}^{n-p}\) can be represented in the form (60) with a unique triple \((I_0, I_+, I_-)\). The dimension of \(C(I_0, I_+, I_-)\) is \(n - |I_0|\), where \(|I_0|\) is the cardinality of \(I_0\).

In the rest of this subsection, we use the above characterization to simplify the definition of \(\Lambda_N(z)(h)\) in (34). The formula (34) enumerates cells in the normal manifold first by their dimensions and then in an arbitrary order within groups of equal dimensions. Here, we will enumerate the cells by enumerating elements of the family \(\Psi\). Indeed, when \(z \in \mathbb{R}^n\) and \(N \in \mathbb{N}\) are given, we only need to include elements of the following subfamily

\[\Psi(z, 1/g(N)) = \{ (I_0, I_+, I_-) \in \Psi \mid d(z, C(I_0, I_+, I_-)) < 1/g(N) \}\]

in the computation of \(\Lambda_N(z)(h)\).

Now, let \(C(I_0, I_+, I_-)\) be a given cell. By the remarks following the proof of Theorem 5.2, the \(B\)-derivative \(d\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(z)\) is the same, for all points \(z\) in the relative interior of \(C(I_0, I_+, I_-)\). We denote this function by \(\Psi_{I_0, I_+, I_-}\). The following lemma provides a precise formula for \(\Psi_{I_0, I_+, I_-}\).

**Lemma 6.1** Let \((I_0, I_+, I_-) \in \Psi\), \(z \in \text{ri} C(I_0, I_+, I_-)\), and let \(\Psi_{I_0, I_+, I_-} = d\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(z)\). Then, for each \(h \in \mathbb{R}^n\),

\[
(\Psi_{I_0, I_+, I_-}(h))_i = (d\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(z)(h))_i = \begin{cases} h_i & \text{if } i \in I_0 \text{ and } h_i \geq 0, \\ 0 & \text{if } i \in I_0 \text{ and } h_i \leq 0, \\ h_i & \text{if } i \in I_+, \\ 0 & \text{if } i \in I_-, \\ h_i & \text{if } i \in \{p+1, \ldots, n\}. \end{cases}
\] (61)

**Proof.** The components of \(\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(x)\) for each \(x \in \mathbb{R}^n\) are given by

\[
(\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(x))_i = \begin{cases} \max(x_i, 0) & \text{if } i = 1, \ldots, p, \\ x_i & \text{if } i = p+1, \ldots, n. \end{cases}
\]

The fact \(z \in \text{ri} C(I_0, I_+, I_-)\) implies that

\[z_i = 0, i \in I_0; \ z_i > 0, i \in I_+ \text{ and } z_i < 0, i \in I_-.
\]

Thus,

\[(\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(z))_i = 0, i \in I_0 \cup I_- \quad \text{and} \quad (\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(z))_i = z_i, i \in I_+ \cup \{p+1, \ldots, n\}.\]

Let \(h \in \mathbb{R}^n\), and let \(\epsilon\) be a positive scalar. As long as \(\epsilon\) is sufficiently small, one has \(z_i + \epsilon h_i > 0\) for each \(i \in I_+ \) and \(z_i + \epsilon h_i < 0\) for each \(i \in I_-\), which implies that

\[(\Pi_{\mathbb{R}^p_+ \times \mathbb{R}^{n-p}}(z + \epsilon h))_i = \begin{cases} \epsilon h_i & \text{if } i \in I_0 \text{ and } h_i \geq 0, \\ 0 & \text{if } i \in I_0 \text{ and } h_i \leq 0, \\ z_i + \epsilon h_i & \text{if } i \in I_+, \\ 0 & \text{if } i \in I_-, \\ z_i + \epsilon h_i & \text{if } i \in \{p+1, \ldots, n\}. \end{cases}
\]

This proves (61). \(\square\)

In the following theorem, we replace the function \(\Psi_I^k\) in (34) by \(\Psi_{I_0, I_+, I_-}\), and use Lemma 6.1 to derive a substantially simplified, and highly structured, formula for \(\Lambda_N(z)(\cdot)\). As shown by this theorem, \(\Lambda_N(z)(\cdot)\) is a piecewise linear function from \(\mathbb{R}^n\) to \(\mathbb{R}^n\). Moreover, on each set of the form \(Y \times \mathbb{R}^{n-p}\) with \(Y\) being an orthant of \(\mathbb{R}^p\), \(\Lambda_N(z)(\cdot)\) coincides with a linear function represented by a diagonal matrix. Each of the first \(p\) diagonal elements of these matrices may take one of two possible values, and the remaining \(n - p\) diagonal elements all equal one. Thus, one only needs to compute \(2p\) values to fully determine \(\Lambda_N(z)(\cdot)\).

**Theorem 6.1** For each \(N \in \mathbb{N}\) and each \(z \in \mathbb{R}^n\), define two \(p\)-dimensional vectors \(v_N^+(z)\) and \(v_N^-(z)\) as

\[
(v_N^+(z))_i = \sum_{(I_0, I_+, I_-) \in \Psi(z, 1/g(N))} 1_{I_0 \cup I_+}[i] \frac{1}{1/g(N) - d(z, C(I_0, I_+, I_-))} \left[ 1 - |I_0| \right]^{n-|I_0|} \frac{1}{1/g(N) - d(z, C(I_0, I_+, I_-))} \left[ 1 - |I_0| \right]^{n-|I_0|} 
\]

\[
(v_N^-(z))_i = \sum_{(I_0, I_+, I_-) \in \Psi(z, 1/g(N))} 1_{I_0 \cup I_-}[i] \frac{1}{1/g(N) - d(z, C(I_0, I_+, I_-))} \left[ 1 - |I_0| \right]^{n-|I_0|} \frac{1}{1/g(N) - d(z, C(I_0, I_+, I_-))} \left[ 1 - |I_0| \right]^{n-|I_0|} 
\] (62)
and
\[
(v_N^\pi(z))_i = \frac{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} 1_{I_+} \{i\} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|}}{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|}}
\]
for each \(i = 1, \ldots, p\). Let \(S = \mathbb{R}_+^p \times \mathbb{R}^{n-p}\) and \(\Lambda_N\) be defined as (34). Then, \(\Lambda_N(z)(\cdot)\) is a piecewise linear function from \(\mathbb{R}^n\) to \(\mathbb{R}^n\). Moreover, on each set of the form \(Y \times \mathbb{R}^{n-p}\), with \(Y\) being an orthant of \(\mathbb{R}^p\), \(\Lambda_N(z)(\cdot)\) is represented by the following \(n \times n\) diagonal matrix:
\[
\Lambda_N(z,Y) = \begin{bmatrix}
\lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \lambda_p & 0 & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 1
\end{bmatrix},
\]
where for each \(i = 1, \ldots, p\),
\[
\lambda_i = \begin{cases}
(v_N^\pi(z))_i & \text{if the } i\text{th components of vectors in } Y \text{ are nonnegative}, \\
(v_N^\pi(z))_i & \text{if the } i\text{th components of vectors in } Y \text{ are nonpositive}.
\end{cases}
\]

**Proof.** Let \(h \in \mathbb{R}^n\). Specializing (34) to the situation in which \(S = \mathbb{R}_+^p \times \mathbb{R}^{n-p}\), we find
\[
\Lambda_N(z)(h) = \frac{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} 1_{I_+} \{i\} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|} \Psi_{I_0,I_+,I_-}(h)}{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|}}.
\]
Let \(i \in \{1, \ldots, p\}\) be given. By Lemma 6.1, if \(h_i \geq 0\) then
\[
(\Lambda_N(z)(h))_i = \frac{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} 1_{I_+} \{i\} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|} h_i}{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|}} = (v_N^\pi(z))_i h_i,
\]
and if \(h_i \leq 0\) then
\[
(\Lambda_N(z)(h))_i = \frac{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} 1_{I_+} \{i\} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|} h_i}{\sum_{(I_0,I_+,I_-) \in \Psi(z,1/g(N))} [1/g(N) - d(z,C(I_0,I_+,I_-))]^{n-|I_0|}} = (v_N^\pi(z))_i h_i.
\]
On the other hand, for each \(i \in \{p+1, \ldots, n\}\), we have
\[
(\Lambda_N(z)(h))_i = h_i.
\]
To summarize, for each \(h \in \mathbb{R}^n\), we have
\[
(\Lambda_N(z)(h))_i = \begin{cases}
(v_N^\pi(z))_i h_i & \text{if } i \in \{1, \ldots, p\} \text{ and } h_i \geq 0, \\
(v_N^\pi(z))_i h_i & \text{if } i \in \{1, \ldots, p\} \text{ and } h_i \leq 0, \\
h_i & \text{if } i \in \{p+1, \ldots, n\}.
\end{cases}
\]
The theorem follows. \(\square\)

Using Theorem 6.1, we can specialize (44) to obtain a formula for \(\Phi_N(z)(\cdot)\), for situations in which \(S = \mathbb{R}_+^p \times \mathbb{R}^{n-p}\). See Corollary 6.1 below.

**Corollary 6.1** Let \(N \in \mathbb{N}\), \(S = \mathbb{R}_+^p \times \mathbb{R}^{n-p}\), and \(z \in \Pi^3_S(O)\). Suppose that the function \(F(\cdot,\xi(\omega))\) is continuously differentiable on \(O\) for almost every \(\omega \in \Omega\). Let \(f_N\) be as defined in (4), and let \(\Phi_N\) be as defined in (44). Then, for almost every \(\omega \in \Omega\), \(\Phi_N(z)(\cdot)\) is a piecewise linear function from \(\mathbb{R}^n\) to \(\mathbb{R}^n\), and is represented by the following matrix on each set of the form \(Y \times \mathbb{R}^{n-p}\) with \(Y\) being an orthant of \(\mathbb{R}^p\):
\[
df_N(\Pi_S(z))\Lambda_N(z,Y) + I_n - \Lambda_N(z,Y).
\]
In the above corollary, $I_n$ represents the $n$-dimensional identity matrix, and $\Lambda_N(z, Y)$ is as defined in Theorem 6.1. The proof of this corollary is straightforward and is omitted.

A final remark about the computation of $d(z, C(I_0, I_+, I_-))$. As commented at the end of the introductory section, any norm can be used to define this distance. In particular, using the $\| \cdot \|_\infty$ norm simplifies computation and leads to the following formula,

$$d(z, C(I_0, I_+, I_-)) = \max\{|z_i|, i \in I_0, \max(-z_i, 0), i \in I_+, \max(z_i, 0), i \in I_-\}.$$  

### 6.2 A numerical example.

In this subsection, we consider a stochastic linear complementarity problem, in which $n = 2$, $d = 6$, and $F : \mathbb{R}^2 \times \mathbb{R}^6 \to \mathbb{R}^2$ is defined by

$$F(x, \xi) = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} x_1 + \begin{bmatrix} \xi_5 \\ \xi_6 \end{bmatrix}. \quad (64)$$

Let

$$M_0 = \begin{bmatrix} E[\xi_1], & E[\xi_2] \\ E[\xi_3], & E[\xi_4] \end{bmatrix}, \quad b_0 = \begin{bmatrix} E[\xi_5] \\ E[\xi_6] \end{bmatrix},$$

and

$$M_N = N^{-1} \sum_{i=1}^N \begin{bmatrix} \xi_i^1 & \xi_i^2 \\ \xi_i^3 & \xi_i^4 \end{bmatrix}, \quad b_N = N^{-1} \sum_{i=1}^N \begin{bmatrix} \xi_i^5 \\ \xi_i^6 \end{bmatrix},$$

where $\xi^1, \cdots, \xi^N$ is a random sample of $N$ realizations of $\xi$. Then we have

$$f_0(x) = E[F(x, \xi)] = M_0 x + b_0 \text{ and } f_N(x) = M_N x + b_N \text{ for each } x \in \mathbb{R}^2.$$

The stochastic linear complementarity problem is

$$0 \in M_0 x + b_0 + N_{\mathbb{R}^2_+}(x), \quad (65)$$

and the corresponding SAA problem is

$$0 \in M_N x + b_N + N_{\mathbb{R}^2_+}(x). \quad (66)$$

Suppose now that $\xi$ follows the uniform distribution over a box in $\mathbb{R}^6$,

$$\{\xi \in \mathbb{R}^6 \mid (0, 0, 0, 0, -1, -1) \leq \xi \leq (2, 1, 2, 4, 1, 1)\}. \quad (67)$$

Thus we have

$$M_0 = \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix}, \quad b_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The matrix $M_0$ is a $P$-matrix, because all of its principal minors are positive (see, e.g., [3, p. 300] for the definition of $P$ matrices). It is easy to check that $x_0 = 0$ is a solution to (65); accordingly, $z_0 = x_0 - f_0(x_0) = 0$. In this example, the set $K$ defined in Assumption 3.2 equals $\mathbb{R}^2_+$, and the function $L$ defined in the same assumption is represented by the matrix $M_0$. The fact that $M_0$ is a $P$-matrix implies that the normal map $L_K$ is a homeomorphism from $\mathbb{R}^2$ to $\mathbb{R}^2$; see, e.g., [3, p.361-363]. It is not hard to check that Assumptions 3.1, 3.3 and 3.4 hold as well. Therefore, the conclusions of Theorems 3.1 and 5.1 hold for this example.

Indeed, linear complementarity problems defined by $P$-matrices have unique solutions globally, see, e.g., [3, Proposition 3.5.10]. Thus, $x_0 = 0$ is the unique solution for (65) in $\mathbb{R}^2$, and the SAA problems (66) almost surely have globally unique solutions, when $N$ is sufficiently large. In other words, in applying Theorem 5.1 to this example, we can let the neighborhood $Z$ defined in this theorem be $\mathbb{R}^2$.

#### 6.2.1 Scatter plots of SAA solutions.

Before applying the main methodology developed in this paper to generate confidence regions for $z_0$, we illustrate graphically how the SAA solutions $z_N$ are distributed. First, we fix the sample size to be $N = 10$, and choose different random seeds to generate 200 random samples of size 10. Each sample yields a SAA problem (66), which we solve using the PATH solver of GAMS, to find the SAA solution $z_{10}$. Each of these $z_{10}$ is represented by one of the 200 points in the scatter plot in Figure 1(a). Similarly, Figure 1(b) plots 200 scatter points, each representing a SAA solution $z_{30}$. 
In both graphs, the scatter points appear to match patterns of the regions well, demonstrating that shrinking the region for $\alpha = 0$ for each of which is enclosed by a fraction of an ellipse centered around $\alpha$ in the corresponding orthants. Consequently, for each given $\alpha$, the set (68) is a union of four pieces, each of which is enclosed by a fraction of an ellipse centered around $z_0 = 0$. Boundaries of these sets for $\alpha = 0, 0.1, 0.2, \cdots , 0.9$ are shown in both graphs of Figure 1, where the outermost curves correspond to $\alpha = 0.1$ and enclose a region of probability 0.9 for $z_N$, and so on. As is implied by the expression (68), shrinking the region for $z_{10}$ by a factor of $1/\sqrt{3}$ gives exactly the probability region for $z_{30}$ at the same level. In both graphs, the scatter points appear to match patterns of the regions well, demonstrating that the $\chi^2$ distribution with $n$ degrees of freedom is a good approximation for the real distribution of (69).

6.2.2 Confidence regions for the true solution. In this subsection, we implement the method developed in this paper, to establish confidence regions for the true solution $z_N$.

Figure 1 also displays the region

$$R = \{ z \in \mathbb{R}^2 \mid N[d(f_0)_{R^2_+}(z_0)(z - z_0)]^T \Sigma_0^{-1} [d(f_0)_{R^2_+}(z_0)(z - z_0)] \leq \chi^2_N(\alpha) \}. \tag{68}$$

Probability that $z_N$ belongs to the set $R$ above is approximately $(1 - \alpha)$ since

$$N[d(f_0)_{R^2_+}(z_N - z_0)]^T \Sigma_0^{-1} [d(f_0)_{R^2_+}(z_0)(z_N - z_0)] \tag{69}$$

asymptotically (as $N \to \infty$) follows the $\chi^2$ distribution with $n$ degrees of freedom. In this example, the covariance matrix of $F(x_0, \xi) = (\xi_5, \xi_6)$ is given by

$$\Sigma_0 = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/3 \end{bmatrix}.$$  

$d\Pi_{R^2_+}(z_0)$ is a piecewise linear function represented by matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

in orthants $\mathbb{R}^2_+, \mathbb{R}_+ \times \mathbb{R}_-, \mathbb{R}_- \times \mathbb{R}_+$ and $\mathbb{R}^2$ respectively, and $d(f_0)_{R^2_+}(z_0) = M_0 \circ d\Pi_{R^2_+}(z_0) + I - d\Pi_{R^2_+}(z_0)$ is a piecewise linear function represented by matrices

$$\begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1/2 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

in the corresponding orthants. Consequently, for each given $\alpha$, the set (68) is a union of four pieces, each of which is enclosed by a fraction of an ellipse centered around $z_0 = 0$. Boundaries of these sets for $\alpha = 0.1, 0.2, \cdots , 0.9$ are shown in both graphs of Figure 1, where the outermost curves correspond to $\alpha = 0.1$ and enclose a region of probability 0.9 for $z_N$, and so on. As is implied by the expression (68), shrinking the region for $z_{10}$ by a factor of $1/\sqrt{3}$ gives exactly the probability region for $z_{30}$ at the same level. In both graphs, the scatter points appear to match patterns of the regions well, demonstrating that the $\chi^2$ distribution with $n$ degrees of freedom is a good approximation for the real distribution of (69).

We start by letting $N = 10$ and simulate $\xi_1, \cdots \xi_N$ according to a Uniform distribution on the box in (67). The sample average function we obtain is

$$f_{10}(x) = M_{10}x + b_{10} = \begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix}x + \begin{bmatrix} -0.1319 \\ -0.2906 \end{bmatrix}.$$
The SAA solution is $x_{10} = (0.0782, 0.1097)$, with $z_{10} = x_{10} - f_{10}(x_{10}) = (0.0782, 0.1097)$. The sample covariance matrix of $F(x_{10}, \xi)$ is given by

$$\Sigma_{10} = \begin{bmatrix} 0.4169 & 0.0137 \\ 0.0137 & 0.1865 \end{bmatrix}. $$

We choose $g(N) = N^{1/3}$. Using formulas in Theorem 6.1, we find

$$v_{10}^+(z_{10}) = \begin{bmatrix} 0.7948 \\ 0.8024 \end{bmatrix} \quad \text{and} \quad v_{10}^-(z_{10}) = \begin{bmatrix} 0.2269 \\ 0.2448 \end{bmatrix}. $$

Accordingly, $\Lambda_{10}(z_{10})(\cdot)$ is a piecewise linear map represented by matrices

$$\begin{bmatrix} 0.7948 & 0 \\ 0 & 0.8024 \end{bmatrix}, \begin{bmatrix} 0.7948 & 0 \\ 0 & 0.2448 \end{bmatrix}, \begin{bmatrix} 0.2269 & 0 \\ 0 & 0.8024 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0.2269 & 0 \\ 0 & 0.2448 \end{bmatrix}$$

in orthants $\mathbb{R}_+^2$, $\mathbb{R}_+ \times \mathbb{R}_-$, $\mathbb{R}_- \times \mathbb{R}_+$ and $\mathbb{R}_-^2$ respectively. Applying the formula in Corollary 6.1, we find $\Phi_{10}(z_{10})(\cdot)$ to be a piecewise linear map represented by matrices

$$\begin{bmatrix} 0.9437 & 0.4333 \\ 0.5989 & 1.8915 \end{bmatrix}, \begin{bmatrix} 0.9437 & 0.1322 \\ 0.5989 & 1.2720 \end{bmatrix}, \begin{bmatrix} 0.9839 & 0.4333 \\ 0.1710 & 1.8915 \end{bmatrix}, \begin{bmatrix} 0.9839 & 0.1322 \\ 0.1710 & 1.2720 \end{bmatrix} $$

in corresponding orthants. Figure 2(a) shows confidence regions for $z_0$ given by (14). These regions are centered around $z_{10}$, marked by ‘$\times$’ in the graph. From the innermost to the outermost, the curves correspond to boundaries of confidence regions at levels $0.1, \cdots, 0.9$ respectively. The true solution $z_0$, known to be 0 for this example, is marked by ‘$+$’ and lies within the 90% confidence region.

Figure 2(b) shows confidence regions for $z_0$ obtained from a different SAA problem, in which $N = 30,

$$f_{30}(x) = M_{30}x + b_{30} = \begin{bmatrix} 0.8893 & 0.5736 \\ 0.9314 & 2.1423 \end{bmatrix} x + \begin{bmatrix} 0.0483 \\ 0.0114 \end{bmatrix},$$

$x_{30} = 0$ and $z_{30} = (-0.0483, -0.0114)$. These regions are smaller than the corresponding regions in Figure 2(a), as expected. The true solution $z_0$ lies within the 20% region.

For the present example, we have the exact expression of $f_0$ and the true values of $z_0$ and $\Sigma_0$, so sets (12) are directly computable (recall that $L_0$ is exactly $d(f_0)|_{\mathbb{R}_2^2}(z_0)$). These sets can be considered as the ideal confidence regions for $z_0$. Figure 3 shows these regions centered around the above SAA solutions $z_{10}$ and $z_{30}$; note that these regions can also be obtained by rotating those in Figure 1 by 180 degrees and then shifting them properly. A comparison between Figures 2 and 3 shows that confidence regions obtained by the proposed method appear to have similar shapes and sizes as the corresponding
ideal regions, and that such similarity is improved as the sample size increases. For example, the areas
enclosed by the outermost curves in Figures 2(a) and 3(a) are 0.3014 and 0.3988 respectively, with their
difference being about 24% of 0.3988. The corresponding areas in 2(b) and 3(b) are 0.1193 and 0.1329
respectively, which are different by about 10% of 0.1329.

6.2.3 Evaluation of normality. Confidence regions built in the preceding subsection is based on
the fact proved in Theorem 3.1, that the expression
\[ \sqrt{N} \Sigma_N^{-1/2} \Phi_N(z_N)(z_N - z_0) \]
(70)
asymptotically follows the \( n \)-dimensional standard normal distribution. This subsection uses \( \chi^2 \) plots to
assess the normality numerically.

We use the same SAA problems that were used to generate scatter points in Figure 1. For each of the
200 SAA problems with sample size \( N = 10 \), we compute the squared distance
\[ N \left[ \Phi_N(z_N)(z_N - z_0) \right]^T \Sigma_N^{-1} \left[ \Phi_N(z_N)(z_N - z_0) \right], \]
and order these distances from smallest to largest as \( d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(200)}^2 \). For each \( j = 1, \cdots, 200 \),
let \( q_{c,2}(j - 1/2)/200 \) be the \( 100(j - 1/2)/200 \) quantile of the \( \chi^2 \) distribution with 2 degrees of freedom.
We then graph the pairs \((q_{c,2}(j - 1/2)/n), d_{(j)}^2\) for \( j = 1, \cdots, 200 \) in Figure 4(a), in which the horizontal
axis is for quantiles and the vertical axis is for squared distances. Figure 4(b) is obtained similarly, from
the 200 SAA problems with sample size \( N = 30 \).

The points in Figure 4(a) nearly follow a straight line through the origin, but the slope of this line
appears to be larger than 1, indicating that the distances tend to be too large. The situation is much
improved with \( N = 30 \), as is shown in Figure 4(b). Points here are nearly on a straight line through
the origin with slope 1. In addition, 91 among the 200 distances, a proportion of 45.5%, are less than or
equal to \( q_{c,2}(0.5) = 1.3863 \). These indicate that, for the present example, (70) nearly follows the standard
normal distribution when \( N = 30 \).

Appendix A. Proof of Lemma 3.1. Let the normed space \( P \) in Theorem 2.1 be \( C^1(X, \mathbb{R}^n) \), and
let \( O_1 \) be an open and convex neighborhood of \( x_0 \) in \( X \). Define
\[ h(x, f) = f(x) \]
(71)
for each \( x \in O_1 \) and \( f \in C^1(X, \mathbb{R}^n) \). We need to verify the hypotheses of Theorem 2.1. We have
\[ \| h(x, f) - h(x, g) \| = \| f(x) - g(x) \| = \| (f - g)(x) \| \leq \| f - g \|_{1,X} \]
for each \( x \in O_1 \) and \( f, g \in C^1(X, \mathbb{R}^n) \). It follows that for each \( x \in O_1 \) the function \( h(x, \cdot) \) is Lipshitz on
\( C^1(X, \mathbb{R}^n) \) with modulus 1.
By the definition of $h$, its partial derivative in $x$ at $(x_0, f_0)$ is $L = df_0(x_0)$. Next, we show that this partial derivative is strong. For each $x, x' \in O_1$ and $f \in C^1(X, \mathbb{R}^n)$, we have
\[
|h(x, f) - h(x', f) - L(x - x')| \\
= |f(x) - f(x') - L(x - x')| \\
= \left| \int_0^1 [df'(x + t(x - x'))(x - x') - L(x - x')] dt \right| \\
\leq \|x - x'\| \int_0^1 \|df'(x + t(x - x')) - L\| dt.
\]
Consequently, for each $\epsilon > 0$, there exist neighborhoods $U$ of $x_0$ in $O_1$ and $V$ of $f_0$ in $C^1(X, \mathbb{R}^n)$ such that
\[
|h(x, f) - h(x', f) - L(x - x')| \\
= \left| \int_0^1 [df'(x + t(x - x'))(x - x') - L(x - x')] dt \right| \\
\leq \epsilon \|x - x'\|
\]
whenever $x$ and $x'$ belong to $U$ and $f$ belongs to $V$.

Finally, it follows from the equality
\[
h(x_0, f) - h(x_0, f_0) = f(x_0) - f_0(x_0) = (f - f_0)(x_0)
\]
that $h$ has a partial F-derivative $dfh(x_0, f_0)$ in $f$ at $(x_0, f_0)$, given by
\[
d_fh(x_0, f_0)(g) = g(x_0).
\]

Thus we have verified all hypotheses of Theorem 2.1. Conclusions of the present lemma follow directly from this theorem.

**Appendix B. Proof of Theorem 4.3.** Corollary 7.17 of [1] shows that, under Assumption 3.1 $\sqrt{N}(f_N - f_0)$ converges in distribution in $C(X, \mathbb{R}^n)$ to a $\tilde{Y}$ such that for each finite subset $\{x_1, x_2, \cdots, x_m\}$ of $X$, the random vector $(\tilde{Y}(x_1), \cdots, \tilde{Y}(x_m))$ has a multivariate normal distribution with zero mean and the same covariance matrix as that of
\[
(F(x_1, \xi), \cdots, F(x_m, \xi)).
\]
Furthermore, from the same result, it follows that $\sqrt{n}(d_xf_N - df_0)$ converges in distribution in $C(X, \mathbb{R}^{n \times n})$ to some random variable $\tilde{U}$. These two statements show that the collection of $C^1(X, \mathbb{R}^n)$ valued random variables $\sqrt{n}(f_N - f_0)$ is tight. Also any weak limit point $Y^*$ of $\sqrt{n}(f_N - f_0)$ in $C^1(X, \mathbb{R}^n)$ has the property that for each finite subset $\{x_1, x_2, \ldots, x_m\}$ of $X$, the random vector $(Y^*(x_1), \ldots, Y^*(x_m))$ has the same distribution as that of the random vector $(\tilde{Y}(x_1), \ldots, \tilde{Y}(x_m))$. Since finite dimensional distributions determine the probability law of a $C^1(X, \mathbb{R}^n)$ valued random variable we now have that $\sqrt{n}(f_N - f_0)$ converges weakly in $C^1(X, \mathbb{R}^n)$, to the (unique in law) random variable $Y$ with the property that for each finite subset $\{x_1, x_2, \ldots, x_m\}$ of $X$, the random vector $(Y(x_1), \ldots, Y(x_m))$ has a multivariate normal distribution as in the statement of the theorem. The result follows.

\section*{Acknowledgments.} Research of the first author was partially supported by National Science Foundation (DMS-0807893 and DMS-1109099). Research of the second author was partially supported by the Army Research Office (Grant W911NF-10-1-0158), National Science Foundation (DMS-1004418 and DMS-1016441), and the US-Israel Binational Science Foundation (Grant 2008466). The first author thanks Professor Stephen M. Robinson for introducing this research topic to her. We thank the two referees for helpful suggestions.

\section*{References}

\begin{enumerate}
\item Mert C. Demir. \textit{Asymptotics and confidence regions for stochastic variational inequalities}. Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin–Madison, Madison, WI, 2000.
\end{enumerate}


