

# SYMMETRIC CONFIDENCE REGIONS AND CONFIDENCE INTERVALS FOR NORMAL MAP FORMULATIONS OF STOCHASTIC VARIATIONAL INEQUALITIES

SHU LU\*

**Abstract.** Stochastic variational inequalities (SVI) model a large class of equilibrium problems subject to data uncertainty, and are closely related to stochastic optimization problems. The SVI solution is usually estimated by a solution to a sample average approximation (SAA) problem. This paper considers the normal map formulation of an SVI, and proposes a method to build asymptotically exact confidence regions and confidence intervals for the solution of the normal map formulation, based on the asymptotic distribution of SAA solutions. The confidence regions are single ellipsoids with high probability. We also discuss the computation of simultaneous and individual confidence intervals.

**Key words.** confidence region, confidence interval, stochastic variational inequality, sample average approximation, stochastic optimization, normal map

**1. Introduction.** This paper considers a stochastic variational inequality (SVI), in which the function that defines the variational inequality is an expectation function. 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  $\mathbb{R}^q$ , and  $F$  be a measurable function from  $O \times \Xi$  to  $\mathbb{R}^q$ , such that for each  $x \in O$  the expectation  $f_0(x) = E[F(x, \xi)]$  is well defined. Let  $S$  be a polyhedral convex set in  $\mathbb{R}^q$ . The SVI problem is to find a point  $x \in S \cap O$  such that

$$0 \in f_0(x) + N_S(x), \quad (1.1)$$

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

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

We use  $\langle \cdot, \cdot \rangle$  to denote the scalar product of two vectors of the same dimension. Here and elsewhere, the symbol  $\subset$  stands for set inclusion and allows the two sets compared to coincide.

The function  $f_0$  defined above is a function from  $O$  to  $\mathbb{R}^q$ . We are interested in situations in which  $f_0$  does not have a closed form expression and is approximated by a sample average function. Let  $\xi^1, \dots, \xi^n$  be independent and identically distributed (i.i.d.) random variables with distribution same as that of  $\xi$ . Define the sample average function  $f_n : O \times \Omega \rightarrow \mathbb{R}^q$  by

$$f_n(x, \omega) = n^{-1} \sum_{i=1}^n F(x, \xi^i(\omega)). \quad (1.2)$$

The sample average approximation (SAA) problem is to find a point  $x \in S \cap O$  such that

$$0 \in f_n(x, \omega) + N_S(x). \quad (1.3)$$

In the rest of this paper, we will write  $f_n(x, \omega)$  as  $f_n(x)$  when clear from the context. We will consider the *normal map* formulations (to be defined below) of both (1.1) and (1.3). The major objective is to

---

\*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 355 Hanes Building, CB#3260, Chapel Hill, NC 27599-3260 ([shulu@email.unc.edu](mailto:shulu@email.unc.edu)).

develop a method to build confidence regions and confidence intervals for the solution of the normal map formulation of (1.1). We will explain how to obtain confidence regions and intervals for the solution of (1.1) in Section 6. For brevity, we refer to a solution to (1.1) or its normal map formulation as a true solution, and a solution to (1.3) or its normal map formulation as an SAA solution.

Under certain regularity conditions, the SAA solutions almost surely converge to a true solution as the sample size  $n$  goes to infinity; see Gürkan, Özge and Robinson [9], King and Rockafellar [11], and Shapiro, Dentcheva and Ruszczyński [35, Section 5.2.1]. [11, Theorem 2.7] and [35, Section 5.2.2] provided the asymptotic distribution of SAA solutions. Xu [41] showed that the SAA solutions converge to the set of true solutions in probability at an exponential rate under some assumptions on the moment generating functions of certain random variables; see [36] for related results on the exponential convergence rate. For work on stability of stochastic optimization problems, see [3, 4, 30, 34] and numerous references therein.

As discussed in Pflug [21], there are two basic approaches to constructing confidence regions for the true solution of a stochastic optimization problem. The first approach is based on the asymptotic distribution of SAA solutions, and the second is based on properties about boundedness in probability with known tail behavior. The latter approach was used in [21] to construct universal confidence sets for the true solution, among other results. Vogel [40] continued that approach and developed a general method for universal confidence sets, taking account for random feasible sets and nonunique optimal solutions.

The method in this paper belongs to the first approach in the above classification. Our goal is to construct asymptotically exact confidence regions that are convenient to compute along with confidence intervals, under assumptions that guarantee the SVI and its SAA approximations to have locally unique solutions. Our work is closely related to Demir [2]. In that dissertation, Demir considered the normal map formulation of the SVI and obtained in [2, Equation (36)] an expression for confidence regions of the true solution. That expression depended on quantities not directly computable. [2, Theorem 3.16] modified it, resulting in a formula for the set called  $S^n$  in [2]. That formula looks similar to (4.2) of the present paper, but it uses  $d(f_0)_S(z_n)(z_n - z)$  (in our notation) instead of  $d(f_n)_S(z_n)(z - z_n)$  in (4.2). We show in this paper that the set (4.2) is an asymptotically exact confidence region of the true solution, by proving that the probability for it to contain the true solution converges to the prescribed level of confidence. The development in [2] did not justify its method with an asymptotic exactness result, aside from using some restrictive assumptions.

The starting point of the asymptotic analysis in this paper is Theorem 3.1, which is proved in [16] and related to results in [2, 11, 35]. Equations (3.4) and (3.5) in that theorem describe the asymptotic distribution of the solution (denoted by  $z_n$ ) to the normal map formulation of the SAA problem (1.3) in terms of a piecewise linear function  $L_K$  and a normal random vector  $Y_0$ . Because  $L_K$  depends on the true solution  $z_0$  and is unknown before  $z_0$  is found, we need to replace it by a suitable estimator in order to establish a confidence region for  $z_0$ . For the general situations considered in this paper, the dependence of  $L_K$  on the location of  $z_0$  is *discontinuous*, due to the nonsmooth structure of  $S$ . Such discontinuity is a major issue to be addressed for establishing computable confidence regions, and this paper handles this issue using a different approach from those in [15, 16].

In [15, 16],  $L_K$  is replaced by estimators designed to converge to  $L_K$  in probability. The consistency of those estimators relies on the exponential convergence rate of  $z_n$  to  $z_0$ . When  $L_K$  is piecewise linear with multiple pieces, its estimators constructed in [15, 16] have multiple pieces with high probability, producing asymmetric confidence regions that are factions of ellipsoids pieced together. In the present paper, we use a property of general piecewise affine functions (Theorem 2.2) to show that for large  $n$

the two vectors  $-d(f_n)_S(z_n)(z_0 - z_n)$  and  $L_K(z_n - z_0)$  are close to each other even though  $d(f_n)_S(z_n)$  may be a very different function from  $L_K$ . Accordingly, we can use  $-d(f_n)_S(z_n)(z_0 - z_n)$  to replace  $L_K(z_n - z_0)$  in (3.5) without losing the limiting property (Theorem 3.3). This leads to a computable confidence region for  $z_0$  given in (4.2) or (4.6) without the need of constructing a consistent estimator for  $L_K$ . The main advantage with the new method comes from Proposition 3.5, which implies that the confidence region built from this method is with high probability a single ellipsoid (or can be approximated by a degenerate ellipsoid in the singular case). This brings much efficiency for describing confidence regions or computing simultaneous confidence intervals. Another advantage of the new method is that it does not rely on the exponential convergence rate of  $z_n$ . Lastly, we mention that with this method the confidence region obtained from a particular  $z_n$  may be very different from that obtained from another  $z_n$ , since  $d(f_n)_S(z_n)$  can be very different for different  $z_n$ . This does not conflict with the asymptotic exactness of the confidence region.

Given the results on confidence regions and simultaneous confidence intervals, a natural question is whether individual confidence intervals for components of  $z_0$  can be computed in a parallel manner, by using  $d(f_n)_S(z_n)$  to replace  $L_K$  in (3.4). Since  $d(f_n)_S(z_n)$  is an invertible linear function with high probability, the individual confidence interval given by this approach can be expressed by a closed-form formula (5.17). Theorem 5.1 shows that the probability for that interval to contain the  $j$ th component of  $z_0$  converges to a quantity related to the random variable  $\Gamma = (L_K)^{-1}(Y_0)$ , and that quantity equals the desired confidence level  $1 - \alpha$  when the condition in (5.5) holds. Based on that convergence result and given the simple format of (5.17), one can use (5.17) as an approximative confidence interval for  $(z_0)_j$  to compare with confidence intervals obtained from other methods (such as methods developed in [13] to compute asymptotically exact individual confidence intervals using estimators in [15, 16]), which generally require more computation when  $L_K$  has more than two pieces.

Examples of stochastic variational inequalities of the form (1.1) include stochastic Nash equilibrium problems in which players' cost functions are expected values of certain random variables, such as the energy market problem studied in [9, 10]. Stochastic variational inequalities also arise as first-order conditions of optimization problems whose objective functions are expectations. If the exact values of some coefficients of the objective function are unknown, and are estimated using sample data average, then the solution obtained for such a problem is essentially an SAA solution. The method of this paper can be applied to such problems to provide a quantitative measure on the effect of sample variations on solutions obtained for those problems. We have applied this method to a type of statistical learning problems [17].

Below we briefly introduce the normal map formulation of variational inequalities. The normal map induced by the function  $f_0 : O \rightarrow \mathbb{R}^q$  and the polyhedron  $S \subset \mathbb{R}^q$  is defined to be a function  $(f_0)_S : \Pi_S^{-1}(O) \rightarrow \mathbb{R}^q$ , with

$$(f_0)_S(z) = f_0(\Pi_S(z)) + (z - \Pi_S(z)) \text{ for each } z \in \Pi_S^{-1}(O), \quad (1.4)$$

where  $\Pi_S(z)$  denotes the Euclidean projection of  $z$  on  $S$ , and  $\Pi_S^{-1}(O)$  is the set of points  $z \in \mathbb{R}^q$  such that  $\Pi_S(z) \in O$ . If a point  $x \in S \cap O$  satisfies (1.1), then the point  $z = x - f_0(x)$  satisfies  $\Pi_S(z) = x$  and

$$(f_0)_S(z) = 0. \quad (1.5)$$

Conversely, if  $z$  satisfies (1.5), then  $x = \Pi_S(z)$  satisfies  $x - f_0(x) = z$  and solves (1.1). Thus, equation (1.5) is an equivalent formulation for (1.1), and is referred to as the normal map formulation of (1.1).

In general, for any function  $g$  from (a subset of)  $\mathbb{R}^q$  to  $\mathbb{R}^q$  and any closed and convex set  $C$  in  $\mathbb{R}^q$ , one can define the normal map induced by  $g$  and  $C$ , denoted by  $g_C$ , in the same way as (1.4) with  $g$  in place of  $f$  and  $C$  in place of  $S$ . For example, the normal map induced by the sample average function  $f_n$  and the set  $S$  is a function  $(f_n)_S : O \rightarrow \mathbb{R}^q$  defined as

$$(f_n)_S(z) = f_n(\Pi_S(z)) + (z - \Pi_S(z)) \text{ for each } z \in \Pi_S^{-1}(O).$$

The following is the normal map formulation for the SAA problem (1.3):

$$(f_n)_S(z) = 0. \tag{1.6}$$

Equation (1.6) is related to (1.3) in the same way as (1.5) is to (1.1).

The above definition for the normal map is from Robinson [26, 27]. The normal map concept is closely related to the Minty parametrization [18], in the sense that the variable  $z$  can be considered as the parameter in the Minty parametrization of the graph of  $N_S$ : the mapping  $z \rightarrow (\Pi_S(z), z - \Pi_S(z))$  is one-to-one from  $\mathbb{R}^q$  onto the graph of  $N_S$ , and the equation (1.5) is a reformulation of (1.1) in terms of  $z$  using that parametrization.

Because  $S$  is a polyhedral convex set by assumption, the Euclidean projector  $\Pi_S$  is a piecewise affine function (see Section 2 below for the precise definition of piecewise affine functions). It coincides with an affine function on each of a family of finitely many  $q$ -dimensional polyhedral convex sets. This family is called the normal manifold of  $S$ , and each set in this family is called an  $q$ -cell, where the symbol  $q$  refers to the dimension of those sets. The union of all the  $q$ -cells is  $\mathbb{R}^q$ . Any two distinct  $q$ -cells are either disjoint, or meet at a common proper face of them. (A face of a convex set  $P$  in  $\mathbb{R}^q$  is defined to be a convex subset  $F$  of  $P$  such that if  $x_1$  and  $x_2$  belong to  $P$  and  $\lambda x_1 + (1 - \lambda)x_2 \in F$  for some  $\lambda \in (0, 1)$ , then  $x_1$  and  $x_2$  actually belong to  $F$ , see, e.g., [28].  $F$  is a proper face of  $P$ , if it is a nonempty face of  $P$  and is not  $P$  itself.) For detailed discussions on the normal manifold and properties of piecewise affine functions, see [22, 23, 26, 32].

To illustrate the normal manifold concept, consider the example when  $S = \mathbb{R}_+^q$ , the nonnegative orthant in  $\mathbb{R}^q$ . For this example, the projector  $\Pi_S$  coincides with an affine function when restricted to each fixed orthant of  $\mathbb{R}^q$ , and that affine function is different for a different orthant: for example we have  $\Pi_S(z) = z$  for points  $z \in \mathbb{R}_+^q$ ,  $\Pi_S(z) = 0$  for  $z \in \mathbb{R}_-^q$ , and  $\Pi_S(z) = (z_1, 0, \dots, 0)$  for  $z \in \mathbb{R}_+ \times \mathbb{R}_-^{q-1}$ , so  $\Pi_S$  coincides with the identity function, the zero function, or the projector onto the  $z_1$  axis, when restricted on  $\mathbb{R}_+^q$ ,  $\mathbb{R}_-^q$  or  $\mathbb{R}_+ \times \mathbb{R}_-^{q-1}$  respectively. Accordingly, the normal manifold of  $\mathbb{R}_+^q$  is the family of all orthants in  $\mathbb{R}^q$ .

Below we introduce some terminology and notation. A subset  $K$  of  $\mathbb{R}^q$  is called a cone if  $\mu x \in K$  whenever  $x \in K$  and  $\mu$  is a positive real number. For a set  $C \subset \mathbb{R}^q$ ,  $\text{int } C$  denotes its interior, and  $\text{cone } C$  is the smallest cone that contains it. (This definition for cone  $C$  is equivalent to the one given in [32, Section 2.1.1] for polyhedral convex sets  $C$ . In this paper we will only apply the definition to such sets.) 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  $\mathcal{N}(0, \Sigma)$  to denote a Normal random vector with covariance matrix  $\Sigma$ . Weak convergence of  $k$ -dimensional random variables  $Y_n$  to  $Y$  will be denoted as  $Y_n \Rightarrow Y$ , which means that  $Ef(Y_n)$  converges to  $Ef(Y)$  for all bounded continuous functional  $f$  on  $\mathbb{R}^k$ . Following [24], a locally Lipschitz function  $g : \mathbb{R}^q \rightarrow \mathbb{R}^m$  is said to be *B-differentiable* at a point  $x_0 \in \mathbb{R}^q$  if there is a positively homogeneous function  $G : \mathbb{R}^q \rightarrow \mathbb{R}^m$ , such that

$$g(x_0 + v) = g(x_0) + G(v) + o(v). \tag{1.7}$$

(Recall that a function  $G$  is positively homogeneous, if  $G(\lambda v) = \lambda G(v)$  for each nonnegative real number  $\lambda$  and each  $v \in \mathbb{R}^q$ .) Such a function  $G$  is called the B-derivative of  $g$  at  $x_0$  and is denoted as  $dg(x_0)$ . Note that  $dg(x_0)(h)$  is exactly the directional derivative of  $g$  at  $x_0$  along the direction  $h$ . Indeed, as pointed out in [33], for a locally Lipschitz function in finite-dimensional spaces, B-differentiability (called the directional differentiability in the sense of Fréchet in [33]) of the function is equivalent to directional differentiability of the function for all directions. If  $g$  is differentiable at  $x_0$ , then the B-derivative  $dg(x_0)$  coincides with the standard Fréchet derivative.

The rest of this paper is organized as follows. Section 2 is a discussion on general piecewise affine functions. Section 3 establishes the main asymptotic distribution results. Section 4 provides methods to build confidence regions and simultaneous confidence intervals. Section 5 discusses computation of individual confidence intervals. Section 6 concludes the paper with numerical examples.

**2. Piecewise affine functions.** This section discusses general piecewise affine functions, and proves some properties to be used in subsequent development. The notation in this section is independent of the notation in the rest of this paper.

By definition, a continuous function  $f$  from  $\mathbb{R}^q$  to  $\mathbb{R}^m$  is called *piecewise affine* if there exists a finite family of affine functions  $f_j : \mathbb{R}^q \rightarrow \mathbb{R}^m, j = 1, \dots, k$ , such that the inclusion  $f(x) \in \{f_1(x), \dots, f_k(x)\}$  holds for each  $x \in \mathbb{R}^q$  [32]. The affine functions  $f_j, j = 1, \dots, k$  are called selection functions of  $f$ . If all  $f_j$ 's are linear functions then  $f$  is called *piecewise linear*.

A closely related concept is the *polyhedral subdivision* [5, 32]. A polyhedral subdivision of  $\mathbb{R}^q$  is a finite collection of polyhedral convex sets in  $\mathbb{R}^q, \mathfrak{P} = \{P_1, \dots, P_l\}$ , that satisfies the following conditions:

1. Each  $P_i$  is a polyhedral convex set of dimension  $q$ .
2. The union of all  $P_i$  is  $\mathbb{R}^q$ .
3. The intersection of each two  $P_i$  and  $P_j, 1 \leq i \neq j \leq l$ , is either empty or a common proper face of both  $P_i$  and  $P_j$ .

If each  $P_i \in \mathfrak{P}$  is a polyhedral convex cone, then  $\mathfrak{P}$  is a *conical subdivision*. It was shown in [32, Proposition 2.2.3] that for any piecewise affine function  $f$  there corresponds a polyhedral subdivision  $\mathfrak{P}$  of  $\mathbb{R}^q$  such that  $f$  coincides with an affine function on each  $P \in \mathfrak{P}$ . If  $f$  is piecewise linear, then the corresponding  $\mathfrak{P}$  is a conical subdivision. For example, if  $S = \mathbb{R}_+^q$ , then the Euclidean projector  $\Pi_S$  is piecewise linear, and the family of all orthants in  $\mathbb{R}^q$  is the corresponding conical subdivision. In general, for any polyhedral convex set  $S$  the Euclidean projector  $\Pi_S$  is piecewise affine, with the normal manifold of  $S$  being the corresponding polyhedral subdivision.

In the rest of this section, let  $f : \mathbb{R}^q \rightarrow \mathbb{R}^m$  be a piecewise affine function with the corresponding subdivision  $\mathfrak{P}$ . Clearly, if  $f$  is represented by different selection functions on  $P_1 \in \mathfrak{P}$  and  $P_2 \in \mathfrak{P}$ , and  $x \in P_1 \cap P_2$ , then  $f$  is nondifferentiable at  $x$ . However, it is well known that  $f$  is B-differentiable at any point in  $\mathbb{R}^q$ ; below we explain a formula for the B-derivative of  $f$  at a point  $x \in \mathbb{R}^q$ . For the derivation of this formula see [5] and [32, Proposition 2.2.6]. Let

$$\mathfrak{P}(x) = \{P \in \mathfrak{P} \mid x \in P\}$$

be the subfamily of  $\mathfrak{P}$  that consists of elements in  $\mathfrak{P}$  containing  $x$ . We use  $|\mathfrak{P}(x)|$  to denote the union of all  $P \in \mathfrak{P}(x)$ , called the *underlying set* of  $\mathfrak{P}(x)$  in algebraic topology [19]. We caution the reader that  $|\mathfrak{P}(x)|$  here does not mean the cardinality of  $\mathfrak{P}(x)$ . It is obvious that  $x$  belongs to the interior of  $|\mathfrak{P}(x)|$ . For each  $x \in \mathbb{R}^q$ , define the following family of polyhedral convex cones:

$$\mathfrak{P}'(x) = \{\text{cone}(P - x) \mid P \in \mathfrak{P}(x)\}.$$

The family  $\mathfrak{P}'(x)$  is a conical subdivision of  $\mathbb{R}^q$ . The B-derivative  $df(x)$  of  $f$  at  $x$  is a piecewise linear function from  $\mathbb{R}^q$  to  $\mathbb{R}^m$ , whose corresponding subdivision is exactly  $\mathfrak{P}'(x)$ . If  $f$  coincides with the affine function  $Ax + b$  on the polyhedral convex set  $P \in \mathfrak{P}(x)$ , then

$$df(x)(h) = Ah \text{ for each } h \in \text{cone}(P - x). \quad (2.1)$$

A consequence of (2.1) is that the selection functions of  $df(x)$  are exactly the linear parts of selection functions of  $f$  on elements of  $\mathfrak{P}(x)$ . In particular, if  $x$  is contained in the interior of some  $P \in \mathfrak{P}$ , then  $df(x)$  is a linear map.

While for a fixed point  $x \in \mathbb{R}^q$  the B-derivative  $df(x)$  is a continuous function on  $\mathbb{R}^q$ , the dependence of  $df(x)$  on  $x$  is discontinuous, because  $df(x)$  changes abruptly to a very different function as  $x$  moves from the interior of some  $P \in \mathfrak{P}$  to its boundary. For example consider  $\Pi_S$  with  $S = \mathbb{R}_+^q$  again. As a piecewise linear function,  $\Pi_S$  is B-differentiable at any point  $z \in \mathbb{R}^q$ . At any  $z$  in the interior of  $\mathbb{R}_+^q$ , the B-derivative  $d\Pi_S(z)$  is the identity map. At any  $z$  with  $z_1 = 0$  and  $z_i > 0$  for  $i = 2, \dots, q$ ,  $d\Pi_S(z)$  is a piecewise linear map with two pieces:

$$\text{For each } h \in \mathbb{R}^q, \quad d\Pi_S(z)(h) = \begin{cases} h, & \text{if } h_1 \geq 0, \\ (0, h_2, \dots, h_q), & \text{if } h_1 \leq 0. \end{cases}$$

As will become clear, the discontinuity of  $df(x)$  with respect to  $x$  is a major issue to be addressed to develop methods for confidence regions in this paper and in [15, 16]. In this paper, we handle this issue by utilizing a ‘‘symmetry’’ property between the B-derivatives  $df(x)$  and  $df(y)$  for two points  $x$  and  $y$  that belong to a common set in  $\mathfrak{P}$ , shown in Theorem 2.2 below. Theorem 2.2 will be used to establish our main result in Theorem 3.3. The proof of Theorem 2.2 uses the following lemma, which gives two equivalent statements for the condition  $y \in |\mathfrak{P}(x)|$ .

LEMMA 2.1. *Let  $x$  and  $y$  be two points in  $\mathbb{R}^q$ . The following are equivalent.*

- (1)  $y \in |\mathfrak{P}(x)|$ .
- (2)  $x \in |\mathfrak{P}(y)|$ .
- (3) There exists  $P \in \mathfrak{P}$  such that both  $x$  and  $y$  belong to  $P$ .

*Proof.* Suppose  $y \in |\mathfrak{P}(x)|$ . Then there exists  $P \in \mathfrak{P}(x)$  such that  $y \in P$ . The fact that  $P \in \mathfrak{P}(x)$  implies  $x \in P$ . This proves the direction (1) $\Rightarrow$ (3).

Now suppose (3) holds. Then  $P \in \mathfrak{P}(x)$ . Since  $y \in P$ , we have  $y \in |\mathfrak{P}(x)|$ . This proves (3) $\Rightarrow$ (1). It follows that (3) and (1) are equivalent. Similarly we can prove (3) and (2) are equivalent. The equivalence between (1) and (2) follows.  $\square$

THEOREM 2.2. *Let  $x \in \mathbb{R}^q$ , and let  $y \in |\mathfrak{P}(x)|$ . Then*

$$df(x)(y - x) = -df(y)(x - y).$$

*Proof.* By Lemma 2.1, there exists  $P \in \mathfrak{P}$  that contains both  $x$  and  $y$ . Let  $A$  be the matrix for the linear part of the selection function of  $f$  on  $P$ . The fact that  $y \in P$  implies  $y - x \in \text{cone}(P - x)$ , so by (2.1) we have  $df(x)(y - x) = A(y - x)$ . The fact  $x \in P$  implies  $x - y \in \text{cone}(P - y)$ . Again by (2.1) we have  $df(y)(x - y) = A(x - y)$ .  $\square$

Proposition 2.3 below shows that  $\mathfrak{P}(x)$  is a superset of  $\mathfrak{P}(y)$  for all  $y$  sufficiently close to  $x$ . Recall from comments below (2.1) that the selection functions of  $df(x)$  are exactly the linear parts of selection functions of  $f$  on elements of  $\mathfrak{P}(x)$ . Therefore, a consequence of Proposition 2.3 is that the family of

selection functions of  $df(x)$  contains the family of selection functions of  $df(y)$ , for all  $y$  sufficiently close to  $x$ .

**PROPOSITION 2.3.** *Let  $x \in \mathbb{R}^q$ . There exists a neighborhood  $X$  of  $x$ , such that each  $y \in X$  satisfies  $\mathfrak{P}(y) \subset \mathfrak{P}(x)$ .*

*Proof.* Recall that  $x$  belongs to the interior of  $|\mathfrak{P}(x)|$ . Let  $X$  be a neighborhood of  $x$  in the interior of  $|\mathfrak{P}(x)|$ , and let  $y \in X$ . Since  $y$  belongs to the interior of  $|\mathfrak{P}(x)|$ , there exist finitely many polyhedrons  $P_1, \dots, P_k$  in  $\mathfrak{P}(x)$ , such that  $y \in P_i$  for each  $i = 1, \dots, k$  and  $y$  belongs to the interior of  $\cup_{i=1}^k P_i$ . Now, let  $P \in \mathfrak{P}(y)$ ; we shall prove that  $P = P_i$  for some  $i = 1, \dots, k$ , by showing that it is impossible for a set  $P \in \mathfrak{P}$  that is different from any of those  $P_i$ 's to meet  $\cup_{i=1}^k P_i$ . Suppose for the purpose of contradiction that  $P$  is different from any  $P_i$ ,  $i = 1, \dots, k$ . Then, for each  $i$  the intersection  $P \cap P_i$  is a proper face of  $P$ . Consequently, the intersection between  $P$  and  $\cup_{i=1}^k P_i$  is the union of finitely many polyhedrons of dimensions less than  $q$ . On the other hand, since  $P$  is of dimension  $q$  and  $\cup_{i=1}^k P_i$  contains  $y$  in its interior, the intersection between  $P$  and  $\cup_{i=1}^k P_i$  contains a full-dimensional convex set. This leads to a contradiction. We have thereby proved that  $P = P_i$  for some  $i = 1, \dots, k$ . It follows that  $\mathfrak{P}(y) \subset \mathfrak{P}(x)$ .  $\square$

**3. Limiting properties.** This section provides some limiting properties of solutions to (1.6). Those properties will be used to develop methods on confidence region and simultaneous confidence intervals.

We make two sets of assumptions. Assumption 1 below is to obtain nice properties of  $f_0$  and  $f_n$  about integrability and convergence. Following that, Assumption 2 is to guarantee the existence, local uniqueness and stability of the true solution. The notation  $T_S(x)$  that appears in Assumption 2 denotes the tangent cone to  $S$  at a point  $x \in S$ . Since  $S$  is a polyhedral convex set, the following definition applies:

$$T_S(x) = \{v \in \mathbb{R}^q \mid \text{there exists } t > 0 \text{ such that } x + tv \in S\}. \quad (3.1)$$

**ASSUMPTION 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  $E\|d_x F(x, \xi)\|^2 < \infty$  for all  $x \in O$ .

(c) There exists a square integrable random variable  $C$  such that

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

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

The notation  $d_x F(x, \xi(\omega))$  here stands for the partial derivative of  $F$  w.r.t.  $x$ , a  $q \times q$  matrix. The norms used in Assumption 1 can be any norms in the  $\mathbb{R}^q$  or  $\mathbb{R}^{q \times q}$  space, since all norms in a finite-dimensional space are equivalent to each other. A consequence of Assumption 1 is the continuous differentiability of  $f_0$  on  $O$ . Moreover, for any nonempty compact subset  $X$  of  $O$ , let  $C^1(X, \mathbb{R}^q)$  be the Banach space of continuously differentiable mappings  $f : X \rightarrow \mathbb{R}^q$ , equipped with the norm

$$\|f\|_{1,X} = \sup_{x \in X} \|f(x)\| + \sup_{x \in X} \|df(x)\|. \quad (3.2)$$

Under Assumption 1, the sample average function  $f_n$  converges to  $f_0$  almost surely as an element of  $C^1(X, \mathbb{R}^q)$ , see, e.g., [35, Theorems 7.44, 7.48 and 7.52] and [16, Theorem 3].

**ASSUMPTION 2.** Suppose that  $x_0$  solves the variational inequality (1.1). 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  $K$ , defined as  $L_K(h) = L(\Pi_K(h)) + h - \Pi_K(h)$  for each  $h \in \mathbb{R}^q$ , is a homeomorphism from  $\mathbb{R}^q$  to  $\mathbb{R}^q$ .

We assume in the above assumption that (1.1) has a solution  $x_0$ . To guarantee the existence of such a solution, one would need to put additional conditions on  $f_0$  and  $S$ . For example, if  $f_0$  is strongly monotone on  $S$  then (1.1) must have a (globally unique) solution. Detailed discussions and more general solution existence conditions for variational inequalities can be found in [6, Chapter 2], [29, Chapter 12] and references therein. The set  $K$  defined in Assumption 2 is called the critical cone to  $S$  associated with  $z_0$ . The homeomorphism condition on  $L_K$  guarantees that  $x_0$  is a locally unique solution of (1.1), and that (1.1) continues to have a locally unique solution around  $x_0$  under small perturbation of  $f_0$ , see [16, Lemma 1] and the original result in [27]. Being the normal map induced by a linear map and a polyhedral convex cone,  $L_K$  is a piecewise linear function. It was shown in [26] that  $L_K$  is a homeomorphism if and only if it is coherently oriented (a piecewise linear function is coherently oriented if the determinants of matrices representing its selection functions all have the same nonzero sign), see also [23, 31] for shortened proofs. A special case in which the coherent orientation condition holds is when the restriction of  $L$  on the linear span of  $K$  is positive definite. In particular, if  $f_0$  is strongly monotone on  $O$ , then the entire matrix  $L$  is positive definite and  $L_K$  is a global homeomorphism.

The normal map  $L_K$  is related to the normal map  $(f_0)_S$  in (1.4) in the following way. Because  $f_0$  is differentiable at  $x_0$  and  $\Pi_S$  is B-differentiable, the normal map  $(f_0)_S$  is B-differentiable at  $z_0$  by the chain rule of B-differentiability, with

$$d(f_0)_S(z_0)(h) = df_0(x_0)(d\Pi_S(z_0)(h)) + h - d\Pi_S(z_0)(h). \quad (3.3)$$

It was shown in [27] that  $L_K$  is exactly  $d(f_0)_S(z_0)$ , the B-derivative of  $(f_0)_S$  at  $z_0$ .

The following theorem is adapted from [16, Theorem 7]. Here and hereafter, we use  $\Sigma_0$  to denote the covariance matrix of  $F(x_0, \xi)$ . Equation (3.4) in the theorem looks similar to the equation in [11, Theorem 2.7] and [35, Equation (5.74)]. Those two references are about the asymptotic distribution of  $x_n$ , a solution to (1.3), while (3.4) here describes the asymptotic distribution of  $z_n$ , a solution to (1.6).

**THEOREM 3.1.** *Suppose that Assumptions 1 and 2 hold. Let  $Y_0$  be a normal random vector in  $\mathbb{R}^q$  with zero mean and covariance matrix  $\Sigma_0$ . Then there exist neighborhoods  $X_0$  of  $x_0$  in  $O$  and  $Z$  of  $z_0$  in  $\mathbb{R}^q$  such that the following hold. For almost every  $\omega \in \Omega$ , there exists an integer  $N_\omega$ , such that for each  $n \geq N_\omega$ , the equation (1.6) has a unique solution  $z_n$  in  $Z$ , and the variational inequality (1.3) has a unique solution in  $X_0$  given by  $x_n = \Pi_S(z_n)$ . Moreover,  $\lim_{n \rightarrow \infty} z_n = z_0$  and  $\lim_{n \rightarrow \infty} x_n = x_0$  almost surely,*

$$\sqrt{n}(z_n - z_0) \Rightarrow (L_K)^{-1}(Y_0), \quad (3.4)$$

and

$$\sqrt{n}L_K(z_n - z_0) \Rightarrow Y_0. \quad (3.5)$$

From (3.5), the random variable  $n[L_K(z_n - z_0)]^T \Sigma_0^{-1} [L_K(z_n - z_0)]$  weakly converges to a  $\chi^2$  random variable with  $q$  degrees of freedom (assuming  $\Sigma_0$  is nonsingular). This leads to an expression for confidence regions of  $z_0$ . However, that expression includes the normal map  $L_K$  in it, which is unknown unless  $z_0$  is known because both  $L$  and  $K$  depend on  $z_0$  or  $x_0$ . Therefore, to obtain a computable confidence region, one needs to substitute  $L_K$  by a function that depends only on  $z_n$  and  $x_n$ , without losing the limiting property.

Recall that  $L_K$  is exactly  $d(f_0)_S(z_0)$ , the B-derivative of the normal map  $(f_0)_S$  at  $z_0$ . A natural estimator of  $L_K$  is therefore  $d(f_n)_S(z_n)$ . However, from the discontinuity of B-derivatives of piecewise



affine functions discussed above Lemma 2.1 and given that  $\Pi_S$  is a piecewise affine function, the B-derivative  $d\Pi_S(z_n)$  is not guaranteed to converge to  $d\Pi_S(z_0)$  even though  $z_n$  converges to  $z_0$  almost surely. From (3.3) it is clear that  $d(f_n)_S(z_n)$  is not guaranteed to converge to  $L_K$ . To provide an alternative estimator of  $L_K$ , the papers [15] and [16] designed some functions to converge to  $L_K$ , by utilizing the exponential convergence rate of  $z_n$  to  $z_0$  in probability obtained under additional assumptions.

The approach taken here is different from methods in [15, 16]. Instead of designing functions that converge to  $L_K$ , we directly use  $d(f_n)_S(z_n)$  to replace  $L_K$  in (3.5), knowing that  $d(f_n)_S(z_n)$  is not guaranteed to converge to  $L_K$  and may be very different with different  $z_n$ . The method is based on Theorem 3.3, which says that for large  $n$  the vector  $d(f_n)_S(z_n)(z_0 - z_n)$  is close to  $-L_K(z_n - z_0)$  with high probability, and that using  $d(f_n)_S(z_n)$  to replace  $L_K$  in (3.5) along with some sign changes keeps the weak convergence result to hold. As mentioned earlier, the proof of Theorem 3.3 relies on a ‘‘symmetry’’ property between the B-derivatives at two different points given in Theorem 2.2. From Theorems 3.3 we derive Theorems 4.1 and 4.2, to show that the set (4.2), or (4.6) for the singular case, is an asymptotically exact confidence region for  $z_0$ : that is, the set contains  $z_0$  with probability converging to the prescribed level of confidence.

Comparing to methods in [15, 16], the main advantage with the new method is given by Proposition 3.5, which states that  $d(f_n)_S(z_n)$  is with high probability an invertible linear function. This implies that the set (4.2) is with high probability a single ellipsoid, when  $\Sigma_0$  is nonsingular. When  $\Sigma_0$  is singular one could use  $R_{n,0}$  in (4.12) to approximate  $R_{n,\epsilon}$  in (4.6), and  $R_{n,0}$  is with high probability a degenerate ellipsoid. In contrast, when  $L_K$  is piecewise linear with multiple pieces, the estimators designed in [15, 16] have multiple pieces with high probability, providing confidence regions that are factions of ellipsoids pieced together. Clearly, it is much easier to describe a single ellipsoid, or to find the minimal enclosing box of a single ellipsoid to obtain simultaneous confidence intervals. Additionally, the new method does not rely on the assumptions for the exponential convergence rate of  $z_n$ , another advantage comparing to [15, 16].

Before proceeding to the proofs, recall that the Euclidean projector  $\Pi_S$  coincides with an affine function on each  $q$ -cell in the normal manifold of  $S$ . We use  $\mathfrak{P}$  to denote the normal manifold of  $S$ , which is the polyhedral subdivision of  $\mathbb{R}^q$  corresponding to  $\Pi_S$ . As in Section 2 we use  $\mathfrak{P}(z_0)$  to denote the family of  $q$ -cells containing  $z_0$ . Lemma 3.2 below will be used in the proof of Theorem 3.3.

LEMMA 3.2. *Suppose that Assumptions 1 and 2 hold. Then for almost every  $\omega \in \Omega$  there exists an integer  $N_\omega$  such that the following equality holds for each  $n \geq N_\omega$ :*

$$d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n) = 0. \quad (3.6)$$

*Proof.* Recall that  $z_0$  belongs to the interior of  $|\mathfrak{P}(z_0)|$ . Since  $z_n$  converges to  $z_0$  w.p. 1, for almost every  $\omega \in \Omega$  there exists an integer  $N_\omega$  such that  $z_n$  belongs to  $|\mathfrak{P}(z_0)|$  for each  $n \geq N_\omega$ . It follows from Theorem 2.2 that each such  $z_n$  satisfies (3.6).  $\square$

Theorem 3.3 below is the main result of this paper.

THEOREM 3.3. *Suppose that Assumptions 1 and 2 hold. Then for each  $\epsilon > 0$  we have*

$$\lim_{n \rightarrow \infty} \text{Prob}\{\sqrt{n}\|d(f_0)_S(z_0)(z_n - z_0) + d(f_n)_S(z_n)(z_0 - z_n)\| > \epsilon\} = 0. \quad (3.7)$$

Consequently,

$$-\sqrt{n}d(f_n)_S(z_n)(z_0 - z_n) \Rightarrow Y_0. \quad (3.8)$$

*Proof.* Recall from Assumption 2 and Theorem 3.1 that  $x_0 = \Pi_S(z_0)$  and  $x_n = \Pi_S(z_n)$  are solutions to (1.1) and (1.3) respectively. From (3.3) we have

$$d(f_0)_S(z_0)(z_n - z_0) = df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) + z_n - z_0 - d\Pi_S(z_0)(z_n - z_0). \quad (3.9)$$

Similarly,

$$d(f_n)_S(z_n)(z_0 - z_n) = df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n)) + z_0 - z_n - d\Pi_S(z_n)(z_0 - z_n). \quad (3.10)$$

It follows that  $\sqrt{n}\|df_0)_S(z_0)(z_n - z_0) + d(f_n)_S(z_n)(z_0 - z_n)\|$  is bounded from above by the sum of the following two terms.

Term (a):  $\sqrt{n}\|df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) + df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n))\|$ .

Term (b):  $\sqrt{n}\|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\|$ .

By Lemma 3.2, term (b) converges to 0 almost surely, so it converges to 0 in probability. It remains to show that Term (a) converges to 0 in probability. Term (a) is bounded from above by the sum of the following two terms.

Term (c):  $\sqrt{n}\|df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) - df_n(x_n)(d\Pi_S(z_0)(z_n - z_0))\|$ .

Term (d):  $\sqrt{n}\|df_n(x_n)(d\Pi_S(z_0)(z_n - z_0)) + df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n))\|$ .

Because  $df_n(x_n)(\cdot)$  is a linear map, by Lemma 3.2 term (d) converges to 0 almost surely and therefore converges to 0 in probability. To prove the theorem it suffices to show that term (c) converges to 0 in probability.

Term (c) is bounded from above by the following

$$\|df_0(x_0) - df_n(x_n)\| \|\sqrt{n}(d\Pi_S(z_0)(z_n - z_0))\| \quad (3.11)$$

where  $\|df_0(x_0) - df_n(x_n)\|$  is the norm of the linear operator  $df_0(x_0) - df_n(x_n)$ , which is bounded from above by

$$\|df_0(x_0) - df_0(x_n)\| + \|df_0(x_n) - df_n(x_n)\|.$$

Since  $x_n$  converges to  $x_0$  almost surely and  $f_0$  is continuously differentiable under Assumption 1,  $\|df_0(x_0) - df_0(x_n)\|$  converges to 0 almost surely. Assumption 1 also guarantees  $f_n$  to converge to  $f_0$  almost surely as an element of  $C^1(X, \mathbb{R}^q)$  for any compact subset  $X$  of  $O$ . Let  $X$  be a compact subset of  $O$  that contains  $x_0$  in its interior; we have  $x_n \in X$  for all sufficiently large  $n$ . It follows that  $\|df_0(x_n) - df_n(x_n)\|$  converges to zero almost surely. This proves that  $\|df_0(x_0) - df_n(x_n)\|$  converges to zero almost surely. On the other hand, since  $d\Pi_S(z_0)$  is positively homogenous, we can rewrite  $\sqrt{n}(d\Pi_S(z_0)(z_n - z_0))$  as  $d\Pi_S(z_0)(\sqrt{n}(z_n - z_0))$ . By (3.4),  $\sqrt{n}(z_n - z_0)$  converges in distribution to a random variable, so it is (uniformly) tight; see, e.g., [39, Theorem 2.4]. It follows that  $d\Pi_S(z_0)(\sqrt{n}(z_n - z_0))$  is uniformly tight, and that the quantity (3.11) converges to 0 in probability. This proves that term (c) converges to 0 in probability, and thereby proves (3.7). Equation (3.8) is a result of (3.5) and (3.7).  $\square$

Proposition 3.5 below shows the function  $d(f_n)_S(z_n)$  that appears in (3.8) is a linear invertible function with high probability. Its proof uses the following lemma.

LEMMA 3.4. *Under Assumptions 1 and 2, there exist a neighborhood  $Z'$  of  $z_0$  in  $\mathbb{R}^q$  and a neighborhood  $\mathbb{L}$  of the matrix  $L$  in  $\mathbb{R}^{q \times q}$ , such that for each  $z' \in Z'$  and  $L' \in \mathbb{L}$  the map  $\Upsilon(z', L') : \mathbb{R}^q \rightarrow \mathbb{R}^q$  defined as*

$$\Upsilon(z', L')(h) = L'd\Pi_S(z')(h) + h - d\Pi_S(z')(h) \text{ for each } h \in \mathbb{R}^q \quad (3.12)$$

is a global homeomorphism from  $\mathbb{R}^q$  to  $\mathbb{R}^q$ .

*Proof.* First, note that for each  $z' \in \mathbb{R}^q$  the B-derivative  $d\Pi_S(z')$  is exactly the Euclidean projector onto the critical cone to  $S$  associated with  $z'$  [20, 25]. Hence, the map  $\Upsilon(z', L')$  is exactly the normal map induced by  $L'$  and the latter critical cone. Recall that such a normal map is a global homeomorphism if and only if it is coherently oriented (see the discussion below Assumption 2). In the rest of this proof, we find neighborhoods  $Z'$  and  $\mathbb{L}$  to guarantee  $\Upsilon(z', L')$  to be coherently oriented for  $z' \in Z'$  and  $L' \in \mathbb{L}$ .

Since  $K$  is the critical cone to  $S$  associated with  $z_0$ , we have  $d\Pi_S(z_0) = \Pi_K$  and  $\Upsilon(z_0, L) = L_K$ . Because  $L_K$  is a global homeomorphism by assumption, it is coherently oriented. The fact that  $K$  is a polyhedral convex cone implies that  $L_K$  is a piecewise linear function. Let  $M_1, \dots, M_k$  be the matrices that represent the selection functions of  $L_K$ . Since  $L_K$  is coherently oriented, the determinants of  $M_1, \dots, M_k$  have the same nonzero sign (see the definition of the coherent orientation condition below Assumption 2). By choosing  $\mathbb{L}$  to be a sufficiently small neighborhood of  $L$ , we can guarantee that the determinants of matrices representing selection functions of  $\Upsilon(z_0, L')$  to have the same nonzero sign for each  $L' \in \mathbb{L}$ . This proves that  $\Upsilon(z_0, L')$  is coherently oriented for each  $L' \in \mathbb{L}$ .

By Proposition 2.3, there exists a neighborhood  $Z'$  of  $z_0$  such that  $\mathfrak{P}(z') \subset \mathfrak{P}(z_0)$  for each  $z' \in Z'$ . As noted in the remark above Proposition 2.3, the family of the selection functions of  $d\Pi_S(z_0)$  includes that of  $d\Pi_S(z')$  for each  $z' \in Z'$ . Thus, for each  $L' \in \mathbb{L}$  and  $z' \in Z'$ , any selection function of  $\Upsilon(z', L')$  is also a selection function of  $\Upsilon(z_0, L')$ , and the coherent orientation of  $\Upsilon(z', L')$  follows from the coherent orientation of  $\Upsilon(z_0, L')$ .  $\square$

PROPOSITION 3.5. *Under Assumptions 1 and 2,*

$$\lim_{n \rightarrow \infty} \text{Prob}\{d(f_n)_S(z_n) \text{ is an invertible linear map}\} = 1.$$

*Proof.* By the chain rule,

$$d(f_n)_S(z_n)(h) = df_n(x_n)(d\Pi_S(z_n)(h)) + h - d\Pi_S(z_n)(h) \text{ for each } h \in \mathbb{R}^q.$$

The linearity of  $d(f_n)_S(z_n)$  depends on the linearity of  $d\Pi_S(z_n)$ . The latter function is linear whenever  $z_n$  belongs to the interior of a  $q$ -cell in the normal manifold of  $S$ . Let  $R$  be the union of boundaries of all the  $q$ -cells; it follows that  $R$  is the union of finitely many polyhedral convex sets of dimensions less than  $q$ . Consider the tangent cone  $T_R(z_0)$ , which is defined in the same way as  $T_S(x)$  in (3.1) if  $z_0 \in R$  and is the empty set if  $z_0 \notin R$ . Let  $Z_0$  be a neighborhood of  $z_0$  such that  $Z_0 \cap R = Z_0 \cap (z_0 + T_R(z_0))$ ; we have

$$\sqrt{n}(z - z_0) \in T_R(z_0) \text{ for each } z \in Z_0 \cap R \text{ and each integer } n. \quad (3.13)$$

The fact that  $z_n$  converges to  $z_0$  almost surely implies

$$\lim_{n \rightarrow \infty} \text{Prob}\{z_n \notin Z_0\} = 0.$$

On the other hand, (3.13) implies

$$\text{Prob}\{z_n \in Z_0 \cap R\} \leq \text{Prob}\{\sqrt{n}(z_n - z_0) \in T_R(z_0)\}.$$

It follows from (3.4) and the fact that  $L_K$  is a piecewise linear homeomorphism that

$$\lim_{n \rightarrow \infty} \text{Prob}\{\sqrt{n}(z_n - z_0) \in T_R(z_0)\} \leq \text{Prob}\{(L_K)^{-1}(Y_0) \in T_R(z_0)\} = 0.$$

Because  $\text{Prob}\{z_n \in R\} \leq \text{Prob}\{z_n \notin Z_0\} + \text{Prob}\{z_n \in Z_0 \cap R\}$ , we have proved

$$\lim_{n \rightarrow \infty} \text{Prob}\{z_n \in R\} = 0, \quad (3.14)$$

which implies that the probability for  $d(f_n)_S(z_n)$  to be a linear function converges to 1 as  $n$  goes to  $\infty$ .

Next, choose neighborhoods  $Z'$  of  $z_0$  and  $\mathbb{L}$  of  $L$  as in Lemma 3.4. Since  $z_n$  converges to  $z_0$  almost surely,  $z_n$  belongs to  $Z'$  almost surely for sufficiently large  $n$ . It was shown in the proof of Theorem 3.3 that  $df_n(x_n)$  almost surely converges to  $df_0(x_0)$ . Consequently,  $df_n(x_n)$  belongs to  $\mathbb{L}$  almost surely for sufficiently large  $n$ . By Lemma 3.4,  $d(f_n)_S(z_n)$  is a global homeomorphism almost surely for sufficiently large  $n$ , so the probability for  $d(f_n)_S(z_n)$  to be an invertible function converges to 1 as  $n$  goes to  $\infty$ . The conclusion of the proposition follows by combining the linearity and invertibility.  $\square$

The random variable  $Y_0$  in (3.8) has covariance matrix  $\Sigma_0$ , which depends on the true solution  $x_0$ . In computing confidence regions and intervals we will replace  $\Sigma_0$  by  $\Sigma_n$ , the sample covariance matrix of  $\{F(x_n, \xi^i)\}_{i=1}^n$ . The lemma below supports such a replacement.

LEMMA 3.6. *Suppose that Assumptions 1 and 2 hold. The matrix  $\Sigma_n$  converges to  $\Sigma_0$  almost surely as  $n \rightarrow \infty$ .*

*Proof.* Under the assumptions,  $\frac{1}{n} \sum_{i=1}^n F(x_n, \xi^i(\omega))F(x_n, \xi^i(\omega))^T$  converges to  $E[F(x, \xi)F(x, \xi)^T]$  almost surely uniformly on each compact subset  $X$  of  $O$  (see, e.g., [35, Theorem 7.48]). Since  $x_n$  converges to  $x_0$  almost surely, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n F(x_n, \xi^i(\omega))F(x_n, \xi^i(\omega))^T = E[F(x_0, \xi)F(x_0, \xi)^T] \text{ almost surely.}$$

Similarly  $f_n(x_n)$  converges to  $f_0(x_0)$  almost surely as  $n \rightarrow \infty$ . This proves the lemma.  $\square$

**4. Confidence regions and simultaneous confidence intervals.** This section provides formulas for confidence regions of  $z_0$  that are computable from  $z_n$ , and provides a method to compute simultaneous confidence intervals for components of  $z_0$ . We treat two cases separately: Theorem 4.1 considers situations in which  $\Sigma_0$  (the covariance matrix of  $F(x_0, \xi)$ ) is nonsingular, and Theorem 4.2 handles situations in which  $\Sigma_0$  is singular. The confidence region is given in (4.2) or (4.6) for the two cases respectively. We use  $\chi_l^2$  to denote a  $\chi^2$  random variable with  $l$  degrees of freedom, and use  $\chi_l^2(\alpha)$  to denote the number that satisfies  $P(\chi_l^2 > \chi_l^2(\alpha)) = \alpha$  for  $\alpha \in [0, 1]$ . For the rest of this section, let  $\alpha \in [0, 1]$  be fixed.

THEOREM 4.1. *Suppose that Assumptions 1 and 2 hold, and that  $\Sigma_0$  is nonsingular. For almost every  $\omega \in \Omega$ , there exists an integer  $N_\omega$  such that  $\Sigma_n$  is nonsingular for  $n \geq N_\omega$ . Moreover,*

$$-\sqrt{n}\Sigma_n^{-1/2}[d(f_n)_S(z_n)(z_0 - z_n)] \Rightarrow \mathcal{N}(0, I_q), \quad (4.1)$$

and the probability for  $z_0$  to belong to the set

$$\left\{ z \in \mathbb{R}^q \mid n[d(f_n)_S(z_n)(z - z_n)]^T \Sigma_n^{-1} [d(f_n)_S(z_n)(z - z_n)] \leq \chi_q^2(\alpha) \right\} \quad (4.2)$$

converges to  $1 - \alpha$  as  $n \rightarrow \infty$ .

*Proof.* Since  $\Sigma_n$  converges to  $\Sigma_0$  almost surely (Lemma 3.6), it is nonsingular for sufficiently large  $n$  almost surely. Equation (4.1) follows from (3.8). From (4.1) the random variable

$$n[d(f_n)_S(z_n)(z_0 - z_n)]^T \Sigma_n^{-1} [d(f_n)_S(z_n)(z_0 - z_n)]$$

weakly converges to a  $\chi_q^2$  random variable, so the set (4.2) contains  $z_0$  with probability converging to  $1 - \alpha$ .  $\square$

**THEOREM 4.2.** *Suppose that Assumptions 1 and 2 hold, and that  $\Sigma_0$  is singular. Let  $\rho > 0$  be the minimum of all positive eigenvalues of  $\Sigma_0$ , and let  $l$  be the number of positive eigenvalues of  $\Sigma_0$  counted with regard to their algebraic multiplicities. Let  $\rho_0$  satisfy  $0 < \rho_0 < \rho$ . Decompose  $\Sigma_n$  as*

$$\Sigma_n = U_n^T \Delta_n U_n \quad (4.3)$$

where  $U_n$  is an orthogonal  $q \times q$  matrix, and  $\Delta_n$  is a diagonal matrix with monotonically decreasing elements. Let  $D_n$  be the upper-left submatrix of  $\Delta_n$  whose diagonal elements are at least  $\rho_0$ . Let  $l_n$  be the number of rows in  $D_n$ ,  $(U_n)_1$  be the submatrix of  $U_n$  that consists of its first  $l_n$  rows, and  $(U_n)_2$  be the submatrix that consists of the remaining rows of  $U_n$ . Then for almost every  $\omega$  the equality  $l_n = l$  holds for sufficiently large  $n$ . Moreover,

$$n[d(f_n)_{S(z_n)}(z_0 - z_n)]^T (U_n)_1^T D_n^{-1} (U_n)_1 [d(f_n)_{S(z_n)}(z_0 - z_n)] \Rightarrow \chi_l^2 \quad (4.4)$$

and

$$n[d(f_n)_{S(z_n)}(z_0 - z_n)]^T (U_n)_2^T (U_n)_2 [d(f_n)_{S(z_n)}(z_0 - z_n)] \Rightarrow 0. \quad (4.5)$$

For each  $\epsilon > 0$ , the set

$$R_{n,\epsilon} = \left\{ z \in \mathbb{R}^q \left| \begin{array}{l} n[d(f_n)_{S(z_n)}(z - z_n)]^T (U_n)_1^T D_n^{-1} (U_n)_1 [d(f_n)_{S(z_n)}(z - z_n)] \leq \chi_l^2(\alpha) \\ \|\sqrt{n}(U_n)_2 [d(f_n)_{S(z_n)}(z - z_n)]\|_\infty \leq \epsilon \end{array} \right. \right\} \quad (4.6)$$

contains  $z_0$  with probability converging to  $1 - \alpha$ .

*Proof.* Let  $\rho$  and  $l$  be as defined in this theorem, and conduct an eigen-decomposition of  $\Sigma_0$  as

$$\Sigma_0 = U_0^T \begin{bmatrix} D_0 & 0 \\ 0 & 0 \end{bmatrix} U_0 = [(U_0)_1^T \quad (U_0)_2^T] \begin{bmatrix} D_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (U_0)_1 \\ (U_0)_2 \end{bmatrix} \quad (4.7)$$

where  $U_0$  is orthogonal,  $D_0$  is diagonal with monotonically decreasing positive diagonal elements, and  $(U_0)_1$  and  $(U_0)_2$  contains the first  $l$  and the last  $q - l$  rows of  $U_0$  respectively. From (3.8) we have

$$-\sqrt{n} \begin{bmatrix} D_0^{-1/2} & 0 \\ 0 & I_{q-l} \end{bmatrix} U_0 [d(f_n)_{S(z_n)}(z_0 - z_n)] \Rightarrow \mathcal{N}(0, I_l) \times 0. \quad (4.8)$$

From Lemma 3.6, for a.e.  $\omega \in \Omega$  there exists an integer  $N_\omega^1$ , such that each  $n \geq N_\omega^1$  satisfies

$$(\Delta_n)_{ii} > \rho_0 \text{ for each } i = 1, \dots, l \text{ and } (\Delta_n)_{ii} < \rho_0 \text{ for each } i = l + 1, \dots, q. \quad (4.9)$$

It follows that that  $l_n = l$  for  $n \geq N_\omega^1$  and that  $D_n$  converges to  $D_0$  almost surely.

For  $n \geq N_\omega^1$ , define

$$\hat{\Sigma}_n = U_n^T \begin{bmatrix} D_n & 0 \\ 0 & 0 \end{bmatrix} U_n \text{ and } \hat{\Sigma}_n^+ = U_n^T \begin{bmatrix} D_n^{-1} & 0 \\ 0 & 0 \end{bmatrix} U_n = (U_n)_1^T D_n^{-1} (U_n)_1.$$

Because of (4.9),  $\hat{\Sigma}_n$  is the unique best approximation of  $\Sigma_n$  in Frobenius norm among all matrices of rank  $\leq l$ ; see, e.g., [1, 7, 38]. Since the unique best approximation of a matrix depends continuously on that matrix (by an application of [29, Theorem 1.17]), and the best approximation of  $\Sigma_0$  is itself,  $\hat{\Sigma}_n$  almost surely converges to  $\Sigma_0$ . As the pseudo-inverse of  $\hat{\Sigma}_n$ ,  $\hat{\Sigma}_n^+$  almost surely converges to the pseudo-inverse of  $\Sigma_0$  (by an application of [37, Corollary 3.5]). This and (4.8) imply (4.4).

Finally, an application of [14, Theorem 3.1] implies that the angle between the row spaces of  $(U_0)_2$  and  $(U_n)_2$  almost surely converges to zero as  $n \rightarrow \infty$ . In view of [14, Equation (2.4)] and [8, Theorem 2.6.1], the matrix  $(U_n)_2^T (U_n)_2$  converges to  $(U_0)_2^T (U_0)_2$  almost surely. This and (4.8) imply (4.5).

Note that the choice of the matrix  $U_n$  in the decomposition of  $\Sigma_n$  is non-unique, when  $\Sigma_n$  has repeated eigenvalues. However, when (4.9) holds the matrices  $(U_n)_1^T D_n^{-1} (U_n)_1$  and  $(U_n)_2^T (U_n)_2$  that appear in (4.4) and (4.5) are uniquely determined by  $\Sigma_n$ , and depend continuously on  $\Sigma_n$ .

Since  $l_n = l$  almost surely for sufficiently large  $n$ , from (4.4) we have

$$\lim_{n \rightarrow \infty} \text{Prob} \left\{ n [d(f_n)_S(z_n)(z_0 - z_n)]^T (U_n)_1^T D_n^{-1} (U_n)_1 [d(f_n)_S(z_n)(z_0 - z_n)] \leq \chi_{l_n}^2(\alpha) \right\} = 1 - \alpha. \quad (4.10)$$

By (4.5) we have

$$\lim_{n \rightarrow \infty} \text{Prob} \left\{ n [d(f_n)_S(z_n)(z_0 - z_n)]^T (U_n)_2^T (U_n)_2 [d(f_n)_S(z_n)(z_0 - z_n)] \leq \epsilon \right\} = 1$$

for each  $\epsilon > 0$ . Since all norms are equivalent in a finite-dimensional space, we can rewrite the above equality as

$$\lim_{n \rightarrow \infty} \text{Prob} \left\{ \sqrt{n} \|(U_n)_2 [d(f_n)_S(z_n)(z_0 - z_n)]\|_\infty \leq \epsilon \right\} = 1, \quad (4.11)$$

where  $\|\cdot\|_\infty$  denotes the  $\infty$ -norm. In writing (4.11), we assume that  $U_n$  is a measurable function of  $\Sigma_n$  so that the set in consideration is a measurable set in  $\Omega$ . Equations (4.10) and (4.11) imply that the set  $R_{n,\epsilon}$  contains  $z_0$  with probability converging to  $1 - \alpha$ .  $\square$

Since for each  $\epsilon > 0$  the set  $R_{n,\epsilon}$  is an asymptotically exact  $(1 - \alpha)100\%$  confidence region for  $z_0$ , it is natural to ask whether the same is true about the set

$$R_{n,0} = \left\{ z \in \mathbb{R}^q \left| \begin{array}{l} n [d(f_n)_S(z_n)(z - z_n)]^T (U_n)_1^T D_n^{-1} (U_n)_1 [d(f_n)_S(z_n)(z - z_n)] \leq \chi_{l_n}^2(\alpha) \\ \sqrt{n} (U_n)_2 [d(f_n)_S(z_n)(z - z_n)] = 0 \end{array} \right. \right\}. \quad (4.12)$$

We cannot prove that  $R_{n,0}$  contains  $z_0$  with probability converging to  $1 - \alpha$ , because the quantity

$$\lim_{n \rightarrow \infty} \text{Prob} \left\{ \sqrt{n} (U_n)_2 [d(f_n)_S(z_n)(z_0 - z_n)] = 0 \right\}$$

may not equal 1. However, Proposition 4.3 below shows for a fixed  $n$  that the set-valued map  $R_{n,\epsilon}$  is Lipschitz continuous on an interval  $[0, \bar{\epsilon}]$  for some positive real number  $\bar{\epsilon}$ . Accordingly, any open set that contains  $R_{n,0}$  must include  $R_{n,\epsilon}$  for a sufficiently small  $\epsilon$ . Since  $R_{n,0}$  has a simpler geometric structure comparing to  $R_{n,\epsilon}$ , one may choose to use  $R_{n,0}$  to approximate  $R_{n,\epsilon}$  for computational efficiency.

**PROPOSITION 4.3.** *Let  $d(f_n)_S(z_n)$  be a fixed invertible linear function,  $U_n$  a fixed orthogonal matrix, and  $D_n$  a fixed  $l_n \times l_n$  diagonal matrix with positive diagonal elements. There exist real numbers  $\bar{\epsilon} > 0$  and  $\sigma > 0$  such that*

$$R_{n,\epsilon'} \subset R_{n,\epsilon} + \sigma |\epsilon - \epsilon'| \mathbb{B}$$

whenever  $0 \leq \epsilon \leq \bar{\epsilon}$  and  $0 \leq \epsilon' \leq \bar{\epsilon}$ , where  $\mathbb{B}$  denotes the closed unit ball in  $\mathbb{R}^q$ .

*Proof.* After the coordinate transformation  $U_n d(f_n)_S(z_n)$ , one can view  $R_{n,\epsilon}$  as a cartesian product of a fixed ellipsoid in  $\mathbb{R}^{l_n}$  and a polyhedron in  $\mathbb{R}^{q-l_n}$ , with  $\epsilon$  being the right hand side of the linear constraints defining the polyhedron. The Lipschitz continuity can be seen in the transformed space.  $\square$

It is shown in Proposition 3.5 that  $d(f_n)_S(z_n)$  is an invertible linear function with high probability. When this holds, the set in (4.2) and  $R_{n,0}$  in (4.12) are ellipsoids in  $\mathbb{R}^q$ , and therefore can be described using their centers, principal directions, and semi-axes. It is often desirable to provide simultaneous confidence intervals for components of  $z_0$ , as intervals are more convenient to describe, visualize, and interpret. We can compute these intervals by finding the minimum enclosing box of a confidence region, i.e., by computing the maximal and minimal values of  $z_i$  for each  $i = 1, \dots, q$  over the confidence region.

Since  $\Sigma_0$  is unknown in practice, it is often difficult to directly check if  $\Sigma_0$  is nonsingular or not. Instead, we conduct an eigen-decomposition of the sample covariance matrix  $\Sigma_n$  as in (4.3), and partition matrices  $U_n$  and  $\Delta_n$  based on positive or zero eigenvalues. If all eigenvalues are strictly positive (larger than a prescribed tolerance), then we use (4.2) as the confidence region. Otherwise, we use  $R_{n,\epsilon}$  in (4.6) for some positive  $\epsilon$  or  $R_{n,0}$  as the confidence region. In the latter case, the confidence region is still bounded because  $d(f_n)_S(z_n)$  is invertible, and is flat in the directions represented by rows of  $(U_n)_2[d(f_n)_S(z_n)]$ . In particular, the set  $R_{n,0}$  is a degenerate ellipsoid. One can still find the minimal enclosing box of the confidence region to obtain the simultaneous confidence intervals. Numerical tests with singular covariance matrices are conducted in [12]. Lastly, we mention that the only requirement on  $\Sigma_n$  for Theorems 4.1 and 4.2 to hold is its almost sure convergence to  $\Sigma_0$ . Hence, any estimator of  $\Sigma_0$  that converges to it almost surely can be used as  $\Sigma_n$ , not necessarily the sample covariance matrix.

**5. Individual confidence intervals.** This section discusses computation of individual confidence intervals for  $z_0$ . An individual confidence interval for the  $j$ th component of  $z_0$  is an interval in  $\mathbb{R}$  that contains  $(z_0)_j$  with a prescribed level of confidence. It is generally narrower than the simultaneous confidence interval for the same level. The difference between widths of the two intervals are substantial in problems with moderate or large dimensions. In such problems, the minimum enclosing box of a confidence region can be much larger than the region itself, resulting in simultaneous confidence intervals too wide to be useful. The sizes of individual confidence intervals are less affected by the dimension of the problem. There are also problems where only confidence intervals of selected components in the true solution are of interest.

Computation of individual confidence intervals requires knowledge on the distribution of each individual component of  $z_n$ . As shown in (3.4),  $\sqrt{n}(z_n - z_0)$  weakly converges to the distribution of a random variable  $\Gamma = (L_K)^{-1}(Y_0)$ . Given results in the preceding two sections, it is natural to ask whether we can substitute  $L_K = d(f_0)_S(z_0)$  with  $d(f_n)_S(z_n)$  in computing individual confidence intervals for  $z_0$ . Such a direct substitution results in a confidence interval in (5.17), where  $r_{n,j}$  is defined in (5.3). The equation (5.4) expresses the limiting probability for such an interval to contain  $(z_0)_j$  in terms of  $\Gamma$ . That limiting probability equals the desired confidence level  $1 - \alpha$  when the condition in (5.5) holds.

Let  $\mathfrak{P}(z_0) = \{P_1, \dots, P_k\}$  be the set of  $q$ -cells in the normal manifold of  $S$  that contains  $z_0$ . For each  $i = 1, \dots, k$ , let  $K_i = \text{cone}(P_i - z_0)$ ,  $A_i$  be the matrix representing  $d\Pi_S(z)$  for points  $z$  in the interior of  $P_i$ , and  $T_i = d(f_0)_S(z_0)(K_i)$  be the image of  $K_i$  under the function  $d(f_0)_S(z_0)$ . From (3.3) we can see that  $d(f_0)_S(z_0)$  coincides with  $dL_S(z_0)$ , where  $L_S$  is the normal map induced by the linear operator  $L$  and the set  $S$ . On the cone  $K_i$ , the map  $d(f_0)_S(z_0)$  is represented by the matrix

$$M_i = LA_i + I - A_i.$$

Under Assumption 2 the map  $d(f_0)_S(z_0)$  is a global homeomorphism, so all the matrices  $M_i, i = 1, \dots, k$  are nonsingular.

As above let  $\Gamma = (L_K)^{-1}(Y_0)$ , and for each  $i = 1, \dots, k$  define a random variable  $\Gamma^i = M_i^{-1}(Y_0)$ . The fact that  $Y_0$  is a multivariate normal random variable implies that each  $\Gamma^i$  is a multivariate normal random variable with covariance matrix  $M_i^{-1}\Sigma_0 M_i^{-T}$ . Accordingly, each component of  $\Gamma^i$  is a normal random variable. If we define a number

$$r_j^i = \sqrt{(M_i^{-1}\Sigma_0 M_i^{-T})_{jj}}$$

for each  $i = 1, \dots, k$  and  $j = 1, \dots, q$ , then for each real number  $\alpha \in (0, 1)$  we have

$$\text{Prob}\left(|\Gamma_j^i| \leq \sqrt{\chi_1^2(\alpha)} r_j^i\right) = 1 - \alpha. \quad (5.1)$$

Let us discuss a relation between  $\Gamma$  and  $\Gamma^i$ . Since  $d(f_0)_S(z_0)$  is represented by the matrix  $M_i$  on the cone  $K_i$ , we have

$$d(f_0)_S(z_0)^{-1}(y) = M_i^{-1}(y) \text{ if } y \in T_i.$$

For each measurable set  $W \subset K_i$  we have

$$\text{Prob}(\Gamma \in W) = \text{Prob}(Y_0 \in M_i(W)) = \text{Prob}(\Gamma^i \in W). \quad (5.2)$$

**THEOREM 5.1.** *Suppose that Assumptions 1 and 2 hold, and suppose that  $\Sigma_0$  has a positive determinant. Let  $P_i, K_i, T_i, A_i, M_i, \Gamma^i, \Gamma, r_j^i$  be defined as above. For each integer  $n$  and each  $j = 1, \dots, q$ , define a number*

$$r_{nj} = \begin{cases} \sqrt{(d(f_n)_S(z_n)^{-1}\Sigma_n d(f_n)_S(z_n)^{-T})_{jj}} & \text{if } d(f_n)_S(z_n) \text{ is an invertible linear map,} \\ 0 & \text{otherwise.} \end{cases} \quad (5.3)$$

Then for each real number  $\alpha \in (0, 1)$  and for each  $j = 1, \dots, q$ ,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{Prob}\left(\frac{\sqrt{n}|(z_n - z_0)_j|}{r_{nj}} \leq \sqrt{\chi_1^2(\alpha)}\right) \\ &= \sum_{i=1}^k \text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in K_i\right) = \sum_{i=1}^k \text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma \in K_i\right). \end{aligned} \quad (5.4)$$

Moreover, suppose for a given  $j = 1, \dots, q$  that the following equality

$$\text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in K_i\right) = \text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)}\right) \text{Prob}(\Gamma^i \in K_i) \quad (5.5)$$

holds for each  $i = 1, \dots, k$ . Then for each real number  $\alpha \in (0, 1)$ ,

$$\lim_{n \rightarrow \infty} \text{Prob}\left\{|(z_n - z_0)_j| \leq \frac{\sqrt{\chi_1^2(\alpha)} r_{nj}}{\sqrt{n}}\right\} = 1 - \alpha. \quad (5.6)$$



*Proof.* Let  $\alpha \in (0, 1)$  be fixed. Recall from Lemma 3.6 that  $\Sigma_n$  converges to  $\Sigma_0$  almost surely, and from the proof of Theorem 3.3 that  $df_n(x_n)$  converges to  $L = df_0(x_0)$  almost surely. Let  $Z'$  be a neighborhood of  $z_0$  in the interior of  $|\mathfrak{P}(z_0)|$  with  $Z' \cap P_i = Z' \cap (z_0 + K_i)$  for each  $i = 1, \dots, k$ . We have

$$\lim_{n \rightarrow \infty} \text{Prob}(z_n \in Z') = 1. \quad (5.7)$$

Writing  $Z'$  as the following union,

$$Z' = \cup_{i=1}^k (Z' \cap (z_0 + \text{int } K_i)) \cup (Z' \cap (z_0 + \mathbb{R}^q / \cup_{i=1}^k \text{int } K_i)),$$

and noting from (3.14) in the proof of Proposition 3.5 that

$$\lim_{n \rightarrow \infty} \text{Prob}(z_n \in Z' \cap (z_0 + \mathbb{R}^q / \cup_{i=1}^k \text{int } K_i)) = 0,$$

we find

$$\lim_{n \rightarrow \infty} \sum_{i=1}^k \text{Prob}(z_n \in Z' \cap (z_0 + \text{int } K_i)) = 1. \quad (5.8)$$

It is not hard to see  $Z' \cap (z_0 + \text{int } K_i) = Z' \cap \text{int } P_i$  for each  $i = 1, \dots, k$ . When  $z_n \in Z' \cap \text{int } P_i$ ,  $d(f_n)_S(z_n)$  is a linear map given by

$$d(f_n)_S(z_n) = df_n(x_n)A_i + I - A_i.$$

For each integer  $n$ ,  $j = 1, \dots, q$  and  $i = 1, \dots, k$ , define a quantity

$$\hat{r}_{nj}^i = r_{nj} 1_{z_n \in Z' \cap \text{int } P_i} + r_j^i 1_{z_n \notin Z' \cap \text{int } P_i}. \quad (5.9)$$

By Proposition 3.5, and from the fact that  $df_n(x_n)$  and  $\Sigma_n$  converge to  $df_0(x_0)$  and  $\Sigma_0$  almost surely respectively,  $\hat{r}_{nj}^i$  converges to  $r_j^i$  in probability as  $n \rightarrow \infty$ .

Now, let us first consider the situations in which  $k \geq 2$ . These are the situations in which  $z_0$  lies on the boundary of some  $q$ -cell. For each  $i = 1, \dots, k$  and  $j = 1, \dots, q$ , choose an  $q$ -dimensional vector  $\bar{h}^{ij}$  such that  $\bar{h}^{ij}$  does not belong to  $K_i$  and its  $j$ th component  $\bar{h}_j^{ij}$  satisfies  $|\bar{h}_j^{ij}| > r_j^i \sqrt{\chi_1^2(\alpha)}$ . Define a random variable  $h_n^{ij} \in \mathbb{R}^q$  by

$$h_n^{ij} = \sqrt{n}(z_n - z_0) 1_{z_n \in Z' \cap \text{int } P_i} + \bar{h}^{ij} 1_{z_n \notin Z' \cap \text{int } P_i}, \quad (5.10)$$

and define another  $q$ -dimensional random variable  $\hat{\Gamma}^{ij}$  by

$$\hat{\Gamma}^{ij} = \Gamma^i 1_{\Gamma^i \in \text{int } K_i} + \bar{h}^{ij} 1_{\Gamma^i \notin \text{int } K_i}. \quad (5.11)$$

Let  $W$  be a measurable subset of  $\text{int } K_i$  with  $\text{Prob}(\Gamma \in \partial W) = 0$ , where  $\partial W$  stands for the boundary of  $W$ . The above definition of  $h_n^{ij}$  and the fact that  $\bar{h}^{ij}$  does not belong to  $K_i$  imply

$$\text{Prob}(h_n^{ij} \in W) = \text{Prob}(\sqrt{n}(z_n - z_0) \in W \text{ and } z_n \in Z' \cap \text{int } P_i).$$

Recalling that  $Z' \cap \text{int } P_i = Z' \cap (z_0 + \text{int } K_i)$  for each  $i = 1, \dots, k$ , we find

$$\text{Prob}(\sqrt{n}(z_n - z_0) \in W \text{ and } z_n \in Z' \cap \text{int } P_i) = \text{Prob}(\sqrt{n}(z_n - z_0) \in W \text{ and } z_n \in Z').$$

Combining the above two equalities with (5.7), we have

$$\lim_{n \rightarrow \infty} \text{Prob}(h_n^{ij} \in W) = \lim_{n \rightarrow \infty} \text{Prob}(\sqrt{n}(z_n - z_0) \in W).$$

By the asymptotic distribution (3.4), the definition of  $\Gamma$  and equation (5.2), we have

$$\lim_{n \rightarrow \infty} \text{Prob}(\sqrt{n}(z_n - z_0) \in W) = \text{Prob}(\Gamma \in W) = \text{Prob}(\Gamma^i \in W).$$

From the definition of  $\hat{\Gamma}^{ij}$  in (5.11) and the facts  $W \subset \text{int } K_i$  and  $\bar{h}^{ij} \notin K_i$  we have

$$\text{Prob}(\Gamma^i \in W) = \text{Prob}(\hat{\Gamma}^{ij} \in W).$$

Combining the above three equalities together, we find

$$\lim_{n \rightarrow \infty} \text{Prob}(h_n^{ij} \in W) = \text{Prob}(\hat{\Gamma}^{ij} \in W)$$

for each measurable set  $W \subset \text{int } K_i$  with  $\text{Prob}(\Gamma \in \partial W) = 0$ . Because the set  $\text{int } K_i$  itself satisfies  $\text{Prob}(\Gamma \in \partial(\text{int } K_i)) = 0$ , the above equality holds with  $\text{int } K_i$  in place of  $W$ . Since  $h_n^{ij}$  and  $\hat{\Gamma}^{ij}$  only take values in  $\text{int } K_i \cup \{\bar{h}^{ij}\}$ , we have  $\lim_{n \rightarrow \infty} \text{Prob}(h_n^{ij} = \bar{h}^{ij}) = \text{Prob}(\hat{\Gamma}^{ij} = \bar{h}^{ij})$ . Also, for each  $W \subset \text{int } K_i$  we have  $\text{Prob}(\Gamma \in \partial W) = \text{Prob}(\hat{\Gamma}^{ij} \in \partial W)$ . It is not hard to see

$$h_n^{ij} \Rightarrow \hat{\Gamma}^{ij}.$$

Since  $\hat{r}_{nj}^i$  converges in probability to the fixed number  $r_j^i$ , which is strictly positive under the assumption in this theorem about  $\Sigma_0$ , we have

$$\frac{(h_n^{ij})_j}{\hat{r}_{nj}^i} \Rightarrow \frac{\hat{\Gamma}_j^{ij}}{r_j^i},$$

where  $(h_n^{ij})_j$  and  $\hat{\Gamma}_j^{ij}$  are the  $j$ th components of  $(h_n^{ij})$  and  $\hat{\Gamma}^{ij}$  respectively. It follows that

$$\lim_{n \rightarrow \infty} \text{Prob} \left( \left| \frac{(h_n^{ij})_j}{\hat{r}_{nj}^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) = \text{Prob} \left( \left| \frac{\hat{\Gamma}_j^{ij}}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right), \quad (5.12)$$

because the probability for  $\frac{\hat{\Gamma}_j^{ij}}{r_j^i}$  to lie on the boundary of  $[-\sqrt{\chi_1^2(\alpha)}, \sqrt{\chi_1^2(\alpha)}]$  is zero. The way  $\hat{\Gamma}^{ij}$  is defined in (5.11) and the fact that  $|\bar{h}_j^{ij}| > r_j^i \sqrt{\chi_1^2(\alpha)}$  imply

$$\text{Prob} \left( \left| \frac{\hat{\Gamma}_j^{ij}}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) = \text{Prob} \left( \left| \frac{\Gamma_j^i}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in \text{int } K_i \right). \quad (5.13)$$

The facts that  $|\bar{h}_j^{ij}| > r_j^i \sqrt{\chi_1^2(\alpha)}$  and that  $\hat{r}_{nj}^i$  almost surely converges to  $r_j^i$  imply

$$\lim_{n \rightarrow \infty} \text{Prob} \left( \left| \frac{\bar{h}_j^{ij}}{\hat{r}_{nj}^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) = 0.$$

We are now ready to put all pieces together to prove (5.4) for the case  $k \geq 2$ . By the definition of  $h_n^{ij}$  in (5.10) and the above equality, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{Prob} \left( \left| \frac{(h_n^{ij})_j}{\hat{r}_{nj}^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) \\ &= \lim_{n \rightarrow \infty} \text{Prob} \left( \frac{\sqrt{n} |(z_n - z_0)_j|}{\hat{r}_{nj}^i} \leq \sqrt{\chi_1^2(\alpha)} \text{ and } z_n \in Z' \cap \text{int } P_i \right). \end{aligned} \quad (5.14)$$

By the definition of  $\hat{r}_{nj}^i$  in (5.9), we can replace it by  $r_{nj}$  in the right hand side of (5.14). Combining (5.12), (5.13) and (5.14) together, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{Prob} \left( \frac{\sqrt{n} |(z_n - z_0)_j|}{r_{nj}} \leq \sqrt{\chi_1^2(\alpha)} \text{ and } z_n \in Z' \cap \text{int } P_i \right) \\ &= \text{Prob} \left( \left| \frac{\Gamma_j^i}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in \text{int } K_i \right) \\ &= \text{Prob} \left( \left| \frac{\Gamma_j^i}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in K_i \right) \end{aligned} \quad (5.15)$$

where the second equality follows from the fact that the probability for  $\Gamma^i$  to belong to the boundary of  $K_i$  is 0. Recalling that  $Z' \cap (z_0 + \text{int } K_i) = Z' \cap \text{int } P_i$ , we can combine the above equality with (5.8) to prove (5.4).

Under the assumption (5.5), we have

$$\begin{aligned} & \sum_{i=1}^k \text{Prob} \left( \left| \frac{\Gamma_j^i}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in K_i \right) \\ &= \sum_{i=1}^k \text{Prob} \left( \left| \frac{\Gamma_j^i}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) \text{Prob}(\Gamma^i \in K_i) \\ &= \sum_{i=1}^k \text{Prob} \left( \left| \frac{\Gamma_j^i}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) \text{Prob}(\Gamma \in K_i) = 1 - \alpha, \end{aligned} \quad (5.16)$$

where the third equality uses (5.2) and the fourth equality uses (5.1) and the fact that  $\sum_{i=1}^k \text{Prob}(\Gamma \in K_i) = 1$ . This proves (5.6) for the case  $k \geq 2$ .

The case  $k = 1$  is much simpler. In this case,  $z_0$  lies in the interior of a single  $q$ -cell  $P_1$ ,  $K_1 = \mathbb{R}^q$ ,  $d(f_0)_S(z_0)$  is an invertible linear map, and  $\Gamma$  is equal to  $\Gamma^1$ . Since  $z_n$  belongs to the interior of  $P_1$  for

all sufficiently large  $n$ , the number  $r_{nj}$  converges almost surely to  $r_j^1$  for each  $j = 1, \dots, q$ . Equation (5.4) follows from the fact that

$$\frac{\sqrt{n}(z_n - z_0)_j}{r_{nj}} \Rightarrow \frac{\Gamma_j^1}{r_j^1},$$

and equation (5.6) follows from the equality (5.1) with  $i = 1$ .  $\square$

Below, we discuss two situations in which the equality (5.5) holds.

The first situation is when  $k \leq 2$ . Obviously, when  $k = 1$  (that is, when  $z_0$  lies in the interior of an  $q$ -cell in the normal manifold of  $S$ ), the cone  $K_1$  is the entire space  $\mathbb{R}^q$ , and the assumption (5.5) automatically holds. It was noted by Michael Lamm that (5.5) also holds when  $k = 2$ . In the latter case, the cone  $K_i$  is a half-space for  $i = 1, 2$ . Since  $\Gamma^i$  is a multivariate normal random variable with mean zero,  $-\Gamma^i$  and  $\Gamma^i$  have the same distribution. It follows that  $\text{Prob}(\Gamma^i \in K_i) = 1/2$ , and that  $\text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in K_i\right)$  and  $\text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } -\Gamma^i \in K_i\right)$  are both equal to  $\frac{1}{2} \text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)}\right)$ . This proves (5.5) when  $k = 2$ .

The second situation is when  $S$  is a box and each  $\Gamma^i$  has a diagonal covariance matrix. In this case, each  $K_i$  is of the form  $\{h \in \mathbb{R}^q \mid h_i \geq 0 \text{ for each } i \in I_+, h_i \leq 0 \text{ for each } i \in I_-\}$  for some disjoint subsets  $I_+$  and  $I_-$  of  $\{1, \dots, q\}$ . Since each  $\Gamma^i$  has a diagonal covariance matrix, its components  $\Gamma_j^i$  are independent of each other. From such independence, and by the symmetry of a mean-0 normal random variable in  $\mathbb{R}$  with respect to the origin, it is not hard to see that (5.5) holds for every  $j$  and  $i$ .

Whenever (5.5) holds, we have (5.6), which means that the interval

$$\left[ (z_n)_j - \frac{\sqrt{\chi_1^2(\alpha)} r_{nj}}{\sqrt{n}}, (z_n)_j + \frac{\sqrt{\chi_1^2(\alpha)} r_{nj}}{\sqrt{n}} \right] \quad (5.17)$$

is an asymptotically exact  $(1 - \alpha)$  confidence interval for  $(z_0)_j$ .

The numerical examples in Section 6 with  $z_0 = 0$  do not belong to either of the above two cases, yet the coverage rates of individual confidence intervals obtained from this method are still reasonable. From those examples we observe that even if the difference between the two sides of (5.5) is large for some  $i$ , the difference tends to be reduced after the summation over all  $i$ 's. As a result, the quantity

$$\sum_{i=1}^k \text{Prob}\left(\left|\frac{\Gamma_j^i}{r_j^i}\right| \leq \sqrt{\chi_1^2(\alpha)} \text{ and } \Gamma^i \in K_i\right)$$

may not be very far from  $1 - \alpha$ . Finally, Theorem 5.1 assumes that  $\Sigma_0$  is nonsingular, which implies the nonsingularity of  $\Sigma_n$  for large  $n$ . Even if  $\Sigma_n$  is singular, one can still use (5.17) to compute a confidence interval, with the caution that the coverage probability for the true solution may not be close to the prescribed level of confidence if  $\Sigma_n$  is very different from  $\Sigma_0$ .

**6. Numerical results.** Before applying the proposed method to numerical examples, we summarize how to use this method in practice.

Among the assumptions, Assumption 1 is standard. For Assumption 2, instead of directly checking if the normal map  $L_K$  is a homeomorphism, we can check if one of its sufficient conditions holds (see the discussion below Assumption 2). Likewise, we can check if (5.5) holds by determining if the

problem belongs to one of the two case discussed below Theorem 5.1. In checking those assumptions, the asymptotical results in [15, 16] can be used to estimate  $d(f_0)_S(z_0)$  and  $K$ .

Confidence regions of  $z_0$  are given by the set (4.2) (if  $\Sigma_0$  is nonsingular) or the set  $R_{n,\epsilon}$  in (4.6) (if  $\Sigma_0$  is singular), and the latter set can be approximated by  $R_{n,0}$ . According to Proposition 3.5, sets (4.2) and  $R_{n,0}$  are ellipsoids with high probability. Computation of simultaneous confidence intervals for  $z_0$  is done by finding the minimal bounding box of its confidence region. Individual confidence intervals of each component of  $z_0$  can be computed using the formula (5.17), provided that (5.5) holds. The quantity  $r_{nj}$  defined in (5.3) depends on  $\Sigma_n$  and  $d(f_n)_S(z_n)$ , and the latter is a nonsingular matrix with high probability.

To convert confidence regions and intervals of  $z_0$  into those of  $x_0$ , the key is to use the equality  $x_0 = \Pi_S(z_0)$ . Suppose the set  $A \subset \mathbb{R}^q$  is a  $(1 - \alpha)100\%$  confidence region for  $z_0$ , then the image of  $A$  under the operator  $\Pi_S$ , denoted by  $\Pi_S(A)$ , contains  $x_0$  with probability at least  $1 - \alpha$ . If  $S$  is a box, then one can easily project the simultaneous confidence intervals of  $z_0$  to obtain simultaneous confidence intervals of  $x_0$ . The latter intervals are conservative, as the probability for the product of all those intervals to contain  $x_0$  is at least  $1 - \alpha$ . When  $S$  is a box, each component of  $x_0$  is the projection of a component of  $z_0$  onto an interval (the product of all such intervals is  $S$ ), and one can project the individual confidence intervals of  $z_0$  to obtain individual confidence intervals of  $x_0$ . For problems in which  $S$  is not a box, the above projection method would not be easy to implement in general, and one would need to exploit special structure in those problems to obtain confidence regions and intervals of  $x_0$ .

**An example with  $q = 2$ .** Here, we apply the method to the same example used in [15, 16]. In this example,  $q = 2$ ,  $d = 6$ ,  $S = \mathbb{R}_+^2$ ,  $F : \mathbb{R}^2 \times \mathbb{R}^6 \rightarrow \mathbb{R}^2$  is defined by

$$F(x, \xi) = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi_5 \\ \xi_6 \end{bmatrix}, \quad (6.1)$$

and the random vector  $\xi$  follows the uniform distribution over the box  $[0, 2] \times [0, 1] \times [0, 2] \times [0, 4] \times [-1, 1] \times [-1, 1]$ . The true problem is

$$0 \in \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} x + N_{\mathbb{R}_+^2}(x). \quad (6.2)$$

The solution to (6.2) is  $x_0 = 0$ , and the solution of the corresponding normal map formulation is  $z_0 = x_0 - E[F(x_0, \xi)] = 0$ . The covariance matrix of  $F(x_0, \xi) = (\xi_5, \xi_6)$  is given by

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

and the B-derivative  $d(f_0)_{\mathbb{R}_+^2}(z_0)$  is a piecewise linear function represented by matrices

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

in orthants  $\mathbb{R}_+^2$ ,  $\mathbb{R}_+ \times \mathbb{R}_-$ ,  $\mathbb{R}_- \times \mathbb{R}_+$  and  $\mathbb{R}_-^2$  respectively. Accordingly, if we define random variables  $\Gamma^i$  as in Section 5, then the covariance matrices of them are

$$\begin{bmatrix} 0.6296 & -0.3704 \\ -0.3704 & 0.2963 \end{bmatrix}, \begin{bmatrix} 0.3333 & -0.3333 \\ -0.3333 & 0.6667 \end{bmatrix}, \begin{bmatrix} 0.3542 & -0.0417 \\ -0.0417 & 0.0833 \end{bmatrix} \text{ and } \begin{bmatrix} 0.3333 & 0 \\ 0 & 0.3333 \end{bmatrix}$$

respectively. An SAA problem with  $n = 10$  is given by

$$0 \in \begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix} x + \begin{bmatrix} -0.1319 \\ -0.2906 \end{bmatrix} + N_{\mathbb{R}_+^2}(x).$$

The SAA solution is  $x_{10} = (0.0782, 0.1097)$ ,  $z_{10} = (0.0782, 0.1097)$ , and the sample covariance matrix of  $F(x_{10}, \xi)$  is

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

The B-derivative  $d\Pi_{\mathbb{R}_+^2}(z_{10})$  is exactly the identity map on  $\mathbb{R}^2$ , and the B-derivative  $d(f_{10})_{\mathbb{R}_+^2}(z_{10})$  is the linear map represented by the matrix

$$\begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.9292 & 0.5400 \\ 0.7536 & 2.1111 \end{bmatrix}.$$

The confidence regions in (4.2) are given by

$$\{z \in \mathbb{R}^2 \mid 10(z - z_{10})^T \begin{bmatrix} 4.8810 & 9.3398 \\ 9.3398 & 24.2564 \end{bmatrix} (z - z_{10}) \leq \chi_2^2(\alpha)\}.$$

Figure 6.1(a) shows boundaries of the above confidence regions. The center of these regions is  $z_{10}$ , marked by ‘×’ in the graph. From the innermost to the outermost, the curves correspond to boundaries of confidence regions for  $z_0$  at levels 10%,  $\dots$ , 90% respectively. The point  $z_0$  is marked by ‘+’ and lies just beyond the 90% confidence region. The dashed rectangle shown in the figure is the minimum enclosing box of the 90% region. Figure 6.1(b) shows confidence regions for  $z_0$  obtained from a different SAA problem with sample size  $n = 30$ ,  $x_{30} = 0$  and  $z_{30} = (-0.0483, -0.0114)$ . Table 6.1 shows the 90% simultaneous and individual confidence intervals for  $z_0$  obtained from the above two SAA problems.

TABLE 6.1  
Confidence intervals of level 90% in the example  $q = 2$

	$n = 10$			$n = 30$		
	Est	Sim CI	Ind CI	Est	Sim CI	Ind CI
$(z_0)_1$	0.08	[-0.52, 0.68]	[-0.38, 0.54]	-0.05	[-0.27, 0.17]	[-0.21, 0.12]
$(z_0)_2$	0.11	[-0.16, 0.38]	[-0.10, 0.32]	-0.01	[-0.23, 0.21]	[-0.18, 0.16]

To test the coverage of confidence intervals obtained from the proposed method, we generate 200 SAA problems with  $n = 10$  and 200 SAA problems with  $n = 30$  from different random seeds, solve them using the PATH solver of GAMS, and compute simultaneous and individual confidence intervals for  $z_0$  of levels 90%, 95% and 99% from the solution for each SAA problem. We count how many times the 2-dimensional box formed by simultaneous confidence intervals cover  $z_0$ , and record the numbers in the row labeled by  $z_0$  in Table 6.2. For example, the 90% simultaneous confidence intervals obtained from 171 SAA problems with  $n = 10$  cover  $z_0$  jointly. We also count how many times each component of  $z_0$  is contained in the corresponding individual confidence intervals, and record the numbers in the remaining rows of Table 6.2. For example, the 90% individual confidence intervals for  $(z_0)_1$  obtained from 164 SAA problems with  $n = 10$  cover the true value  $(z_0)_1 = 0$ .

FIG. 6.1. Confidence regions for  $z_0$  in the example  $q = 2$  at levels 10%,  $\dots$ , 90%

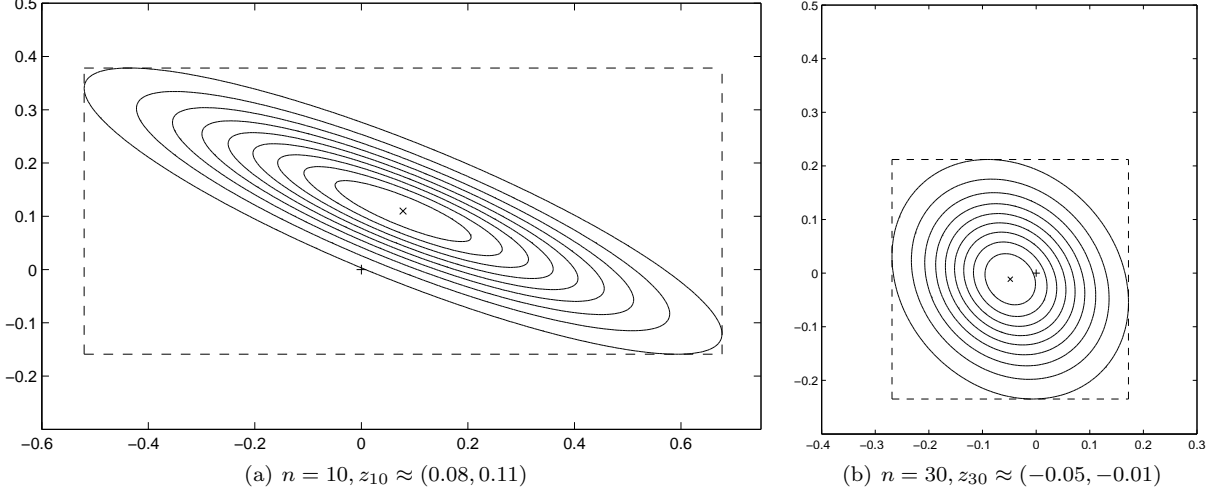


TABLE 6.2

True solution coverage by confidence intervals in the example  $q = 2$  from 200 SAA problems

	$n = 10$			$n = 30$		
	$\alpha = 0.1$	0.05	0.01	$\alpha = 0.1$	0.05	0.01
$z_0$	171	180	187	184	192	197
$(z_0)_1$	164	185	194	172	188	198
$(z_0)_2$	159	175	191	176	186	196

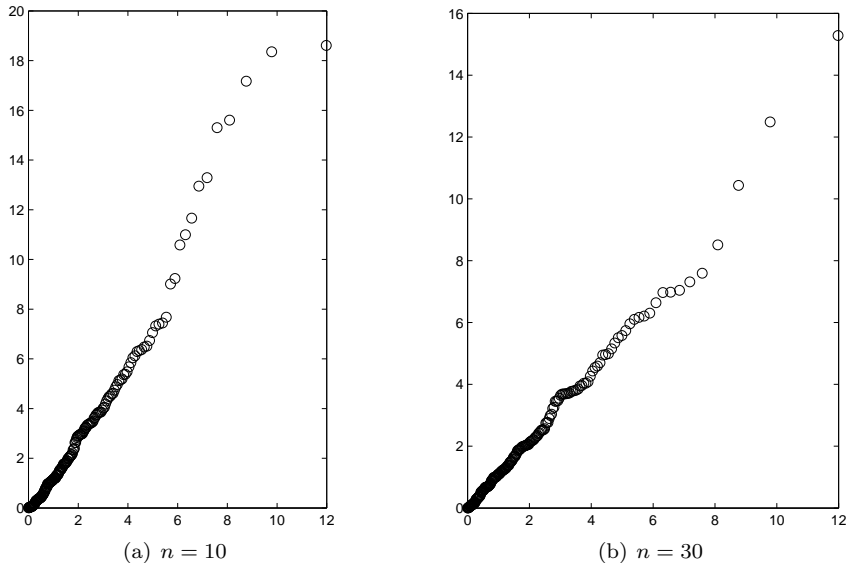
The proposed method generates confidence regions and simultaneous confidence intervals based on the asymptotic distribution in (4.1) (or (4.4) for singular cases). We evaluate how closely  $z_n$  follow the asymptotic distribution using  $\chi^2$  plots. We use the same SAA problems generated above for coverage tests. For each SAA problem with  $n = 10$  we compute the squared distance

$$n[d(f_n)_S(z_n)(z_0 - z_n)]^T \Sigma_n^{-1} [d(f_n)_S(z_n)(z_0 - z_n)],$$

and order these distances from smallest to largest as  $d_{(1)}^2 \leq d_{(2)}^2 \leq \dots \leq d_{(200)}^2$ . For each  $j = 1, \dots, 200$ , 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  $(q_{c,2}((j-1/2)/200), d_{(j)}^2)$  for  $j = 1, \dots, 200$  in Figure 6.2(a), in which the horizontal axis is for quantiles and the vertical axis is for squared distances. Figure 6.2(b) is obtained similarly, from the 200 SAA problems with sample size  $n = 30$ . In each figure the points nearly follow a straight line through the origin with slope around 1, and the slope of the line in Figure 6.2(b) is closer to 1. This suggests that the expression on the left hand side of (4.1) approximately follows the standard normal distribution.

**Examples with  $q = 10$ .** We let  $q = 10$ ,  $d = 110$ ,  $S = \mathbb{R}_+^{10}$ , and  $F : \mathbb{R}^{10} \times \mathbb{R}^{110} \rightarrow \mathbb{R}^{10}$  be defined as  $F(x, \xi) = \Lambda(\xi)x + b(\xi)$ , where  $\Lambda(\xi)$  is a  $10 \times 10$  matrix whose entries are the first 100 components

FIG. 6.2.  $\chi^2$  plots in the example  $q = 2$



of  $\xi$ , and  $b(\xi) \in \mathbb{R}^{10}$  consists of the last 10 components of  $\xi$ . Each diagonal entry of  $\Lambda(\xi)$  is uniformly distributed on the interval  $[0,4]$ , each entry above the main diagonal is uniformly distributed on  $[0,3]$ , and each entry below the main diagonal is uniformly distributed on  $[0,2]$ . Thus,  $E[\Lambda(\xi)] = \Lambda_0$  with  $(\Lambda_0)_{ii} = 2$ ,  $(\Lambda_0)_{ij} = 1.5$  for  $i < j$  and  $(\Lambda_0)_{ij} = 1$  for  $i > j$ . We consider three different choices for the uniform distribution of  $b(\xi)$ , to obtain three different examples. In example 1, each component of  $b(\xi)$  is uniformly distributed on  $[-1,1]$ , so  $E[b(\xi)]_i = 0$  for each  $i$ . In example 2, the first five components of  $b(\xi)$  are uniformly distributed on  $[-1,0.8]$  and the last five components uniformly distributed on  $[-1,1]$ . In example 3, each component of  $b(\xi)$  is uniformly distributed on  $[-1,0.8]$ . The true solution  $z_0$  is given by

$$\begin{aligned} z_0 &= 0 \in \mathbb{R}^{10} \text{ in example 1,} \\ z_0 &= [0.21, 0.43, 0.85, 1.7, 3.4, -6.6, -6.6, -6.6, -6.6, -6.6]^T \times 10^{-2} \text{ in example 2,} \\ z_0 &= [0.01, 0.01, 0.03, 0.05, 0.1, 0.21, 0.42, 0.83, 1.67, 3.34]^T \times 10^{-2} \text{ in example 3.} \end{aligned}$$

For each example, we generate 200 SAA problems with  $n = 50$ , compute the SAA solutions, and obtain simultaneous and individual confidence intervals for  $z_0$  of levels 90%, 95% and 99% from each SAA solution. Table 6.3 lists the averages of the 90% confidence intervals for  $(z_0)_1$  and  $(z_0)_{10}$  obtained from the 200 SAA problems, for each example. Table 6.4 is analogous to Table 6.2. It summarizes the joint coverage of the true solution by simultaneous confidence intervals obtained from the 200 SAA problems for each example, and the coverage of  $(z_0)_1$  and  $(z_0)_{10}$  by the corresponding individual confidence intervals.

**Acknowledgments.** Research of the author was supported by National Science Foundation under the grant DMS-1109099. The author thanks Amarjit Budhiraja, Michael Lamm, Yufeng Liu, Stephen M.



TABLE 6.3  
Average 90% confidence intervals for  $(z_0)_1$  and  $(z_0)_{10}$  in examples  $q = 10$  over 200 SAA problems

	Example 1		Example 2		Example 3	
	Sim CI	Ind CI	Sim CI	Ind CI	Sim CI	Ind CI
$(z_0)_1$	[-0.49, 0.30]	[-0.26, 0.07]	[-0.40, 0.30]	[-0.19, 0.09]	[-0.44, 0.28]	[-0.23, 0.07]
$(z_0)_{10}$	[-0.41, 0.28]	[-0.20, 0.08]	[-0.46, 0.25]	[-0.25, 0.04]	[-0.33, 0.30]	[-0.15, 0.11]

TABLE 6.4  
True solution coverage by confidence intervals in examples with  $q = 10$  from 200 SAA problems

	Example 1			Example 2			Example 3		
	$\alpha = 0.1$	0.05	0.01	$\alpha = 0.1$	0.05	0.01	$\alpha = 0.1$	0.05	0.01
$z_0$	198	199	200	196	198	200	197	197	198
$(z_0)_1$	156	178	194	172	186	192	164	179	195
$(z_0)_{10}$	173	183	195	176	188	197	172	184	195

Robinson and Liang Yin for helpful discussions related to this research, and the referees and the associate editor for comments that have improved the presentation of this paper. Michael Lamm observed that (5.5) holds when  $k = 2$  and provided the idea for the current proof of Proposition 4.3.

#### REFERENCES

- [1] J. S. CHIPMAN, ‘Proofs’ and proofs of the Eckart-Young theorem, in Stochastic Processes and Functional Analysis: in Celebration of M. M. Rao’s 65th Birthday, Jerome A. Goldstein, Neil E. Gretskey, and J. J. Uhl Jr., eds., Marcel Dekker, New York, 1997.
- [2] M. C. DEMIR, *Asymptotics and confidence regions for stochastic variational inequalities*, Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin–Madison, Madison, WI, 2000.
- [3] D. DENTCHEVA AND W. RÖMISCH, *Differential stability of two-stage stochastic programs*, SIAM Journal on Optimization, 11 (2000), pp. 87–112.
- [4] J. DUPAČOVÁ AND R. WETS, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, The annals of statistics, 16 (1988), pp. 1517–1549.
- [5] B. C. EAVES AND U. G. ROTHBLUM, *Relationship of properties of piecewise affine maps over ordered fields*, Linear Algebra and Its Applications, 132 (1990), pp. 1–63.
- [6] F. FACCHINEI AND J. S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Series in Operations Research, Springer-Verlag, New York, 2003. Published in two volumes, paginated continuously.
- [7] G. H. GOLUB, A. HOFFMAN, AND G. W. STEWART, *A generalization of the Eckart-Young-Mirsky matrix approximation theorem*, Linear Algebra and its Applications, 88-89 (1987), pp. 317 – 327.
- [8] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 3d ed., 1996.
- [9] G. GÜRKAN, A. YONCA ÖZGE, AND S. M. ROBINSON, *Sample-path solution of stochastic variational inequalities*, Mathematical Programming, 84 (1999), pp. 313–333.
- [10] A. HAURIE, C. ZACCOUR, J. LEGRAND, AND Y. SMEERS, *A stochastic dynamic Nash-Cournot model for the European gas market*, Technical Report G-87-24, GERAD, École de Hautes Etudes Commerciales, Montréal, Québec, Canada, 1987.
- [11] A. J. KING AND R. T. ROCKAFELLAR, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Mathematics of Operations Research, 18 (1993), pp. 148–162.
- [12] M. LAMM AND S. LU, *Confidence interval computation for stochastic variational inequalities and stochastic Cournot-Nash equilibria*, (2014). Working paper.
- [13] M. LAMM, S. LU, AND A. BUDHIRAJA, *Individual confidence intervals for true solutions to stochastic variational*

- inequalities*, (2014). Working paper.
- [14] R. C. LI, *Relative perturbation theory:  $l_1$  eigenspace and singular space variations*, SIAM Journal on Matrix Analysis and Applications, 20 (1999), pp. 471–492.
  - [15] S. LU, *A new method to build confidence regions for solutions of stochastic variational inequalities.*, Optimization, (2012). Published online before print at <http://www.tandfonline.com/doi/abs/10.1080/02331934.2012.727556>.
  - [16] S. LU AND A. BUDHIRAJA, *Confidence regions for stochastic variational inequalities*, Mathematics of Operations Research, 38 (2013), pp. 545–568.
  - [17] S. LU AND Y. LIU, *Confidence intervals and regions for the Lasso using stochastic variational inequality techniques in optimization.*, (2014). Working paper.
  - [18] G. J. MINTY, *Monotone (nonlinear) operators in hilbert space*, Duke Mathematical Journal, 29 (1962), pp. 341–346.
  - [19] J. R. MUNKRES, *Elements of Algebraic Topology*, Westview Press, 1996.
  - [20] J. S. PANG, *Newton's method for B-differentiable equations*, Mathematics of Operations Research, 15 (1990), pp. 311–341.
  - [21] G. C. PFLUG, *Stochastic optimization and statistical inference*, in Stochastic Programming, A. Ruszczyński and A. Shapiro, eds., vol. 10 of Handbooks in Operations Research and Management Science, Elsevier, 2003.
  - [22] D. RALPH, *A new proof of Robinson's homeomorphism theorem for PL-normal maps*, Linear Algebra and Its Applications, 178 (1993), pp. 249–260.
  - [23] ———, *On branching numbers of normal manifolds*, Nonlinear Analysis: Theory, Methods, and Applications, 22 (1994), pp. 1041–1050.
  - [24] S. M. ROBINSON, *Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity*, Mathematical Programming Studies, 30 (1987), pp. 45–66.
  - [25] ———, *An implicit-function theorem for a class of nonsmooth functions*, Mathematics of Operations Research, 16 (1991), pp. 292–309.
  - [26] ———, *Normal maps induced by linear transformations*, Mathematics of Operations Research, 17 (1992), pp. 691–714.
  - [27] ———, *Sensitivity analysis of variational inequalities by normal-map techniques*, in Variational Inequalities and Network Equilibrium Problems, F. Giannessi and A. Maugeri, eds., New York, 1995, Plenum Press, pp. 257–269.
  - [28] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
  - [29] R. T. ROCKAFELLAR AND R. WETS, *Variational Analysis*, vol. 317 of A Series of Comprehensive Studies in Mathematics, Springer-Verlag, Berlin, 2009.
  - [30] W. RÖMISCH, *Stability of stochastic programming*, in Stochastic Programming, A. Ruszczyński and A. Shapiro, eds., vol. 10 of Handbooks in Operations Research and Management Science, Elsevier, 2003.
  - [31] S. SCHOLTES, *A proof of the branching number bound for normal manifolds*, Linear Algebra and its Applications, 246 (1996), pp. 83–95.
  - [32] ———, *Introduction to Piecewise Differentiable Equations*, SpringerBriefs in Optimization, Springer, 2012.
  - [33] A. SHAPIRO, *On concepts of directional differentiability*, Journal of Optimization Theory and Applications, 66 (1990), pp. 477–487.
  - [34] ———, *Asymptotic behavior of optimal solutions in stochastic programming*, Mathematics of Operations Research, 18 (1993), pp. 829–845.
  - [35] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Lectures on Stochastic Programming: Modeling and Theory*, MPS-SIAM Series on Optimization, SIAM and MPS, Philadelphia, PA, 2009.
  - [36] A. SHAPIRO AND H. XU, *Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation*, Optimization, 57 (2008), pp. 395–418.
  - [37] G. W. STEWART, *On the perturbation of pseudo-inverses, projections and linear least squares problems*, SIAM Review, 19 (1977), pp. 634–662.
  - [38] ———, *On the early history of the singular value decomposition*, SIAM Review, 35 (1993), pp. 551–566.
  - [39] A. W. VAN DER VAART, *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK, 1998.
  - [40] S. VOGEL, *Universal confidence sets for solutions of optimization problems*, SIAM Journal on Optimization, 19 (2008), pp. 1467–1488.
  - [41] H. XU, *Sample average approximation methods for a class of stochastic variational inequality problems*, Asia-Pacific Journal of Operational Research, 27 (2010), pp. 103–119.