.

Individual confidence intervals for true solutions to stochastic variational inequalities

Michael Lamm · **Shu Lu** · **Amarjit Budhiraja.**

Abstract Stochastic variational inequalities (SVI) provide a means for modeling various optimization and equilibrium problems where data are subject to uncertainty. Often it is necessary to estimate the true SVI solution by the solution of a sample average approximation (SAA) problem. This paper proposes three methods for building confidence intervals for components of the true solution, and those intervals are computable from a single SAA solution. The first two methods use an "indirect approach" that requires initially computing asymptotically exact confidence intervals for the solution to the normal map formulation of the SVI. The third method directly constructs confidence intervals for the true SVI solution; intervals produced with this method meet a minimum specified level of confidence in the same situations for which the first two methods are applicable. We justify the three methods theoretically with weak convergence results, discuss how to implement these methods, and test their performance using two numerical examples.

M. Lamm
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, B05 Hanes Hall, CB#3260, Chapel Hill, NC 27599-3260
E-mail: mlamm@email.unc.edu

S. Lu
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 355 Hanes Hall, CB#3260, Chapel Hill, NC 27599-3260
E-mail: shulu@email.unc.edu

A. Budhiraja
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 357 Hanes Hall, CB#3260, Chapel Hill, NC 27599-3260
E-mail: budhiraja@email.unc.edu

1 Introduction

This paper considers the problem of building individual confidence intervals for components of the true solution to a stochastic variational inequality (SVI). An SVI is defined as follows. Let $(\Omega, \mathscr{F}, P)$ be a probability space, and ξ be a d-dimensional random vector defined on Ω and supported on a closed subset Ξ of $\mathbb{R}^d$. Let O be an open subset of $\mathbb{R}^n$, and F be a measurable function from $O \times \Xi$ to $\mathbb{R}^n$, such that for each $x \in O$, $E\|F(x, \xi)\| < \infty$. Let S be a polyhedral convex set in $\mathbb{R}^n$. The SVI problem is to find a point $x \in S \cap O$ such that

$$0 \in f_0(x) + N_S(x), \tag{1}$$

where $f_0(x) = E[F(x, \xi)]$ and $N_S(x) \subset \mathbb{R}^n$ denotes the normal cone to S at x:

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

Here $\langle \cdot, \cdot \rangle$ denotes the scalar product of two vectors of the same dimension.

Variational inequalities provide a means for modeling a variety of optimization and equilibrium problems, see [3, Chapter 1]. Stochastic variational inequalities allow for the incorporation of uncertainty in the model data. As an expectation function, f_0 often does not have a closed form expression and is difficult to evaluate. In such circumstances the problem (1) is replaced by a suitable approximation. This paper considers the case when a sample average approximation (SAA) is used. The SAA method takes independent and identically distributed (i.i.d) random vectors $\xi^1, \xi^2, \ldots, \xi^N$ with the same distribution as ξ and constructs a sample average function $f_N : O \times \Omega \to \mathbb{R}^n$ as

$$f_N(x, \omega) = N^{-1} \sum_{i=1}^{N} F(x, \xi^i(\omega)). \tag{2}$$

The SAA problem is to find for given $\omega \in \Omega$ a point $x \in O \cap S$ such that

$$0 \in f_N(x, \omega) + N_S(x). \tag{3}$$

We will use x_0 to denote a solution to (1) and refer to it as the true solution, and use x_N to denote a solution to (3) and call it an SAA solution; the formal definitions of x_0 and x_N will be given in Assumption 2 and Theorem 1 respectively.

A natural question to ask is how well the SAA solutions approximate the true solution. An answer to this question depends on the convergence behavior of SAA solutions. Under certain regularity conditions, SAA solutions are known to converge almost surely to a true solution as the sample size N goes to infinity, see Gürkan, Özge and Robinson [6], King and Rockafellar [7], and Shapiro, Dentcheva and Ruszczyński [18, Section 5.2.1]. Xu [21] showed the convergence of SAA solutions to the set of true solutions in probability at an exponential rate under some assumptions on the moment generating functions of certain random variables; related results on the exponential convergence rate are given in [19]. Working with the exponential rate of convergence of SAA solutions, Anitescu and Petra in [1] developed confidence intervals for the optimal value of stochastic programming problems

using bootstrapping. The asymptotic distribution of SAA solutions was obtained in King and Rockafellar [7, Theorem 2.7] and Shapiro, Dentcheva and Ruszczyński [18, Section 5.2.2]. For random approximations to deterministic optimization problems, universal confidence sets for the true solution set were developed by Vogel in [20] using concentration of measure results.

The objective of this paper is to provide methods to compute confidence intervals for each individual component of the true solution x_0 from a single SAA solution x_N, based on the asymptotic distribution of SAA solutions. To our knowledge, this line of work started from the dissertation [2] of Demir. By considering the *normal map formulation* (to be defined formally in §2) of variational inequalities, Demir used the asymptotic distribution to obtain an expression for confidence regions of the solution to the normal map formulation of (1), which we denote by z_0 (the formal definition of z_0 is in Assumption 2). Because some quantities in that expression depend on the true solutions x_0 and z_0 and are not computable, Demir proposed a substitution method to make that expression computable. He did not, however, justify why that substitution method preserves the weak convergence property needed for the asymptotic exactness of the confidence regions. The general nonsmooth structure of S creates issues related to discontinuity of certain quantities, which prevents standard techniques from being applicable for the required justification.

In [10] Lu and Budhiraja continued to consider the normal map formulations of both (1) and (3). They provided and justified a new method of constructing asymptotically exact confidence regions for z_0, computable from a solution to the normal map formulation of a single SAA problem (3); the latter solution is denoted by z_N and is formally defined in Theorem 1. The approach in [10] was to combine the asymptotic distribution of z_N with its exponential rate of convergence, and its computation involved calculating a weighted-sum of a family of functions. The method was later simplified by Lu in [8] by using a single function from the family. When z_N does not asymptotically follow a normal distribution, confidence regions generated from [8] and [10] are fractions of multiple ellipses pieced together. Lu [9] proposed a different method to construct asymptotically exact confidence regions, by using only the asymptotic distribution and not the exponential convergence rate. The method in [9] has the advantage that the confidence region generated from it is with high probability a single ellipse, even when the asymptotic distribution of z_N is not normal, and is therefore easier to use. Nonetheless, methods in [8,10] provide valuable information beyond confidence regions. In the present paper we will show how to use such information to compute individual confidence intervals for the true solutions. Even with the estimators from [8,10] in place, it is not straightforward to obtain asymptotically exact individual confidence intervals, due to the piecewise linear structure that underlies the asymptotic distributions of z_N and x_N. How to reduce the computational burden related to that piecewise linear structure is another challenge. Those difficulties are what we aim to overcome in this paper.

Compared to confidence regions, component-wise confidence intervals are usually more convenient to visualize and interpret. By finding the axis-aligned minimal bounding box of a confidence region of z_0 (or x_0), one can find simultaneous confidence intervals, that jointly contain z_0 (or x_0) with a probability no less than a prescribed confidence level. However, individual confidence intervals that can be

obtained by using confidence regions are too conservative for any practical use, especially for large scale problems. Individual confidence intervals provide a quantitative measure of the uncertainty of each individual component, and therefore carry important information not covered by simultaneous intervals. Lu [9] proposed a method to construct individual confidence intervals for z_0, but that method relies on some restrictive assumptions to guarantee the specified level of confidence is met. The methods we develop in this paper are shown to achieve the guaranteed confidence levels in more general situations.

As noted above, the confidence region/interval methods in [2, 8–10] are mainly designed for z_0. The points z_0 and x_0 are related by the equality $x_0 = \Pi_S(z_0)$. From a confidence region of z_0, one can obtain a confidence region for x_0, by projecting the confidence region of z_0 onto S. The resulting set will cover x_0 with a rate at least as large as the coverage rate of the original confidence region for z_0. Simultaneous confidence intervals of x_0 can then be obtained from the minimum bounding box of its confidence region. When S is a box, individual confidence intervals of x_0 can also be obtained from projecting the individual confidence intervals of z_0. We shall refer to such approaches as "indirect approaches." The indirect approaches are convenient to implement when the set S is a box, or has a similar structure that facilitates taking (individual) projections. Beyond those situations, it would be hard to use the indirect approaches for finding confidence intervals for x_0. Another contribution of the present paper is to provide a *direct approach* to finding confidence intervals for x_0.

Altogether, this paper presents three new methods for constructing individual confidence intervals, justifies them with weak convergence results, discusses how to implement these methods, and provides numerical examples. The first two methods belong to the aforementioned indirect approaches. They produce confidence intervals for z_0 from a single z_N, and the asymptotic level of confidence can be specified under general situations. The third method is a direct approach that produces individual confidence intervals for x_0. The intervals produced by the third method meet a specified minimum level of confidence in the same situations for which the first two methods are applicable. While our main interest in this paper is on stochastic variational inequalities and their normal map formulations, the ideas of the first two methods work for general piecewise linear homeomorphisms. We outline the ideas below, and leave formal definitions and proofs to Sections 2 and 3. Throughout, we use $\mathcal{N}(\nu, \Sigma)$ to denote a Normal random vector with mean ν and covariance matrix Σ, and use $Y_n \Rightarrow Y$ to denote the weak convergence of random variables Y_n to Y. For a vector $v \in \mathbb{R}^n$, $(v)_j$ will denote the jth coordinate. Similarly for a function $f : \mathbb{R}^n \to \mathbb{R}^n$, $(f)_j$ will denote the jth component function. We use $\|\cdot\|$ to denote the norm of an element in a normed space; unless a specific norm is stated it can be any norm, as long as the same norm is used in all related contexts.

For the first two methods, suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is a piecewise linear homeomorphism with a family of selection functions $\{M_1, \ldots, M_l\}$ and the corresponding conical subdivision $\{K_1, \ldots, K_l\}$, so f is represented by the linear map M_i when restricted to K_i. Suppose z_N is an n-dimensional random vector such that $\sqrt{N}(z_N - z_0) \Rightarrow f^{-1}(Z)$, where $z_0 \in \mathbb{R}^n$ is an unknown parameter, $Z \sim \mathcal{N}(0, I_n)$, and I_n is the $n \times n$ identity matrix. Our objective is to obtain a confidence interval for $(z_0)_j$, $j = 1, \cdots, n$. The idea of the first method is to look for a number a such that $\Pr(|(f^{-1})_j(Z)| \le a)$

equals a prescribed confidence level, and then use $[(z_N)_j - aN^{-1/2}, (z_N)_j + aN^{-1/2}]$ as the interval. For situations considered in this paper, z_0 and z_N are solutions to the normal map formulations of (1) and (3) respectively, and the function f is unknown and is substituted by an estimator obtained from approaches in [8] and [10]. Such a substitution does not affect the asymptotic exactness of confidence intervals computed from this method, as we show in Theorem 4. In addition, to allow for some choice in where the interval is centered, we introduce a parameter r and consider the probability $\Pr(|(f^{-1})_j(Z) - r| \le a)$.

A challenge that arises with the first method is that when the function f is piecewise linear we lack a closed form expression for the value of a. The computation of $\Pr(|(f^{-1})_j(Z) - r| \le a)$ for fixed a and r requires enumerating all pieces of f^{-1}, and for each such piece one needs to compute the probability for some normal random vector to belong to a certain polyhedron. Thus, the calculations necessary to find a confidence interval increase with the number of pieces in f. These limitations lead to the consideration of upper bounds for interval half-widths, presented in §4, and the development of the second method in this paper.

The second method uses the idea of conditioning. Suppose that for each $\omega \in \Omega$ we can identify a cone $K(\omega)$, such that with high probability $K(\omega)$ belongs to the family $\{K_1, \ldots, K_l\}$ and contains $z_N - z_0$ in its interior; for situations in this paper this can be done using an approach in [8]. For the fixed ω we then look for a number $a(\omega)$ such that the following conditional probability

$$\frac{\Pr\left(|(f^{-1})_j(Z)| \le a(\omega),\, f^{-1}Z \in K(\omega)\right)}{\Pr\left(f^{-1}Z \in K(\omega)\right)}$$

equals a prescribed confidence level, and use $[(z_N)_j - a(\omega)N^{-1/2}, (z_N)_j + a(\omega)N^{-1/2}]$ as a confidence interval for $(z_0)_j$. We will again use an estimator to replace the unknown f, and justify the method with a convergence result (Theorem 5). The second method dramatically reduces the computation needed for the first method, by focusing on a single cone $K(\omega)$ and avoiding the enumeration of all pieces of f.

The third method also uses the idea of conditioning, but it is a direct approach and is different from the second method. In general, one cannot apply the first two methods or the method in [9] directly to compute individual confidence intervals for x_0, because if one would put the asymptotic distribution of x_N in the form $\sqrt{N}(x_N - x_0) \Rightarrow f(Z)$ for some function f then f is generally non-invertible. Such non-invertibility relates to a fact that there is possibly a nonzero probability for some components of x_N and x_0 to coincide, a situation that does not occur when considering solutions to the normal map formulations. The third method handles that non-invertibility by looking into the exact cause of such non-invertibility, and produces intervals that meet a minimum specified level of confidence in the same situations for which the first two methods are shown to be asymptotically exact. In the proof of the convergence result for this method (Theorem 6), we see that the intervals it produces exceed the specified level of confidence only if the corresponding components of x_N and x_0 have a nonzero probability to coincide. When the latter situation happens, the third method returns a point estimate with a nonzero probability.

The organization of this paper is as follows. Section 2 reviews pertinent background material on piecewise linear functions, the normal map formulation and pre-

vious asymptotics results. The main theoretical results of this paper are presented in §3, and §4 presents computational approaches for calculating intervals using these results. The paper concludes in §5 with two numerical examples.

2 Background

In this section we discuss the normal map formulation of a variational inequality, pertinent properties of piecewise linear functions, the notion of B-differentiability and previous works on this topic.

For f_0 and S defined as above the *normal map* induced by f_0 and S is a function $(f_0)_S : \Pi_S^{-1}(O) \to \mathbb{R}^n$, defined as

$$(f_0)_S(z) = f_0(\Pi_S(z)) + (z - \Pi_S(z)). \tag{4}$$

Here Π_S denotes the Euclidian projector onto the set S and $\Pi_S^{-1}(O)$ is the set of all points $z \in \mathbb{R}^n$ such that $\Pi_S(z) \in O$. One can check that $x \in S \cap O$ is a solution to (1) only if $z = x - f_0(x)$ satisfies

$$(f_0)_S(z) = 0. \tag{5}$$

When the above equality is satisfied, one also has $\Pi_S(z) = x$. We refer to (5) as the normal map formulation of (1).

The normal map induced by f_N and S is similarly defined to be a function on $\Pi_S^{-1}(O)$:

$$(f_N)_S(z) = f_N(\Pi_S(z)) + (z - \Pi_S(z)). \tag{6}$$

The normal map formulation of the SAA problem (3) is then

$$(f_N)_S(z) = 0, \tag{7}$$

where (7) and (3) are related in the same manner as (5) and (1). In general for a function G mapping from a subset D of $\mathbb{R}^n$ back into $\mathbb{R}^n$, the normal map G_S induced by G and S is a map from $\Pi_S^{-1}(D)$ into $\mathbb{R}^n$ with $G_S(z) = G(\Pi_S(z)) + z - \Pi_S(z)$.

Since S is a polyhedral convex set, the Euclidian projector Π_S is a piecewise affine function. A continuous function $f : \mathbb{R}^n \to \mathbb{R}^k$ is piecewise affine if there exists a finite family of affine functions $f_j : \mathbb{R}^n \to \mathbb{R}^k$, $j = 1, \dots, m$, such that for all $x \in \mathbb{R}^n$ $f(x) \in \{f_1(x), \dots, f_m(x)\}$. The affine functions f_i, $i = 1, \dots, m$, are referred to as the selection functions of f. When each selection function is linear the function f is called piecewise linear.

Closely related to piecewise affine functions is the concept of a polyhedral subdivision. A polyhedral subdivision of $\mathbb{R}^n$ is defined to be a finite collection of convex polyhedra, $\Gamma = \{\gamma_1, \dots, \gamma_m\} \subset \mathbb{R}^n$, satisfying the following three conditions:

1. Each γ_i is of dimension n.
2. The union of all the γ_i is $\mathbb{R}^n$.
3. The intersection of any two γ_i and γ_j, $1 \le i \ne j \le m$, is either empty or a common proper face of both γ_i and γ_j.

If each of the γ_i is additionally a cone, then Γ is referred to as a conical subdivision. As seen in [17, Proposition 2.2.3], for every piecewise affine function f there is a corresponding polyhedral subdivision of $\mathbb{R}^n$ such that the restriction of f to each γ_i is an affine function. When f is piecewise linear the corresponding subdivision is conical, and the restriction of f to each cone of the subdivision a linear function.

We now return to the special case of the Euclidian projector onto a polyhedral convex set $S \subset \mathbb{R}^n$, a thorough discussion of which can be found in [17, Section 2.4]. Let $\mathscr{F}$ be the finite collection of all nonempty faces of S. On the relative interior of each nonempty face $F \in \mathscr{F}$ the normal cone to S is a constant cone, denoted as $N_S(\mathrm{ri}F)$, and $C_F = F + N_S(\mathrm{ri}F)$ is a polyhedral convex set of dimension n. The collection of all such sets C_F form the polyhedral subdivision of $\mathbb{R}^n$ corresponding to Π_S. This collection of sets is also referred to as the normal manifold of S, with each C_F called an n-cell in the normal manifold. Each k-dimensional face of an n-cell is called a k-cell in the normal manifold for $k = 0, 1, \ldots, n$. The relative interiors of all cells in the normal manifold of S form a partition of $\mathbb{R}^n$.

Next we introduce the concept of B-differentiability. A function $h : \mathbb{R}^n \to \mathbb{R}^m$ is said to be B-differentiable at a point $x \in \mathbb{R}^n$ if there exists a positive homogeneous function, $H : \mathbb{R}^n \to \mathbb{R}^m$, such that

$$h(x+v) = h(x) + H(v) + o(v).$$

Recall that a function G is positive homogeneous if for all positive numbers $\lambda \in \mathbb{R}$ and points $x \in \mathbb{R}^n$ $G(\lambda x) = \lambda G(x)$. The function H is referred to as the B-derivative of h at x and will be denoted $dh(x)$. When in addition to $dh(x)$ being positive homogeneous it is also linear, $dh(x)$ is the classic Fréchet derivative (F-derivative). A function $h : U \times V \to Z$ is partially B-differentiable in x at $(x_0, y_0) \in U \times V$, if the function $h(\cdot, y_0)$ is B-differentiable at x_0. The partial B-derivative is denoted by $d_x h(x_0, y_0)$.

A piecewise affine function f, while not F-differentiable at all points, is B-differentiable everywhere. More precisely, let Γ be the polyhedral subdivision associated with f. At points x in the interior of a polyhedra $\gamma_i \in \Gamma$, $df(x)$ is a linear function equal to $df_i(x)$, the F-derivative of the corresponding selection function f_i. When x lies in the intersection of two or more polyhedra let $\Gamma(x) = \{\gamma_i \in \Gamma | x \in \gamma_i\}$, $I = \{i | \gamma_i \in \Gamma(x)\}$ and $\Gamma'(x) = \{\mathrm{cone}(\gamma_i - x) | i \in I\}$. That is, $\Gamma(x)$ is the collection of elements in Γ that contain x, and $\Gamma'(x)$ is the "globalization" of $\Gamma(x)$ along with a shift of the origin. With this notation, $df(x)$ is piecewise linear with a family of selection functions $\{df_i(x) | i \in I\}$ and the corresponding conical subdivision $\Gamma'(x)$.

The following four assumptions are used to prove pertinent asymptotic properties of SAA solutions.

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. $\omega \in \Omega$.
(c) There exists a square integrable random variable C such that for all $x, x' \in O$

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

for a.e. $\omega \in \Omega$.

From Assumption 1 it follows that f_0 is continuously differentiable on O, see, e.g., [18, Theorem 7.44]. For any nonempty compact subset X of O, let $C^1(X,\mathbb{R}^n)$ be the Banach space of continuously differentiable mappings $f : X \to \mathbb{R}^n$, equipped with the norm

$$\|f\|_{1,X} = \sup_{x \in X} \|f(x)\| + \sup_{x \in X} \|df(x)\|. \tag{8}$$

Then in addition to providing nice integrability properties for f_N, as shown in [18, Theorem 7.48] Assumption 1 will guarantee the almost sure convergence of the sample average function f_N to f_0 as an element of $C^1(X,\mathbb{R}^n)$ and that $df_0(x) = E[d_x F(x,\xi)]$.

Before stating the second assumption we must first define two sets related to the geometric structure of the set S. For polyhedral convex S, the tangent cone to S at a point $x \in S$ is defined to be

$$T_S(x) = \{v \in \mathbb{R}^n \,|\, \text{there exists } t > 0 \text{ such that } x + tv \in S\},$$

and the critical cone to S at a point $z \in \mathbb{R}^n$ is

$$K(z) = T_S(\Pi_S(z)) \cap \{z - \Pi_S(z)\}^\perp.$$

Assumption 2 *Suppose that x_0 solves the variational inequality (1). Let $z_0 = x_0 - f_0(x_0)$, $L = df_0(x_0)$, $K_0 = T_S(x_0) \cap \{z_0 - x_0\}^\perp$, and assume that the normal map L_{K_0} induced by L and K_0 is a homeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$.*

Assumption 3 *Let Σ_0 denote the covariance matrix of $F(x_0,\xi)$. Suppose that the determinant of Σ_0 is strictly positive.*

Assumption 4 *(a) For each $t \in \mathbb{R}^n$ and $x \in X$, let*

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

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

1. *There exists $\zeta > 0$ such that $M_x(t) \le \exp\left\{\zeta^2 \|t\|^2 / 2\right\}$ for every $x \in X$ and every $t \in \mathbb{R}^n$.*
2. *There exists a nonnegative random variable κ such that*

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

   for all $x, x' \in O$ and almost every $\omega \in \Omega$.
3. *The moment generating function of κ is finite valued in a neighborhood of zero.*

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

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

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

1. *There exists $\varsigma > 0$ such that $\mathscr{M}_x(T) \le \exp\left\{\varsigma^2 \|T\|^2 / 2\right\}$ for every $x \in X$ and every $T \in \mathbb{R}^{n \times n}$.*

2. *There exists a nonnegative random variable ν such that*

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

for all $x, x' \in O$ and almost every $\omega \in \Omega$.
3. *The moment generating function of ν is finite valued in a neighborhood of zero.*

Assumptions 1 and 2 ensure that the variational inequality (1) has a locally unique solution under sufficiently small perturbations of f_0 in $C^1(X, \mathbb{R}^n)$, see [10, Lemma 1] and the original result in [16]. It is worth mentioning the relation between the normal map L_{K_0} in Assumption 2 and the normal map $(f_0)_S$. As a piecewise affine function, Π_S is B-differentiable. If we use Γ to denote the normal manifold of S, then Γ is also the polyhedral subdivision for Π_S. Following the discussion of B-differentiability above Assumption 1, $\Gamma'(z_0)$ denotes the conical subdivision that corresponds to $d\Pi_S(z_0)$. Since f_0 is differentiable from Assumption 1, the chain rule of B-differentiability implies $(f_0)_S$ to be B-differentiable, with its B-derivative at z_0 given by

$$d(f_0)_S(z_0)(h) = df_0(x_0)(d\Pi_S(z_0)(h)) + h - d\Pi_S(z_0)(h). \tag{9}$$

The conical subdivision for $d(f_0)_S(z_0)$ is also $\Gamma'(z_0)$. Moreover, as shown in [14, Corollary 4.5] and [13, Lemma 5], for any point $z \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$ sufficiently small the equality

$$\Pi_S(z + h) = \Pi_S(z) + \Pi_{K(z)}(h) \tag{10}$$

holds, which implies

$$d\Pi_S(z) = \Pi_{K(z)} \text{ for any } z \in \mathbb{R}^n. \tag{11}$$

Applying (11) to z_0, one can see the normal map L_{K_0} is exactly $d(f_0)_S(z_0)$, a result that first appeared in [15]. Finally, note that the B-derivative for the normal map $(f_N)_S$, denoted by $d(f_N)_S(\cdot)$, will take an analogous form to (9).

We shall use Σ_N to denote the sample covariance matrix of $\{F(x_N, \xi_i)\}_{i=1}^N$, where x_N is an SAA solution to be formally defined in Theorem 1. Under Assumptions 1 and 2, Σ_N converges almost surely to Σ_0, see [9, Lemma 3.6]. This combined with Assumption 3 implies that for almost every $\omega \in \Omega$ there exists an N_ω such that $\Sigma_N(\omega)$ is invertible for $N \ge N_\omega$.

From Assumption 4 it follows that f_N converges to f_0 in probability at an exponential rate, as shown in [10, Theorem 4] which is based on a general result [18, Theorem 7.67]. That is, there exist positive real numbers β_1, μ_1, M_1 and σ_1, such that the following holds for each $\varepsilon > 0$ and N:

$$\Pr\left(\|f_N - f_0\|_{1,X} \ge \varepsilon\right) \le \beta_1 \exp\{-N\mu_1\} + \frac{M_1}{\varepsilon^n} \exp\left\{-\frac{N\varepsilon^2}{\sigma_1}\right\}. \tag{12}$$

Finally, note that Assumption 4 implies all conditions in Assumption 1; we put Assumption 1 as a separate assumption because some intermediate results do not require the stronger Assumption 4.

The following theorem is adapted from [10, Theorem 7] and provides results relating to the asymptotic distribution of solutions to (3) and (5).

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Let Y_0 be a normal random vector in $\mathbb{R}^n$ with zero mean and covariance matrix Σ_0. Then there exist neighborhoods X_0 of x_0 and Z of z_0 such that the following hold. For almost every $\omega \in \Omega$, there exists an integer N_ω, such that for each $N \geq N_\omega$, the equation (7) has a unique solution z_N in Z, and the variational inequality (3) has a unique solution in X_0 given by $x_N = \Pi_S(z_N)$. Moreover, $\lim_{N \to \infty} z_N = z_0$ and $\lim_{N \to \infty} x_N = x_0$ almost surely,*

$$\sqrt{N}(z_N - z_0) \Rightarrow (L_{K_0})^{-1}(Y_0), \tag{13}$$

$$\sqrt{N} L_{K_0}(z_N - z_0) \Rightarrow Y_0, \tag{14}$$

and

$$\sqrt{N}(\Pi_S(z_N) - \Pi_S(z_0)) \Rightarrow \Pi_{K_0} \circ (L_{K_0})^{-1}(Y_0). \tag{15}$$

Suppose in addition that Assumption 4 holds. Then there exist positive real numbers $\varepsilon_0, \beta_0, \mu_0, M_0$ and σ_0, such that the following holds for each $\varepsilon \in (0, \varepsilon_0]$ and each N:

$$\begin{aligned}
\Pr\left(\|x_N - x_0\| < \varepsilon\right) &\geq \Pr\left(\|z_N - z_0\| < \varepsilon\right) \\
&\geq 1 - \beta_0 \exp\{-N\mu_0\} - \frac{M_0}{\varepsilon^n} \exp\left\{\frac{-N\varepsilon^2}{\sigma_0}\right\}.
\end{aligned} \tag{16}$$

The asymptotic distributions in (13), (14) and (15) depend on z_0 through Σ_0, $L_{K_0} = d(f_0)_S(z_0)$ and $\Pi_{K_0} = d\Pi_S(z_0)$. How to estimate these functions using z_N and the sample data requires special attention, since $d\Pi_S(z_N)$ does not always converge to $d\Pi_S(z_0)$. While $d\Pi_S(\cdot)$ is the same function for all points in the relative interior of a cell in the normal manifold of S [10, Section 5.2], the function changes dramatically across different cells. In particular, if $z_0 \in \mathrm{ri}\, C_j$, where C_j is a k-cell in the normal manifold of S with $k < n$, then $d\Pi_S(z_0)$ is piecewise linear with multiple pieces. In contrast, as shown in [9, Proposition 3.5], under Assumptions 1 and 2, the probability of $d\Pi_S(z_N)$ and $d(f_N)_S(z_N)$ being linear maps goes to one as the sample size N goes to infinity. Thus, as long as z_0 does not belong to the interior of an n-cell in the normal manifold of S, $d\Pi_S(z_N)$ does *not* converge to $d\Pi_S(z_0)$.

In [8] this issue was addressed by choosing a point near z_N, but not necessarily z_N itself, to use in the estimate for $d\Pi_S(\cdot)$. To choose such a point, for each cell C_i in the normal manifold of S define a function $d_i : \mathbb{R}^n \to \mathbb{R}$ by

$$d_i(z) = d(z, C_i) = \min_{x \in C_i} \|x - z\|, \tag{17}$$

and a function $\Psi_i : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Psi_i(\cdot) = d\Pi_S(z)(\cdot) \text{ for any } z \in \mathrm{ri}\, C_i. \tag{18}$$

In (17) any norm for vectors in $\mathbb{R}^n$ can be chosen, and in (18) any $z \in \mathrm{ri}\, C_i$ can be chosen since $d\Pi_S(z)$ is the same function on the relative interior of a cell. Next, choose a function $g : \mathbb{N} \to \mathbb{R}$ satisfying

1. $g(N) > 0$ for each $N \in \mathbb{N}$.
2. $\lim_{N \to \infty} g(N) = \infty$.

3. $\lim\limits_{N\to\infty} \frac{N}{g(N)^2} = \infty$.

4. $\lim\limits_{N\to\infty} g(N)^n \exp\left\{-\sigma_0 \frac{N}{(g(N))^2}\right\} = 0$ for $\sigma_0 = \min\left\{\frac{1}{4\sigma_0}, \frac{1}{4\sigma_1}, \frac{1}{4\sigma_0(E[C])^2}\right\}$, where σ_0, and σ_1 are as in (12) and (16) respectively and C as in Assumption 1.

5. $\lim\limits_{N\to\infty} \frac{N^{n/2}}{g(N)^n} \exp\left\{-\sigma g(N)^2\right\} = 0$ for each positive real number σ.

Note that $g(N) = N^p$ for any $p \in (0, 1/2)$ satisfies 1–5.

Now for each integer N and any point $z \in \mathbb{R}^n$, choose an index i_0 by letting C_{i_0} be a cell that has the smallest dimension among all cells C_i such that $d_i(z) \le 1/g(N)$. Then define functions $\Lambda_N(z) : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Lambda_N(z)(h) = \Psi_{i_0}(h), \tag{19}$$

and $\Phi_N : \Pi_S^{-1}(O) \times \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ by

$$\Phi_N(z, h, \omega) = df_N(\Pi_S(z))(\Lambda_N(z)(h)) + h - \Lambda_N(z)(h). \tag{20}$$

Moving forward we will be interested in $\Phi_N(z_N(\omega), h, \omega)$, which for convenience we will express as $\Phi_N(z_N)(h)$ with the ω suppressed. We shall use z_N^* to denote a point in the relative interior of the cell C_{i_0} associated with (N, z_N). With this notation it follows that $d\Pi_S(z_N^*) = \Psi_{i_0}$ and

$$\Phi_N(z_N)(h) = df_N(\Pi_S(z_N))(d\Pi_S(z_N^*)(h)) + h - d\Pi_S(z_N^*)(h). \tag{21}$$

As shown by Theorem 2 below, the function $\Lambda_N(z_N)$, which is the same as $d\Pi_S(z_N^*)$, provides a nice estimate for $d\Pi_S(z_0)$. The reason behind this result is the following. From (16), there is a high probability for the collection of cells in the normal manifold of S that are within a distance of $1/g(N)$ from z_N to coincide with the collection of cells that contain z_0. Whenever this happens, C_{i_0} is the cell that contains z_0 in its relative interior, and the two points z_N^* and z_0 belong to the relative interior of the same cell C_{i_0}, with $d\Pi_S(z_N^*) = d\Pi_S(z_0)$. This observation will be used in the proofs of Theorems 5 and 6 (with the definition of event A_N in (35)). Theorem 2 below was proved in [8, Corollaries 3.2 and 3.3].

Theorem 2 *Suppose that Assumptions 2 and 4 hold. For each $N \in \mathbb{N}$, let Λ_N and Φ_N be as defined in (19) and (20). Then*

$$\lim_{N\to\infty} \Pr\left[\Lambda_N(z_N)(h) = d\Pi_S(z_0)(h) \text{ for all } h \in \mathbb{R}^n\right] = 1, \tag{22}$$

and there exists a positive real number θ, such that

$$\lim_{N\to\infty} \Pr\left[\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Phi_N(z_N)(h) - d(f_0)_S(z_0)(h)\|}{\|h\|} < \frac{\theta}{g(N)}\right] = 1. \tag{23}$$

Moreover suppose Assumption 3 holds, and let Σ_N be as defined above. Then

$$\sqrt{N}\Sigma_0^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow \mathcal{N}(0, I_n),$$

and

$$\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow \mathcal{N}(0, I_n). \tag{24}$$

In contrast to (13) and (14), the quantities in (24) are computable using only the sample data, providing a basis for building confidence regions of z_0. Additionally, (11) and (22) suggest the use of Λ_N as an estimate for Π_{K_0} when developing methods for building confidence intervals for x_0. Similar results were shown in [10] but with Λ_N taken to be a weighted average of all the functions Ψ_i satisfying $d_i(z_N) \leq 1/g(N)$.

In [9] a different tack was taken on constructing confidence regions. Instead of estimating functions that converge to $d(f_0)_S(z_0)$, it was shown that under Assumptions 1 and 2, the difference of $-\sqrt{N}d(f_N)_S(z_N)(z_0-z_N)$ and $\sqrt{N}d(f_0)_S(z_0)(z_N-z_0)$ converges to zero in probability, and consequently that

$$-\sqrt{N}d(f_N)_S(z_N)(z_0 - z_N) \Rightarrow Y_0.$$

Because $d(f_N)_S(z_N)$ is a linear function with high probability, even when $d(f_0)_S(z_0)$ is piecewise linear, the above expression provides an easier method to calculate confidence regions and simultaneous confidence intervals.

As noted earlier, confidence regions do not directly lead to useful individual confidence intervals. The papers [8] and [10] did not discuss how to compute individual confidence intervals, while [9] provided a method for such computation. Below we briefly introduce the latter method.

With the notation used above (9), let Γ denote the normal manifold of S and $\Gamma'(z_0)$ denote the conical subdivision that corresponds to $d\Pi_S(z_0)$, which is also the conical subdivision for $d(f_0)_S(z_0)$. Suppose $\Gamma'(z_0) = \{K_1, \ldots, K_k\}$. Then for each $i = 1, \ldots, k$, the restriction of $d(f_0)_S(z_0)$ on K_i, which we denote by $d(f_0)_S(z_0)|_{K_i}$, coincides with a linear function; let M_i be the matrix representing that linear function. Moreover under Assumption 2, $d(f_0)_S(z_0)$ is a global homeomorphism so each matrix M_i is invertible. We then define $Y^i = M_i^{-1}Y_0$. Since Y_0 is a multivariate normal random vector each Y^i is a multivariate normal random vector with covariance matrix $M_i^{-1}\Sigma_0 M_i^{-T}$.

We define the number $r_j^i = \sqrt{(M_i^{-1}\Sigma_0 M_i^{-T})_{jj}}$ for each $i = 1, \ldots, k$ and $j = 1, \ldots, n$. Finally for each $\alpha \in (0,1)$ let $\chi_1^2(\alpha)$ be the $(1-\alpha)^{\text{th}}$ percentile of a χ^2 random variable with one degree of freedom. It then follows that

$$\Pr\left(|(Y^i)_j| \leq r_j^i\sqrt{\chi_1^2(\alpha)}\right) = 1-\alpha.$$

The following theorem on individual confidence intervals for components of z_0 was proven in [9, Theorem 5.1].

Theorem 3 *Suppose that Assumptions 1, 2 and 3 hold. Let K_i, M_i, Y^i and r_j^i be defined as above. For each integer N with $d(f_N)_S(z_N)$ being an invertible linear map, define a number*

$$r_{Nj} = \sqrt{(d(f_N)_S(z_N)^{-1}\Sigma_N d(f_N)_S(z_N)^{-T})_{jj}}$$

for each $j = 1, \ldots, n$. Let $r_{Nj} = 0$ if $d(f_N)_S(z_N)$ is not an invertible linear map. Then for each real number $\alpha \in (0,1)$ and for each $j = 1, \ldots, n$,

$$\lim_{N\to\infty} \Pr\left(\frac{\sqrt{N}|(z_n - z_0)_j|}{r_{Nj}} \leq \sqrt{\chi_1^2(\alpha)}\right)$$

$$= \sum_{i=1}^{k} \Pr\left(\left| \frac{(Y^i)_j}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \ \textit{and} \ Y^i \in K_i \right) \tag{25}$$

Moreover, suppose for a given $j = 1, \ldots, n$ that the equality

$$\Pr\left(\left| \frac{(Y^i)_j}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \ \textit{and} \ Y^i \in K_i \right) = \Pr\left(\left| \frac{(Y^i)_j}{r_j^i} \right| \leq \sqrt{\chi_1^2(\alpha)} \right) \Pr\left(Y^i \in K_i \right)$$

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

$$\lim_{N \to \infty} \Pr\left(|(z_N - z_0)_j| \leq \frac{\sqrt{\chi_1^2(\alpha)} r_{Nj}}{\sqrt{N}} \right) = 1 - \alpha.$$

We see in (25) that this method of constructing individual confidence intervals, while easily computable using only the sample data, produces intervals whose asymptotic level of confidence is dependant on the true solution, unless the condition below (25) is satisfied. The latter condition is satisfies, when $d(f_0)_S(z_0)$ is a linear function or has only two selection functions, in which case the intervals computed from this method will be asymptotically exact. In general, however, the level of confidence for such intervals cannot be guaranteed. This limitation motivates the development of methods proposed in the following section.

3 New methods for building individual confidence intervals

In this section we present three new methods for building individual confidence intervals. The first two methods produce intervals for $(z_0)_j$, that have a specified level of confidence for situations more general than the method examined in Theorem 3. Those two methods rely on the estimate $\Phi_N(z_N)$; when $\Phi_N(z_N)$ is a linear function, they return the same interval as the method examined in Theorem 3. The methods differ when $\Phi_N(z_N)$ is piecewise linear. The first method (given in Theorem 4) uses all selection functions of $\Phi_N(z_N)$ to calculate an interval. The second (given in Theorem 5) uses z_N to determine a subset of selection functions to be used in an interval's computation. When the set S is a box these intervals can be projected onto S to produce intervals that cover $(x_0)_j$ at a rate at least as large as the coverage rate of $(z_0)_j$ by the initial intervals.

The third method (given in Theorem 6) considers the computation of individual confidence intervals for x_0 directly. This method estimates the function that appears in the right-hand of (15) by using both the function Λ_N as defined in (19) and the function $\Phi_N(z_N)$. Initially these two functions are considered separately, and the relation between x_N and z_N is used to emulate the approach of the second method. When calculating an interval's length, with high probability one only need to consider a single selection function of the estimate constructed from Λ_N and $\Phi_N(z_N)$.

3.1 The first method (an indirect approach)

In this method, we compute confidence intervals for $(z_0)_j$, for each $j = 1, \cdots, n$, based on equation (31) in Theorem 4. In that equation, r is an arbitrarily chosen real number, and $a^r(\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j)$ returns a number determined by the jth component of the function $\Phi_N^{-1}(z_N)\Sigma_N^{1/2}$. In the following, we start with the definition of $a^r(\cdot)$.

Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and $Z \sim \mathcal{N}(0, I_n)$. Suppose that $\Pr(\psi(Z) = b) = 0$ for all b and $\Pr(\beta_1 < \psi(Z) < \beta_2) > 0$ for all $\beta_1 < \beta_2$. Then given any $\alpha \in (0, 1)$ and $r \in \mathbb{R}$ there exists a unique point $a^r(\psi) \in (0, \infty)$ such that

$$\Pr(-a^r(\psi) \le \psi(Z) - r \le a^r(\psi)) = 1 - \alpha.$$

Let $\alpha \in (0, 1)$ be fixed. For any function $f : \mathbb{R}^n \to \mathbb{R}$, define

$$a^r(f) = \inf\{l \ge 0 \,|\, \Pr(-l \le f(Z) - r \le l) \ge 1 - \alpha\}. \tag{26}$$

It then follows that

1. $a^r(f) < \infty$.
2. $\Pr(-a^r(f) \le f(Z) - r \le a^r(f)) \ge 1 - \alpha$.
3. $\Pr(-(a^r(f) - \delta) \le f(Z) - r \le a^r(f) - \delta) < 1 - \alpha$ for all $\delta > 0$.

In the proof of Theorem 4 we use the following two lemmas.

Lemma 1 *Let ψ be as above and $\{\psi_N\}_{N=1}^{\infty}$ be a sequence of functions that converges pointwise to ψ. Then for any $r \in \mathbb{R}$, $\lim_{N\to\infty} a^r(\psi_N) = a^r(\psi)$.*

Proof Note $\sup_N a^r(\psi_N) < \infty$. This follows from the fact that $\psi_N(Z)$ converges to $\psi(Z)$ a.s. and so $\{\psi_N(Z)\}_{N=0}^{\infty}$ is tight. Next fix a subsequence, again indexed by N, along which $a^r(\psi_N) \to a^*$. It suffices to show $a^* = a^r(\psi)$.

Note that $a^* \ne 0$. If this were the case then for every $\varepsilon > 0$

$$1 - \alpha \le \lim_{N\to\infty} \Pr(-\varepsilon \le \psi_N(Z) - r \le \varepsilon) = \Pr(-\varepsilon \le \psi(Z) - r \le \varepsilon).$$

Since ε is arbitrary this would imply $\Pr(\psi(Z) = r) \ge 1 - \alpha$, a contradiction.

Assume now without loss of generality that $\inf_N a^r(\psi_N) > 0$. Then

$$1 - \alpha \le \lim_{N\to\infty} \Pr\left(-1 \le \frac{\psi_N(Z) - r}{a^r(\psi_N)} \le 1\right) = \Pr\left(-1 \le \frac{\psi(Z) - r}{a^*} \le 1\right). \tag{27}$$

Applying the same argument for all $0 < \delta < \inf_N a^r(\psi_N)$ we see that

$$\Pr\left(-1 \le \frac{\psi(Z) - r}{(a^* - \delta)} \le 1\right) \le 1 - \alpha.$$

Sending δ to 0 we obtain $\Pr(-a^* \le \psi(Z) - r \le a^*) \le 1 - \alpha$, which combined with (27) gives

$$\Pr(-a^* \le \psi(Z) - r \le a^*) = 1 - \alpha.$$

Thus $a^* = a^r(\psi)$, and $\lim_{N\to\infty} a^r(\psi_N) = a^r(\psi)$.

$\square$

Let $C(\mathbb{R}^n, \mathbb{R})$ denote the space of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}$. Equipped with the local uniform topology, this is a Polish space.

Lemma 2 *Let $\{\psi_N\}_{N=1}^{\infty}$ be a sequence of $C(\mathbb{R}^n, \mathbb{R})$ valued random variables which converges in distribution to ψ. Also let $\{Z_N\}_{N=1}^{\infty}$ be a sequence of $\mathbb{R}^n$ valued random variables converging in distribution to Z. Then for any $r \in \mathbb{R}$,*

$$\Pr\left(-a^r(\psi_N) \le \psi_N(Z_N) - r \le a^r(\psi_N)\right) \to 1 - \alpha.$$

Proof By Lemma 1 and the convergence of ψ_N to ψ, it follows that $a^r(\psi_N) \to a^r(\psi)$ in probability. Also since $a^r(\psi) > 0$,

$$\frac{1}{a^r(\psi_N)} \mathbb{1}_{a^r(\psi_N) > 0} \to \frac{1}{a^r(\psi)}$$

in probability, where $\mathbb{1}_{a^r(\psi_N) > 0}$ is the indicator random variable for the event $a^r(\psi_N) > 0$. Let A_N denote the event that $a^r(\psi_N) > 0$. Then

$$\Pr\left(-a^r(\psi_N) \le \psi_N(Z_N) - r \le a^r(\psi_N)\right) = \Pr\left(A_N; -1 \le \frac{\psi_N(Z_N) - r}{a^r(\psi_N)} \le 1\right)$$
$$+ \Pr\left(A_N^c; -a^r(\psi_N) \le \psi_N(Z_N) - r \le a^r(\psi_N)\right).$$

By $a^r(\psi_N) \to a^r(\psi)$ in probability and $a^r(\psi) > 0$, it follows that $\Pr(A_N) \to 1$. Therefore,

$$\Pr\left(A_N^c; -a^r(\psi_N) \le \psi_N(Z_N) - r \le a^r(\psi_N)\right) \to 0 \text{ as } N \to \infty.$$

Let B_N be the event that $-1 \le \frac{\psi_n(Z_N) - r}{a^r(\psi_N)} \mathbb{1}_{a^r(\psi_N) > 0} \le 1$. By the convergence of ψ_N to ψ and Z_N to Z, we have that $\psi_N(Z_N) \Rightarrow \psi(Z)$ in distribution, and thus

$$\Pr(B_N) \to \Pr\left(-1 \le \frac{\psi(Z) - r}{a^r(\psi)} \le 1\right) = \Pr\left(-a^r(\psi) \le \psi(Z) - r \le a^r(\psi)\right) = 1 - \alpha.$$

Consequently, $\Pr\left(-a^r(\psi_N) \le \psi_N(Z_N) - r \le a^r(\psi_N)\right) \to 1 - \alpha.$

$\square$

The application of these lemmas to our problem of interest is facilitated by the following two propositions.

Proposition 1 *(a) Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a piecewise linear function and $\{f_N\}_{N=1}^{\infty}$ a sequence of piecewise linear functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ with*

$$\sup_{h \in \mathbb{R}^n, h \ne 0} \frac{\|f_N(h) - f(h)\|}{\|h\|} \to 0. \tag{28}$$

Suppose that there exists a conical subdivision $\Gamma = \{\gamma_1, \gamma_2 \dots \gamma_m\}$ of $\mathbb{R}^n$ such that for all N sufficiently large $f_N|_{\gamma_i} = A_{N,i}$ and $f|_{\gamma_i} = A_i$ are linear functions for each γ_i. Then

$$\sup_{h \in \mathbb{R}^n, h \ne 0} \frac{\|A_{N,i}h - A_i h\|}{\|h\|} \to 0 \text{ for } i = 1, \dots, m. \tag{29}$$

(b) Suppose in addition that f is a homeomorphism. Then for all N sufficiently large f_N is a homeomorphism and f_N^{-1} converges uniformly on compacts to f^{-1}.

Proof By (28), $\sup_{h\in\gamma_i, h\neq 0} \frac{\|A_{N,i}h - A_ih\|}{\|h\|}$ converges to 0 as $N \to \infty$, for each $i = 1,\ldots,m$. As Γ is a conical subdivision of $\mathbb{R}^n$, γ_i is of dimension n which means that it contains a ball in $\mathbb{R}^n$. The fact that $\|A_{N,i}h - A_ih\|$ converges to 0 for all h in a ball implies that the matrix $A_{N,i}$ converges to A_i, giving (29).

To prove (b) first note that since f is a homeomorphism, A_i^{-1} is well defined for each i and $\{A_1^{-1}, A_2^{-1}, \ldots, A_m^{-1}\}$ provides a family of selection functions for f^{-1} [17, Proposition 2.3.2]. Moreover we have that f^{-1} is Lipschitz continuous with Lipschitz constant

$$\delta = \max_{1\leq i\leq m} \left(\|A_i^{-1}\|\right) < \infty.$$

Similarly for N sufficiently large the functions $f_N - f$ will be piecewise linear with a family of selection functions given by $\{A_{N,1} - A_1, \ldots, A_{N,m} - A_m\}$, and thus Lipschitz continuous with Lipschitz constant

$$\rho_N = \max_{1\leq i\leq m} \left(\|A_{N,i} - A_i\|\right)$$

From part (a) we have $\lim_{N\to\infty}\|A_{N,i} - A_i\| = 0$ for each i, so for all N sufficiently large $\rho_N < \delta^{-1}$. From [14, Lemma 3.1] it then follows that f_N is a homeomorphism for N sufficiently large.

To obtain $f_N^{-1} \to f^{-1}$ uniformly on compacts, note first from $\lim_{N\to\infty} A_{N,i}^{-1} = A_i^{-1}$ it follows that $\{f_N^{-1}\}_{N=v}^{\infty}$ is uniformly Lipschitz continuous for v large enough. Then for any compact set C and any subsequence of f_N^{-1} there exists a further subsequence, $f_{N_k}^{-1}$ that converges uniformly on C to some function g. To prove part (b) it then suffices to show that $g(x) = f^{-1}(x)$.

To see that this holds let $x \in C, \alpha_k = f_{N_k}^{-1}(x)$, and $\alpha = g(x)$. By $\alpha_k \to \alpha$ and $f_{N_k} \to f$ it follows that $f_{N_k}(\alpha_k) \to f(\alpha)$. Also for each k

$$f_{N_k}(\alpha_k) = f_{N_k}(f_{N_k}^{-1}(x)) = x.$$

Thus $x = f(\alpha) = f(g(x))$, or $g(x) = f^{-1}(x)$, the desired result.

$\square$

Proposition 2 *Suppose that Assumptions 2, 3 and 4 hold, and for each $N \in \mathbb{N}$ let $\Phi_N(z_N)$ be as in (21). Then $\Phi_N^{-1}(z_N)\Sigma_N^{1/2}$ converges to $d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}$ in probability, uniformly on compacts.*

Proof As previously noted, when Assumption 4 holds the conditions of Assumption 1 are satisfied, and under Assumptions 1 and 2 Σ_N converges almost surely to Σ_0. Convergence of Σ_N to Σ_0 and (23) imply that for all $\varepsilon > 0$

$$\lim_{N\to\infty} \Pr\left(\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Sigma_N^{-1/2}\Phi_N(z_N)(h) - \Sigma_0^{-1/2}d(f_0)_S(z_0)(h)\|}{\|h\|} < \varepsilon\right) = 1. \quad (30)$$

By a standard subsequential argument we can assume without loss of generality that almost surely

$$\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Sigma_N^{-1/2}\Phi_N(z_N)(h) - \Sigma_0^{-1/2}d(f_0)_S(z_0)(h)\|}{\|h\|} \to 0.$$

In order to show almost sure convergence of $\Phi_N^{-1}(z_N)\Sigma_N^{1/2}$ to $d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}$ we will apply Proposition 1. It suffices then that for a.e. ω, with $f_N = \Phi_N^{-1}(z_N(\omega))\Sigma_N^{1/2}(\omega)$ and $f = d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}$, conditions of Proposition 1 are satisfied.

To this end recall the expressions for $d(f_0)_S(z_0)$ given in (9), $\Phi_N(z_N)$ given in (20) and $\Lambda_N(z_N)$ given in (19). From these it is clear that the conditions in part (a) of Proposition 1 will be satisfied if we can find a conical subdivision Γ such that for every $\gamma_i \in \Gamma$ and $z \in \mathbb{R}^n$, $d\Pi_S(z)|_{\gamma_i}$ is equal to a linear function.

Let $C_1, \ldots, C_l$ be all of the k-cells in the normal manifold of S, $k = 0, 1, \ldots, n$. Then for every $z \in \mathbb{R}^n$, $z \in \mathrm{ri}\,C_j$ for some j, and $d\Pi_S(z)(\cdot) = \Psi_j(\cdot)$ for Ψ_j defined as in (18). The desired subdivision Γ can be constructed by taking the collection of all cones with non-empty interior of the form $\gamma = \cap_{k=1}^l \gamma_k$ where each γ_k is from a conical subdivision of Ψ_k.

Finally by Assumptions 2 and 3, $\Sigma_0^{-1/2}d(f_0)_S(z_0)$ is a homeomorphism, satisfying the condition in part (b) of Proposition 1. The result follows.

$\square$

At this point we are able to present the main result for our first method on computation of asymptotically exact individual confidence intervals.

Theorem 4 *Suppose that Assumptions 2, 3 and 4 hold. Let $\alpha \in (0,1)$, $r \in \mathbb{R}$, and let $a^r(\cdot)$ be as defined in (26). Then for every $j = 1, \ldots, n$,*

$$\lim_{N \to \infty} \Pr\left(\left| \sqrt{N}(z_N - z_0)_j - r \right| \leq a^r\left((\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j \right) \right) = 1 - \alpha. \qquad (31)$$

Proof By Proposition 2, $(\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j$ converges to $(L_K^{-1}\Sigma_0^{1/2})_j$ in $C(\mathbb{R}^n, \mathbb{R})$, in probability. Since $L_K^{-1}\Sigma_0^{1/2}$ is a piecewise linear homeomorphism it follows that for $Z \sim N(0, I_n)$ and each $j = 1, \ldots, n$,

$$\Pr\left((L_K^{-1}\Sigma_0^{1/2})_j(Z) = b \right) = 0 \text{ for all } b$$

and

$$\Pr\left(\beta_1 < (L_K^{-1}\Sigma_0^{1/2})_j(Z) < \beta_2 \right) > 0 \text{ for all } \beta_1 < \beta_2.$$

Taking $Z_N = \sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0)$, by Theorem 2 (see (24)) Z_N converges in distribution to Z. Then with $\psi_N = (\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j$, and $\psi = (L_K^{-1}\Sigma_0^{1/2})_j$, it follows from Lemma 2 that

$$\begin{aligned}
& \Pr\left(-a^r(\psi_N) \leq \psi_N(Z_N) - r \leq a^r(\psi_N) \right) \\
= \ & \Pr\left(-a^r(\psi_N) \leq (\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j\left(\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N - z_0) \right) - r \leq a^r(\psi_N) \right) \\
= \ & \Pr\left(-a^r(\psi_N) \leq \sqrt{N}(\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j\left(\Sigma_N^{-1/2}\Phi_N(z_N - z_0) \right) - r \leq a^r(\psi_N) \right) \\
= \ & \Pr\left(-a^r\left((\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j \right) \leq \sqrt{N}(z_N - z_0)_j - r \leq a^r\left((\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j \right) \right)
\end{aligned}$$

converges to $1 - \alpha$ as $N \to \infty$.

$\square$

While Theorem 4 proves the asymptotic exactness of intervals for a general choice of r, (31) and (26) indicate how the choice of r will affect both an interval's center and length. Additionally, when $\Phi_N(z_N)$ is piecewise linear evaluating $a^r\left(\left(\Phi_N^{-1}(z_N)\Sigma_N^{1/2}\right)_j\right)$ requires working with each selection function, which can pose a computational challenge if the number of selection functions is large. The second method limits the computational burden of working with a piecewise linear function by considering only a subset of selection functions indicated by z_N.

3.2 The second method (an indirect approach)

In this method, we compute confidence intervals for $(z_0)_j$, for each $j = 1, \cdots, n$, based on equation (34) in Theorem 5, in which $\eta_j^\alpha(\cdot, \cdot)$ replaces $a^r(\cdot)$ in the first method to determine an interval's width. Below we give the definition of $\eta_j^\alpha(\cdot, \cdot;)$. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a piecewise linear homeomorphism with a family of selection functions $\{M_1, \ldots, M_l\}$, and the corresponding conical subdivision $\{K_1, \ldots, K_l\}$. As before, let $(f)_j$ denote the jth component function of f. For any choice of cone K_i, $i = 1, \ldots, l$, component $j = 1, \ldots, n$ and $\alpha \in (0,1)$ we first define $\eta_j^\alpha(f, x)$ for points $x \in \text{int} K_i$ as the unique and strictly positive number satisfying

$$\Pr\left(\left|\left(f^{-1}(Z)\right)_j\right| \le \eta_j^\alpha(f, x),\ f^{-1}(Z) \in K_i\right) = (1-\alpha)\Pr\left(f^{-1}(Z) \in K_i\right). \quad (32)$$

Note that $\eta_j^\alpha(f, x)$ is the same number for all $x \in \text{int} K_i$, since nothing in the above definition depends on the exact location of x, except that K_i has to be the cone containing x in its interior. Because f is a homeomorphism we can rewrite (32) as

$$\Pr\left(\left|\left(M_i^{-1}Z\right)_j\right| \le \eta_j^\alpha(f, x),\ M_i^{-1}Z \in K_i\right) = (1-\alpha)\Pr\left(M_i^{-1}Z \in K_i\right). \quad (33)$$

For points $x \in \bigcap_{s=1}^k K_{i_s}$ define $\eta_j^\alpha(f, x) = \max_{s=1,\ldots,k} \eta_j^\alpha(f, x_{i_s})$ where $x_{i_s} \in \text{int} K_{i_s}$.

The following Lemma will play a similar role in the proof of Theorem 5 as Lemma 1 did in the proof of Theorem 4.

Lemma 3 *Let $\{f_m\}_{m=1}^\infty$ be a sequence of piecewise linear functions such that for all m sufficiently large f_m and f have a common conical subdivision $\{K_1, \ldots, K_l\}$, and*

$$\sup_{h\in\mathbb{R}^n, h\ne0} \frac{\|f_m(h) - f(h)\|}{\|h\|} \to 0.$$

Then for all m sufficiently large f_m will be a homeomorphism and for all $\alpha \in (0,1)$, $x \in \mathbb{R}^n$ and $j = 1, \ldots, n$ one has $\eta_j^\alpha(f_m, x) \to \eta_j^\alpha(f, x)$.

Proof From Proposition 1 it follows that f_m will be a homeomorphism for all m sufficiently large. The convergence of $\eta_j^\alpha(f_m, x)$ to $\eta_j^\alpha(f, x)$ can be shown using an argument analogous to the one used in the proof of Lemma 1 and is therefore omitted.

□

In the proof of Theorem 5 we make use of the notation introduced before Theorem 3. With this notation $\Gamma'(z_0) = \{K_1, \ldots, K_k\}$ is the conical subdivision associated with $d(f_0)_S(z_0)$ such that $d(f_0)_S(z_0)|_{K_i} = M_i$ and $K_i = \text{cone}(P_i - z_0)$ where $P_1, \ldots, P_k$ are all n-cells in the normal manifold of S that contain z_0. Note that for $i = 1, \ldots, k$ we can write $Y^i = M_i^{-1} \Sigma_0^{1/2} Z$ and $Y_0 = \Sigma_0^{1/2} Z$ where $Z \sim \mathscr{N}(0, I)$. Finally we define $Y^* = d(f_0)_S^{-1}(z_0) \Sigma_0^{1/2} Z$, and note that $Y^* \mathbb{1}_{Y^* \in K_i} = Y^i \mathbb{1}_{Y^i \in K_i}$.

Theorem 5 *Let Assumptions 2, 3 and 4 hold. Then with $\Phi_N(z_N)(\cdot)$ and z_N^* as defined in (21) one has that for all $j = 1, \ldots, n$ and $\alpha \in (0, 1)$,*

$$\Pr\left(\sqrt{N} |(z_N - z_0)_j| \leq \eta_j^\alpha(\Sigma_N^{-1/2} \Phi_N(z_N), z_N - z_N^*)\right) \to 1 - \alpha. \tag{34}$$

Proof Let C_i, $i = 1, \ldots, l$ be all of the cells in the normal manifold of S, and for each N define the event

$$A_N = \left\{ \omega \,\middle|\, \{i \,|\, d_i(z_N(\omega)) \leq 1/g(N)\} = \{i \,|\, z_0 \in C_i\} \right\}. \tag{35}$$

By the remarks below (21), if $\omega \in A_N$ then the two points z_N^* and z_0 belong to the relative interior of the same cell in the normal manifold of S, with $\Gamma'(z_0) = \Gamma'(z_N^*(\omega))$ and $d(f_0)_S(z_0)$ and $\Phi_N(z_N(\omega))$ sharing the conical subdivision $\{K_1, \ldots, K_k\}$. Moreover as shown in [8, Theorem 3.1] $\lim_{N \to \infty} \Pr(A_N) = 1$, so it follows from (30)

$$\lim_{N \to \infty} \Pr\left(A_N; \sup_{h \in \mathbb{R}^n, h \neq 0} \frac{\|\Sigma_N^{-1/2} \Phi_N(z_N)(h) - \Sigma_0^{-1/2} d(f_0)_S(z_0)(h)\|}{\|h\|} < \varepsilon\right) = 1. \tag{36}$$

Combining this with Lemma 3 it follows that for all fixed x, $\eta_j^\alpha(\Sigma_N^{-1/2} \Phi_N(z_N), x)$ converges in probability to $\eta_j^\alpha(\Sigma_0^{-1/2} d(f_0)_S(z_0), x)$.

Next let B be a fixed neighborhood of z_0 such that $B \cap (z_0 + K_i) = B \cap P_i$ for $i = 1, \ldots, k$. We then have

$$\lim_{N \to \infty} \Pr\left(\sqrt{N} |(z_N - z_0)_j| \leq \eta_j^\alpha(\Sigma_N^{-1/2} \Phi_N(z_N), z_N - z_N^*)\right)$$

$$= \lim_{N \to \infty} \Pr\left(\sqrt{N} |(z_N - z_0)_j| \leq \eta_j^\alpha(\Sigma_N^{-1/2} \Phi_N(z_N), z_N - z_N^*); A_N\right)$$

$$= \lim_{N \to \infty} \sum_{i=1}^k \Pr\left(\sqrt{N} |(z_N - z_0)_j| \leq \eta_j^\alpha(\Sigma_N^{-1/2} \Phi_N(z_N), z_N - z_N^*); A_N; z_N \in B \cap \text{int} P_i\right)$$

$$= \lim_{N \to \infty} \sum_{i=1}^k \Pr\left(\sqrt{N} |(z_N - z_0)_j| \leq \eta_j^\alpha(\Sigma_N^{-1/2} \Phi_N(z_N), x_i); A_N; z_N \in B \cap \text{int} P_i\right)$$

where x_i is any point in $\text{int} K_i$. The first equality above follows from $\lim_{N \to \infty} \Pr(A_N) = 1$, and the second from $\lim_{N \to \infty} \Pr\left(z_N \in \mathbb{R}^n \setminus \cup_{i=1}^k B \cap \text{int} P_i\right) = 0$ as shown in [9, Proposition 3.5]. For the final equality, recall that $\omega \in A_N$ implies that z_N^* and z_0 belong to the relative interior of the same cell in the normal manifold. Since the latter cell

is a face of each P_i, $i = 1, \cdots, k$, by the additional requirement $z_N \in \mathrm{int}P_i$ one has $z_N - z_N^* \in \mathrm{cone}(\mathrm{int}P_i - z_N^*)$ and the latter set is exactly $\mathrm{cone}(\mathrm{int}P_i - z_0)$, namely $\mathrm{int}K_i$.

When $k = 1$, z_0 is contained in the interior of an n-cell P_1 and $K_1 = \mathbb{R}^n$. In this case $Y^* \sim \mathcal{N}\left(0, M_1^{-1}\Sigma_0 M_1^{-T}\right)$, and (34) follows from,

$$\frac{\sqrt{N}(z_N - z_0)_j}{\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), x_1)} \Rightarrow \frac{(Y^*)_j}{\eta_j^\alpha(\Sigma_0^{-1/2}d(f_0)_S(z_0), x_1)}.$$

Next we consider the case when $k \geq 2$. For all $j = 1, \ldots, n$ and $i = 1, \ldots, k$ let $\vec{v}^{i,j} \in \mathbb{R}^n$ be such that $\vec{v}^{i,j} \notin K_i$ and $|(\vec{v}^{i,j})_j| > \eta_j^\alpha(\Sigma_0^{-1/2}d(f_0)_S(z_0), x_i)$. Define random variables

$$v_N^{i,j} = \sqrt{N}(z_N - z_0)\mathbb{1}_{z_N \in B \cap \mathrm{int}P_i} + \vec{v}^{i,j}\mathbb{1}_{z_N \notin B \cap \mathrm{int}P_i},$$

$$\hat{Y}^{i,j} = Y^i\mathbb{1}_{Y^i \in \mathrm{int}K_i} + \vec{v}^{i,j}\mathbb{1}_{Y^i \notin \mathrm{int}K_i},$$

$$\hat{\eta}_N^{i,j} = \eta_j^\alpha\left(\Sigma_N^{-1/2}\Phi_N(z_N), x_i\right)\mathbb{1}_{z_N \in B \cap \mathrm{int}P_i} + \eta_j^\alpha\left(\Sigma_0^{-1/2}d(f_0)_S(z_0), x_i\right)\mathbb{1}_{z_N \notin B \cap \mathrm{int}P_i},$$

and note that

$$\hat{\eta}_N^{i,j} \Rightarrow \eta_j^\alpha\left(\Sigma_0^{-1/2}d(f_0)_S(z_0), x_i\right).$$

Next, for all Borel sets $W \subset \mathrm{int}K_i$,

$$\Pr\left(v_N^{i,j} \in W\right) = \Pr\left(\sqrt{N}(z_N - z_0) \in W,\, z_N \in B \cap \mathrm{int}P_i\right)$$
$$= \Pr\left(\sqrt{N}(z_N - z_0) \in W,\, z_N \in B\right),$$

and hence

$$\lim_{N \to \infty} \Pr\left(v_N^{i,j} \in W\right) = \lim_{N \to \infty} \Pr\left(\sqrt{N}(z_N - z_0) \in W, z_N \in B\right)$$
$$= \lim_{N \to \infty} \Pr\left(\sqrt{N}(z_N - z_0) \in W\right)$$
$$= \Pr\left(Y^* \in W\right) = \Pr\left(Y^i \in W\right) = \Pr\left(\hat{Y}^{i,j} \in W\right). \qquad (37)$$

Since $z_N \to z_0$ in probability and $\mathrm{int}K_i = \mathrm{cone}(\mathrm{int}P_i - z_0)$, it follows that as $N \to \infty$,

$$\Pr\left(\sqrt{N}(z_N - z_0) \in (\mathrm{int}K_i)^c,\, z_N \in B \cap \mathrm{int}P_i\right) \to 0,$$

and

$$\Pr\left(z_N \notin B \cap \mathrm{int}P_i\right) \to \Pr\left(Y^* \notin \mathrm{int}K_i\right) = \Pr\left(Y^i \notin \mathrm{int}K_i\right) = \Pr\left(\hat{Y}^{i,j} \notin \mathrm{int}K_i\right).$$

Thus for any Borel set D in $\mathbb{R}^n$,

$$\lim_{N \to \infty} \Pr\left(v_N^{i,j} \in D \cap (\mathrm{int}K_i)^c\right)$$
$$= \lim_{N \to \infty} \mathbb{1}_{D \cap (\mathrm{int}K_i)^c}(\vec{v}^{i,j})\Pr\left(z_N \notin B \cap \mathrm{int}P_i\right)$$
$$= \mathbb{1}_{D \cap (\mathrm{int}K_i)^c}(\vec{v}^{i,j})\Pr\left(\hat{Y}^{i,j} \notin \mathrm{int}K_i\right)$$
$$= \Pr\left(\hat{Y}^{i,j} \in D \cap (\mathrm{int}K_i)^c\right). \qquad (38)$$

Combining (37) with (38) and since $\hat{\eta}_N^{i,j}$ and $\eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right)$ are strictly positive under our assumptions we have that

$$\frac{v_N^{i,j}}{\hat{\eta}_N^{i,j}} \Rightarrow \frac{\hat{Y}^{i,j}}{\eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right)},$$

and thus

$$\lim_{N\to\infty} \ \mathrm{Pr}\left(\left| \frac{(v_N^{i,j})_j}{\hat{\eta}_N^{i,j}} \right| \leq 1 \right) = \mathrm{Pr}\left(\left| \frac{(\hat{Y}^{i,j})_j}{\eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right)} \right| \leq 1 \right)$$

$$= \mathrm{Pr}\left(\left| \frac{(Y^i)_j}{\eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right)} \right| \leq 1, \ Y^i \in \mathrm{int}K_i \right),$$

where we used the fact $|(\bar{v}^{i,j})_j| > \eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right)$. The latter fact also implies $\lim_{N\to\infty} \mathrm{Pr}\left(\left| \frac{(\bar{v}^{i,j})_j}{\hat{\eta}_N^{i,j}} \right| \leq 1 \right) = 0$, so it follows that

$$\lim_{N\to\infty} \mathrm{Pr}\left(\sqrt{N} \frac{|(z_N - z_0)_j|}{\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), x_i)} \leq 1; A_N; z_N \in B \cap \mathrm{int}P_i \right)$$

$$= \lim_{N\to\infty} \mathrm{Pr}\left(\sqrt{N} \frac{|(z_N - z_0)_j|}{\hat{\eta}_N^{i,j}} \leq 1, z_N \in B \cap \mathrm{int}P_i \right) = \lim_{N\to\infty} \mathrm{Pr}\left(\frac{|(v_N^{i,j})_j|}{\hat{\eta}_N^{i,j}} \leq 1 \right)$$

$$= \mathrm{Pr}\left(\left| \frac{(Y^i)_j}{\eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right)} \right| \leq 1, \ Y^i \in \mathrm{int}K_i \right)$$

$$= \mathrm{Pr}\left(|(M_i^{-1}\Sigma_0^{1/2}Z)_j| \leq \eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right), M_i^{-1}\Sigma_0^{1/2}Z \in K_i \right)$$

$$= \mathrm{Pr}\left(|(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z)_j| \leq \eta_j^\alpha \left(\Sigma_0^{-1/2} d(f_0)_S(z_0), x_i \right), d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i \right)$$

$$= (1 - \alpha) \mathrm{Pr}\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i \right).$$

Finally, since on A_N we have $z_N - z_N^* \in \mathrm{int}K_i$,

$$\lim_{N\to\infty} \mathrm{Pr}\left(\sqrt{N}|(z_N - z_0)_j| \leq \eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*) \right)$$

$$= \lim_{N\to\infty} \sum_{i=1}^k \mathrm{Pr}\left(\sqrt{N} \frac{|(z_N - z_0)_j|}{\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), x_i)} \leq 1; A_N; z_N \in B \cap \mathrm{int}P_i \right)$$

$$= \sum_{i=1}^k (1 - \alpha) \mathrm{Pr}\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i \right)$$

$$= (1 - \alpha) \sum_{i=1}^k \mathrm{Pr}\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i \right) = 1 - \alpha.$$

$\square$

Comparing the above two methods, computation of $\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)$ is more efficient than that of $a^r\left((\Phi_N^{-1}(z_N)\Sigma_N^{1/2})_j\right)$, as it with high probability restricts the computation to a single cone in the conical subdivision of $\Phi_N(z_N)$, namely the cone that contains $z_N - z_N^*$ in its interior (the same cone also contains $z_N - z_0$ in its interior whenever the event A_N in (35) holds).

3.3 The third method (a direct approach)

Comparing the asymptotic distributions for z_N and x_N, as given by (13) and (15) respectively, we see that the latter distribution has Π_{K_0} in it, the projector onto the critical cone to S at z_0. Since Π_{K_0} is generally non-invertible, neither of the methods presented in §3.1 and §3.2 can be used to directly construct intervals for $(x_0)_j$. Both methods require the invertibility of the function appearing in the asymptotic distribution either in the construction of an interval or the proof of the interval's exactness.

The non-invertibility of Π_{K_0} also leads us to change our focus from asymptotically exact intervals to intervals meeting a specified minimum level of confidence for the following reason. If the function $\Pi_{K_0} \circ (L_{K_0})^{-1}(\cdot)$ appearing in (15) has a selection function whose matrix representation contains a row of zeros (say the jth row), then there is a non-zero probability for $(x_N)_j$ to equal $(x_0)_j$. In this case any reasonable method for constructing individual confidence intervals of $(x_0)_j$ will have a lower bound on its performance: no matter how narrow the interval is, the probability for it to contain $(x_0)_j$ is no less than the probability for $(x_0)_j$ and $(x_N)_j$ to coincide.

The method to be presented below determines the interval width based on equation (41) in Theorem 6, in which $h_j^\alpha(\cdot, \cdot, \cdot)$ replaces $\eta_j^\alpha(\cdot, \cdot)$ in the previous method. Below we introduce the definition of $h_j^\alpha(f, g, x)$, where f and g are piecewise linear functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ that share a common conical subdivision, $\{K_1, \ldots, K_k\}$, with g invertible. For any choice of cone K_i, $i = 1, \ldots, k$, component $j = 1, \ldots, n$ and $\alpha \in (0,1)$ we first define $h_j^\alpha(f, g, x)$ for points $x \in \text{int}K_i$ to be

$$h_j^\alpha(f, g, x) = \inf\left\{ l \geq 0 \,\middle|\, \frac{\Pr\left(|\left(f(g^{-1}(Z))\right)_j| \leq l \text{ and } g^{-1}(Z) \in K_i\right)}{\Pr\left(g^{-1}(Z) \in K_i\right)} \geq (1-\alpha) \right\}.$$

Denoting the matrix representations of the selection functions on each cone as $f|_{K_i} = Q_i$ and $g|_{K_i} = M_i$, for all points $x \in \text{int}K_i$ the function $h_j^\alpha(f, g, x)$ will take the same value and the above definition is equivalent to

$$h_j^\alpha(f, g, x) = \inf\left\{ l \geq 0 \,\middle|\, \frac{\Pr\left(|(Q_i)_j M_i^{-1} Z| \leq l \text{ and } M_i^{-1} Z \in K_i\right)}{\Pr\left(M_i^{-1} Z \in K_i\right)} \geq (1-\alpha) \right\}. \quad (39)$$

For points $x \in \bigcap_{s=1}^v K_{i_s}$ define $h_j^\alpha(f, g, x) = \max_{s=1,\ldots,v} h_j^\alpha(f, g, x_{i_s})$ where $x_{i_s} \in \text{int}K_{i_s}$. As shown in the following lemma we can identify when $h_j^\alpha(f, g, x) = 0$ based on x and the matrix representations for the appropriate selection functions of f.

Lemma 4 *For any point $x \in \bigcap_{s=1}^{v} K_{i_s}$, $j = 1, \ldots, n$ and $\alpha \in (0,1)$, $h_j^{\alpha}(f,g,x) = 0$ if and only if $(Q_{i_s})_j$ is the zero vector for all $s = 1, \ldots, v$.*

Proof It suffices to prove the result for $x \in \mathrm{int} K_i$. If $h_j^{\alpha}(f,g,x) = 0$,

$$0 < (1-\alpha)\Pr\left(M_i^{-1}Z \in K_i\right) \leq \Pr\left(|(Q_i)_j M_i^{-1}Z| \leq 0 \text{ and } M_i^{-1}Z \in K_i\right),$$

and hence,

$$0 < \Pr\left((Q_i)_j M_i^{-1}Z = 0 \text{ and } M_i^{-1}Z \in K_i\right) \leq \Pr\left((Q_i)_j M_i^{-1}Z = 0\right). \qquad (40)$$

Since $(Q_i)_j M_i^{-1}Z \sim \mathcal{N}\left(0, \|(Q_i)_j M_i^{-1}\|^2\right)$, where $\|\cdot\|$ denotes the Euclidian norm, (40) implies that $\|(Q_i)_j M_i^{-1}\| = 0$, and thus $(Q_i)_j$ is a vector of zeroes. The reverse implication follows immediately.

$\square$

When using $h_j^{\alpha}(f,g,x)$ to construct confidence intervals for solutions to (1) we will be interested in

$$f = \Pi_{K_0} \text{ and } g = \Sigma_0^{-1/2} d(f_0)_S(z_0)$$

and their estimates

$$f_N = \Lambda_N(z_N) = d\Pi_S(z_N^*) \text{ and } g_N = \Sigma_N^{-1/2} \Phi_N(z_N).$$

From (11) and (22) it follows that the probability of all four functions sharing a common conical subdivision and f_N equalling f goes to one as the sample size goes to infinity. We therefore take this to be the setting for the following lemma.

Lemma 5 *Let $f,g : \mathbb{R}^n \to \mathbb{R}^n$ be piecewise linear functions with g a homeomorphism. Suppose that $\{f_N\}_{N=1}^{\infty}$ and $\{g_N\}_{N=1}^{\infty}$ are two sequences of piecewise linear functions such for all N sufficiently large*

1. *$f_N = f$.*
2. *f, g and g_N all share a common conical subdivision $\{K_1, \ldots, K_k\}$.*
3. *$\sup\limits_{h \in \mathbb{R}^n, h \neq 0} \frac{\|g_N(h) - g(h)\|}{\|h\|} \to 0$.*

Then for all N sufficiently large g_N will be a homeomorphism and $h_j^{\alpha}(f_N, g_N, x) \to h_j^{\alpha}(f,g,x)$ for all $x \in \mathbb{R}^n$, $\alpha \in (0,1)$ and $j = 1, \ldots, n$.

Proof From Proposition 1 it follows that for all N sufficiently large g_N is a homeomorphism and that g_N^{-1} converges uniformly on compacts to g^{-1}. Next take v to be large enough so that for all $N \geq v$ the functions g_N are invertible, $f_N = f$ and f, g and g_N all share common conical subdivision $\{K_1, \ldots, K_k\}$. To prove the remainder of the Lemma's claim it suffices to show that for any $x \in \mathrm{int} K_i$, $i = 1, \ldots, k$, $h_j^{\alpha}(f, g_N, x) \to h_j^{\alpha}(f,g,x)$.

When $x \in \mathrm{int} K_i$ and $h_j^{\alpha}(f,g,x) = 0$, it follows from Lemma 4 that $h_j^{\alpha}(f, g_N, x) = 0$. In the case of $x \in \mathrm{int} K_i$ and $h_j^{\alpha}(f,g,x) > 0$, the convergence can be shown using an argument analogous to the proof of Lemma 1 and Lemma 3 and is therefore omitted.

□

The main result of this section, Theorem 6, can now be proven. We will use the same notation used in Theorem 5 where $\Gamma'(z_0) = \{K_1,\ldots,K_k\}$ is the conical subdivision associated with $d(f_0)_S(z_0)$ such that $d(f_0)_S(z_0)|_{K_i} = M_i$ and $K_i = \text{cone}(P_i - z_0)$, where $P_1,\ldots,P_k$ are all n-cells in the normal manifold of S that contain z_0. We additionally denote $\Pi_{K_0}|_{K_i} = Q_i$ and define the following random variables:

$$Y^i = M_i^{-1}\Sigma_0^{1/2}Z, \ Y_0 = \Sigma_0^{1/2}Z \text{ and } Y^* = d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z.$$

Theorem 6 *Let Assumptions 2, 3 and 4 hold. Let $\Phi_N(z_N)(\cdot)$ and z_N^* be as defined in (21). For all $j = 1,\ldots,n$ and $\alpha \in (0,1)$,*

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(x_N - x_0)_j| \leq h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right) \geq 1 - \alpha. \quad (41)$$

Proof As in the proof of Theorem 5 we begin by letting C_i, $i = 1,\ldots,l$ denote the cells in the normal manifold of S and for each N let the event A_N be as defined in (35). Now for $\omega \in A_N$ the equality $\Pi_{K_0} = d\Pi_S(z_N^*)$ holds, and $\{K_1,\ldots,K_k\}$ provides a common conical subdivision for Π_{K_0}, $d(f_0)_S(z_0)$ and $\Phi_N(z_N(\omega))$. From (36) and Lemma 5 it follows that for all fixed u, $h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u)$ converges in probability to $h_j^\alpha(\Pi_{K_0}, \Sigma_0^{-1/2}d(f_0)_S(z_0), u)$.

Next let B be a fixed neighborhood of z_0 such that $B \cap (z_0 + K_i) = B \cap P_i$ for $i = 1,\ldots,k$. We then have

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(x_N - x_0)_j| \leq h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right)$$

$$= \lim_{N\to\infty} \Pr\left(\sqrt{N}|\left(\Pi_S(z_N) - \Pi_S(z_0)\right)_j| \leq h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*); A_N\right)$$

$$= \lim_{N\to\infty} \sum_{i=1}^k \Pr\left(\sqrt{N}|\left(\Pi_{K_0}(z_N - z_0)\right)_j| \leq h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*); A_N; z_N \in B \cap \text{int}P_i\right)$$

$$= \sum_{i=1}^k \lim_{N\to\infty} \Pr\left(\sqrt{N}|(Q_i)_j(z_N - z_0)| \leq h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i); A_N; z_N \in B \cap \text{int}P_i\right) \quad (42)$$

where u_i is any point in $\text{int}K_i$. The first equality uses the relation between solutions to a variational inequality and its normal map formulation, while the second equality combines the almost sure convergence of z_N to z_0 with (10). The final equality uses the fact that for $\omega \in A_N$ and $z_N \in \text{int}P_i$ both $z_N - z_0$ and $z_N - z_N^*$ will be contained in $\text{int}K_i$ and thus $z_N - z_N^*$ may be replaced with u_i and $\Pi_{K_0}(z_N - z_0) = Q_i(z_N - z_0)$.

Evaluating each term in (42) depends on $(Q_i)_j$. If $(Q_i)_j$ is the zero vector for some i, then

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(Q_i)_j(z_N - z_0)| \leq h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i); A_N; z_N \in B \cap \text{int}P_i\right)$$

$$= \lim_{N\to\infty} \Pr\left(\sqrt{N}(z_N - z_0) \in \text{int}K_i\right) = \Pr(Y^* \in \text{int}K_i)$$

$$= \Pr\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i\right). \quad (43)$$

On the other hand, if $(Q_i)_j$ is a nonzero vector (i.e., it contains at least one nonzero element) for some i, we define a vector $\vec{v}^{i,j}$ to be such that $\vec{v}^{i,j} \notin K_i$ and $|(Q_i)_j \vec{v}^{i,j}| > h_j^\alpha(\Pi_{K_0}, \Sigma_0^{-1/2} d(f_0)_S(z_0), u_i)$. With these we define random vectors

$$
\begin{aligned}
v_N^{i,j} &= \sqrt{N}(z_N - z_0)\mathbb{1}_{z_N \in B \cap \mathrm{int}P_i} + \vec{v}^{i,j}\mathbb{1}_{z_N \notin B \cap \mathrm{int}P_i}, \\
\hat{Y}^{i,j} &= Y^i \mathbb{1}_{Y^i \in \mathrm{int}K_i} + \vec{v}^{i,j}\mathbb{1}_{Y^i \notin \mathrm{int}K_i}, \\
\hat{h}_N^{i,j} &= h_j^\alpha\left(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i\right)\mathbb{1}_{z_N \in B \cap \mathrm{int}P_i} + h_j^\alpha\left(\Pi_{K_0}, \Sigma_0^{-1/2}d(f_0)_S(z_0), u_i\right)\mathbb{1}_{z_N \notin B \cap \mathrm{int}P_i}.
\end{aligned}
$$

Using the same arguments as in Theorem 5 it follows that

$$
\frac{v_N^{i,j}}{\hat{h}_N^{i,j}} \Rightarrow \frac{\hat{Y}^{i,j}}{h_j^\alpha\left(\Pi_{K_0}, \Sigma_0^{-1/2}d(f_0)_S(z_0), u_i\right)}
$$

and

$$
\lim_{N \to \infty} \mathrm{Pr}\left(\sqrt{N}\frac{|(Q_i)_j(z_N - z_0)|}{h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i)} \le 1; A_N; z_N \in B \cap \mathrm{int}P_i\right)
$$
$$
= (1-\alpha)\mathrm{Pr}\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i\right). \tag{44}
$$

Combining (43) and (44), with the fact that $z_N - z_N^* \in \mathrm{int}K_i$ on A_N, we have

$$
\lim_{N \to \infty} \mathrm{Pr}\left(\sqrt{N}|(x_N - x_0)_j| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right)
$$
$$
= \lim_{N \to \infty}\sum_{i=1}^k \mathrm{Pr}\left(\sqrt{N}|(Q_i)_j(z_N - z_0)| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i); A_N; z_N \in B \cap \mathrm{int}P_i\right)
$$
$$
\ge (1-\alpha)\sum_{i=1}^k \mathrm{Pr}\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i\right) = 1 - \alpha.
$$

$\square$

An important fact seen in the proof of Theorem 6 is that

$$
\lim_{N \to \infty} \mathrm{Pr}\left(\sqrt{N}|(x_N - x_0)_j| \le h_j^\alpha(\Pi_{K_N}, \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right) > 1 - \alpha
$$

if and only if there exists a cone K_i in the conical subdivision of Π_{K_0} such that the jth component of $\Pi_{K_0}|_{K_i}$ is zero. When this is the case, we have

$$
(x_N(\omega) - x_0)_j = \left(\Pi_S(z_N(\omega)) - \Pi_S(x_0)\right)_j = (Q_i)_j(z_N(\omega) - z_0) = 0,
$$

as long as $z_N(\omega) - z_0 \in \mathrm{int}K_i$ and z_N is sufficiently close to z_0. If additionally $\omega \in A_N$, then $d\Pi_S(z_N^*) = \Pi_{K_0}$ and we have by Lemma 4

$$
h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*) = 0,
$$

meaning that the method of Theorem 6 returns the correct point estimate $(x_N(\omega))_j = (x_0)_j$. Recalling that $\lim_{N \to \infty}\mathrm{Pr}(A_N) = 1$ and that z_N converges to z_0 almost surely,

we see that the potentially conservative asymptotic level of confidence is not the result of using unnecessarily long intervals, but instead follows from the fact that for sufficiently large sample sizes the proposed method will return the correct point estimate with a nonzero probability.

While (31), (34) and (41) provide computable intervals with the desired asymptotic properties, in general $a^r(\cdot)$, $\eta_j^\alpha(\cdot,\cdot)$ and $h_j^\alpha(\cdot,\cdot,\cdot)$ lack closed form expressions. In the next section we consider the computation of these quantities. For ease of exposition, moving forward we will suppress the arguments of a^r, η_j^α and h_j^α.

4 Interval Computation

This section considers the computation of a^r, η_j^α and h_j^α, and discusses how to find upper bounds for these quantities. Before presenting a general method for computing a^r, η_j^α and h_j^α we consider special cases when either closed form expressions exist or less burdensome techniques can be used. For each of these discussions we begin by considering a^r with the results for η_j^α and h_j^α following in a similar fashion.

The first case we consider is when $\Sigma_N^{-1/2}\Phi_N(z_N)$ and $d\Pi_S(z_N^*)$ are linear functions with matrix representations M_N and Q_N respectively. Since $(M_N^{-1}Z)_j$ is a mean zero Normal random variable for each coordinate j, it is natural to set $r = 0$ for a^r. Then from basic properties of Normal random vectors,

$$a^0 = \eta_j^\alpha = \sqrt{\chi_1^2(\alpha)\|(M_N^{-1})_j\|^2} \quad \text{and} \quad h_j^\alpha = \sqrt{\chi_1^2(\alpha)\|(Q_N)_j M_N^{-1}\|^2}$$

where $\|\cdot\|$ is the Euclidian norm. Note in this case both intervals for $(z_0)_j$ are the same as the interval considered in Theorem 3.

In the piecewise linear case let $\phi_{N,j}$ denote the jth component function of $\Phi_N^{-1}(z_N)\Sigma_N^{1/2}$. Finding a^r requires a search over values of $l > 0$ and evaluating $\Pr(|\phi_{N,j}(Z) - r| \leq l)$. To evaluate this probability we rewrite it in terms of the selection functions of $\phi_{N,j}$.

To this end, let $\Gamma = \{K_1, \ldots, K_k\}$ be the common conical subdivision for $\Sigma_N^{-1/2}\Phi_N(z_N)$ and $d\Pi_S(z_N^*)$, and let $\{M_{N,1}, \ldots, M_{N,k}\}$ and $\{Q_{N,1}, \ldots, Q_{N,k}\}$ be the matrix representations for the respective selection functions. Then with

$$T_i = \Sigma_N^{-1/2}\Phi_N(z_N)(K_i) = M_{N,i}(K_i),$$

$\{T_1, \ldots, T_k\}$ provides a conical subdivision for $\phi_{N,j}$ such that $\phi_{N,j}|_{T_i} = (M_{N,i}^{-1})_j$. Due to the high probability of $\Sigma_N^{-1/2}\Phi_N(z_N)$ and $d(f_0)_S(z_0)$ sharing a common conical subdivision we have used the same notation K_i, $i = 1, \ldots, k$, as was introduced before Theorem 3. In the discussion that follows it is not necessary for the functions to share a common conical subdivision. Additionally, any assumptions made about the value of k will refer to the number of selection functions for a particular realization of $\Sigma_N^{-1/2}\Phi_N(z_N)$ unless otherwise stated.

For any two cones $T_v, T_u \in \Gamma'$ with $v \neq u$, their intersection is either empty or a proper face of both cones, and hence $\Pr(Z \in T_v \cap T_u) = 0$. The probability we need to

evaluate can thus be rewritten as

$$\Pr\left(|\phi_{N,j}(Z) - r| \leq l\right) = \sum_{i=1}^{k} \Pr\left(|\phi_{N,j}(Z) - r| \leq l \text{ and } Z \in T_i\right)$$

$$= \sum_{i=1}^{k} \Pr\left(|(M_{N,i}^{-1})_j Z - r| \leq l \text{ and } Z \in T_i\right). \qquad (45)$$

Note the connection between (45) and what must be considered to find η_j^α. Finding η_j^α requires us to evaluate

$$\Pr\left(|(M_{N,i}^{-1})_j Z| \leq l \text{ and } M_{N,i}^{-1} Z \in K_i\right) = \Pr\left(|(M_{N,i}^{-1})_j Z| \leq l \text{ and } Z \in T_i\right), \qquad (46)$$

for different values of l, but only for those indices i such that $z_N - z_N^* \in K_i$. At this point we see the computational benefits of η_j^α over a^r. Recall from the proof of Theorem 5 that

$$\lim_{N \to \infty} \sum_{i=1}^{k} \Pr\left(A_N \text{ and } z_N \in B \cap \text{int} P_i\right) = 1,$$

where k is the number of selection functions for $d(f_0)_S(z_0)$, A_N is as defined in (35) and $K_i = \text{cone}(P_i - z_0)$. Moreover when A_N holds and $z_N \in B \cap \text{int} P_i$ it was argued that $z_N - z_N^* \in \text{int} K_i$. Therefore with high probability each value of l we consider when finding η_j^α will involve evaluating (46) for a single index i. In contrast, (45) involves a similar calculation for every cone in the subdivision. Finding h_j^α will with high probability also require considering only a single index i, but with the quantity evaluated being $\Pr\left(|(Q_{N,i})_j M_{N,i}^{-1} Z| \leq l \text{ and } Z \in T_i\right)$.

The question of finding a^r, η_j^α and h_j^α in the piecewise linear case now becomes how to evaluate

$$\Pr\left(|b_{N,i}^T Z - r| \leq l \text{ and } Z \in T_i\right), \qquad (47)$$

where $b_{N,i}^T = (M_{N,i}^{-1})_j$ when finding a^r and η_j^α, and $b_{N,i}^T = (Q_{N,i})_j M_{N,i}^{-1}$ when finding h_j^α. When $k = 2$ and $r = 0$ evaluating (47) is simplified by observing that the two cones in Γ' satisfy $T_1 = -T_2$ and the fact that Z and $-Z$ have the same distribution. It then follows that

$$\Pr\left(|b_{N,i}^T Z| \leq l \text{ and } Z \in T_i\right) = 1/2 \Pr\left(|b_{N,i}^T Z| \leq l\right)$$
$$= \Pr\left(Z \in T_i\right) \Pr\left(|b_{N,i}^T Z| \leq l\right).$$

In this case no search is necessary for η_j^α and h_j^α. Finding a^0 may still require a search over different values of l but this search can refer to the cumulative distribution function of a standard Normal random variable to evaluate the necessary probabilities.

When $k > 2$ our approach to evaluating (47) is to rewrite it as the probability of a Normal random vector being in a box with possibly infinite endpoints. Once formulated in this manner the probability can be evaluated either using the numerical techniques of [12] or the Monte Carlo and Quasi-Monte Carlo methods of [4, Chapter 4], both of which are implemented in R package mvtnorm [4,5]. Comparisons of the methods for different problem sizes can be found in [11]. The method in [12] requires

the Normal random vector of interest to have a non-singular covariance matrix, so we first consider a class of SVIs for which this condition holds when finding a^r and η_j^α.

When the SVI is a complementarity problem with $S = \mathbb{R}^m \times \mathbb{R}_+^{n-m}$, where $\mathbb{R}_+^k$ denotes the positive orthant, each of the polyhedral cones $K_i \in \Gamma$ can be expressed as an n-dimensional box,

$$K_i = [l_1^i, u_1^i] \times \ldots \times [l_n^i, u_n^i]$$

with 0, ∞ or $-\infty$ as endpoints. Additionally by $\Phi_N^{-1}(z_N)\Sigma_N^{1/2}$ a homeomorphism it follows that for each $i = 1, \ldots, k$ and $x \in \mathbb{R}^n$

$$x \in T_i \Leftrightarrow \Phi_N^{-1}(z_N)\Sigma_N^{1/2}(x) \in K_i \Leftrightarrow M_{N,i}^{-1}x \in K_i.$$

Therefore we can write

$$\Pr\left(|(M_{N,i}^{-1})_j Z - r| \le l \text{ and } Z \in T_i\right) = \Pr\left(r - l \le (M_{N,i}^{-1})_j Z \le r + l \text{ and } M_{N,i}^{-1}Z \in K_i\right)$$

$$= \Pr\left(M_{N,i}^{-1}Z \in [l_1^i, u_1^i] \times \ldots \times [\max(l_j^i, r-l), \min(u_j^i, r+l)] \times \ldots \times [l_n^i, u_n^i]\right)$$

$$= \Pr\left(\tilde{Z} \in [l_1^i, u_1^i] \times \ldots \times [\max(l_j^i, r-l), \min(u_j^i, r+l)] \times \ldots \times [l_n^i, u_n^i]\right)$$

where $\tilde{Z} \sim \mathcal{N}\left(0, M_{N,i}^{-1}M_{N,i}^{-T}\right)$. It follows that $\tilde{Z}$ has a non-singular covariance matrix, and either method of evaluating the probability can be used. Note that this approach cannot be used to find h_j^α due to the additional consideration of $(Q_{N,i})_j$.

In general to compute a^r, η_j^α and h_j^α we can use the structure of T_i being a polyhedral cone. In this case we express the cone as a system of linear inequalities,

$$T_i = \{x \in \mathbb{R}^n | C_i x \le 0_v\}$$

for C_i some $v \times n$ matrix and 0_v the v-dimensional zero vector. We then rewrite

$$\Pr\left(|b_{N,i}^T Z - r| \le l \text{ and } C_i Z \le 0_v\right) = \Pr\left(\bar{Z} \in (-\infty, 0] \times \ldots \times (-\infty, 0] \times [r-l, r+l]\right)$$

$$\text{where } \bar{Z} \sim \mathcal{N}\left(0_{v+1}, D_i D_i^T\right) \text{ and } D_i = \begin{bmatrix} C_i \\ b_{N,i}^T \end{bmatrix}.$$

When the covariance matrix of $\bar{Z}$ is singular only the methods of [4, Chapter 4] may be employed.

The potential of having to search over values of l when finding a^r leads us to consider the question finding an upper bound for a^r. Since for linear functions a^r is easily found one might hope that for ψ piecewise linear with family of selection functions expressed as n dimensional row vectors $\{b^1, \ldots, b^k\}$, with $\|b^1\| \le \|b^2\| \le \ldots \le \|b^k\|$, that $a^r(\psi) \le a^r(b^k)$. This need not be true.

For example take

$$b^1 = \begin{bmatrix} 1/5 & 7/5 \end{bmatrix}, b^2 = \begin{bmatrix} 7/5 & 1/5 \end{bmatrix}, b^3 = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

and $\gamma_i = \{x \in \mathbb{R}^2 | C_i x \le 0\}$ for $i = 1, \ldots, 5$, where

$$C_1 = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} -1 & 1 \\ -1 & 2 \end{bmatrix}, C_3 = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix},$$

$$C_4 = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \text{and } C_5 = \begin{bmatrix} -1 & 1 \end{bmatrix}.$$

Note both $\{\gamma_1, \gamma_2, \gamma_3\}$ and $\{\gamma_4, \gamma_5\}$ are conical subdivisions of $\mathbb{R}^2$. Define ψ_1 and ψ_2 to be piecewise linear functions such that $\psi_1|_{\gamma_i} = b^i$ for $i = 1, 2, 3$, $\psi_2|_{\gamma_4} = b^1$ and $\psi_2|_{\gamma_5} = b^2$. It follows that $a^0(b^i) = a^0(\psi_2) = \sqrt{2\chi_1^2(\alpha)}$, $i = 1, 2, 3$. Next let

$$
\begin{aligned}
R_1 &= \left\{ z \in \mathbb{R}^2 \mid -a^0(\psi_2) \leq \psi_1(z) \leq a^0(\psi_2) \right\}, \\
R_2 &= \left\{ z \in \mathbb{R}^2 \mid -a^0(\psi_2) \leq \psi_2(z) \leq a^0(\psi_2) \right\}.
\end{aligned}
$$

As shown in Figure 1, the set R_2 includes R_1 as a subset with $D = R_2 \setminus R_1$ having a non-empty interior. Thus $\Pr(Z \in R_1) < \Pr(Z \in R_2)$ and $a^0(\psi_2) < a^0(\psi_1)$, showing that $\max a^0(b^i)$ is not an upper bound for $a^0(\psi_1)$.
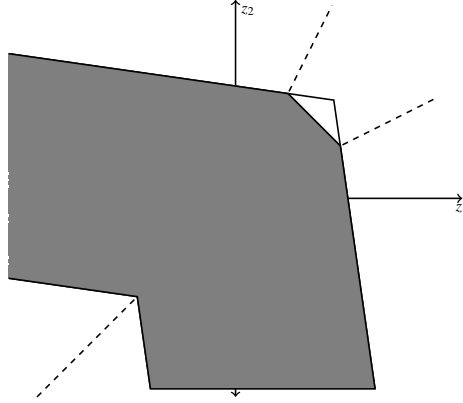

Fig. 1: Sets R_1 (shaded) and R_2 for $\alpha = .05$

To construct a valid upper bound for an interval's half width we will use the following Lemma.

Lemma 6 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a piecewise linear function with family of selection functions given by n dimensional row vectors $\{b_1, \ldots, b_m\}$ and corresponding conical subdivision $\Gamma = \{\gamma_1, \ldots, \gamma_m\}$. Let $Z \sim \mathcal{N}(0, I_n)$, $c_j = \Pr(Z \in \gamma_j)$, and $u > 0$ be such that*

$$
\Pr(|b_j Z - r| \leq u) \geq 1 - c_j \alpha
$$

for $j = 1, \ldots, m$, $\alpha \in (0, 1)$. Then $\Pr(-u \leq f(Z) - r \leq u) \geq 1 - \alpha$.

Proof Let E_j be the event that $\{|b_j Z - r| \leq u \text{ and } Z \in \gamma_j\}$. As argued previously $\Pr(|f(Z) - r| \leq u) = \sum_{j=1}^{m} \Pr(E_j)$. Next note

$$
\begin{aligned}
\Pr(E_j^c) &\leq \Pr(Z \in \gamma_j^c) + \Pr(|b_j Z - r| > u) \\
&\leq 1 - c_j + c_j \alpha = 1 - (1 - \alpha) c_j.
\end{aligned}
$$

Thus $\Pr(E_j) \geq (1-\alpha)c_j$ and

$$\Pr(|f(Z) - r| \leq u) = \sum_{j=1}^{m} \Pr(E_j) \geq (1-\alpha) \sum_{j=1}^{m} c_j = 1 - \alpha.$$

$\square$

Corollary 1 *Let $\alpha_j = \alpha \Pr(Z \in \gamma_j)$, then $u_j = \|b_j\| \sqrt{\chi_1^2(\alpha_j)}$ will satisfy $\Pr(|b_j Z| \leq u_j)$*
$= 1 - \alpha_j$, *and* $u = \max\limits_{1 \leq j \leq m} u_j$ *satisfies* $\Pr(|f(Z)| \leq u) \geq 1 - \alpha.$

Note that while Corollary 1 provides an upper bound for $a^0(f)$, Lemma 6 can similarly be used to find upper bounds for a^r when $r \neq 0$. Additionally upper bounds for η_j^α and h_j^α can be found as in Corollary 1 by considering only the subset of cones γ_i indicated by $z_N - z_N^*$.

5 Numerical Examples

This section applies the proposed methods and the method of Theorem 3 to two numerical examples. The half-width of intervals produced using the method of Theorem 3 will be denoted by v_j^α. When calculating a^r, η_j^α or h_j^α for a function with three or more selection functions, the approach used throughout the examples is to perform a binary search with probabilities calculated as in §4 using the methods of [4, Chapter 4]. This search terminates when either the distance between the upper and lower bounds or the probability of the value being tested are within specified tolerance levels.

In each example we are able to find the true solution allowing us to examine the coverage rates for the different methods. For each example we generate 2,000 SAA problems at each sample size of N=50, 100, 200 and 2,000. For each sample the value of r used for a^r is chosen by generating i.i.d. $Z_v \sim \mathcal{N}(0, I_n)$, calculating

$$r_N = 10^{-3} \sum_{v=1}^{10^3} \Phi_N^{-1}(z_N) \Sigma_N^{1/2}(Z_v),$$

and taking the appropriate coordinate of this vector. The use of this procedure will be indicated with the notation a^{r_N}.

Example 1

For the first example, we consider a non-complementarity problem with

$$S = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{bmatrix} .5 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \text{ and } F(x, \xi) = \begin{bmatrix} 4 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},$$

where ξ is uniformly distributed over the box $[-1, 1] \times [-2, 2]$. In this case

$$f_0(x) = \begin{bmatrix} 4 & 0 \\ 3 & 2 \end{bmatrix},$$

and the SVI and its corresponding normal map formulation have true solutions $x_0 = z_0 = 0$. The function $d(f_0)_S(z_0)$ is then piecewise linear, with the family of selection functions given by matrices

$$\begin{bmatrix} 4 & 0 \\ 3 & 2 \end{bmatrix}, \begin{bmatrix} 1.6 & 1.2 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 3.4 & 1.2 \\ 2.8 & 2.4 \end{bmatrix}$$

and the corresponding conical subdivision $\{K_1, K_2, K_3 \text{ and } K_4\}$ given by $K_i = \{x \in \mathbb{R}^2 | C_i x \leq 0\}$ with

$$C_1 = \begin{bmatrix} .5 & -1 \\ -2 & 1 \end{bmatrix} C_2 = \begin{bmatrix} 2 & -1 \\ -.5 & -1 \end{bmatrix} C_3 = \begin{bmatrix} .5 & 1 \\ -2 & 1 \end{bmatrix} \text{ and } C_4 = \begin{bmatrix} -2 & -1 \\ -.5 & 1 \end{bmatrix}.$$

With this information we evaluate (25) for $\alpha = .05$ and observe values of .9454 and .9461 for $j = 1$ and 2 respectively.

In Tables 1 and 2 we summarize the coverage rates of $(z_0)_1$ and $(z_0)_2$ for each interval determined by v_j^α, a^{r_N} and η_j^α. We see that the overall performance of the

Table 1: Coverage rates $(z_0)_1$ $\alpha = .05$

	v_1^α	a^{r_N}	η_1^α
N=50	94.25%	94.75%	94.2%
N=100	94.55%	94.95 %	94.9%
N=200	94.1 %	94.55 %	94.85 %
N=2,000	94.7%	95.35%	95.45%

Table 2: Coverage rates $(z_0)_2$ $\alpha = .05$

	v_2^α	a^{r_N}	η_2^α
N=50	93.8%	95.95%	93.65 %
N=100	94.15%	95.5%	93.65%
N=200	94.2%	95.25 %	94.95%
N=2,000	94.9%	95.45 %	95.4%

three approaches is generally comparable and in line with the specified 95% level of confidence and (25).

Differences between the methods become apparent in Figure 2 where for the samples of size 2,000 we plot the length of intervals for $(z_0)_2$ by which K_i contains $z_N - z_0$. These differences are further illustrated in Table 3 where we break down the coverage of $(z_0)_2$ and average interval length by which K_i contains $z_N - z_0$.
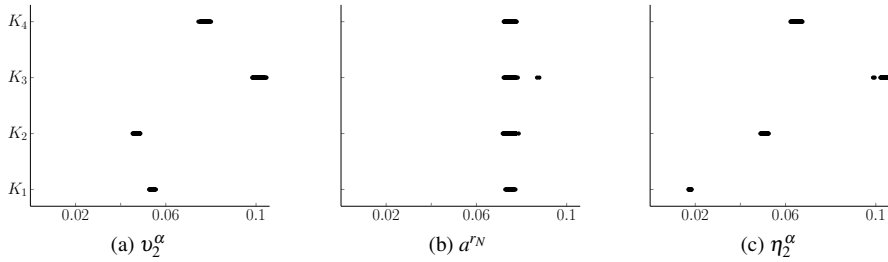

Fig. 2: Interval length for $(z_0)_2$ by cone, $N = 2,000$

Table 3: Coverage of $(z_0)_2$ and half-width by cone, $N = 2{,}000$, $\alpha = .05$

Cone (samples in cone)	Coverage rate			Average length		
	υ_2^α	a^{r_N}	η_2^α	υ_2^α	a^{r_N}	η_2^α
$K_1(80)$	100%	100%	90%	.0541	.0750	.0177
K_2 (689)	92.31%	98.84%	95.21%	.0471	.0749	.0508
K_3 (824)	95.39%	90.29%	96.24 %	.1012	.0749	.1051
K_4 (407)	97.3%	99.26%	95.09%	.0775	.0749	.0649

The consistent value of a^{r_N} across samples is to be expected given Lemma 1 and Proposition 1. Note that values of a^{r_N} that deviate slightly from this pattern correspond to the two samples for which z_N^* was not contained in the relative interior of the same k-cell as z_0. Across cones the performance of the intervals varies, but this is accounted for in the definition of a^{r_N}. Compare this with the intervals with half-width υ_2^α. This approach does not directly account for the effect $d(f_0)_S(z_0)$ being piecewise linear has on the asymptotic distribution of SAA solutions, and therefore the performance of the intervals. While in this example we can calculate (25) and observe that the intervals have an asymptotic level of confidence close to the desired 95%, in general the varying performance across cones is not accounted for and the method may be unreliable. The value of η_2^α also varies across cones, but its use of $z_N - z_N^*$ and $\Phi_N(z_N)$ to estimate the location of $z_N - z_0$ in the conical subdivision of $d(f_0)_S(z_0)$ allows for a level of confidence to be specified with less restrictive conditions. Additionally, the benefit of allowing η_2^α to vary in a systematic way is seen in the more consistent performance of this approach across the four cones.

Table 4: Coverage of $(x_0)_i$, $\alpha = .05$

	$(x_0)_1$	$(x_0)_2$
N=50	96.05%	96.2%
N=100	97%	97.25%
N=200	97.1%	97.15%
N=2,000	97%	95.33%

We next examine the performance of confidence intervals for $(x_0)_j$. For any real numbers $l \le u$ neither $\Pi_S(\mathbb{R} \times [l, u])$ nor $\Pi_S([l, u] \times \mathbb{R})$ result in sets that yield meaningful confidence intervals for $(x_0)_1$ or $(x_0)_2$. Therefore the indirect approach of projecting confidence intervals for $(z_0)_j$ onto S cannot be used and only the direct approach proposed in §3.3 is applicable. Combining (41) and the fact $S \subset \mathbb{R}_+^2$ we consider $\left[\max\{0, (x_N)_j - N^{-1/2} h_j^\alpha\}, (x_N)_j + N^{-1/2} h_j^\alpha \right]$ as the confidence interval for $(x_0)_j$. In Table 4 we summarize the coverage of $(x_0)_1$ and $(x_0)_2$ at each sample size with $\alpha = .05$, and in Table 5 we examine the performance and length of the intervals for the samples of size 2,000 broken down by the location of $z_N - z_0$. Since the selection function corresponding to $\Pi_{K_0}|_{K_3}$ is represented by the zero matrix when $z_N - z_0 \in K_3$ the correct point $(x_N)_j = (x_0)_j = 0$ is returned and as a result we see that the intervals for each component of x_0 outperform the specified confidence level of 95%.

Table 5: Intervals for $(x_0)_i$ by cone, $N = 2{,}000$, $\alpha = .05$

Cone (samples in cone)	Coverage rate		Average length	
	$(x_0)_1$	$(x_0)_2$	$(x_0)_1$	$(x_0)_2$
$K_1(80)$	88.75%	90%	.0104	.0132
K_2 (689)	95.36%	95.36%	.0089	.0177
K_3 (824)	100%	100%	0	0
K_4 (407)	95.33%	95.33%	.0073	.0036

Example 2

For the second example we let $S = \mathbb{R}_+^5$,

$$F(x,\xi) = \begin{bmatrix} \xi_1 & 1.5 & .5 & .75 & .9 \\ 1.5 & \xi_2 & 0 & .8 & 1.5 \\ .5 & 0 & \xi_3 & .75 & 1.7 \\ .75 & .8 & .75 & \xi_4 & 1 \\ .9 & 1.5 & 1.7 & 1 & \xi_5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} + \begin{bmatrix} \xi_6 \\ \xi_7 \\ \xi_8 \\ \xi_9 \\ \xi_{10} \end{bmatrix},$$

with ξ uniformly distributed over the box

$$[2,4] \times [0,4] \times [0,3] \times [2,6] \times [-1,6] \times [-1,1] \times [-.5,.5] \times [-2,2] \times [-.75,.75] \times [-1,1].$$

The SVI and its normal map formulation have solutions $x_0 = z_0 = 0$. Moreover $\Pi_{\mathbb{R}_+^5} = d\Pi_{\mathbb{R}_+^5}(z_0)$ with

$$d\Pi_{\mathbb{R}_+^5}(z_0)(x) = \begin{bmatrix} h_1 & 0 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 & 0 \\ 0 & 0 & h_3 & 0 & 0 \\ 0 & 0 & 0 & h_4 & 0 \\ 0 & 0 & 0 & 0 & h_5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \quad \text{where } h_i = \begin{cases} 0 \text{ if } x_i \leq 0, \\ 1 \text{ if } x_i \geq 0, \end{cases}$$

so $d(f_0)_{\mathbb{R}_+^5}(z_0)(\cdot)$ is piecewise linear with a family of thirty-two selection functions. Taking $\alpha = .05$ we first consider confidence intervals for $(z_0)_j$. Evaluating (25) for each value of $j = 1,\ldots,5$ we observe that the intervals for $(z_0)_j$ considered in Theorem 3 have asymptotic levels of confidence of 93.85%, 93.33%, 94.38%, 93.39% and 92.96% respectively.

Table 6: Coverage rates for $(z_0)_3$

	υ_4^α	a^{r_N}	η_4^α
$N = 50$	93.05%	96.3 %	93.3%
$N = 100$	92.85%	99.95 %	92.8%
$N = 200$	94%	94.7 %	94.95%
$N = 2{,}000$	94.35%	94.6 %	94.8%

Coverage rates of the confidence intervals are largely in line with the specified level of confidence or as indicated by (25), with the coverage rates of $(z_0)_3$ summarized in Table 6 for each approach and sample size considered. The performance of

the different methods broken down by where $z_N - z_0$ falls in the conical subdivision associated with $d(f_0)_S(z_0)$ cannot be as well observed, given the large number of cones relative to the number of samples. What we are able to observe is the consistent values of a^{r_N} across samples as compared to the values of v_j^α and η_j^α, shown in Figure 3 for $(z_0)_3$ and $N = 2,000$. Note there are two samples for which the value
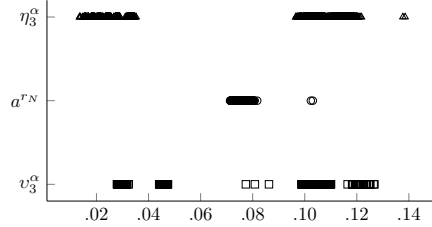

Fig. 3: Intervals Lengths for $(z_0)_3, N = 2,000$

of a^{r_N} deviate from this pattern, and as in the previous example they correspond to samples for which z_N^* and z_0 are not contained in the relative interior of the same k-cell. In this example the computational benefits of η_j^α are clear. For almost all of the samples calculating a^{r_N} required working with a piecewise linear function with thirty two selection functions, whereas for all of the samples calculating η_j^α only involved a single selection function, leading to a dramatic reduction in the necessary computation.

With this example we also examine how upper bounds satisfying the conditions of Lemma 6 compare to the actual half-widths. In Table 7 we summarize average and median ratio of bound to actual half-width for samples of size $N = 2,000$. While easier to compute we see that the bounds can be quite conservative. This is in large part due to their dependance on estimates of $\Pr(Z \in T_i)$, especially in the case of bounds for a^{r_N} which require considering each T_i.

Table 7: Ratio of upper bound to interval half-width

		a^{r_N}		η_j^α	
		Average ratio	Median ratio	Average ratio	Median ratio
	$(z_0)_1$	6.20	6.33	3.04	2.18
	$(z_0)_2$	15.53	13.44	3.58	2.92
$N = 2,000$	$(z_0)_3$	4.00	3.49	2.25	1.55
	$(z_0)_4$	5.27	5.26	3.69	2.37
	$(z_0)_5$	9.20	8.04	2.80	2.12

When computing intervals for $(x_0)_j$ note that since $S = \mathbb{R}_+^5$ each selection function of $d\Pi_S(z_N^*)$ is represented by a diagonal matrix with values of zero and one along the diagonal. When $z_N - z_N^*$ falls into a cone for which the jth diagonal element of the selection function's matrix representation is one from (33) and (39) we see that η_j^α and h_j^α will be equal. The interval for $(x_0)_j$ produced using the approach of §3.3

would then be the same as the projection onto S of the interval for $(z_0)_j$ produced using the approach of §3.2. If jth diagonal element is zero the method of §3.3 returns the correct point estimate $(x_N)_j = 0$, whereas the projection onto S of the interval for $(z_0)_j$ produced using the approach of §3.2 is given by $\left[0, \max\left\{0, (z_N)_j + N^{-1/2}\eta_j^{\alpha}\right\}\right]$. The
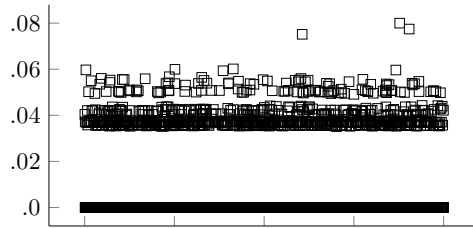

Fig. 4: Bounds for $h_3^{\alpha}, N = 2,000$ by sample

two approaches therefore produce intervals that cover $(x_0)_j = 0$ at an identical rate with the approach of §3.3 returning the correct point estimate more often. Moreover the bound for h_j^{α} provided by Corollary 1 will have similar properties since these bounds consider adjusting only the value of α and not the selection functions used. Therefore when the jth diagonal element of the indicated selection function is one the bounds for h_j^{α} and η_j^{α} will be the same, and if the jth diagonal element is zero the bound for h_j^{α} is also zero. This is illustrated in Figure 4 where we have plotted the bounds for h_3^{α} for each sample of size 2,000.

References

1. Anitescu, M., Petra, C.: Higher-order confidence intervals for stochastic programming using bootstrapping. Tech. Rep. ANL/MCS-P1964-1011, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL (2011)
2. Demir, M.C.: Asymptotics and confidence regions for stochastic variational inequalities. Ph.D. thesis, University of Wisconsin, Madison (2000)
3. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I. Springer, New York (2003)
4. Genz, A., Bretz, F.: Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics. Springer-Verlag, Heidelberg (2009)
5. Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T.: mvtnorm: Multivariate Normal and t Distributions (2013). URL http://CRAN.R-project.org/package=mvtnorm. R package version 0.9-9996
6. Gürkan, G., Yonca Özge, A., Robinson, S.M.: Sample-path solution of stochastic variational inequalities. Mathematical Programming **84**, 313–333 (1999)
7. King, A.J., Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming. Mathematics of Operations Research **18**, 148–162 (1993)

8. Lu, S.: 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
9. Lu, S.: Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities. SIAM Journal on Optimization (2014). Forthcoming
10. Lu, S., Budhiraja, A.: Confidence regions for stochastic variational inequalities. Mathematics of Operations Research **38**, 545–568 (2013)
11. Mi, X., Miwa, T., Hothorn, T.: mvtnorm: New numerical algorithm for multivariate normal probabilities. The R Journal **1**(1), 37–39 (2009)
12. Miwa, T., Hayter, A., Kuriki, S.: The evaluation of general non-centred orthant probabilities. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **65**(1), 223–234 (2003)
13. Pang, J.: Newton's method for b-differentiable equations. Mathematics of Operations Research **15**, 311–341 (1990)
14. Robinson, S.M.: An implicit-function theorem for a class of nonsmooth functions. Mathematics of Operations Research **16**(2), 292–309 (1991)
15. Robinson, S.M.: Normal maps induced by linear transformations. Mathematics of Operations Research **17**(3), 691–714 (1992)
16. Robinson, S.M.: Sensitivity analysis of variational inequalities by normal-map techniques. In: F. Giannessi, A. Maugeri (eds.) Variational Inequalities and Network Equilibrium Problems, pp. 257–269. Plenum Press, New York (1995)
17. Scholtes, S.: Introduction to Piecewise Differentiable Equations. Springer (2012)
18. Shapiro, A., Dentcheva, D., Ruszczyński, A.P.: Lectures on Stochastic Programming: Modeling and Theory. Society for Industrial and Applied Mathematics and Mathematical Programming Society (2009)
19. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. Optimization **57**, 395–418 (2008)
20. Vogel, S.: Universal confidence sets for solutions of optimization problems. SIAM Journal on Optimization **19**(3), 1467–1488 (2008)
21. Xu, H.: Sample average approximation methods for a class of stochastic variational inequality problems. Asia-Pacific Journal of Operational Research **27**(1), 103–119 (2010)