CONFIDENCE INTERVALS FOR STOCHASTIC VARIATIONAL INEQUALITES

Michael Lamm

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2015

Approved by:

Shu Lu

Amarjit Budhiraja

Vidyadhar Kulkarni

Gabor Pataki

Scott Provan

ABSTRACT

MICHAEL LAMM: Confidence intervals for solutions to stochastic variational inequalities
(Under the direction of Shu Lu)

This dissertation examines the effects of uncertain data on a general class of optimization and equilibrium problems. The common framework used for modeling these problems is a stochastic variational inequality. Variational inequalities can be used to model conditions that characterize an equilibrium state, or describe necessary conditions for solutions to constrained optimization problems. For example, Cournot-Nash equilibrium problems and the Karush-Kuhn-Tucker conditions for nonlinear programming problems both fit in the framework of a variational inequality. Uncertain model data can be incorporated into a variational inequality through the use of an expectation function. A variational inequality defined in this manner is referred to as a stochastic variational inequality (SVI).

For many problems of interest the SVI cannot be solved directly. This can be due to limited distributional information or an expectation function that lacks a closed form expression and is difficult to evaluate. When this is the case, the SVI must be replaced with a suitable approximation. A common approach is to solve a sample average approximation (SAA). The SAA problem is a variational inequality with the expectation function replaced by a function that depends on a sample of the uncertain data. A natural question is then how the solution of the SAA problem compares to the true solution of the SVI. To address this question, this dissertation examines the construction of simultaneous and individual confidence intervals for the true solution of an SVI.

To my Uncle Mike

ACKNOWLEDGEMENTS

First and foremost I must recognize my advisor Shu Lu. It has been a pleasure learning from and working with you. If not for the time and effort you have invested in me I would not be at this point. While I cannot fully express my gratitude for all that you have done in this space, I hope the work in this dissertation is able to provide some reflection of its effects. Next, I must thank Professor Amarjit Budhiraja. I have benefited greatly from the time I have had to work with you, and your impact on me and this dissertation are greatly appreciated.

I would also like to recognize the members of my dissertation committee, Professors Vidyadhar Kulkarni, Gabor Pataki, and Scott Proven, along with the other instructors I have had at UNC, Professors Nilay Argon, Edward Carlstein, Vladas Pipiras, and Serhan Ziya. All of you have been wonderful and have made my time as a graduate student a truly rewarding experience that has exceeded all of my expectations. Also, I would like to thank Alison Kieber and the rest of the staff in the STOR department for helping to keep me blissfully unaware of all of the administrative work that must have accompanied my time as a graduate student.

To my fellow students here at UNC, in particular those in my cohort and Sean Skwerer, thank you for all of your support. Your advice, input, and willingness to be sounding boards for everything from teaching, research, practice talks, and avoiding work has always been appreciated. I feel very lucky to have overlapped with all of you here in the STOR department and wish all of you the best moving forward.

Finally, I must thank my family for their love and support. To my parents Maureen and David, I always remember the standards you have set and modeled for me, to do good, be kind, be happy, and take pride in my work. Thank you Erin for your support and messages of encouragement that, without fail, came before every milestone. Kristin and Dan, I am not really sure how to characterize all that you have done for me and Melissa. All I can say

is that the time we spent living across the street from you will always be cherished. Melissa, I cannot imagine how much more difficult graduate school would have been without you. I have always cherished my time with you and look forward to the next stage of our lives together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

SAA	Sample average approximation
SVI	Stochastic variational inequality
$a^r(\cdot)$	Width of an individual confidence interval for $(z_0)_j$ using the first method
$C^1(X, \mathbb{R}^n)$	The space of continuously differentiable mappings $f : X \to \mathbb{R}^n$
$\eta_j^\alpha(\cdot, \cdot)$	Width of an individual confidence interval for $(z_0)_j$ for the second method
$\tilde{\eta}_j^{\alpha_2}(\cdot, \cdot)$	Width of a relaxed individual confidence interval for $(z_0)_j$
f_S^{nor}	The normal map induced by f and S
$h_j^\alpha(\cdot, \cdot, \cdot)$	Width of an individual confidence interval for $(x_0)_j$
$\tilde{h}_j^{\alpha_2}(\cdot, \cdot, \cdot)$	Width of a relaxed individual confidence interval for $(x_0)_j$
$K(z)$	The critical cone to S at z
Π_S	The Euclidian projector onto a set S
N	Sample size
$\mathbf{N}_S(x)$	The normal cone to S at x
$\mathbb{R}^n$	Set of n dimensional real valued vectors
$\text{ri}(S)$	The relative interior of a convex set S
$T_S(x)$	The tangent cone to S at x
$w_{N,j}^\epsilon$	Width of jth component of a simultaneous confidence interval for $(z_0)_j$
x_N	A solution to an SAA problem
x_0	A solution to an SVI
z_N	A solution to the normal map formulation of an SAA problem
z_0	A solution to the normal map formulation of an SVI

CHAPTER 1
Introduction

Variational inequalities model a general class of equilibrium problems and also arise as first-order necessary conditions of optimization problem, see (Attouch et al., 2009; Facchinei and Pang, 2003; Ferris and Pang, 1997a,b; Giannessi and Maugeri, 1995; Giannessi et al., 2001; Harker and Pang, 1990; Pang and Ralph, 2009). A variational inequality, defined formally in §1.1, is characterized by a set S and a function f. In many problems of interest, data defining the function are subject to uncertainty. One way to handle such uncertainty is to treat f as an expectation function, and this gives rise to a stochastic variational inequality (SVI). For many problems the expectation function lacks a closed form expression and a numerical approximation is generally required. Such approximations usually make use of some sampling procedure. Based on how sampling is incorporated into the approximation scheme, SVI algorithms can be classified into stochastic approximation (SA) methods and sample average approximation (SAA) methods. SA methods as introduced in (Robbins and Monro, 1951) update iterate points with samples taken at each step. The application of SA methods to SVIs have been studied in (Chen et al., 2014; Jiang and Huifu, 2008; Juditsky et al., 2011; Koshal et al., 2013; Nemirovski et al., 2009a) and references therein. For the development of SA methods in stochastic optimization, see (Nemirovski et al., 2009b; Polyak, 1990; Polyak and Juditsky, 1992) and references therein.

In this dissertation we consider the case when a sample average approximation (SAA) is used. The SAA method uses a single sample to estimate the unknown function f with a sample average function, defined formally in §1.1. A solution to the SAA problem is a solution to the variational inequality defined by the sample average function and set S. A natural question to consider is how the solution to the SAA problem compares to the solution of the original SVI. 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 et al., 1999; King and Rockafellar, 1993; Shapiro et al., 2009). Xu (Xu, 2010) 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 (Shapiro and Xu, 2008). Working with the exponential rate of convergence of SAA solutions, Anitescu and Petra in (Anitescu and Petra, 2011) developed confidence intervals for the optimal value of stochastic programming problems using bootstrapping. Limiting distributions for SAA solutions were obtained in (King and Rockafellar, 1993, Theorem 2.7) and (Shapiro et al., 2009, Section 5.2.2). For random approximations to deterministic optimization problems, universal confidence sets for the true solution set were developed by Vogel in (Vogel, 2008) using concentration of measure results.

The major contribution of this dissertation is the development methods for the efficient calculation of confidence intervals for the true solution to an SVI from a single SAA solution, based on the asymptotic distribution of SAA solutions. To our knowledge, the computation of confidence sets for an SVI's solution based upon the asymptotic distribution of SAA solutions started from the dissertation of Demir (Demir, 2000). By considering the *normal map formulation* (to be defined formally in §1.1) of variational inequalities, Demir used the asymptotic distribution to obtain an expression for confidence regions of the solution to the normal map formulation of an SVI. Because some quantities in that expression depend on the true solutions 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. Standard techniques for the required justification cannot be used due to the general nonsmooth structure of S and the discontinuities this creates in certain quantities.

In (Lu and Budhiraja, 2013) Lu and Budhiraja provided and justified a new method of constructing asymptotically exact confidence regions for z_0. The confidence regions were computable from a solution to the normal map formulation of a single SAA problem (1.3); the latter solution is denoted by z_N and is formally defined in Theorem 1. The approach in (Lu and Budhiraja, 2013) was to combine the asymptotic distribution of z_N with its expo-

nential rate of convergence, and involved calculating a weighted-sum of a family of functions. The method was later simplified by Lu in (Lu, 2012) by using a single function from the family. Due to the potentially piecewise linear structure that underlies the asymptotic distribution of SAA solutions, the methods in (Lu, 2012; Lu and Budhiraja, 2013) may require working with piecewise linear transformations of normal random vectors. Lu in (Lu, 2014) 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 (Lu, 2014) is easier to use since it has the advantage of working (with high probability) with linear transformations of normal random vectors, even when the asymptotic distribution of z_N is not normal.

Component-wise confidence intervals for the true solution are generally easier to visualize and interpret compared to confidence regions. 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. Additionally, individual confidence intervals provide a quantitative measure of the uncertainty in each individual component, and therefore cary important information not covered by larger confidence sets. Individual confidence intervals that can be obtained by using confidence regions are too conservative for any practical use, especially for large scale problems. A method to construct individual confidence intervals for z_0 using linear estimates was analyzed in (Lu, 2014). While computationally efficient, the method requires some restrictive assumptions to guarantee that the specified level of confidence is met.

The methods for computing individual confidence intervals we develop in this dissertation are shown to achieve the guaranteed confidence levels in more general situations. This attribute differentiates our methods from existing approaches that consider the specialized case when the asymptotic distribution is Gaussian or generate conservative confidence sets based on error bounds. The methods we develop are also able to limit the computational burden of working with the possibly piecewise linear transformations. Another contribution of this dissertation is to provide a *direct approach* to finding individual confidence intervals for components of x_0. As noted above, the confidence region/interval methods in (Demir, 2000; Lu, 2012, 2014; Lu and Budhiraja, 2013) 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 set of z_0, one can obtain a confidence set for x_0, by projecting the confidence set 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 set for z_0. When S is a box, individual confidence intervals of x_0 can be obtained from projecting the individual confidence intervals of z_0 onto S. 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.

In Section 1.1 the SVI and SAA problems are formally defined along with their normal map formulations. Pertinent properties of piecewise affine functions are reviewed in §1.2 along with the notion of B-differentiability. Previous works on the relationship between the SVI and SAA problems are summarized in §1.3, and §1.4 outlines the methods for interval computation discussed in remainder of this dissertation.

1.1 Stochastic variational inequalities

An SVI is defined as follows. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and ξ be a 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 $E\|F(x,\xi)\| < \infty$ for each $x \in O$. 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

$$0 \in f_0(x) + \mathbf{N}_S(x), \tag{1.1}$$

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

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

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

It is often the case that the function f_0 does not have a closed form expression and is difficult to evaluate, in which case an SAA problem may be solved instead. The SAA

method takes independent and identically distributed (i.i.d) random variables $\xi^1, \xi^2, \ldots, \xi^N$ with the same distribution as ξ and constructs a sample average function. The sample average function $f_N : O \times \Omega \to \mathbb{R}^n$ is defined by

$$f_N(x, \omega) = N^{-1} \sum_{i=1}^{N} F(x, \xi^i(\omega)). \tag{1.2}$$

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

$$0 \in f_N(x, \omega) + \mathbf{N}_S(x). \tag{1.3}$$

Solutions of (1.1) are referred to as true solutions, whereas solutions of (1.3) are refereed to as SAA solutions.

The formulations of the SVI and SAA problems as given in (1.1) and (1.3) involve the set valued mapping $\mathbf{N}_S(\cdot)$. In their normal map formulations the set valued mapping is removed and solutions are identified as the zeros of single-valued non-smooth functions. For the SVI, the function is the normal map induced by f_0 and S, $f_{0,S}^{\text{nor}} : \Pi_S^{-1}(O) \to \mathbb{R}^n$, defined as

$$f_{0,S}^{\text{nor}}(z) = f_0 \circ \Pi_S(z) + (z - \Pi_S(z)). \tag{1.4}$$

Here Π_S denotes the Euclidian projector onto the set S, $\Pi_S^{-1}(O)$ is the set of all points $z \in \mathbb{R}^n$ such that $\Pi_S(z) \in O$, and $f_0 \circ \Pi_S$ is the composite function of f_0 and Π_S. The normal map formulation of (1.1) is to find a point $z \in \Pi_S^{-1}(O)$ such that

$$f_{0,S}^{\text{nor}}(z) = 0. \tag{1.5}$$

The two formulations are related by the fact that $x \in O \cap S$ solves (1.1) only if $z = x - f_0(x)$ satisfies (1.5). Moreover when this equality is satisfied it additionally holds that $\Pi_S(z) = x$.

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

$$f_{N,S}^{\text{nor}}(z) = f_N \circ \Pi_S(z) + (z - \Pi_S(z)). \tag{1.6}$$

The normal map formulation of the SAA problem is then to find $z \in \Pi_S^{-1}(O)$ such that

$$f_{N,S}^{\text{nor}}(z) = 0, \tag{1.7}$$

where (1.7) and (1.3) are related in the same manner as (1.5) and (1.1). In general, for a function G mapping from a subset D of $\mathbb{R}^n$ back into $\mathbb{R}^n$, the normal map induced by G and S is a map defined on $\Pi_S^{-1}(D)$ with $G_S^{\text{nor}}(z) = G \circ \Pi_S(z) + z - \Pi_S(z)$.

By assumption, S is a polyhedral convex set, so the Euclidian projector Π_S is a piecewise affine function. In the next section we provide a summary of pertinent properties of piecewise affine functions, in particular the notion of B-differentiability.

1.2 Piecewise affine functions

A continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ is piecewise affine if there exists a finite collection of affine functions f_j, $j = 1, \ldots, l$, such that for all $x \in \mathbb{R}^n$ the inclusion $f(x) \in \{f_1(x), \ldots, f_l(x)\}$ holds. The affine functions f_j are refereed to as the selection functions of f. When each f_j is a linear 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, $\Gamma = \{P_1, \ldots, P_l\}$, satisfying the following three conditions:

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

If each of the P_i is additionally a cone, then Γ is referred to as a conical subdivision. As seen in (Scholtes, 2012, 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 P_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 next consider the special case of the Euclidian projector onto a polyhedral convex set S, a thorough discussion of which can be found in (Scholtes, 2012, Section 2.4). Let $\mathcal{F}$ be the finite collection of all nonempty faces of S. On the relative interior of each nonempty face $F \in \mathcal{F}$ the normal cone to S is a constant cone, denoted as $\mathbf{N}_S(\mathrm{ri}F)$, and the set addition $C_F = F + \mathbf{N}_S(\mathrm{ri}F)$ results in 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 H is positive homogeneous if $H(\lambda x) = \lambda H(x)$ for all positive numbers $\lambda \in \mathbb{R}$ and points $x \in \mathbb{R}^n$. The function H is referred to as the B-derivative of h at x and will be denoted $dh(x)$. When $dh(x)$ is also linear, $dh(x)$ is the classic Fréchet derivative (F-derivative).

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 polyhedron $P_i \in \Gamma$, $df(x)$ 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 $\Gamma(x) = \{P_i \in \Gamma | x \in P_i\}$, $I = \{i \,| P_i \in \Gamma(x)\}$ and $\Gamma'(x) = \{K_i = \mathrm{cone}(P_i - x) | i \in I\}$. 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 the family of selection functions given by $\{df_i(x) | i \in I\}$ and the corresponding conical subdivision given by $\Gamma'(x)$.

The relation between the normal manifold of S and Π_S extends to the form of the B-derivative $d\Pi_S(x)$. First we define the tangent cone to a polyhedral convex set S at $x \in S$

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 $z \in \mathbb{R}^n$ to be

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

As shown in (Robinson, 1991, Corollary 4.5) and (Pang, 1990, Lemma 5), for any point $z \in \mathbb{R}^n$ and any sufficiently small $h \in \mathbb{R}^n$ the equality

$$\Pi_S(z + h) = \Pi_S(z) + \Pi_{K(z)}(h) \tag{1.8}$$

holds, which implies

$$d\Pi_S(z) = \Pi_{K(z)}. \tag{1.9}$$

The connection to the normal manifold of S follows from the fact that for all points z in the relative interior of a k-cell the critical cone $K(z)$ is a constant cone; see (Lu and Budhiraja, 2013, Theorem 8), and thus $d\Pi_Z(z)(\cdot)$ is the same function for all z in the relative interior of a k-cell. For points z and z' in the relative interior of different k-cells $d\Pi_S(z)(\cdot)$ and $d\Pi_S(z')(\cdot)$ can be quite different, and as a result small changes in the choice of z can result in significant changes in the form of $d\Pi_S(z)(\cdot)$.

To illustrate these concepts we end this section with an example. Take $S = \mathbb{R}^2_+$, where $\mathbb{R}_+ = \{x \in \mathbb{R},\, x \geq 0\}$. The set S has four nonempty faces with $\mathfrak{F} = \left\{\mathbb{R}^2_+,\, \mathbb{R}_+ \times \{0\},\, \{0\} \times \{0\},\, \{0\} \times \mathbb{R}_+\right\}$. The corresponding 2-cells in the normal manifold of S are the orthants $\mathbb{R}^2_+$, $\mathbb{R}_+ \times \mathbb{R}_-$, $\mathbb{R}^2_-$ and $\mathbb{R}_- \times \mathbb{R}_+$. There are five k-cells with $k < n$. Four 1-cells are the half-lines defined by the positive and negative axes, $\mathbb{R}_+ \times \{0\}$, $\{0\} \times \mathbb{R}_+$, $\mathbb{R}_- \times \{0\}$, $\{0\} \times \mathbb{R}_-$, and the fifth k-cell with $k = 0$ is the origin $\{0\} \times \{0\}$.

The restriction of Π_S to each 2-cell is a linear function, with the functions represented by the matrices

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

At $x = (0,1) \in \text{ri}\,(\{0\} \times \mathbb{R}_+)$, Π_S is not F-differentiable but has B-derivative $d\Pi_S(x)(\cdot)$

$$d\Pi_s(x)(h) = \begin{bmatrix} v & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \quad \text{where} \quad v = \begin{cases} 1 & \text{if} \quad h_1 \geq 0, \\ 0 & \text{if} \quad h_1 \leq 0. \end{cases}$$

In contrast, for a point $x' = (\epsilon, 1) \in \text{ri}\left(\mathbb{R}^2_+\right)$ for $\epsilon > 0$, the B-derivative $d\Pi_S(x')(\cdot)$ is a linear function represented by the identity matrix.

1.3 Background

In this section we discuss previous work on the computation of confidence sets for the true solution to an SVI. This section begins with a review of conditions under which the SAA solutions will have the required asymptotic properties. These properties include the almost sure convergence of the SAA solutions to a true solution, an exponential rate for the convergence in probability, and the weak convergence of SAA solutions.

The following notation will be used throughout this section and the remainder of this dissertation. Let x_0 and x_N denote solutions to the true SVI and SAA problems (1.1) and (1.3). We use Σ_0 to denote the covariance matrix of $F(x_0, \xi)$, and Σ_N to denote the sample covariance matrix of $\{F(x_N, \xi^i)\}_{i=1}^N$. A normal random vector with mean μ and covariance matrix Σ shall be denoted by $\mathcal{N}(\mu, \Sigma)$. A χ^2 random variable with l degrees of freedom will be denoted by χ^2_l. Weak convergence of random variables Y_n to Y will be denoted as $Y_n \Rightarrow Y$.

Assumption 1. *(a) $E\|F(x,\xi)\|^2 < \infty$ for all $x \in O$.*
(b) The map $x \mapsto F(x, \xi(\omega))$ is continuously differentiable on O for a.e. $\omega \in \Omega$, and $E\|d_x F(x,\xi)\|^2 < \infty$ for all $x \in O$.
(c) There exists a square integrable random variable C such that for all $x, x' \in O$

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

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

From Assumption 1 it follows that f_0 is continuously differentiable on O, see, e.g., (Shapiro et al., 2009, 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

$$\|f\|_{1,X} = \sup_{x \in X} \|f(x)\| + \sup_{x \in X} \|df(x)\|. \tag{1.10}$$

Then in addition to providing nice integrability properties for f_N, as shown in (Shapiro et al., 2009, Theorem 7.48) Assumption 1 will guarantee the almost sure convergence of the sample average approximation function f_N to f_0 as an element of $C^1(X, \mathbb{R}^n)$ and that $df_0(x) = E\left[d_x F(x, \xi)\right]$.

Assumption 2. *Suppose that x_0 solves the variational inequality (1.1). Let $z_0 = x_0 - f_0(x_0)$, $L = df_0(x_0)$, $K_0 = T_S(x_0) \cap \{z_0 - x_0\}^\perp$, and assume that $L^{nor}_{K_0}$ is a homeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$, where $L^{nor}_{K_0}$ is the normal map induced by L and K_0.*

Assumption 2 guarantees that x_0 is a locally unique solution and that (1.1) has a locally unique solution under sufficiently small perturbations of f_0 in $C^1(X, \mathbb{R}^n)$, see (Lu and Budhiraja, 2013, Lemma 1) and the original result in (Robinson, 1995). Since the critical cone K_0 is a polyhedral convex cone, $L^{\text{nor}}_{K_0}$ is a piecewise linear function. It was shown in (Robinson, 1992) that $L^{\text{nor}}_{K_0}$ is a homeomorphism if and only if the determinants of the matrices representing its selections functions all have the same nonzero sign. Shorter proof of this result can be found in (Ralph, 1994) and (Scholtes, 1996). A piecewise linear function with this property is said to be coherently oriented. A special case in which the coherent orientation condition holds is when the restriction of L on the linear span of K_0 is positive definite. In particular, if f_0 is strongly monotone on O, then the entire matrix L is positive definite and $L^{\text{nor}}_{K_0}$ is a global homeomorphism. Another special case is when the cone $K_0 = \mathbb{R}^n_+$, the nonnegative orthant; for such a case the coherent orientation condition on $L^{\text{nor}}_{K_0}$ is equivalent to the requirement that L is a P-matrix.

The normal maps $L^{\text{nor}}_{K_0}$ and $f^{\text{nor}}_{0,S}$ are also related through the B-differentiability of Π_S. Following the discussion of B-differentiability above Assumption 1, let $\Gamma'(z_0)$ denote the conical subdivision that corresponds to $d\Pi_S(z_0)$. Since f_0 is differentiable under Assumption

1, the chain rule of B-differentiability implies that $f_{0,S}^{\text{nor}}$ is B-differentiable, with its B-derivative at z_0 given by

$$df_{0,S}^{\text{nor}}(z_0)(h) = df_0(x_0) \circ d\Pi_S(z_0)(h) + h - d\Pi_S(z_0)(h) \tag{1.11}$$

with corresponding conical subdivision $\Gamma'(z_0)$.

Applying (1.9) to z_0, one can see the normal map $L_{K_0}^{\text{nor}}$ is exactly $df_{0,S}^{\text{nor}}(z_0)$, a result that first appeared in (Robinson, 1992). Note that the B-derivative for the normal map $f_{N,S}^{\text{nor}}$, denoted by $df_{N,S}^{\text{nor}}(\cdot)$, will take an analogous form to (1.11).

The following theorem is adapted from (Lu and Budhiraja, 2013, Theorem 7). It provides the almost sure and weak convergence of the SAA solutions z_N and x_N. Those results are obtained by combining convergence properties of the sample average function f_N with sensitivity analysis techniques originally developed in (Robinson, 1995) for deterministic variational inequalities. Similar results were also shown in (King and Rockafellar, 1993, Theorem 2.7) using the concept of subinvertibility and a set of assumptions that are implied by those used here.

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 (1.7) has a unique solution z_N in Z, and the variational inequality (1.3) has a unique solution in X_0 given by $x_N = \Pi_S(z_N)$. Moreover, $\lim_{N\to\infty} z_N = z_0$ and $\lim_{N\to\infty} x_N = x_0$ almost surely,*

$$\sqrt{N}(z_N - z_0) \Rightarrow (L_{K_0}^{nor})^{-1}(Y_0), \tag{1.12}$$

$$\sqrt{N} L_{K_0}^{nor}(z_N - z_0) \Rightarrow Y_0, \tag{1.13}$$

and

$$\sqrt{N}(\Pi_S(z_N) - \Pi_S(z_0)) \Rightarrow \Pi_{K_0} \circ (L_{K_0}^{nor})^{-1}(Y_0). \tag{1.14}$$

The results of Theorem 1 follow from the convergence of f_N to f_0 in $C^1(X, \mathbb{R}^n)$, and the existence of locally unique solutions to (1.1) for sufficiently small perturbations of f_0 in this same space. In particular, Assumptions 1 provides a sufficient conditions for the weak convergence of $\sqrt{N}(f_N - f_0)$ in $C^1(X, \mathbb{R}^n)$ which combined with Assumption 2 yields the asymptotic distributions in (1.12), (1.13) and (1.14).

In his dissertation (Demir, 2000), Demir developed methods to compute confidence regions for true solutions of SVIs using (1.13). Recognizing that the resulting expression depended on the true solution through both Σ_0 and $L^{\mathrm{nor}}_{K_0}$, he proposed to use Σ_N and $df^{\mathrm{nor}}_{N,S}(z_N)$ in the expression for the confidence regions. He did not, however, justify how such a replacement preserves the weak convergence property needed for the asymptotic exactness of the confidence regions. The discontinuity of $d\Pi_S(z)$ with respect to z, and in particular the fact that $d\Pi_S(z_N)$ does not in general converge to $d\Pi_S(z_0)$, prevents standard techniques from being applicable for such a justification. The issue that arises is that when $d\Pi_S(z_0)$ is piecewise linear the probability of $d\Pi_S(z_N)$ being a linear map goes to one as the sample size N goes to infinity; see (Lu, 2014, Proposition 3.5). While this poses a challenge for establishing the exactness of confidence regions constructed using $df^{\mathrm{nor}}_{N,S}(z_N)$ as an estimate for $L^{\mathrm{nor}}_{K_0}$, it also illustrates the desirability of using such regions since their expression would with high probability involve only linear functions.

To establish the exactness of confidence regions constructed using $df^{\mathrm{nor}}_{N,S}(z_N)$ Lu in (Lu, 2014, Theorem 3.3 and 4.1) examined the relationship between $df^{\mathrm{nor}}_{0,S}(z_0)(z_N - z_0)$ and $-df^{\mathrm{nor}}_{N,S}(z_N)(z_0 - z_N)$ and proved the following results.

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

$$\lim_{N\to\infty} \Pr\{\sqrt{N}\|df^{nor}_{0,S}(z_0)(z_N - z_0) + df^{nor}_{N,S}(z_N)(z_0 - z_N)\| > \epsilon\} = 0. \tag{1.15}$$

Consequently, we have

$$-\sqrt{N}df^{nor}_{N,S}(z_N)(z_0 - z_N) \Rightarrow Y_0. \tag{1.16}$$

Moreover, if Σ_0 is nonsingular, then

$$-\sqrt{N}\Sigma_N^{-1/2} df_{N,S}^{nor}(z_N)(z_0 - z_N) \Rightarrow \mathcal{N}(0, I_n). \tag{1.17}$$

If Σ_0 is singular, let $\rho > 0$ be the minimum of all positive eigenvalues of Σ_0, and let l be the number of positive eigenvalues of Σ_0 counted with regard to their algebraic multiplicity. Decompose Σ_N as

$$\Sigma_N = U_N^T \Delta_N U_N$$

where U_N is an orthogonal $n \times n$ matrix, and Δ_N is a diagonal matrix with monotonically decreasing elements. Let D_N be the upper-left submatrix of Δ_N whose diagonal elements are at least $\rho/2$. Let l_N be the number of rows in D_N, $(U_N)_1$ be the submatrix of U_N that consists of its first l_N rows, and $(U_N)_2$ be the submatrix that consists of the remaining rows of U_N. Then for almost every ω the equality $l_N = l$ holds for sufficiently large N. Moreover,

$$N\big[df_{N,S}^{nor}(z_N)(z_0 - z_N)\big]^T (U_N)_1^T D_N^{-1} (U_N)_1 \big[df_{N,S}^{nor}(z_N)(z_0 - z_N)\big] \Rightarrow \chi_l^2 \tag{1.18}$$

and

$$N df_{N,S}^{nor}(z_N)(z_0 - z_N)^T (U_N)_2^T (U_N)_2 df_{N,S}^{nor}(z_N)(z_0 - z_N) \Rightarrow 0. \tag{1.19}$$

Using (1.17), (1.18) and (1.19) we can give computable expressions for asymptotically exact confidence regions for z_0. To this end, for any $\alpha \in (0,1)$ and integer k let $\chi_k^2(\alpha)$ be the $(1-\alpha)$ percentile of a χ^2 random variable with k degree's of freedom, and let $\|\cdot\|_\infty$ denote the ∞-norm for a vector $x \in \mathbb{R}^n$. Then for any $\epsilon > 0$ and integer N we define sets R_N when Σ_N is nonsingular, and $R_{N,\epsilon}$ when Σ_N is singular, to be

$$R_N = \left\{ z \in \mathbb{R}^n \,\middle|\, N\big[df_{N,S}^{\mathrm{nor}}(z_N)(z - z_N)\big]^T \Sigma_N^{-1} \big[df_{N,S}^{\mathrm{nor}}(z_N)(z - z_N)\big] \le \chi_n^2(\alpha) \right\}, \tag{1.20}$$

$$R_{N,\epsilon} = \left\{ z \in \mathbb{R}^n \,\middle|\, \begin{array}{l} N\big[df_{N,S}^{\mathrm{nor}}(z_N)(z - z_N)\big]^T (U_N)_1^T D_N^{-1} (U_N)_1 \big[df_{N,S}^{\mathrm{nor}}(z_N)(z - z_N)\big] \le \chi_{l_N}^2(\alpha) \\ \\ \|\sqrt{N}(U_N)_2 df_{N,S}^{\mathrm{nor}}(z_N)(z - z_N)\|_\infty \le \epsilon \end{array} \right\}.$$

Depending on if Σ_N is singular or not, by Theorem 2 we will have that either

$$\lim_{N\to\infty} \Pr\{z_0 \in R_N\} = 1-\alpha \text{ or } \lim_{N\to\infty} \Pr\{z_0 \in R_{N,\epsilon}\} = 1-\alpha.$$

Note that the expression for confidence regions in the nonsingular case is the same as that proposed by Demir. Since the nonsingular case can be treated as a specialization of the singular case with $l_N = n$ and $\epsilon = 0$, moving forward we focus on the singular case and consider regions $R_{N,\epsilon}$.

While the regions $R_{N,\epsilon}$ have a specified asymptotic level of confidence, they are not necessarily amenable to easy interpretation and visualization. It was thus suggested in (Lu, 2014) to construct easier to interpret simultaneous confidence intervals by finding the axis-aligned minimal bounding box that contains the region $R_{N,\epsilon}$. We examine questions raised by this approach to building simultaneous confidence intervals in an application to a stochastic Cournot-Nash equilibrium problem of moderate size in Chapter 2.

We now move our focus to the question of computing individual confidence intervals for components of z_0. A first approach would be to use the component interval of the simultaneous confidence intervals considered above, but such intervals are too conservative for any practical use. In (Lu, 2014) a natural expression for individual confidence intervals suggested by (1.17) was analyzed. Recall that (1.17) required the additional assumption that Σ_0 be nonsingular. Since this assumption is used throughout the discussion of individual confidence intervals we formally declare it as

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

The primary purpose of Assumption 3 will be to provide a sufficient condition for certain limits in the convergence results in Theorems 5, 6 and 7 to be well defined. Under Assumptions 1 and 2, the sample covariance matrix Σ_N converges almost surely to Σ_0, see (Lu, 2014, Lemma 3.6). A well conditioned Σ_N will therefore give us high confidence that Assumption 3 is true. Even if Σ_N is ill conditioned, it is possible to relax Assumption 3 in Theorems 5, 6 and 7; the way to relax it differs for each theorem and is noted after the proofs of those theorems. One can inspect if the relaxed conditions hold by checking

properties of Σ_N as well as the locations of $F(x_N, \xi^i)$, $i = 1, \ldots, N$, with respect to the normal manifold of S.

Before summarizing the results for the confidence intervals suggested by (1.17) some notation must be introduced. Let $df^{\text{nor}}_{0,S}(z_0)$ be piecewise linear with l pieces and the corresponding conical subdivision $\{K_1, \ldots, K_l\}$ .Then $df^{\text{nor}}_{0,S}(z_0)|_{K_i} = M_i$ for each $i = 1, \ldots, l$, where M_i stands for the matrix that represents $df^{\text{nor}}_{0,S}(z_0)$ on K_i. Under Assumption 2, $df^{\text{nor}}_{0,S}(z_0)$ 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

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

for each $i = 1, \ldots, l$ and $j = 1, \ldots, n$. Then for each $\alpha \in (0, 1)$ it follows that

$$\Pr\left(|(Y^i)_j| \leq r^i_j \sqrt{\chi^2_1(\alpha)} \right) = 1 - \alpha.$$

With this notation the following theorem was shown in (Lu, 2014, Theorem 5.1)

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

$$r_{Nj} = \sqrt{df^{nor}_{N,S}(z_N)^{-1} \Sigma_N df^{nor}_{N,S}(z_N)^{-T})_{jj}}$$

for each $j = 1, \ldots, n$. Then for each real number $\alpha \in (0, 1)$ and for each $j = 1, \ldots, n$,

$$\begin{aligned} \lim_{N \to \infty} \Pr & \left(\frac{\sqrt{N} |(z_n - z_0)_j|}{r_{Nj}} \leq \sqrt{\chi^2_1(\alpha)} \right) \\ & = \sum_{i=1}^{l} \Pr \left(\left| \frac{(Y^i)_j}{r^i_j} \right| \leq \sqrt{\chi^2_1(\alpha)} \text{ and } Y^i \in K_i \right) \end{aligned} \tag{1.21}$$

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

$$\Pr \left(\left| \frac{(Y^i)_j}{r^i_j} \right| \leq \sqrt{\chi^2_1(\alpha)} \text{ and } Y^i \in K_i \right) = \Pr \left(\left| \frac{(Y^i)_j}{r^i_j} \right| \leq \sqrt{\chi^2_1(\alpha)} \right) \Pr \left(Y^i \in K_i \right)$$

holds for each $i = 1, \ldots, l$. *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 (1.21) that this method of constructing individual confidence intervals, while easily computable using only the sample data, produces intervals whose asymptotic level of confidence is dependent on the true solution, unless the condition below (1.21) is satisfied. The latter condition is satisfied, when $df_{0,S}^{\mathrm{nor}}(z_0)$ is a linear function or has only two selection functions, in which case the intervals computed from this method will be asymptotically exact. However, in general the level of confidence for such intervals cannot be guaranteed. The issue with the linear estimate $df_{N,S}^{\mathrm{nor}}(z_N)$ is that it does not properly account for the possibly piecewise linear structure of $df_{0,S}^{\mathrm{nor}}(z_0)$. This limitation motivates the development of the methods proposed in Chapter 3 and 4. The three methods all produce intervals that maintain their desired asymptotic properties in the general setting by using estimates that capture the possibly piecewise linear structure of $df_{0,S}^{\mathrm{nor}}(z_0)$. To construct such estimates we will need the following additional assumption.

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* $M_x(t) \leq \exp\left\{\zeta^2 \|t\|^2/2\right\}$ *for every* $x \in X$ *and every* $t \in \mathbb{R}^n$.

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

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

   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

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

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

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

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

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

   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.*

First note that when Assumption 4 holds the conditions of Assumption 1 are satisfied. From Assumption 4 it follows that f_N converges to f_0 in probability at an exponential rate, as shown in (Lu and Budhiraja, 2013, Theorem 4) based on a general result (Shapiro et al., 2009, 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 $\epsilon > 0$ and N:

$$\Pr\left(\|f_N - f_0\|_{1,X} \geq \epsilon\right) \leq \beta_1 \exp\left\{-N\mu_1\right\} + \frac{M_1}{\epsilon^n}\exp\left\{-\frac{N\epsilon^2}{\sigma_1}\right\}. \tag{1.22}$$

Revisiting Theorem 1, if one additionally supposes that Assumption 4 holds, then as shown in (Lu and Budhiraja, 2013, Theorem 7), there exist positive real numbers $\epsilon_0, \beta_0, \mu_0, M_0$ and σ_0, such that the following holds for each $\epsilon \in (0, \epsilon_0]$ and each N:

$$\begin{aligned}
&\Pr\left(\|x_N - x_0\| < \epsilon\right) \geq \Pr\left(\|z_N - z_0\| < \epsilon\right) \\
&\geq 1 - \beta_0\exp\left\{-N\mu_0\right\} - \frac{M_0}{\epsilon^n}\exp\left\{\frac{-N\epsilon^2}{\sigma_0}\right\}.
\end{aligned} \tag{1.23}$$

The convergence of SAA solutions to the set of true solutions in probability at an exponential rate was also shown using the concept of subinvertibility in (Xu, 2010) with an assumption similar to Assumption 4.

The exponential rate of convergence as given in (1.23) was used in (Lu and Budhiraja, 2013) to estimate $df^{\text{nor}}_{0,S}(z_0)$ by a weighted-sum of a family of functions. The estimates were later simplified in (Lu, 2012) by using a single function from the family. Due to the computational ease of using a single function we focus our presentation to the estimates for $df^{\text{nor}}_{0,S}(z_0)$ used in (Lu, 2012). In this approach a point near z_N is used in the estimate for $d\Pi_S(z_0)$. More precisely, for each cell C_i in the normal manifold of S define a function $d_i : \mathbb{R}^n \to \mathbb{R}$ by

$$d_i(z) = d(z, C_i) = \min_{x \in C_i} \|x - z\|, \tag{1.24}$$

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

$$\Psi_i(\cdot) = d\Pi_S(z)(\cdot) \text{ for any } z \in \text{ri}C_i. \tag{1.25}$$

In (1.24) any norm for vectors in $\mathbb{R}^n$ can be chosen, and in (1.25) any $z \in \text{ri}C_i$ 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. $\lim_{N\to\infty} \frac{N}{g(N)^2} = \infty$.
4. $\lim_{N\to\infty} g(N)^n \exp\left\{-\sigma_0 \frac{N}{(g(N))^2}\right\} = 0$ 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 (1.22) and (1.23) respectively and C as in Assumption 1.
5. $\lim_{N\to\infty} \frac{N^{n/2}}{g(N)^n} \exp\left\{-\sigma g(N)^2\right\} = 0$ for each positive real number σ.

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

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 C_i is a face of an n-cell

that contains z and $d_i(z) \leq 1/g(N)$. Then define functions $\Lambda_N(z) : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Lambda_N(z)(h) = \Psi_{i_0}(h), \tag{1.26}$$

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

$$\Phi_N(z, h, \omega) = df_N(\Pi_S(z)) \circ \Lambda_N(z)(h) + h - \Lambda_N(z)(h). \tag{1.27}$$

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

$$\Phi_N(z_N)(h) = df_N(\Pi_S(z_N)) \circ d\Pi_S(z_N^*)(h) + h - d\Pi_S(z_N^*)(h). \tag{1.28}$$

We end the review of previous works with the following results shown in (Lu, 2012, Corollaries 3.2 and 3.3).

Theorem 4. *Suppose that Assumptions 2 and 4 hold. For each $N \in \mathbb{N}$, let Λ_N and Φ_N be as defined in (1.26) and (1.27). Then*

$$\lim_{N\to\infty} \Pr\left[\Lambda_N(z_N)(h) = d\Pi_S(z_0)(h) \text{ for all } h \in \mathbb{R}^n\right] = 1, \tag{1.29}$$

and there exists a positive real number θ, such that

$$\lim_{N\to\infty} \Pr\left[\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Phi_N(z_N)(h) - df_{0,S}^{nor}(z_0)(h)\|}{\|h\|} < \frac{\theta}{g(N)}\right] = 1. \tag{1.30}$$

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

$$\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0) \Rightarrow \mathcal{N}(0, I_n). \tag{1.31}$$

1.4 Outline

In the remainder of this dissertation we develop methods to compute confidence intervals for the true solution to an SVI, and apply those methods in stochastic optimization and equilibrium problems. In Chapter 2 we begin by examining the computation of simultaneous confidence intervals from the confidence regions given in (1.20) using the approach suggested in (Lu, 2014). Of particular interest will be the sensitivity of the interval widths and performance to the choice of the two parameters that arises in the case of a degenerate covariance matrix. The sensitivity of the interval's width to the parameters is first examined through a discussion of the interval's computation and the role of the parameters in these computations. The chapter then introduces the framework of stochastic Cournot-Nash equilibrium problems. The procedures for computing confidence regions and intervals are then applied to an example of the European gas market with numerical comparisons and sensitivity analysis results for both the confidence regions and intervals.

In Chapter 3 we propose two methods for constructing individual confidence interval for components of the true solution to the normal map formulation of an SVI. The two methods differ from the approach considered in Theorem 3 in terms of how they estimate the potentially piecewise linear transformation that appears in the asymptotic distribution of SAA solutions. The first method replaces the linear estimate used in Theorem 3 with the potentially piecewise linear estimate (1.27). The method produces intervals that will be asymptotically exact with less restrictive assumptions than those necessary for the method in Theorem 3. This improvement in interval accuracy comes with a computational cost. When an estimate for the transformation is piecewise linear with more than two selection functions, the intervals lack closed form expressions, and the computation necessary for finding an interval's width increases with the number of selection functions. This motivates the development of the second method of Chapter 3. The second method also uses the potentially piecewise linear estimate (1.27) but makes use of the SAA solution to limit the number of selection functions used in an interval's computation. Both of the methods proposed in Chapter 3 belong to the aforementioned indirect approaches. Approaches for computing the intervals are presented in Chapter 3 along with the establishment of upper

bounds for the interval lengths. The chapter ends with a comparison of the different methods using three numerical examples.

In Chapter 4 we propose a direct method for constructing individual confidence intervals for components of the true solution to an SVI as formulated in (1.1). The approaches for constructing confidence intervals for the normal map formulation of an SVI proposed in Chapter 3 cannot be applied to this problem due to the addition of a possibly noninvertible function to the asymptotic distribution. The new function also raises an issue unique to this chapter, namely, the possibility of components of the SAA solutions equaling the corresponding components of the true solution with a nonzero probability. This possibility provides a potential lower bound on the performance of any interval that contains the SAA solution, and therefore shifts the focus from asymptotically exact intervals to intervals for which a lower bound on the level of confidence can be guaranteed. A method for constructing intervals is then presented along with a theoretical justification. The chapter ends with two numerical examples.

In Chapter 5 we consider the computation of individual confidence intervals when the results of Theorem 4 do not hold and $\Phi_N(z_N)$ may not be a consistent estimate of $df^{\text{nor}}_{0,S}(z_0)$. This would allow us to relax the condition that z_N converge to z_0 in probability at an exponential rate, and therefore omit Assumption 4. To do so, we allow for some limited error in the estimation of a selection function of $df^{\text{nor}}_{0,S}(z_0)$. This error is then offset by adjusting the target probabilities used in the second indirect method of Chapter 3 and the direct method of Chapter 4. As a result, the method we propose produces intervals that meet a minimum level of confidence.

CHAPTER 2
Simultaneous confidence intervals

2.1 Construction simultaneous confidence intervals

In this chapter we examine the computation and performance of confidence regions and simultaneous confidence intervals for z_0. We focus on the case when the sample covariance matrix is singular, from which the nonsingular case can then be treated as a specialization. The singular case is of additional interest due to the dependence of the confidence regions on the choice of the parameter ϵ and value l_N. By Theorem 2, the confidence regions $R_{N,\epsilon}$ in (1.20) are asymptotically exact for all $\epsilon > 0$, but it remains to be seen how sensitive their performance is to the choice of ϵ for fixed sample sizes. For the choice of l_N, recall that l_N determines the matrices D_N, $(U_N)_1$, and $(U_N)_2$, and corresponds to the number of eigenvalues of Σ_N that are treated as nonzero. In Theorem 2, the smallest eigenvalue of Σ_0 is used to determine l_N. In practice, since only sample data are available, l_N and the matrices D_N and $(U_N)_1$ must be determined in a different manner.

To compute the simultaneous confidence intervals we use the approach suggested in (Lu, 2014) and find the minimal axis-aligned bounding box that contains the confidence regions $R_{N,\epsilon}$. In the remainder of this section we discuss the computation of simultaneous confidence intervals using this approach. From this discussion follows Proposition 1 which provides a closed form expression for the widths of the component intervals when the estimate $df^{\mathrm{nor}}_{N,S}(z_N)$ is a linear function. In §2.2 we introduce the framework of stochastic Cournot-Nash equilibrium problems and illustrate the procedures of computing confidence intervals using an example of the European gas market. Numerical comparisons and sensitivity analysis results are provided for both the confidence regions and simultaneous confidence intervals.

To begin, finding the left and right endpoints of the simultaneous confidence intervals requires solving

$$\begin{aligned} z_j^r = &\underset{z \in R_{N,\epsilon}}{\text{maximum}} \quad (z)_j \qquad \text{and} \qquad z_j^l = \underset{z \in R_{N,\epsilon}}{\text{minimum}} \quad (z)_j \end{aligned} \tag{2.1}$$

for $j = 1, 2 \ldots, n$, where $(z)_j$ denotes the jth component of the vector z. If $df_{N,S}^{\text{nor}}(z_N)$ is a piecewise linear function with corresponding conical subdivision $\{K_1, \ldots, K_m\}$, then problems in (2.1) needs to be further broken down to the following problems

$$\begin{aligned} z_{i,j}^r = &\underset{z \in R_{N,\epsilon} \cap K_i}{\text{maximum}} \quad (z)_j \qquad \text{and} \qquad z_{i,j}^r = \underset{z \in R_{N,\epsilon} \cap K_i}{\text{minimum}} \quad (z)_j \end{aligned} \tag{2.2}$$

for $j = 1, 2 \ldots, n$ and $i = 1, \ldots, m$, to account for the different expressions for $df_{N,S}^{\text{nor}}(z_N)$ on each K_i. The right and left endpoints for the jth component interval are then $z_j^r = \max\limits_{i=1,\ldots,m} z_{i,j}^r$ and $z_j^l = \min\limits_{i=1,\ldots,m} z_{i,j}^l$.

Computation of the endpoints is greatly simplified when $df_{N,S}^{\text{nor}}(z_N)$ is a linear function. In this case, the simultaneous confidence intervals are given by

$$\big[(z_N)_1 - w_{N,1}^\epsilon, (z_N)_1 + w_{N,1}^\epsilon\big] \times \cdots \times \big[(z_N)_n - w_{N,n}^\epsilon, (z_N)_n + w_{N,n}^\epsilon\big] \tag{2.3}$$

where $w_{N,j}^\epsilon$ is the optimal value of the following problem:

$$\begin{aligned} &\text{maximize} \quad (w)_j \\ &\text{subject to} \quad N\big[df_{N,S}^{\text{nor}}(z_N)(w)\big]^T (U_N)_1^T D_N^{-1} (U_N)_1 \big[df_{N,S}^{\text{nor}}(z_N)(w)\big] \le \chi_{l_N}^2(\alpha) \\ &\qquad\qquad \|\sqrt{N}(U_N)_2 df_{N,S}^{\text{nor}}(z_N)(w)\|_\infty \le \epsilon. \end{aligned} \tag{2.4}$$

We are therefore able to express the confidence intervals in terms of the optimal values of n quadratically constrained convex programs with linear objective functions. Since $df_{N,S}^{\text{nor}}(z_N)$ is with high probability a linear function, regardless of whether $df_{0,S}^{\text{nor}}(z_0)$ is piecewise linear or linear (Lu, 2014, Proposition 3.5), we will expect to experience the computational benefit from the linearity of $df_{N,S}^{\text{nor}}(z_N)$.

Note that both l_N and ϵ are responsible for determining the constraints in (2.4), the problem to find an interval's endpoints. The first constraint in that problem

$$N\big[df^{\text{nor}}_{N,S}(z_N)(w)\big]^T(U_N)_1^T D_N^{-1}(U_N)_1\big[df^{\text{nor}}_{N,S}(z_N)(w)\big] \leq \chi^2_{l_N}(\alpha) \tag{2.5}$$

defines an unbounded set whenever l_N is strictly less than n. With the linear independence between the rows of $(U_N)_1$ and $(U_N)_2$, the second constraint

$$\|\sqrt{N}(U_N)_2 df^{\text{nor}}_{N,S}(z_N)(w)\|_\infty \leq \epsilon \tag{2.6}$$

complements the first constraint to yield a bounded feasible region and therefore a guaranteed finite optimal solution to (2.4). In the following proposition we see that $w^\epsilon_{N,j}$ depends on ϵ as an affine function whose slope and intercept are determined by l_N.

Proposition 1. *Suppose that $df^{nor}_{N,S}(z_N)$ is a linear homeomorphism and Σ_N has decomposition $\Sigma_N = U_N^T \Delta_N U_N$, where U_N is an orthogonal matrix with rows $u_{N,1}, \ldots, u_{N,n}$ and Δ_N is a diagonal matrix with elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. For a choice of l_N with $\lambda_{l_N} > 0$, let D_N be a diagonal matrix with elements $\lambda_1, \ldots, \lambda_{l_N}$,*

$$(U_N)_1 = \begin{bmatrix} u_{N,1} \\ \vdots \\ u_{N,l_N} \end{bmatrix} \quad \textit{and} \quad (U_N)_2 = \begin{bmatrix} u_{N,l_N+1} \\ \vdots \\ u_{N,n} \end{bmatrix}.$$

Then for each $j = 1, \ldots, n$, the optimal value of (2.4) is an affine function of ϵ with

$$w^\epsilon_{N,j} = \sqrt{\frac{\chi^2_{l_N}(\alpha)\sum_{i=1}^{l_N}(c_{N,j}u^T_{N,i})^2\lambda_i}{N}} + \frac{\epsilon}{\sqrt{N}}\sum_{i=l_N+1}^{n}|c_{N,j}u^T_{N,i}|, \tag{2.7}$$

where $c_{N,j}$ is the jth row of $df^{nor}_{N,S}(z_N)^{-1}$.

Proof. Let V_N and T_N be the subspaces spanned by $\left\{u^T_{N,1}, \ldots, u^T_{N,l_N}\right\}$ and $\left\{u^T_{N,l_N+1}, \ldots, u^T_{N,n}\right\}$, respectively. Then V_N is the orthogonal complement of T_N and any

vector $d(f_N)_S(z_N)(w)$ can be decomposed as

$$df^{\text{nor}}_{N,S}(z_N)(w) = v + t$$

with $v \in V_N$ and $t \in T_N$. Denoting the jth row of $df^{\text{nor}}_{N,S}(z_N)^{-1}$ by $c_{N,j}$, (2.4) can be reformulated as

$$\begin{aligned} \text{maximize} \quad & c_{N,j}v + c_{N,j}t \\ \text{subject to} \quad & N[v^T(U_N)_1^T D_N^{-1}(U_N)_1 v] \leq \chi^2_{l_N}(\alpha) \\ & \|\sqrt{N}(U_N)_2 t\|_\infty \leq \epsilon \\ & v \in V_N,\ t \in T_N. \end{aligned} \tag{2.8}$$

By expressing $v \in V_N$ as $v = \sum_{i=1}^{l_N} s_i u_{N,i}^T$ and $t \in W_N$ as $t = \sum_{i=l_N+1}^{n} s_i u_{N,i}^T$ for $s_i \in \mathbb{R}$ we can separate (2.8) into the following two problems

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{l_N} (c_{N,j}u_{N,i}^T)s_i \\ \text{subject to} \quad & N\sum_{i=1}^{l_N} s_i^2 \lambda_i^{-1} \leq \chi^2_{l_N}(\alpha) \end{aligned} \tag{2.9}$$

and

$$\begin{aligned} \text{maximize} \quad & \sum_{i=l_N+1}^{n} (c_{N,j}u_{N,i}^T)s_i \\ \text{subject to} \quad & \frac{-\epsilon}{\sqrt{N}} \leq s_i \leq \frac{\epsilon}{\sqrt{N}} \ \ i = l_N + 1, \ldots, n. \end{aligned} \tag{2.10}$$

It immediately follows that (2.10) has optimal value $\frac{\epsilon}{\sqrt{N}} \sum_{i=l_N+1}^{n} |c_{N,j}u_{N,j}^T|$, and it can be easily checked using KKT conditions for (2.9) that it has optimal value $N^{-1/2}\sqrt{\chi^2_{l_N}(\alpha)\sum_{i=1}^{l_N}(c_{N,j}u_{N,i}^T)^2\lambda_i}$, proving the result. □

From (2.7) we observe that l_N determines the upper index of the summation in the intercept of $w^\epsilon_{N,j}$, the degrees of freedom of the χ^2 random variable in the intercept, and the lower index of the summation in the slope of $w^\epsilon_{N,j}$. Therefore increasing l_N from k to $k+1$ increases $w^\epsilon_{N,j}$ for values of ϵ below some threshold and decreases $w^\epsilon_{N,j}$ for values of ϵ above this threshold. It is possible for the width of the confidence interval to be constant with respect to ϵ for some components j. This occurs only when $(df^{\text{nor}}_{N,S}(z_N)^{-1})_j$ is a linear

combination of the rows of $(U_N)_1$, in which case increasing l_N from k to $k+1$ only increases the value of $w^{\epsilon}_{N,j}$. In the next section we use the expressions for $R_{N,\epsilon}$ and $w^{\epsilon}_{N,j}$ to investigate the sensitivity of the confidence regions and simultaneous confidence regions to the choices of ϵ and l_N.

2.2 Application to a stochastic Cournot-Nash equilibrium problem

In this section, we consider a stochastic equilibrium model of the European natural gas market, compute confidence intervals for the true solution of this model, and examine the sensitivity of the confidence regions and confidence intervals to the choice of l_N and ϵ. The model is adapted from (Gürkan et al., 1999), and is an example of a Cournot-Nash equilibrium problem.

In a Cournot-Nash equilibrium problem, m competitive players are assumed to produce a homogenous product and must simultaneously decide their level of production and how to distribute their production between n markets. In each of the markets, the price the product sells for is a function of the total quantity allocated to that market by all of the players. The uncertainty in the model arises from the dependence of each player's profit function, denoted by Υ_i , on a random vector $\xi \in \mathbb{R}^b$.

Let x_i denote the decision vector of player i, $S_i \subset \mathbb{R}^{d_i}$ denote the set of feasible decisions for player i, and $x = (x_1, \ldots, x_m) \in S_1 \times \cdots \times S_m$ be the concatenation of all players' decisions. With $\phi_{i_0}(x) = E\left[\Upsilon_i(x, \xi)\right]$ denoting the expected profit function for player i, $x^* = (x_1^*, \ldots, x_m^*)$ is a Cournot-Nash equilibrium if

$$x_i^* \in \text{argmax}_{x_i \in S_i} \phi_{i_0}(x_1^*, \ldots, x_{i-i}^*, x_i, x_{i+1}^*, \ldots, x_m^*) \text{ for each } i = 1, \ldots, m.$$

When the expected profit functions are continuously differentiable, a necessary condition for a point to be a Cournot-Nash equilibrium can be expressed as a variational inequality. In the example considered in this chapter, $S_i = \mathbb{R}^{d_i}_+$ for each $i = 1, \cdots, m$, and the first order necessary condition for player i's profit maximization problem is

$$0 \in -\frac{\partial \phi_{i_0}}{\partial x_i}(x) + \mathbf{N}_{\mathbb{R}^{d_i}_+}(x_i).$$

Let $S = \mathbb{R}^{d_1}_+ \times \cdots \times \mathbb{R}^{d_m}_+$ and

$$f_0(x) = \begin{pmatrix} -\frac{\partial \phi_{1_0}}{\partial x_1}(x) \\ \vdots \\ -\frac{\partial \phi_{m_0}}{\partial x_m}(x) \end{pmatrix}.$$

A necessary condition for x^* to be a Cournot-Nash equilibrium is

$$0 \in f_0(x^*) + \mathbf{N}_S(x^*). \tag{2.11}$$

The above condition is sufficient when each of the expected profit functions is concave. For (2.11) to fit the framework of an SVI we require a function $F(x, \xi)$ such that $f_0(x) = E\left[F(x, \xi)\right]$ and $E\|F(x, \xi)\| < \infty$ for all $x \in S$. The natural candidate

$$F(x, \xi) = \begin{pmatrix} -\frac{\partial \Upsilon_1}{\partial x_1}(x, \xi) \\ \vdots \\ -\frac{\partial \Upsilon_m}{\partial x_m}(x, \xi) \end{pmatrix}$$

will meet these criteria if the profit functions Υ_i satisfy the conditions of Assumption 1. In this case the SVI (2.11) gives rise to the SAA problem

$$0 \in f_N(x) + \mathbf{N}_S(x) \tag{2.12}$$

where

$$f_N = N^{-1} \sum_{k=1}^{N} F(x, \xi^k).$$

In the European gas market model that we consider, there are four players, indexed by $i = 1, 2, 3, 4$. These four players represent the gas producing countries Russia, the Netherlands, Norway, and Algeria. There are six European markets, indexed by j, which represent markets of the United Kingdom, the Netherlands, Italy, France, France and Germany (FRGer), and Belgium and Luxembourg (BelLux). Producers decide on the quantity of gas to ship each year during time period t, for $t = 1, 2, 3, 4$, to the six markets. There are 24 decision variables for each producer, denoted by $x^t_{i,j}$, corresponding to the amount of natural gas shipped by producer i to market j each year in time period t. In the model's

formulation, the following parameters are used :

D_j^t: the domestic gas production of market j each year in time period t,

c_i^t: the constant marginal transportation cost of shipping for producer i in time period t,

e_j^t: the price elasticity of demand for natural gas in market j in time period t,

y_t: the number of years in time period t, taken to be 5 years for time periods 1, 2, 3, and 20 years for time period 4.

In time period t, the yearly production cost for producer i is given by

$$G_i(x) = a_i - b_i \ln(X_i - \sum_{j=1}^{6} x_{i,j}^t),$$

where a_i, b_i and X_i are parameters. The parameter X_i provides an upper bound on the yearly production of producer i. Values for the parameters indexed by player i are given in Table 2.1 and values for the parameters indexed by market j are given in Table 2.2.

Table 2.1: Producer parameter values

Producer	a	b	X	c^1	c^2	c^3	c^4
Russia	1.606	51	80	.58	.56	.55	.55
Netherlands	1.212	67	80	.14	.13	.13	.12
Norway	1.507	85	80	.35	.34	.34	.33
Algeria	2.102	96	80	.70	.69	.64	.62

Table 2.2: Values for price elasticity e and demand D

Market	Period 1		Period 2		Period 3		Period 4	
BelLux	-1.07	0.00	-1.26	0.00	-1.34	0.00	-1.42	0.00
FRGer	-1.46	13.70	-1.58	13.80	-1.68	13.80	-1.79	13.80
France	-.81	4.80	-1.19	2.90	-1.57	3.00	-2.01	3.00
Italy	-1.15	10.40	-1.36	10.00	-1.45	10.00	-1.54	10.40
Netherlands	-.94	22.93	-1.13	20.96	-1.29	24.11	-1.45	23.90
UK	-.61	33.70	-.87	35.00	-1.10	37.00	-1.30	38.00

The uncertainty in the problem is associated with the price of natural gas in the different markets. The price of natural gas in market j for time period t is determined by the total amount of natural gas available annually, as well as ξ^t the random price of oil in time period t, and is given by

$$P_j^t(x, \xi^t) = p_j^t(\xi^t) \left(\frac{Q_j^t(x)}{q_j^t(\xi^t)} \right)^{1/e_j^t}$$

In the above equation, $Q_j^t(x) = D_j^t + \sum_{i=1}^{6} x_{i,j}^t$ is the total amount of natural gas available in market j annually throughout time period t. The functions $p_j^t(\xi^t)$ and $q_j^t(\xi^t)$ provide the base price and the base demand for natural gas as a function of the price of oil, and are defined as

$$p_j^t(\xi^t) = p0_j^t\left(\xi^t/or_t\right) \quad \text{and} \quad q_j^t(\xi^t) = q0_j^t\left(\xi^t/or_t\right)^{\eta_t}$$

with parameters:

$p0_j^t$: reference price of natural gas in market j in time period t,

$q0_j^t$: reference demand for natural gas in market j in time period t,

or_t: reference price for oil in time period t,

η_t: the elasticity relating the relative demand for natural gas to the relative price of oil.

We assume that the prices of oil in each time period are independent and uniformly distributed with lower and upper bounds L_t and U_t. The values for the parameters in the base price and demand functions are given in Tables 2.3 and 2.4.

Table 2.3: Reference prices $p0$ and demands $q0$

Market	Period 1		Period 2		Period 3		Period 4	
BelLux	5.12	7.8	2.56	9.4	3.41	9.4	5.12	9.5
FRGer	5.27	40.7	2.64	46.2	3.52	46.5	5.27	44.6
France	5.25	23.6	2.62	28.3	3.50	9.8	5.25	28.5
Italy	5.15	25.3	2.57	34.9	3.43	37.5	5.15	37.2
Netherlands	5.16	28.9	2.58	29.9	3.44	32.2	5.16	29.7
UK	4.54	43.8	2.27	50.3	3.03	56.4	4.54	53.7

Table 2.4: Time period parameters in base price demand function

t	η_t	or_t	L_t	U_t
1	-0.10	30	16	34
2	-0.12	15	12	18
3	-0.24	30	24	36
4	-0.36	35	28	42

To account for the multiple time periods of the model, all income and costs are considered in terms of their present value. Assuming a fixed annual interest rate of $r = 0.1$, for

each time period we use the factor f_t to express the future value of money with

$$f_t = \left(\frac{(1+r)^{y_t}-1}{r}\right)\left(\frac{1}{(1+r)^{\sum_{s=1}^{t} y_s}}\right).$$

The net present value profit function for producer i is then defined to be

$$\Upsilon_i(x,\xi) = \sum_{t=1}^{4} f_t \left[\sum_{j=1}^{6} \left(P_j^t(x,\xi^t) - c_i^t\right) x_{i,j}^t - G_i(x)\right]. \tag{2.13}$$

Taking the expectation of (2.13) reduces to calculating $E\left[P_j^t(x,\xi^t)\right]$ and provides us with an expression for ϕ_{i_0}. Under the assumption that the oil prices are uniformly distributed we have

$$E\left[P_j^t(x,\xi^t)\right] = p0_j^t \left(Q_j^t(x) q0_j^t\right)^{1/e_j^t} or_t^{\eta_t/e_j^t - 1} \left(U_t^{2-\eta_t/e_j^t} - L_t^{2-\eta_t/e_j^t}\right) \frac{1}{(U_t - L_t)(2-\eta_t/e_j^t)}.$$

With expressions for Υ_i and ϕ_{i_0} we are able to obtain explicit formulas for both $f_0(x)$ and $f_N(x)$.

To find solutions to both the true SVI (2.11) and its SAA (2.12) we make use of the fact that $S = \mathbb{R}_+^{96}$. For any $x \in S$, the normal cone to S at x is

$$\mathbf{N}_S(x) = \left\{v \in \mathbb{R}^{96} |\ v_i = 0 \text{ if } x_i > 0 \text{ and } v_i \leq 0 \text{ if } x_i = 0\right\}.$$

Therefore, the variational inequalities (2.11) and (2.12) are equivalent to the mixed complementarity problems (MCPs)

$$0 \leq x \perp f_0(x) \geq 0 \quad \text{and} \quad 0 \leq x \perp f_N(x) \geq 0$$

respectively. To solve the MCPs, we use the PATH solver (Dirkse and Ferris, 1995b) implemented in GAMS (Rosenthal, 2012). With this knowledge of the true solution, we observe that $df_{0,S}^{\text{nor}}(z_0)$ is a linear function and Σ_0 is degenerate. To calculate the confidence regions $R_{N,\epsilon}$ and formulate the problems in (2.1) to find the simultaneous confidence intervals, re-

quires evaluating $f_N(x,\xi)$, $df_N(x,\xi)$, and the B-derivative of the projection onto $S = \mathbb{R}^{96}_+$ at a point z which is equal to

$$d\Pi_S(z)(h) = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{96} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_{96} \end{bmatrix} \quad \text{where } \lambda_i = \begin{cases} 1 & (z)_i > 0, \\ 1 & (z)_i = 0 \text{ and } h_i \geq 0, \\ 0 & (z)_i = 0 \text{ and } h_i \leq 0, \\ 0 & (z)_i < 0. \end{cases}$$

The calculation of confidence regions and simultaneous confidence intervals is done in (MATLAB, 2010) using the MATLAB/GAMS interface (Ferris, 2005) to pass the SAA and true solutions between programs.

To analyze the performance of the confidence regions and corresponding simultaneous confidence intervals, we generate 2,000 replications of the SAA problem at each sample size of N =20, 200, 2,000, and 20,000. For each sample, $df^{\text{nor}}_{N,S}(z_N)$ is linear and the simultaneous confidence intervals take the form of (2.3). To determine l_N, all eigenvalues of Σ_N larger than a threshold ρ_N are treated as nonzero. Three different procedures are considered for choosing the threshold ρ_N. In the first, $\rho_{N,1} = N^{-1/3}$, while in the second and third ρ_N is held constant at $\rho_{N,2} = 10^{-10}$ and $\rho_{N,3} = 0.001$ respectively. Note that the choice of $\rho_{N,1}$ will be asymptotically correct if Σ_N converges to Σ_0 in probability at an exponential rate. This would occur if in Theorem 2 we replace Assumption 1 with Assumption 4. The use of $\rho_{N,1}$ results in four eigenvalues being treated as nonzero across all samples, while the constant thresholds results in values of l_N that vary slightly between samples. When $\rho_{N,2}$ is used l_N equals either eight or nine and when $\rho_{N,3}$ is used l_N equals either four or five. In Table 2.5, we summarize the coverage rates of z_0 by the confidence regions $R_{N,\epsilon}$ for each choice of ρ_N and values of ϵ =0.0001, 0.1, 1, and ∞. For example, with the choices of $\rho_{N,1}$ and $\epsilon = 0.1$, the true value of z_0 is covered by 83.3% of the 95% confidence regions computed from the 2,000 replications at $N = 20$.

For all three methods of determining l_N, we observe extremely poor coverage of z_0 for $\epsilon \leq .0001$, even at large sample sizes. The sensitivity of the confidence regions to the

Table 2.5: Coverage rates of confidence regions for z_0, $\alpha = .05$

		$N = 20$	$N = 200$	$N = 2,000$	$N = 20,000$
$\rho_{N,1}$	$\epsilon = .0001$	0%	0%	0%	0%
	$\epsilon = .1$	57%	78.8 %	95.1%	94.05%
	$\epsilon = 1$	83.3%	94.4 %	95.35%	94.2%
	$\epsilon = \infty$	85.15%	94.4 %	95.35%	94.2%
$\rho_{N,2}$	$\epsilon = .0001$	0%	0.2%	0.75%	7.6%
	$\epsilon = .1$	2.25%	24.75 %	49.05%	73.85%
	$\epsilon = 1$	2.25%	24.75 %	49.05%	73.85%
	$\epsilon = \infty$	2.25%	24.75 %	49.05%	73.85%
$\rho_{N,3}$	$\epsilon = .0001$	0%	0%	0%	0%
	$\epsilon = .1$	56.7%	78.05 %	94.75%	94.5%
	$\epsilon = 1$	81.25%	93.55 %	94.75%	94.5%
	$\epsilon = \infty$	83%	93.55 %	94.75%	94.5%

choice of l_N is seen in the different coverage rates of z_0 for values of $\epsilon \geq 0.1$. In the liberal classification scheme that uses $\rho_{N,2}$, near zero eigenvalues are included in D_N. When D_N is inverted the reciprocals of these near zero eigenvalues offset the increase in the degrees of freedom of the χ^2 random variable on the right hand side of (2.5), resulting in poor coverage of z_0. The classification schemes that use $\rho_{N,1}$ and $\rho_{N,3}$ have a higher threshold for treating eigenvalues as nonzero. As a result these thresholds avoid the inclusion of overly large values in D_N^{-1} and produce regions that perform largely in line with the specified level of confidence. The choice of $\epsilon = \infty$ corresponds to the percentage of samples that satisfy (2.5) and provides an upper bound on the coverage rates.

Next, as we examine the performance of simultaneous confidence intervals we observe that their coverage rates keep increasing as ϵ increases, and eventually reach 100% for ϵ sufficiently large, see Table 2.6. This is consistent with the analytical results in Proposition 1, since for this example each sample and choice or ρ_N results in $w_{N,j}^{\epsilon}$ being an affine function of ϵ with positive slope. In contrast, while the size of the confidence regions also increases with ϵ their coverage of z_0 never reaches 100% due to the constraint (2.5). A further difference between the confidence regions and simultaneous confidence intervals, is the coverage rates of z_0 at small values of ϵ. While the confidence regions largely fail to cover z_0 for values of $\epsilon \leq .0001$, the simultaneous confidence intervals not only cover z_0,

Table 2.6: Coverage rates of simultaneous confidence intervals for z_0, $\alpha = .05$

		$N = 20$	$N = 200$	$N = 2,000$	$N = 20,000$
$\rho_{N,1}$	$\epsilon = 0$	84.05%	85.35%	99.05%	98.9%
	$\epsilon = .01$	88.05%	88.15%	99.55%	99.5%
	$\epsilon = .1$	94.05%	99.35%	99.9%	100%
	$\epsilon = 1$	100%	100%	100%	100%
$\rho_{N,2}$	$\epsilon = 0$	92%	89.9%	99.95%	100%
	$\epsilon = .01$	93.35%	90.15%	99.95%	100%
	$\epsilon = .1$	94.4%	100%	100%	100%
	$\epsilon = 1$	100%	100%	100%	100%
$\rho_{N,3}$	$\epsilon = 0$	84.9%	86.95%	99.55%	99.65%
	$\epsilon = .01$	88.6%	88.9%	99.65%	99.9%
	$\epsilon = .1$	94.05%	99.45%	100%	100%
	$\epsilon = 1$	100%	100%	100%	100%

but for the larger sample sizes do so at a conservative rate. The conservative performance at small values of ϵ is most obvious with the choice of $\rho_{N,2}$. As noted after Proposition 1 treating more eigenvalues as nonzero increases the intercept term of (2.7), which increases the interval's length for ϵ sufficiently small.

Next, we examine the computation of individual confidence intervals, and compare them with the simultaneous confidence intervals. In this example $df^{\text{nor}}_{0,S}(z_0)$ is linear and $df^{\text{nor}}_{0,S}(z_0)^{-1}\Sigma_0 df^{\text{nor}}_{0,S}(z_0)^{-T}$ has nonzero diagonal elements. Therefore, by Theorem 3, the formula (1.21) will provide asymptotically exact intervals for this example. Using this formula we consider individual confidence intervals at both $\alpha = .05$ and with a Bonferroni adjustment of $\alpha' = \frac{.05}{96}$. Below, we refer to intervals produced using the Bonferroni adjustment as adjusted confidence intervals, and will examine their performance as simultaneous confidence intervals.

The individual confidence intervals with $\alpha = .05$ perform largely in line with expectations. At the sample size of $N = 20$, coverage rates of the different components $(z_0)_i$ range from 71.1% to 96.4% with an overall average of 93.1%. For the samples of size $N = 20,000$, the coverage rates range from 94.05% to 96.2% with an overall average of 95.09%. For the adjusted confidence intervals we examine their rates of jointly covering z_0. At the sample sizes of N =20, 200, 2,000 and 20,000, the coverage rates of the adjusted confidence inter-

vals are 88.2%, 88.4%, 99.75%, and 99.75%, respectively. These rates are comparable to the coverage rates of the simultaneous confidence intervals calculated using (2.4) for small values of ϵ as given in Table 2.6.

To observe differences between the adjusted and simultaneous confidence intervals we compare their interval lengths. Table 2.7 summarizes the half widths of the individual, adjusted, and simultaneous confidence intervals for $(z_0)_{59}$ for a single replication at each sample size. Half widths of the individual and adjusted confidence intervals do not depend on ρ_N. However, in Table 2.7 their values are repeated for each choice of ρ_N, to be compared with the corresponding simultaneous confidence intervals. With the choice of $\rho_{N,2}$, even the

Table 2.7: Half-widths of intervals for $(z_0)_{59}$, $\alpha = .05$

				h^{ϵ}_{59}			
		Individual	Adjusted	$\epsilon = 0$	$\epsilon = 0.01$	$\epsilon = 0.1$	$\epsilon = 1$
$\rho_{N,1}$	$N = 20$	0.2879	0.5098	0.4525	0.5213	1.1406	7.3331
	$N = 200$	0.0717	0.1270	0.1127	0.1360	0.3458	2.4435
	$N = 2,000$	0.0224	0.0396	0.0352	0.0420	0.1035	0.7184
	$N = 20,000$	0.0070	0.0124	.0110	0.0133	0.0347	0.2483
$\rho_{N,2}$	$N = 20$	0.2879	0.5098	0.6043	0.6713	1.2742	7.3028
	$N = 200$	0.0717	0.1270	0.1505	0.1731	0.3759	2.4043
	$N = 2,000$	0.0224	0.0396	0.0470	0.0535	0.1124	0.7018
	$N = 20,000$	0.0070	0.0124	.01476	0.0170	0.0378	0.2458
$\rho_{N,3}$	$N = 20$	0.2879	0.5098	0.4888	0.5573	1.1741	7.3416
	$N = 200$	0.0717	0.1270	0.1127	0.1360	0.3458	2.4435
	$N = 2,000$	0.0224	0.0396	0.0380	0.0448	0.1061	0.7190
	$N = 20,000$	0.0070	0.0124	.0118	0.0142	0.0356	0.2492

smallest simultaneous confidence interval with $\epsilon = 0$ contains the adjusted confidence interval at each sample size. This is indeed the case across all components and samples. The simultaneous confidence intervals calculated using $\rho_{N,2}$ therefore contain the conservative Bonferroni adjusted simultaneous confidence intervals, which illustrates the overly conservative interval lengths obtained when using $\rho_{N,2}$. Choosing either $\rho_{N,1}$ or $\rho_{N,3}$ changes this effect, and the adjusted confidence intervals contain the simultaneous confidence intervals with $\epsilon = 0$ across all components and samples. Using (2.7), we calculate the value of ϵ for which the jth components of the simultaneous and adjusted confidence intervals equal one another. This value varies largely depending on the component considered. When $\rho_{N,1}$ is

used, this value of ϵ is between 8.86×10^{-4} and 0.1728, and between 3.57×10^{-4} and 0.1395 when using $\rho_{N,3}$.

As noted after Proposition 1, the choice of l_N determines the degrees of freedom of the χ^2 random variable, as well as the upper index of summation in the intercept, and the lower index of summation in the slope of $w^{\epsilon}_{N,j}$. When comparing interval lengths for different choices of ρ_N and $\epsilon = 0$, the differences are largely the result of changes in the degrees of freedom of the χ^2 random variable. This is seen by comparing the ratio of $w^0_{N,j}$ for two choices of ρ_N to the ratio of the square root of $\chi^2_{l_N}$ for the same choices of ρ_N. The difference between these two ratios is on the order of 10^{-4} across all components and samples.

So far, we have considered only confidence regions and intervals for z_0, the true solution to the normal map formulation. In most problems, the true solution to the variational inequality, namely x_0, has a more direct interpretation and is of greater interest. The relation $\Pi_S(z_0) = x_0$ and the easily observed fact that

$$\Pr\left(z_0 \in I_N(\omega)\right) \leq \Pr\left(\Pi_S(z_0) \in \Pi_S(I_N(\omega))\right), \text{ for any random set } I_N(\omega),$$

provides one indirect approach for obtaining confidence intervals for x_0 that cover the true solution with a rate that is at least as large as the coverage rate of z_0 by $I_N(\omega)$. With $S = \mathbb{R}^{96}_+$, projecting the simultaneous confidence intervals for z_0 onto S reduces to replacing negative endpoints of these intervals with zero. Comparing the coverage rates of x_0, as summarized in Table 2.8, to the coverage rates of z_0, we observe the largest increase for the smaller sample sizes and values of ϵ.

The expression for $w^{\epsilon}_{N,j}$ in (2.7) and the analysis of this example provides useful insights for choosing ϵ and l_N when the sample covariance matrix is singular. When choosing l_N care should be taken to avoid classifying overly small eigenvalues as nonzero. In the case of confidence regions, such care can prevent poor coverage performance due to large elements of D_N^{-1} offsetting the increases to the right hand side of (2.5). For simultaneous confidence intervals, too large a value of l_N is undesirable since it inflates the intercept of $w^{\epsilon}_{N,j}$ and produces excessively long intervals. The choice of ϵ depends on the specific set of interest.

Table 2.8: Coverage rates of simultaneous confidence intervals for x_0, $\alpha = .05$

		$N = 20$	$N = 200$	$N = 2,000$	$N = 20,000$
$\rho_{N,1}$	$\epsilon = 0$	94.75%	99.9%	99.65%	99.65%
	$\epsilon = .01$	94.9%	99.9%	99.95%	99.8%
	$\epsilon = .1$	95.2%	100%	100%	100%
	$\epsilon = 1$	100%	100%	100%	100%
$\rho_{N,2}$	$\epsilon = 0$	95.2%	100%	99.95%	100%
	$\epsilon = .01$	95.2%	100 %	99.95%	100%
	$\epsilon = .1$	95.2%	100 %	100%	100%
	$\epsilon = 1$	100%	100 %	100%	100%
$\rho_{N,3}$	$\epsilon = 0$	94.8%	99.9%	99.9%	99.85%
	$\epsilon = .01$	94.95%	99.95 %	99.95%	99.9%
	$\epsilon = .1$	95.2%	100 %	100%	100%
	$\epsilon = 1$	100%	100 %	100%	100%

When the confidence regions are the primary set of interest, small values of ϵ often lead to poor coverage performance, and there is an upper bound on the coverage rate as ϵ goes to infinity. These properties suggest choosing a larger value of ϵ to obtain the desired level of coverage by the confidence regions. When the confidence regions are to be used to build simultaneous confidence intervals for z_0 or x_0, a small value of ϵ, even the extreme choice of $\epsilon = 0$, appears appropriate. This is based on the expression for $w_{N,j}^{\epsilon}$ in (2.7) and the conservative performance of the simultaneous confidence intervals demonstrated in this numerical example.

CHAPTER 3

Confidence intervals for the normal map solution

3.1 Introduction

This chapter presents two new methods for constructing individual confidence intervals for the normal map formulation of an SVI. For both methods, a level of confidence can be specified under general situations. While our main interest is on SVIs and their normal map formulations, the ideas of those two methods work for general piecewise linear functions. We outline the ideas below, and leave formal definitions and proofs to §3.2 and §3.3. Recalling the notation introduced in Chapter 1, we use $(v)_j$ to denote the jth coordinate of a vector v, and $(M)_j$ to denote the jth row of a matrix M. Similarly for an invertible function $f : \mathbb{R}^n \to \mathbb{R}^n$, $(f)_j$ will denote the jth component function of f and $(f^{-1})_j$ the jth component function of f^{-1}.

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 $\Pr(|(f^{-1})_j(Z)| \leq a)$ 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 chapter, z_0 and z_N are solutions to the normal map formulations of (1.1) and (1.3) respectively, and the unknown function f is substituted by an estimator obtained from approaches in (Lu, 2012) and (Lu and Budhiraja, 2013). Such a substitution does not affect the asymptotic exactness of confidence intervals computed from this method, as we show in Theorem 5. In addition,

to allow for some choice in where the interval is centered, we introduce a parameter r and consider the probability $\Pr(|(f^{-1})_j(Z) - r| \leq a)$.

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 a satisfying $\Pr(|(f^{-1})_j(Z) - r| \leq a)$ for a fixed r requires enumerating all pieces of f^{-1}, and for each 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 §3.4, and the development of the second method in this chapter.

The second method uses the idea of conditioning. For any point $x \in \text{int}K_i$ there exists a number $\eta_j^\alpha(f, x)$ such that the following conditional probability

$$\frac{\Pr\left(|(f^{-1})_j(Z)| \leq \eta_j^\alpha(f,x),\ f^{-1}(Z) \in K_i\right)}{\Pr\left(f^{-1}(Z) \in K_i\right)}$$

equals $1 - \alpha$. If we choose a point x to be contained in the same cone K_i that contains $z_N - z_0$, the interval

$$[(z_N)_j - \eta_j^\alpha(f,x)N^{-1/2}, (z_N)_j + \eta_j^\alpha(f,x)N^{-1/2}]$$

will have a level of confidence equal to $1 - \alpha$. In situations considered in this chapter, we will again use an estimator to replace the unknown f, and follow an approach in (Lu, 2012) to choose x. The method is justified with a convergence result in Theorem 6. The second method avoids the enumeration of all pieces f by conditioning on the cone that contains $z_N - z_0$. The ability to work with a single piece of f provides the second method with a dramatic computational advantage over the first method, and makes it possible to apply the second method to problems with a large number of selection functions. In the third numerical example of §3.5, the number of selection functions we need to handle for some SAA problems is 2^{12}. While the first method failed in those cases, the second method was able to finish the computation very quickly.

3.2 The first method

This section presents the first method to compute individual confidence intervals. This method differs from the approach examined in Theorem 3 in how it estimates the transformation $df_{0,S}^{\mathrm{nor}}(z_0)^{-1}$. The estimate $df_{N,S}^{\mathrm{nor}}(z_N)^{-1}$ used for the approach examined in Theorem 3 has the benefit that it is with high probability a linear function and the interval will have a closed form expression. The limitation of this approach is that it does not account for how the location $z_N - z_0$ in the conical subdivision associated with $df_{0,S}^{\mathrm{nor}}(z_0)$ affects the form of $d\Pi_S(z_N)$ and thus $df_{N,S}^{\mathrm{nor}}(z_N)$. Therefore as seen in (1.21), when $df_{0,S}^{\mathrm{nor}}(z_0)$ is piecewise linear the intervals produced using the linear estimate $df_{N,S}^{\mathrm{nor}}(z_N)$ may have an asymptotic level of confidence different than that indicated by the choice of α.

To guarantee the asymptotic exactness of intervals with less restrictive assumptions than those necessary in Theorem 3, the method proposed in this section uses $\Phi_N(z_N)$ as in (1.28) to estimate $df_{0,S}^{\mathrm{nor}}(z_0)$. The convergence of $\Phi_N(z_N)$ to $df_{0,S}^{\mathrm{nor}}(z_0)$, see Theorem 4 (1.31), allows us to directly account for the effect that $df_{0,S}^{\mathrm{nor}}(z_0)$ being piecewise linear has on the intervals' performance. The cost of using this approach is that when $\Phi_N(z_N)$ is piecewise linear we no longer have a closed form expression for the intervals and the computational costs of determining an interval's width increases with the number of selection functions. The width of an interval produced using the method of this section is determined by (3.1) with the exactness of the intervals proven in Theorem 5, see (3.6), the proof of which uses properties of transformations of normal random vectors.

To begin let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and $Z \sim \mathcal{N}(0, I_n)$. Suppose that $\Pr(f(Z) = b) = 0$ for all b and $\Pr(b_1 < f(Z) < b_2) > 0$ for all $b_1 < b_2$. Then given any $\alpha \in (0, 1)$ and $r \in \mathbb{R}$ there exists a unique point $a^r(f) \in (0, \infty)$ such that

$$\Pr\left(-a^r(f) \leq f(Z) - r \leq a^r(f)\right) = 1 - \alpha.$$

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

$$a^r(g) = \inf\{\ell \geq 0 | \Pr\left(-\ell \leq g(Z) - r \leq \ell\right) \geq 1 - \alpha\}. \tag{3.1}$$

It then follows that

1. $a^r(g) < \infty$.

2. $\Pr\left(-a^r(g) \le g(Z) - r \le a^r(g)\right) \ge 1 - \alpha$.

3. $\Pr\left(-(a^r(g) - \delta) \le g(Z) - r \le a^r(g) - \delta\right) < 1 - \alpha$ for all $\delta > 0$.

In the proof of Theorem 5 we use the following two lemmas. Here is a comment about notation. We use f_N to denote the sample average function (1.2) unless explicitly stated otherwise. In some lemmas and propositions we use f_N for different meanings, which will be made clear in the statements of those results. For example, f_N in Lemma 1 stands for a deterministic function from $\mathbb{R}^n$ to $\mathbb{R}$.

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

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

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

$$1 - \alpha \le \lim_{N\to\infty} \Pr\left(-\epsilon \le f_N(Z) - r \le \epsilon\right) = \Pr\left(-\epsilon \le f(Z) - r \le \epsilon\right).$$

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

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

$$1 - \alpha \le \lim_{N\to\infty} \Pr\left(-1 \le \frac{f_N(Z) - r}{a^r(f_N)} \le 1\right) = \Pr\left(-1 \le \frac{f(Z) - r}{a^*} \le 1\right). \tag{3.2}$$

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

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

Sending δ to 0 we obtain $\Pr(-a^* \leq f(Z) - r \leq a^*) \leq 1 - \alpha$, which combined with (3.2) gives

$$\Pr(-a^* \leq f(Z) - r \leq a^*) = 1 - \alpha.$$

Thus $a^* = a^r(f)$, and $\lim_{N\to\infty} a^r(f_N) = a^r(f)$. □

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 $\{f_N\}_{N=1}^{\infty}$ be a sequence of $C(\mathbb{R}^n, \mathbb{R})$ valued random variables which converges in distribution to f. 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 $r \in \mathbb{R}$,*

$$\Pr(-a^r(f_N) \leq f_N(Z_N) - r \leq a^r(f_N)) \to 1 - \alpha.$$

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

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

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

$$\begin{aligned}\Pr(-a^r(f_N) \leq f_N(Z_N) - r \leq a^r(f_N)) \quad &= \Pr\left(A_N;\ -1 \leq \frac{f_N(Z_N) - r}{a^r(f_N)} \leq 1\right) \\ &+ \Pr(A_N^c;\ -a^r(f_N) \leq f_N(Z_N) - r \leq a^r(f_N)).\end{aligned}$$

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

$$\Pr(A_N^c;\ -a^r(f_N) \leq f_N(Z_N) - r \leq a^r(f_N)) \to 0 \text{ as } N \to \infty.$$

Let B_N be the event that $-1 \leq \frac{f_n(Z_N)-r}{a^r(f_N)} \mathbb{1}_{a^r(f_N)>0} \leq 1$. By the convergence of f_N to f and Z_N to Z, we have $f_N(Z_N) \Rightarrow f(Z)$, and therefore

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

Consequently, $\Pr\left(-a^r(f_N) \leq f_N(Z_N) - r \leq a^r(f_N)\right) \to 1-\alpha$. □

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

Proposition 2. *(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*

$$\sup_{h \in \mathbb{R}^n, h \neq 0} \frac{\|f_N(h) - f(h)\|}{\|h\|} \to 0. \tag{3.3}$$

Suppose that there exists a conical subdivision $\Gamma = \{K_1, K_2 \dots K_l\}$ of $\mathbb{R}^n$ such that for all N sufficiently large the restrictions of f_N and f on each K_i are represented by matrices $M_{N,i}$ and M_i respectively. Then

$$\sup_{h \in \mathbb{R}^n, h \neq 0} \frac{\|M_{N,i}h - M_i h\|}{\|h\|} \to 0 \text{ for } i = 1, \dots, l. \tag{3.4}$$

(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 (3.3), $\sup_{h \in K_i, h \neq 0} \frac{\|M_{N,i}h - M_i h\|}{\|h\|}$ converges to 0 as $N \to \infty$, for each $i = 1, \dots, l$. As Γ is a conical subdivision of $\mathbb{R}^n$, K_i is of dimension n which means that it contains a ball in $\mathbb{R}^n$. The fact that $\|M_{N,i}h - M_i h\|$ converges to 0 for all h in a ball implies that the matrix $M_{N,i}$ converges to M_i, giving (3.4).

To prove (b) first note that since f is a homeomorphism, M_i^{-1} is well defined for each i and $\left\{M_1^{-1}, M_2^{-1}, \dots, M_l^{-1}\right\}$ provides a family of selection functions for f^{-1} (Scholtes, 2012, Proposition 2.3.2). Moreover we have that f^{-1} is Lipschitz continuous with the constant

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

Similarly for N sufficiently large the functions $f_N - f$ will be piecewise linear with a family of selection functions given by $\{M_{N,1} - M_1, \ldots, M_{N,l} - M_l\}$, and therefore Lipschitz continuous with the constant

$$\rho_N = \max_{1 \le i \le m} (\|M_{N,i} - M_i\|)$$

From part (a) we have $\lim_{N\to\infty} \|M_{N,i} - M_i\| = 0$ for each i, so for all N sufficiently large $\rho_N < \delta^{-1}$. From (Robinson, 1991, Lemma 3.1) it then follows that f_N is a homeomorphism for N sufficiently large.

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

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

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

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

Proposition 3. *Suppose that Assumptions 2, 3 and 4 hold, and for each $N \in \mathbb{N}$ let $\Phi_N(z_N)$ be as in (1.28). Then $\Phi_N(z_N)^{-1}\Sigma_N^{1/2}$ converges to $df_{0,S}^{nor}(z_0)^{-1}\Sigma_0^{1/2}$ 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 (1.30) imply that for all $\epsilon > 0$

$$\lim_{N\to\infty} \Pr\left(\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Sigma_N^{-1/2}\Phi_N(z_N)(h) - \Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0)(h)\|}{\|h\|} < \epsilon \right) = 1. \tag{3.5}$$

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

$$\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Sigma_N^{-1/2}\Phi_N(z_N)(h) - \Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0)(h)\|}{\|h\|} \to 0.$$

We will apply Proposition 2 to show the almost sure convergence of $\Phi_N(z_N)^{-1}\Sigma_N^{1/2}$ to $df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}$. To this end, it suffices to show that the conditions of Proposition 2 are satisfied for a.e. ω, with $\Sigma_N^{-1/2}(\omega)\Phi_N(z_N(\omega))$ and $\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0)$ playing the roles of f_N and f in that proposition.

From the expressions for $df_{0,S}^{\mathrm{nor}}(z_0)$ in (1.11), $\Phi_N(z_N)$ and z_N^* in (1.28), it is clear that the conditions in part (a) of Proposition 2 will be satisfied if we can find a conical subdivision Γ' such that $d\Pi_S(z)|_{K_i}$ is equal to a linear function for every $K_i \in \Gamma'$ and $z \in \mathbb{R}^n$.

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 (1.25). The desired subdivision Γ' can be constructed by taking the collection of all cones with non-empty interior of the form $K = \cap_{j=1}^m K_j$ where each K_j is from a conical subdivision of Ψ_j.

Finally, by Assumptions 2 and 3, $\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0)$ is a homeomorphism, satisfying the condition in part (b) of Proposition 2. The result follows. □

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

Theorem 5. *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 (3.1). Then for every $j = 1, \ldots, n$,*

$$\lim_{N\to\infty} \Pr\left(\left|\sqrt{N}(z_N - z_0)_j - r\right| \leq a^r\left((\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j\right)\right) = 1 - \alpha. \tag{3.6}$$

Proof. By Proposition 3, $(\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j$ converges to $((L_{K_0}^{\mathrm{nor}})^{-1}\Sigma_0^{1/2})_j$, in probability. Since $(L_{K_0}^{\mathrm{nor}})^{-1}\Sigma_0^{1/2}$ is a piecewise linear homeomorphism it follows that for $Z \sim N(0, I_n)$ and each $j = 1, \ldots, n$,

$$\Pr\left(\left((L_{K_0}^{\mathrm{nor}})^{-1}\Sigma_0^{1/2}\right)_j(Z) = b\right) = 0 \text{ for all } b \tag{3.7}$$

and

$$\Pr\left(b_1 < \left((L_{K_0}^{\text{nor}})^{-1}\Sigma_0^{1/2}\right)_j(Z) < b_2\right) > 0 \text{ for all } b_1 < b_2. \tag{3.8}$$

Let $Z_N = \sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N)(z_N - z_0)$; by Theorem 4 (see (1.31)) Z_N converges in distribution to Z. Since

$$\begin{aligned} & (\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j\big(\sqrt{N}\Sigma_N^{-1/2}\Phi_N(z_N - z_0)\big) \\ = \quad & \sqrt{N}(\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j\big(\Sigma_N^{-1/2}\Phi_N(z_N - z_0)\big) \\ = \quad & \sqrt{N}(z_N - z_0)_j, \end{aligned}$$

it follows from an application of Lemma 2 with $(\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j$ and $(L_K^{-1}\Sigma_0^{1/2})_j$ playing the roles and f_N and f that

$$\Pr\left(-a^r\big((\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j\big) \le \sqrt{N}(z_N - z_0)_j - r \le a^r\big((\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j\big)\right)$$

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

It is possible to relax Assumption 3 in the proof of Theorem 5. This would require some minor modifications to the definition of a^r. In particular, a^r would need to depend on two separate arguments, one for the estimate for Σ_0 and another for that of $df_{0,S}^{\text{nor}}(z_0)$. The statements of the supporting results would need to be adjusted accordingly. We can then replace Assumption 3 with conditions that guarantee equations (3.7) and (3.8) to hold. These equations ensure that the limit in (3.6) is well defined.

A limitation of this first approach is that evaluating $a^r\big((\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j\big)$ requires working with each selection function of $\Phi_N(z_N)$, making it computationally intractable when there are a large number of selection functions. This is an issue, since the number of selection functions can grow exponentially with the problem size. Additionally, as we shall see in the third example of §3.5, considering each selection function also makes this approach sensitive to errors in the estimation of $d\Pi_S(z_0)$. These limitations motivate the development of the second method. The second method limits the computational burden of working with a piecewise linear function, by restricting the computation to only a subset of

selection functions indicated by $z_N - z_N^*$. This subset will generally consist of only a single selection function, leading to dramatic computational savings.

3.3 The second method

In this section we propose a second method for the construction of asymptotically exact individual confidence intervals. Like the approach of §3.2 the exactness of the intervals will depend on the use of $\Phi_N(z_N)$ as an estimate for $df_{0,S}^{\mathrm{nor}}(z_0)$. Specifically this approach relies upon the fact that with $\Phi_N(z_N)$ one can accurately estimate both the conical subdivision of $df_{0,S}^{\mathrm{nor}}(z_0)$ and the location of $z_N - z_0$ in the subdivision. The calculation of an interval's width with the second method uses only the selection functions indicated by these estimates, reducing the computational burden of working with $\Phi_N(z_N)$.

As will be discussed further in the main result of this section, Theorem 6, the probability of $z_N - z_0$ being in the interior of a cone in the conical subdivision of $df_{0,S}^{\mathrm{nor}}(z_0)$ approaches one as the sample size goes to infinity. Therefore the method proposed in this section will (with high probability) require working with only a single selection function. This leads to the following comparison to the method of constructing confidence intervals considered in Theorem 3. Recall that the limitation of using $df_{N,S}^{\mathrm{nor}}(z_N)$ as an estimate for $df_{0,S}^{\mathrm{nor}}(z_0)$ when computing a confidence interval is that this approach does not account for a possible dependence between how the function $df_{N,S}^{\mathrm{nor}}(z_N)$ is estimated and how intervals produced using the estimate will perform. The method in this section will with high probability calculate an interval's width using a single linear selection function. Since the domain of this section function is restricted by a cone in the conical subdivision, the dependence between when an estimate is used to calculate an interval and the interval's performance can be accounted for by using the idea of conditioning.

The asymptotic exactness of the intervals proposed in this section is proven in Theorem 6, see (3.11). We begin the discussion of the second method by defining what replaces $a^r(\cdot)$ and determines an interval's width. 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 $Z \sim \mathcal{N}(0, I_n)$. 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

$$\Pr\left(|\left(f^{-1}(Z)\right)_j| \leq \eta_j^\alpha(f,x),\ f^{-1}(Z) \in K_i\right) = (1-\alpha)\Pr\left(f^{-1}(Z) \in K_i\right), \tag{3.9}$$

where $\left(f^{-1}(Z)\right)_j$ stands for the jth component of the random variable $f^{-1}(Z)$. 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 (3.9) as

$$\Pr\left(|\left(M_i^{-1}Z\right)_j| \leq \eta_j^\alpha(f,x),\ M_i^{-1}Z \in K_i\right) = (1-\alpha)\Pr\left(M_i^{-1}Z \in K_i\right). \tag{3.10}$$

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 for deterministic functions will play a similar role in the proof of Theorem 6 as Lemma 1 did in the proof of Theorem 5.

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

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

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

Proof. From Proposition 2 it follows that f_N will be a homeomorphism for all N sufficiently large. The convergence of $\eta_j^\alpha(f_N,x)$ 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 6 below, we make use of the notation introduced before Theorem 3. With this notation $\Gamma'(z_0) = \{K_1, \ldots, K_l\}$ is the conical subdivision associated with $df_{0,S}^{\text{nor}}(z_0)$ such that $df_{0,S}^{\text{nor}}(z_0)|_{K_i} = M_i$ and $K_i = \text{cone}(P_i - z_0)$ where $P_1, \ldots, P_l$ are all n-cells in the normal manifold of S that contain z_0. As before, we write $Y_0 = \Sigma_0^{1/2}Z$ and

$Y^i = M_i^{-1}\Sigma_0^{1/2}Z$ for $i = 1, \ldots, l$. Finally we define $Y^* = df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z$, and note that $Y^*\mathbb{1}_{Y^*\in K_i} = Y^i\mathbb{1}_{Y^i\in K_i}$.

Theorem 6. *Let Assumptions 2, 3 and 4 hold, and let $\Phi_N(z_N)(\cdot)$ and z_N^* be as defined in (1.28). For all $j = 1, \ldots, n$ and $\alpha \in (0, 1)$,*

$$\Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right) \to 1 - \alpha. \tag{3.11}$$

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

$$A_N = \left\{\omega \middle| \left\{i|d_i(z_N(\omega)) \le 1/g(N)\right\} = \left\{i|z_0 \in C_i\right\}\right\}. \tag{3.12}$$

By the remarks below (1.28), 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 $df_{0,S}^{\mathrm{nor}}(z_0)$ and $\Phi_N(z_N(\omega))$ share the conical subdivision $\{K_1, \ldots, K_l\}$. Moreover as shown in (Lu, 2012, Theorem 3.1) $\lim_{N\to\infty}\Pr(A_N) = 1$, so it follows from (3.5)

$$\lim_{N\to\infty}\Pr\left(A_N;\ \sup_{h\in\mathbb{R}^n, h\neq 0}\frac{\|\Sigma_N^{-1/2}\Phi_N(z_N)(h) - \Sigma_0^{-1/2}df_{0,S}^{\mathrm{nor}}(z_0)(h)\|}{\|h\|} < \epsilon\right) = 1. \tag{3.13}$$

Combining this with Lemma 3 it follows that $\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), x)$ converges in probability to $\eta_j^\alpha(\Sigma_0^{-1/2}df_{0,S}^{\mathrm{nor}}(z_0), x)$ for all fixed x.

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

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

where x_i in the last expression 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 $\lim_{N\to\infty}\Pr\left(z_N\in\mathbb{R}^n\backslash\cup_{i=1}^{l}B\cap\text{int}P_i\right)=0$ as shown in (Lu, 2014, 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, $i=1,\cdots,l$, by the additional requirement $z_N\in\text{int}P_i$ one has $z_N-z_N^*\in\text{cone}(\text{int}P_i-z_N^*)$ and the latter set is exactly $\text{cone}(\text{int}P_i-z_0)$, namely $\text{int}K_i$.

When $l=1$, z_0 is contained in the interior of an n-cell P_1 and $K_1=\mathbb{R}^n$. In this case $Y^*\sim\mathcal{N}\left(0,M_1^{-1}\Sigma_0M_1^{-T}\right)$, and (3.11) follows from the fact

$$\frac{\sqrt{N}(z_N-z_0)_j}{\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N),x_1)}\Rightarrow\frac{(Y^*)_j}{\eta_j^\alpha(\Sigma_0^{-1/2}df_{0,S}^{\text{nor}}(z_0),x_1)}.$$

Next, we consider the case when $l\geq 2$. For all $j=1,\ldots,n$ and $i=1,\ldots,l$ let $\bar{v}^{i,j}\in\mathbb{R}^n$ be such that $\bar{v}^{i,j}\notin K_i$ and $|(\bar{v}^{i,j})_j|>\eta_j^\alpha(\Sigma_0^{-1/2}df_{0,S}^{\text{nor}}(z_0),x_i)$. Define random variables

$$\begin{aligned}
v_N^{i,j} &= \sqrt{N}(z_N-z_0)\mathbb{1}_{z_N\in B\cap\text{int}P_i}+\bar{v}^{i,j}\mathbb{1}_{z_N\notin B\cap\text{int}P_i},\\
\hat{Y}^{i,j} &= Y^i\mathbb{1}_{Y^i\in\text{int}K_i}+\bar{v}^{i,j}\mathbb{1}_{Y^i\notin\text{int}K_i},\\
\hat{\eta}_N^{i,j} &= \eta_j^\alpha\left(\Sigma_N^{-1/2}\Phi_N(z_N),x_i\right)\mathbb{1}_{z_N\in B\cap\text{int}P_i}+\eta_j^\alpha\left(\Sigma_0^{-1/2}df_{0,S}^{\text{nor}}(z_0),x_i\right)\mathbb{1}_{z_N\notin B\cap\text{int}P_i},
\end{aligned}$$

and note that

$$\hat{\eta}_N^{i,j}\Rightarrow\eta_j^\alpha\left(\Sigma_0^{-1/2}df_{0,S}^{\text{nor}}(z_0),x_i\right).$$

For all Borel sets $W\subset\text{int}K_i$, we have

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

and therefore

$$\begin{aligned}\lim_{N\to\infty}\Pr\left(v_N^{i,j}\in W\right) &= \lim_{N\to\infty}\Pr\left(\sqrt{N}(z_N-z_0)\in W, z_N\in B\right)\\ &= \lim_{N\to\infty}\Pr\left(\sqrt{N}(z_N-z_0)\in W\right)\\ &= \Pr\left(Y^*\in W\right)=\Pr\left(Y^i\in W\right)=\Pr\left(\hat{Y}^{i,j}\in W\right). \quad (3.14)\end{aligned}$$

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

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

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

$$\begin{aligned}&\lim_{N\to\infty}\Pr\left(v_N^{i,j}\in D\cap(\mathrm{int}K_i)^c\right)\\ &=\lim_{N\to\infty}\mathbb{1}_{D\cap(\mathrm{int}K_i)^c}(\bar{v}^{i,j})\Pr\left(z_N\notin B\cap\mathrm{int}P_i\right)\\ &=\mathbb{1}_{D\cap(\mathrm{int}K_i)^c}(\bar{v}^{i,j})\Pr\left(\hat{Y}^{i,j}\notin\mathrm{int}K_i\right)\\ &=\Pr\left(\hat{Y}^{i,j}\in D\cap(\mathrm{int}K_i)^c\right). \quad (3.15)\end{aligned}$$

By combining (3.14) with (3.15), and noting that $\hat{\eta}_N^{i,j}$ and $\eta_j^\alpha\left(\Sigma_0^{-1/2}df_{0,S}^{\mathrm{nor}}(z_0),x_i\right)$ are strictly positive under our assumptions, we find

$$\frac{v_N^{i,j}}{\hat{\eta}_N^{i,j}}\Rightarrow\frac{\hat{Y}^{i,j}}{\eta_j^\alpha\left(\Sigma_0^{-1/2}df_{0,S}^{\mathrm{nor}}(z_0),x_i\right)},$$

and

$$\lim_{N\to\infty} \quad \Pr\left(\Big|\frac{(v_N^{i,j})_j}{\hat{\eta}_N^{i,j}}\Big| \le 1\right) = \Pr\left(\Big|\frac{(\hat{Y}^{i,j})_j}{\eta_j^\alpha\left(\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), x_i\right)}\Big| \le 1\right)$$
$$= \Pr\left(\Big|\frac{(Y^i)_j}{\eta_j^\alpha\left(\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), x_i\right)}\Big| \le 1,\ Y^i \in \mathrm{int}K_i\right),$$

where we have used the fact $|(\bar{v}^{i,j})_j| > \eta_j^\alpha\left(\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), x_i\right)$. The latter fact also implies $\lim_{N\to\infty} \Pr\left(\Big|\frac{(\bar{v}^{i,j})_j}{\hat{\eta}_N^{i,j}}\Big| \le 1\right) = 0$, so it follows that

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}\frac{|(z_N - z_0)_j|}{\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), x_i)} \le 1;\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right)$$
$$= \lim_{N\to\infty} \Pr\left(\sqrt{N}\frac{|(z_N - z_0)_j|}{\hat{\eta}_N^{i,j}} \le 1,\ z_N \in B \cap \mathrm{int}P_i\right) = \lim_{N\to\infty} \Pr\left(\frac{|(v_N^{i,j})_j|}{\hat{\eta}_N^{i,j}} \le 1\right)$$
$$= \Pr\left(\Big|\frac{(Y^i)_j}{\eta_j^\alpha\left(\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), x_i\right)}\Big| \le 1,\ Y^i \in \mathrm{int}K_i\right)$$
$$= \Pr\left(|(M_i^{-1}\Sigma_0^{1/2}Z)_j| \le \eta_j^\alpha\left(\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), x_i\right),\ M_i^{-1}\Sigma_0^{1/2}Z \in K_i\right)$$
$$= \Pr\left(|(df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z)_j| \le \eta_j^\alpha\left(\Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), x_i\right),\ df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z \in K_i\right)$$
$$= (1-\alpha)\Pr\left(df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z \in K_i\right).$$

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

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right)$$
$$= \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(\sqrt{N}\frac{|(z_N - z_0)_j|}{\eta_j^\alpha(\Sigma_N^{-1/2}\Phi_N(z_N), x_i)} \le 1;\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right)$$
$$= \sum_{i=1}^{l}(1-\alpha)\Pr\left(df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z \in K_i\right)$$
$$= (1-\alpha)\sum_{i=1}^{l}\Pr\left(df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z \in K_i\right) = 1-\alpha.$$

□

As in the proof of (3.6), Assumption 3 is used primarily to ensure that the limit in equation (3.11) is well defined. To omit Assumption 3 without affecting the convergence results, two conditions must be satisfied for each selection function of $df^{\text{nor}}_{S,0}(z_0)$. First, for each cone K_i in the conical subdivision of $df^{\text{nor}}_{S,0}(z_0)$ and the corresponding matrix $M_i = df^{\text{nor}}_{S,0}(z_0)|_{K_i}$, the equation (3.10) must have a unique strictly positive solution when Z is replaced by Y_0. Second, for each K_i and all $\ell > 0$ the polyhedra $\left\{x \in \mathbb{R}^n \middle| \; M_i^{-1}x \in K_i,\; |(M_i^{-1}x)_j| \leq \ell\right\}$ must be continuity sets with respect to the random vector Y_0. These two conditions are required to hold for each selection function, so that the convergence is well defined when restricted to each cone. These are similar to the way to relax Assumption 3 for Theorem 5.

Compared to the first method, the second method is computationally much more efficient 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. When the event A_N in (35) holds, that cone also contains $z_N - z_0$ in its interior. In the third example of §3.5, we also observe that the second method is more robust than the first when the sample size is small and A_N does not hold.

While (3.6) and (3.11) provide computable asymptotically exact intervals in general both $a^r(\cdot)$ and $\eta_j^\alpha(\cdot,\cdot)$ lack closed form expressions, an issue addressed in the next section. For ease of exposition moving forward we will suppress the arguments of a^r, η_j^α and υ_j^α, where υ_j^α is the half-width for the intervals considered in Theorem 3.

3.4 Interval computation

This section discusses the computation of a^r and η_j^α, and discusses how to find upper bounds for these quantities. We begin by considering a^r with the results for η_j^α following in a similar fashion. Throughout this section we shall use $\Gamma'(z_N^*) = \{K_1, \ldots, K_l\}$ to denote the conical subdivision for a realization of $\Phi_N(z_N)$. The matrix representations for the selection functions of $\Sigma_N^{-1/2}\Phi_N(z_N)$ on K_i will be denoted by $M_{N,i}$, for $i = 1, \ldots, l$.

Finding a^r requires a search over values of $\ell > 0$ and evaluating

$$\Pr\left(|(\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j(Z) - r| \leq \ell\right).$$

To evaluate this probability we rewrite it in terms of the individual selection functions. The conical subdivision of $\Phi_N(z_N)^{-1}\Sigma_N^{1/2}$ is given by $\{T_1, \ldots, T_l\}$, where

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

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

For any two cones T_v and T_u 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 then be rewritten as

$$\begin{aligned}
&\sum_{i=1}^{l} \Pr\left(|(\Phi_N(z_N)^{-1}\Sigma_N^{1/2})_j(Z) - r| \leq \ell \text{ and } Z \in T_i\right) \\
&= \sum_{i=1}^{l} \Pr\left(|(M_{N,i}^{-1})_j Z - r| \leq \ell \text{ and } Z \in T_i\right). \qquad (3.16)
\end{aligned}$$

Note the connection between (3.16) and what must be considered to find η_j^α. Finding η_j^α requires us to evaluate

$$\Pr\left(|(M_{N,i}^{-1})_j Z| \leq \ell \text{ and } M_{N,i}^{-1} Z \in K_i\right) = \Pr\left(|(M_{N,i}^{-1})_j Z| \leq \ell \text{ and } Z \in T_i\right), \qquad (3.17)$$

for different values of ℓ, but only for those indices i such that $z_N - z_N^* \in K_i$. This difference provides the indirect method of §3.3 with a significant computational advantage over the method of §3.2. Recall from the proof of Theorem 6 that

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

where l was the number of selection functions for $df_{0,S}^{\text{nor}}(z_0)$, B is a certain neighborhood of z_0, A_N is as defined in (3.12) and $K_i = \text{cone}(P_i - z_0)$. When A_N holds and $z_N \in B \cap \text{int}P_i$, it was shown that $z_N - z_N^* \in \text{int}K_i$. Therefore with high probability finding η_j^α will involve evaluating (3.17) for a single index i. In contrast, (3.16) involves a similar calculation for every cone in the subdivision. For this reason, the conditioning based method is scalable with respect to the number of selection functions. Such scalability is very useful, particularly

because the simpler approach given in Theorem 3, which requires the least amount of computation among all methods, cannot be guaranteed to produce asymptotically exact intervals in the piecewise case.

The question of finding a^r and η_j^α now becomes how to evaluate

$$\Pr\left(|b_{N,i}Z - r| \leq \ell \text{ and } Z \in T_i\right), \tag{3.18}$$

where the row vector $b_{N,i}$ is given by $(M_{N,i}^{-1})_j$. When $l \leq 2$ (that is, when $\Phi_N(z_N)$ has no more than 2 selection functions) and $r = 0$, we can evaluate (3.18) using percentiles of standard normal random variables. To see this, consider the case $l = 1$ first. In this case $T_1 = \mathbb{R}^n$ and $M_{N,1}^{-1}Z = \Phi_N(z_N)^{-1}\Sigma_N^{1/2}(Z)$. Then from basic properties of normal random vectors,

$$a^0 = \eta_j^\alpha = \sqrt{\chi_1^2(\alpha)\|(M_{N,1}^{-1})_j\|^2} \quad \text{and} \quad h_j^\alpha = \sqrt{\chi_1^2(\alpha)\|(Q_{N,1})_j M_{N,1}^{-1}\|^2}$$

where $\|\cdot\|$ is the Euclidian norm. In this case both intervals for $(z_0)_j$ are the same as the interval proposed in Theorem 3. Next consider the case $l = 2$; we observe that the two cones satisfy $T_1 = -T_2$ and that Z and $-Z$ have the same distribution. It then follows that

$$\begin{aligned}\Pr\left(|b_{N,i}Z| \leq \ell \text{ and } Z \in T_i\right) &= 1/2\Pr\left(|b_{N,i}Z| \leq \ell\right) \\ &= \Pr\left(Z \in T_i\right)\Pr\left(|b_{N,i}Z| \leq \ell\right).\end{aligned}$$

Thus, when $l = 2$, η_j^α can again be computed using a simple formula; finding a^0 in this case may still require a search over different values of ℓ but the probabilities needed to evaluate for each ℓ can be obtained from the cumulative distribution function of standard normal random variables.

When $l > 2$, our approach to evaluating (3.18) is to rewrite it as the probability of a normal random vector being in a possibly unbounded box. Once formulated in this manner, the probability can be evaluated using the Monte Carlo or Quasi-Monte Carlo methods of (Genz and Bretz, 2009, Chapter 4), both of which are implemented in the R

package mvtnorm (Genz and Bretz, 2009; Genz et al., 2013). Below, we discuss details about this for complementarity problems as well as general SVI's.

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

$$K_i = [l^i_1, u^i_1] \times \cdots \times [l^i_n, u^i_n]$$

with l^i_j and u^i_j taking values in $\{0, \infty, -\infty\}$. Since $\Phi_N(z_N)^{-1}\Sigma_N^{1/2}$ is a homeomorphism, for each $i = 1, \ldots, l$ and $x \in \mathbb{R}^n$ the following equivalences hold:

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

Therefore we can write

$$\begin{aligned}
&\Pr\left(|(M_{N,i}^{-1})_j Z - r| \le \ell \text{ and } Z \in T_i\right) = \Pr\left(r - \ell \le (M_{N,i}^{-1})_j Z \le r + \ell \text{ and } M_{N,i}^{-1} Z \in K_i\right) \\
&= \Pr\left(M_{N,i}^{-1} Z \in [l^i_1, u^i_1] \times \cdots \times [\max(l^i_j, r - \ell), \min(u^i_j, r + \ell)] \times \cdots \times [l^i_n, u^i_n]\right) \\
&= \Pr\left(\tilde{Z} \in [l^i_1, u^i_1] \times \cdots \times [\max(l^i_j, r - \ell), \min(u^i_j, r + \ell)] \times \cdots \times [l^i_n, u^i_n]\right)
\end{aligned}$$

where $\tilde{Z} \sim \mathcal{N}\left(0, M_{N,i}^{-1} M_{N,i}^{-T}\right)$.

For a general SVI, to compute a^r and η^α_j we can use the structure of T_i as a polyhedral cone and express it by linear inequalities,

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

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

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

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

Since finding a^r and η_j^α in the piecewise case requires a search over values of ℓ, we are motivated to look for upper bounds for those quantities that do not require a search to compute. Below we discuss how to find upper bounds for a^r; this idea works similarly for η_j^α.

A natural conjecture related to the upper bound is that $a^r(f) \leq a^r(b^k)$ for a piecewise linear function f with selection functions represented by row vectors $b^1, \ldots, b^k$ with $\|b^1\| \leq \|b^2\| \leq \cdots \leq \|b^k\|$. This conjecture is not true in general. 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 $T_i = \left\{x \in \mathbb{R}^2 | A_i x \leq 0\right\}$ for $i = 1, \ldots, 5$, where

$$A_1 = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 1 \\ -1 & 2 \end{bmatrix}, A_3 = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix},$$
$$A_4 = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \text{and } A_5 = \begin{bmatrix} -1 & 1 \end{bmatrix}.$$

Note both $\{T_1, T_2, T_3\}$ and $\{T_4, T_5\}$ are conical subdivisions of $\mathbb{R}^2$. Define f_1 and f_2 to be

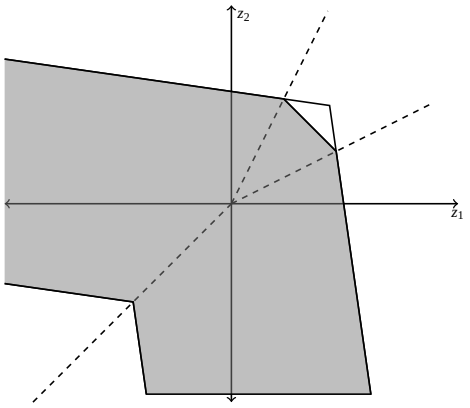

Figure 3.1: Sets R_1 (shaded) and R_2 for $\alpha = .05$

piecewise linear functions such that $f_1|_{T_i} = b^i$ for $i = 1, 2, 3$, $f_2|_{T_4} = b^1$ and $f_2|_{T_5} = b^2$. It

follows that $a^0(b^i) = a^0(f_2) = \sqrt{2\chi_1^2(\alpha)}$, $i = 1, 2, 3$. Next, let

$$\begin{aligned} R_1 &= \left\{ z \in \mathbb{R}^2 | -a^0(f_2) \leq f_1(z) \leq a^0(f_2) \right\}, \\ R_2 &= \left\{ z \in \mathbb{R}^2 | -a^0(f_2) \leq f_2(z) \leq a^0(f_2) \right\}. \end{aligned}$$

As shown in Figure 3.1, 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 $a^0(f_2) < a^0(f_1)$, showing that $\max a^0(b^i)$ is not an upper bound for $a^0(f_1)$.

To construct a valid upper bound for a^r we will use the following Lemma.

Lemma 4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a piecewise linear function with selection functions represented by n dimensional row vectors $b_1, \ldots, b_l$, with the corresponding conical subdivision $\Gamma = \{K_1, \ldots, K_l\}$. Let $Z \sim \mathcal{N}(0, I_n)$, $c_i = \Pr(Z \in K_i)$, $\alpha \in (0, 1)$ and $r \in \mathbb{R}$. Suppose $\ell > 0$ satisfies*

$$\Pr(|b_i Z - r| \leq \ell) \geq 1 - c_i \alpha$$

for $i = 1, \ldots, l$. Then $\Pr(-\ell \leq f(Z) - r \leq \ell) \geq 1 - \alpha$.

Proof. Let E_i be the event that $\{|b_i Z - r| \leq \ell \text{ and } Z \in K_i\}$. As argued previously $\Pr(|f(Z) - r| \leq \ell) = \sum_{i=1}^{l} \Pr(E_i)$. Next note

$$\begin{aligned} \Pr(E_i^c) &\leq \Pr(Z \in K_i^c) + \Pr(|b_i Z - r| > \ell) \\ &\leq 1 - c_i + c_i \alpha = 1 - (1 - \alpha) c_i. \end{aligned}$$

Thus $\Pr(E_i) \geq (1 - \alpha) c_i$ and

$$\Pr(|f(Z) - r| \leq \ell) = \sum_{i=1}^{l} \Pr(E_i) \geq (1 - \alpha) \sum_{i=1}^{l} c_i = 1 - \alpha.$$

□

Corollary 1. *Let f, K_i, b_i, α and Z be as Lemma 4. Let $\alpha_i = \alpha \Pr(Z \in K_i)$. Then $\ell_i = \|b_i\| \sqrt{\chi_1^2(\alpha_i)}$ satisfies $\Pr(|b_i Z| \leq \ell_i) = 1 - \alpha_i$, and $\ell = \max_{1 \leq i \leq l} \ell_i$ satisfies $\Pr(|f(Z)| \leq \ell) \geq 1 - \alpha$.*

While Corollary 1 provides an upper bound for $a^0(f)$, Lemma 4 can be analogously used to find upper bounds for a^r when $r \neq 0$. The similar idea can be used to find upper bounds for η_j^α by considering only cones K_i that contain $z_N - z_N^*$.

3.5 Numerical examples

This section applies the proposed methods and the method of Theorem 3 to three numerical examples. The first example is a complementarity problem used in (Lu, 2012, 2014; Lu and Budhiraja, 2013), the second a complimentarily problem of a slightly larger size. The third example is a nonlinear complementarity problem adapted from (Floudas et al., 1999) and (Dirkse and Ferris, 1995a). When calculating a^r or η_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 §3.4 using the methods of (Genz and Bretz, 2009, Chapter 4). This search terminates when either the distance between the upper and lower bounds for the half-width or the probability of the value being tested is 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 the first two 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}.

The third example is chosen to examine the performance of the proposed methods when the estimates used to compute interval lengths deviate from their asymptotic properties at small sample sizes. In particular, for this example the true solution z_0 lies in the interior of an n-cell of the normal manifold of S, but is close to a number of k-cells of lower dimensions. As a result, the estimates z_N^* obtained from some SAA solutions with small sample sizes does not lie in the relative interior of the same cell that contains z_0 it its relative interior,

so $\Phi_N(z_N)$ has a different structure from the linear function $df^{\text{nor}}_{0,S}(z_0)$. For this example we solve 1,000 SAA problems at sample sizes of N=100 and 3,000 and consider a^r with $r = 0$.

3.5.1 Example 1

For the first example, we consider a complementarity problem with $S = \mathbb{R}^2_+$,

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

and ξ uniformly distributed over the box $[0,2] \times [0,1] \times [0,2] \times [0,4] \times [-1,1] \times [-1,1]$. In this case

$$f_0(x) = \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and the SVI and its corresponding normal map formulation have true solutions $x_0 = z_0 = 0$. The function $df^{\text{nor}}_{0,S}(z_0)$ is then piecewise linear with family of selection functions given by the matrices

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

and corresponding conical subdivision $\mathbb{R}^2_+, \mathbb{R}_+ \times \mathbb{R}_-, \mathbb{R}_- \times \mathbb{R}_+$ and $\mathbb{R}^2_-$. With this information we evaluate (1.21) for $\alpha = .05$ and observe values of .9450 and .9448 for $j = 1$ and 2 respectively. This means that confidence intervals proposed in Theorem 3 will cover $(z_0)_1$ and $(z_0)_2$ with those probabilities in the limit.

In Tables 3.1 and 3.2 we summarize the coverage rates of $(z_0)_1$ and $(z_0)_2$ for each interval determined by υ^α_j, a^{r_N} and η^α_j. We see that the three approaches overall performance is

	υ^α_1	a^{r_N}	η^α_1
N=50	93.65%	94.25%	94.25%
N=100	94.05%	94.85%	94.35%
N=200	94.4%	95%	95.05%
N=2,000	93.65%	94.25%	94.8%

Table 3.1: Coverage rates $(z_0)_1$ $\alpha = .05$

	υ^α_2	a^{r_N}	η^α_2
N=50	93.5%	94.4%	93.7%
N=100	94.4%	94.65%	94.65%
N=200	94.75%	95.35%	95.4%
N=2,000	93.95%	94.5%	94.45%

Table 3.2: Coverage rates $(z_0)_2, \alpha = .05$

generally comparable and in line with the specified 95% level of confidence (for a^{r_N} and η_j^α), or as predicted by the values of (1.21) (for υ_j^α).

Differences between the methods become apparent in Figure 3.2, where the interval lengths for $N = 2,000$ are divided by which cone contains $z_N - z_0$. These differences are also apparent in Table 3.3, where we break down the coverage rates of $(z_0)_2$ and average interval lengths by which K_i contains $z_N - z_0$.

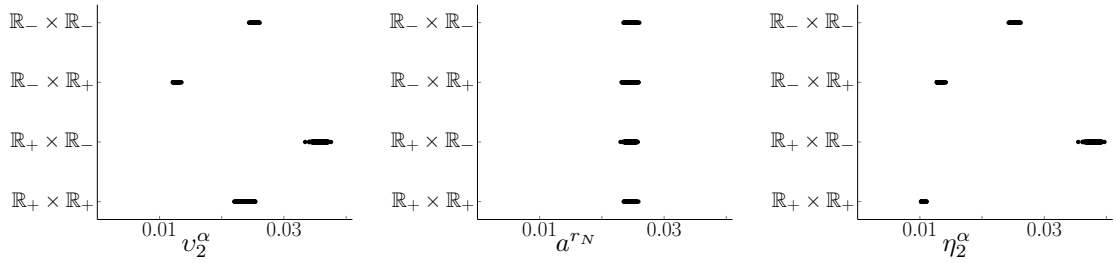

Figure 3.2: Intervals widths for $(z_0)_2$ by cone, $N = 2,000$

Table 3.3: Coverage rates of $(z_0)_2$ and half-widths for $(z_0)_2$ by cone, $N = 2,000$

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
$\mathbb{R}_- \times \mathbb{R}_-(513)$	94.15%	97.66%	93.37%	.0253	.0246	.0253
$\mathbb{R}_- \times \mathbb{R}_+$ (553)	93.85%	99.64%	95.84%	.0127	.0246	.0133
$\mathbb{R}_+ \times \mathbb{R}_-$ (739)	92.29%	87.28%	94.05%	.0358	.0246	.0379
$\mathbb{R}_+ \times \mathbb{R}_+$ (195)	100%	98.97%	98.46%	.0238	.0245	.0106

As shown in Figure 3.2(a) and the column under υ_2^α in Table 3.3, the interval lengths and coverage rates produced by the method of Theorem 3 vary significantly depending on the location of $z_N - z_0$. This is because the linear functions $df_{N,S}^{\text{nor}}(z_N)$ used to calculate υ_2^α are dramatically different when $z_N - z_0$ belongs to different cones. The expression for υ_2^α does not account for the piecewise structure of $df_{0,S}^{\text{nor}}(z_0)$. Because the values of (1.21) are close to to the desired 95% for this example, the overall coverage rates by υ_2^α are only slightly smaller than those of other methods, as shown in Tables 3.1 and 3.2. In general, one cannot expect the overall coverage rates of υ_2^α to be at the desired level, as opposed to our proposed methods based on the estimate $\Phi_N(z_N)$

When $\Phi_N(z_N)$ is piecewise linear, our estimate for the limiting distribution of $z_N - z_0$ has a piecewise structure. Evaluating a^r requires considering each piece of this estimate. Since $\Phi_N(z_N)$ converges to $df_{0,S}^{\text{nor}}(z_0)$, the value of a^r converges to a fixed value. This value

leads to asymptotically exact intervals by averaging out the performance across the different pieces of the limiting distribution. As a result we see intervals of consistent lengths, but with varying performance depending on the location of $z_N - z_0$.

Computation of η_2^α, the second method proposed in this paper, uses the same estimate for the limiting distribution. However, instead of enumerating all pieces of the piecewise distribution, the calculation of η_2^α only requires considering the cone that contains $z_N - z_N^*$ in the conical subdivision of $\Phi_N(z_N)$. Since the probability of z_N^* and z_0 being contained in the relative interior of the same cell goes to one, this approach can accurately condition on which piece of the limiting distribution describes $z_N - z_0$. The definition of η_j^α uses this idea of conditioning to vary the intervals widths and achieve a more consistent coverage rate across the different cones.

3.5.2 Example 2

In this example, $S = \mathbb{R}^5_+$,

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

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

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

so $d(f_0)_{\mathbb{R}^5_+}(z_0)(\cdot)$ is piecewise linear with 32 selection functions. Taking $\alpha = .05$, we first consider confidence intervals for $(z_0)_j$. By evaluating the value of (1.21) for each $j = 1, \ldots, 5$, we find the asymptotic confidence levels of intervals proposed in Theorem 3 to be 93.85%, 93.33%, 94.38%, 93.39% and 92.96% respectively.

Table 3.4: Coverage rates for $(z_0)_3$

	υ_3^α	a^{r_N}	η_3^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 we obtain for this example are largely in line with the specified level of confidence (for a^{r_N} and η_j^α), or as predicted by the values of (1.21) (for υ_j^α). Table 3.4 summarizes the coverage rates of $(z_0)_3$ for each approach and sample size considered. Given the large number of selection functions relative to the number of SAA problems, it is not practical to observe the performances of the different methods broken down by where $z_N - z_0$ falls in the conical subdivision associated with $df_{0,S}^{\text{nor}}(z_0)$. What we are able to observe is the consistency of values of a^{r_N} across samples, as compared to the varied values of υ_j^α and η_j^α, as shown in Figure 3.3 for $(z_0)_3$ and N =2,000.

In this example the computational benefits of η_j^α are clear. For almost all of the samples calculating a^{r_N} requires working with a piecewise linear function with 32 selection functions, whereas for all of the samples calculating η_j^α only involves a single selection function. This difference leads to a dramatic reduction in the computation needed for η_j^α.

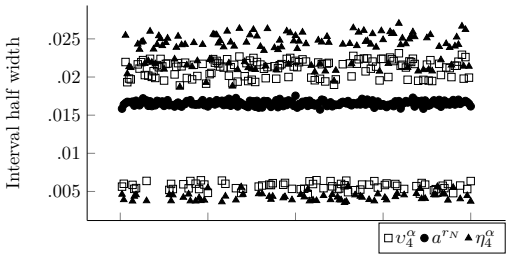

Figure 3.3: Intervals lengths for $(z_0)_3, N = 2,000$

With this example we also examine how upper bounds satisfying the conditions of Lemma 4 compare to the actual interval half-widths. Table 3.5 summarizes the average and median ratios between the bounds and the actual half-widths for sample size $N = 2,000$. While easier to compute, the bounds can be quite conservative, especially those for a^{r_N}.

Table 3.5: Ratios of upper bounds to interval half-widths

		a^{r_N}		η_j^α	
		Average ratio	Median ratio	Average ratio	Median ratio
$N = 2,000$	$(z_0)_1$	6.20	6.33	3.04	2.18
	$(z_0)_2$	15.53	13.44	3.58	2.92
	$(z_0)_3$	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

3.5.3 Example 3

The third example is the invariant capital stock problem from (Dirkse and Ferris, 1995a; Floudas et al., 1999; Hansen and Koopmans, 1972). This problem considers an economy growing over an infinite time horizon. The time horizon is assumed to have discrete periods, and at each time period the economy determines activity levels for the production of capital and consumption goods. The activity levels are constrained by the resources available at the start of each time period and the investment in capital goods made in the previous period. A reward is derived from the consumption goods produced, and the problem is to determine an initial investment of the capital goods that maximizes the sum of the discounted rewards and at the same time results in a constant investment of capital goods over all time periods.

With appropriate conditions on the reward function and constraints, the problem can be solved by finding a solution to the nonlinear complementarity problem,

$$\begin{aligned} 0 \leq \nabla v(q) + (A - \gamma B)^T + C^T u \quad &\perp \quad q \geq 0 \\ 0 \leq (B - A)q \quad &\perp \quad y \geq 0 \\ 0 \leq -Cq + w \quad &\perp \quad u \geq 0. \end{aligned}$$

Here $q \in \mathbb{R}^{10}_+$ denotes the activity levels for the production processes, and A and B denote the capital input and output matrices respectively. The resource input matrix is denoted by C, and w equals the constant amount resources available at the start of each time period. Dual variables for the resource and capital constraints are given by u and y in $\mathbb{R}^2_+$. The nonlinear reward function v is given by

$$v(q) = (q_1 + 2.5q_2)^{0.2}(2.5q_3 + q_4)^{0.2}(2q_5 + 3q_6)^{0.2}$$

and $\gamma \in (0, 1)$ is the discount factor. To formulate this problem as an SVI we assume that the elements of w and the matrices A, B and C are uniformly distributed over intervals of length one; that their expectation equal the quantities in (Floudas et al., 1999); and that for each column of the matrix A, B or C, the components of the vector are dependent with a correlation of one-half. With these assumptions $\xi \in \mathbb{R}^{62}_+$ and the vector of decision variables is given by $x = (q, y, u) \in \mathbb{R}^{14}_+$.

All components of the true solution z_0 are nonzero. Three components of z_0 are less than 0.1 in absolute value, and all are between -0.6575 and 0.6833. While $df^{\text{nor}}_{S,0}(z_0)$ is linear, for moderate sample sizes it is likely that $d\Pi_S(z^*_N)$ and $\Phi_N(z_N)$ are piecewise linear when the function $g(N) = N^{1/3}$ is used to determine z^*_N. This example therefore allows us to examine the performance of the proposed methods when the estimates z^*_N and $\Phi_N(z_N)$ deviate from their asymptotic properties at finite sample sizes.

We generate 1,000 replications of the SAA problem at samples sizes $N = 100$ and $3,000$. For each replication, the SAA solution z_N has components small enough in absolute value to lead to an incorrect estimate z^*_N, in the sense that z^*_N does not lie in the interior of the

Table 3.6: Coverage rates for $(z_0)_j$, $N = 100$ and $N = 3,000$, $\alpha = .05$

Component	$N = 100$		$N = 3,000$		
	υ_j^α	η_j^α	υ_j^α	η_j^α	a^0
$(z_0)_1$	89.6%	93%	95.4%	95.5%	83.7%
$(z_0)_2$	88.8%	93.4%	93.9%	93.9%	83.7%
$(z_0)_3$	89.3%	92.6%	94.2%	94.3%	83.7%
$(z_0)_4$	89.7%	93.1%	94.9%	94.9%	83.7%
$(z_0)_5$	89.8%	91%	95.2%	95.2%	83.7%
$(z_0)_6$	88.3%	91.5%	95.4%	95.4%	83.7%
$(z_0)_7$	89.5%	91.9%	96.1%	96.1%	83.7%
$(z_0)_8$	89.7%	92.6%	95.2%	95.2%	83.6%
$(z_0)_9$	91%	94.2%	95%	95.1%	83.6%
$(z_0)_{10}$	95.1%	96.4%	95.1%	95.2%	83.7%
$(z_0)_{11}$	90.5%	92%	95.3%	95.3%	83.7%
$(z_0)_{12}$	90.7%	93.7%	94.7%	95.1%	83.7%
$(z_0)_{13}$	88.9%	93%	95.4%	95.4%	83.7%
$(z_0)_{14}$	92%	93.3%	93.8%	93.8%	83.7%

n-cell that contains z_0. The performances of the different intervals for $(z_0)_j$ are given in Table 3.6. The method of §3.2 is most sensitive to the use of the incorrect estimate $d\Pi_S(z_N^*)$. For each replication with $N = 100$ the estimate $\Phi_N(z_N)$ has 2^{12} selection functions, and evaluating a^0 is computationally impractical. Even with N =3,000, for about eight percent of the replications no intervals can be computed for a^0. This poor performance is due to the inclusion of all selection functions of $\Phi_N(z_N)$ in the calculation of a^0. The evaluation of a^0 becomes intractable, when the number of selection functions becomes too large to enumerate each piece, or when a selection function has a singular of near singular matrix representation. Moreover, even when a^0 can be evaluated, the performance of those intervals is heavily impacted by the incorrect identification of selection functions, because the method of a^0 is designed to achieve the desired level of confidence by averaging out the performance across the different pieces of the limiting distribution.

In contrast to the poor performance of intervals computed using a^0, the intervals computed using η_j^α (i.e., the second approach) perform well. Even for the cases in which $N = 100$ and $\Phi_N(z_N)$ has 2^{12} selection functions, this conditioning based approach can quickly compute an interval's length. Not only is this approach computationally feasible at this relatively small sample size, its performance is close to the desired level of 95%. The

computation efficiency is due to the fact that only a single selection function of the estimate $\Phi_N(z_N)$ is needed for the computation of an interval's length. This selection function is characterized by its matrix representation and the corresponding cone in the conical subdivision. The matrix representation depends on z_N, while the choice of the cone is determined by z_N^*. Because the incorrect selection functions do not directly affect the computation, they have less impact on the interval length. What is especially noteworthy is that at small sample sizes the intervals computed using η_j^α with "incorrect" choices of z_N^* outperform the intervals using υ_j^α. This may look surprising, because the method using υ_j^α is asymptotically exact for this example. The intuition behind this observation is that the "incorrect" choices of z_N^* is in part a reflection of the difference between the distribution of z_N at small sample sizes and its asymptotic distribution. The computation of η_j^α therefore incorporates these differences in a limited manner, whereas the linear function used to calculate υ_j^α does not capture these differences. As the sample size increases, the asymptotic equivalence of η_j^α and υ_j^α becomes apparent.

CHAPTER 4
Direct confidence intervals

4.1 Motivation

In this chapter we propose a direct method for constructing individual confidence intervals for the solution to an SVI as formulated in (1.1). We begin the chapter with a motivating discussion on how this problem differs from the construction of individual confidence intervals for the solution to the normal map formulation.

To begin, comparing (1.12) and (1.14) we see the difference between the asymptotic distributions for SAA solutions to the normal map and direct formulation is the addition in the latter of Π_{K_0}, the projection onto the critical cone to S at z_0. Since Π_{K_0} is generally non-invertible, neither of the methods presented in Chapter 3 can be used to directly construct intervals for $(x_0)_j$, because both methods require the function that defines the asymptotic distribution to be invertible.

An indirect approach to constructing confidence intervals for $(x_0)_j$ using the methods of Chapter 3 would be projecting confidence intervals for $(z_0)_j$ onto the set S. As noted in §2.2 for a random set $I_N(\omega)$

$$\Pr\left(z_0 \in I_N(\omega)\right) \leq \Pr\left(\Pi_S(z_0) \in \Pi_S(I_N\left(\omega\right))\right) = \Pr\left(x_0 \in \Pi_S(I_N\left(\omega\right))\right).$$

Intervals found in this way will then cover x_0 with a rate that is at least as large as the coverage rate of z_0 by $I_N(\omega)$. 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.

When developing a direct method for calculating confidence intervals for $(x_0)_j$ we would like to emulate the method of §3.3 due to the benefits of working with only a single selection function. As stated above, this method cannot be applied directly since the definition of $\eta_j^\alpha(f,x)$ requires that the piecewise linear function f be invertible. In this definition it is the function f^{-1} that is used primarily in the calculation of $\eta_j^\alpha(f,x)$. The role of f is limited to using its conical subdivision to define a partition of the range of f^{-1}, which combined with x identifies the selection functions of f^{-1} to be considered. An initial attempt to extend this method to general piecewise linear functions would be to identify selection functions by using a partition of the function's range that does not require it to be invertible.

In the case of building individual confidence intervals for $(x_0)_j$ the function of interest is the transformation appearing in the right hand side of (1.14),

$$g = \Pi_{K_0} \circ df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2}(\cdot),$$

whose range is the critical cone K_0. Taking $\Phi_N(z_N)$ and z_N^* to be as in (1.28) we define

$$\Pi_{K_N} = d\Pi_S(z_N^*) = \Lambda_N(z_N), \tag{4.1}$$

and

$$g_N = \Pi_{K_N} \circ \Phi_N(z_N)^{-1}\Sigma_N^{1/2}(\cdot).$$

From (1.9) and Theorem 4 it follows that

$$\lim_{N\to\infty} \Pr\left(\Pi_{K_N} = \Pi_{K_0}\right) = 1 \tag{4.2}$$

suggesting the use of g_N as an estimate for g when constructing confidence intervals. Since K_0 is a polyhedron a natural partition to consider would be the relative interiors of the faces of K_0. The shortcoming of this approach is that there can exist a face of K_0, say C_i, such that multiple selection functions of g map to $\mathrm{ri}C_i$ and $\Pr\left(g(Z) \in \mathrm{ri}C_i\right) > 0$. When such a face C_i exists, we would no longer have the desired property that as the sample size goes to infinity the probability of working with a single selection function of g_N goes to one.

One could attempt to avoid this issue by instead choosing the selection functions based on a point in a function's domain. For the function g of interest such a partition of the domain would depend on the conical subdivision associated with $df^{\text{nor}}_{0,S}(z_0)^{-1}$. When constructing intervals using the estimates g_N, the partition would depend on the conical subdivision associated with $\Phi_N(z_N)^{-1}$. The issue with this approach is that while the probability of $\Phi_N(z_N)$ and $df^{\text{nor}}_{0,S}(z_0)$ sharing a common conical subdivision goes to one as the sample size goes to infinity, such a result will not hold for their inverses. Thus evaluating the performance of intervals produced using this approach would require addressing the case when the point falls in a region that identifies selection functions of g and g_N that do not correspond to one another.

The presence of Π_{K_0} in the asymptotic distribution also complicates the aim of constructing intervals with an exact level of confidence. Consider the following extreme example. If one modifies the SVI used in §3.5.1 by changing the function $F(x, \xi)$ to be

$$F(x,\xi) = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi_5 + 1 \\ \xi_6 + 1 \end{bmatrix},$$

the solution to (1.1) is still $x_0 = (0, 0)$, while the solution to (1.5) becomes $z_0 = (-1, -1)$. With this change the critical cone $K_0 = \{(0, 0)\}$ and Π_{K_0} can be expressed as the zero matrix. Moreover, both ξ_5 and ξ_6 are bounded below by negative one so for all samples $x_N = (0, 0) = x_0$. Therefore any interval containing $(x_N)_j$ will cover $(x_0)_j$ due to the complete lack of variability in SAA solutions.

Removing our modification to the function $F(x, \xi)$ and considering the SVI as given in Section 3.5.1 we see a second less extreme example. In this case with $z_0 = (0, 0)$ one has $K_0 = \mathbb{R}^2_+$ and Π_{K_0} is piecewise linear. Now it is the restriction Π_{K_0} to the negative orthant that can be expressed as the zero matrix, and one can similarly observe that for all samples such that $z_N \in \mathbb{R}^2_-$, $x_N = (0, 0) = x_0$ and any interval containing $(x_N)_j$ will cover $(x_0)_j$. For $z_N \in \mathbb{R}_- \times \mathbb{R}_+$ it is only true that $(x_N)_1 = (x_0)_1$ and the performance of intervals for $(x_0)_2$ centered at $(x_N)_2$ will depend on how the intervals' widths are chosen.

In these examples we see that it is possible for $\Pr\left((x_N)_j = (x_0)_j\right) > 0$, providing a lower bound for any confidence interval that contains $(x_N)_j$. Therefore the aim should not be to construct asymptotically exact confidence intervals, but instead intervals for which a lower bound on the level of confidence is met. In the remainder of this chapter we propose an approach of constructing such intervals, with Theorem 7 providing the theoretical justification for the proposed method. In §4.3 we apply the proposed method to two numerical examples.

4.2 Methodology

To determine the width of an interval the proposed method replaces $\eta_j^\alpha(\cdot,\cdot)$ with a function $h_j^\alpha(f,g,x)$ where f and g are piecewise linear functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ that share a common conical subdivision, $\{K_1,\ldots,K_l\}$, with only g required to be invertible. For any choice of cone K_i, $i = 1,\ldots,l$, 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

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

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

$$h_j^\alpha(f,g,x) = \inf\left\{\ell \geq 0 \;\middle|\; \frac{\Pr\left(|(Q_i)_j M_i^{-1} Z| \leq \ell \text{ and } M_i^{-1}Z \in K_i\right)}{\Pr\left(M_i^{-1}Z \in K_i\right)} \geq (1-\alpha)\right\}. \tag{4.3}$$

where Q_i and M_i are the matrices that satisfy $f|_{K_i} = Q_i$ and $g|_{K_i} = M_i$, and $(Q_i)_j$ denotes the jth row of Q_i. 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 $x_{i_s} \in \text{int}K_{i_s}$. The following lemma shows that the location of x and the selection functions of f determine when $h_j^\alpha(f,g,x) = 0$.

Lemma 5. *Let $(Q_i)_j$ denote the jth row of Q_i, the matrix that satisfies $f|_{K_i} = Q_i$. Then for any $x \in \bigcap_{s=1}^v K_{i_s}$, $j = 1,\ldots,n$, 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 $x \in \text{int}K_i$. If $h_j^\alpha(f,g,x) = 0$,

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

which implies

$$0 < \Pr\left((Q_i)_j M_i^{-1}Z = 0 \text{ and } M_i^{-1}Z \in K_i\right) \leq \Pr\left((Q_i)_j M_i^{-1}Z = 0\right). \tag{4.4}$$

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, (4.4) implies that $\|(Q_i)_j M_i^{-1}\| = 0$, and thus $(Q_i)_j$ is a vector of zeroes. The reverse implication is immediate. □

When using $h_j^\alpha(f,g,x)$ to construct confidence intervals for solutions to (1.1) based on (1.14), Π_{K_0} and $\Sigma_0^{-1/2} df_{0,S}^{\text{nor}}(z_0)$ play the roles of f and g respectively. These functions will be estimated by $d\Pi_S(z_N^*)$ and $\Sigma_N^{-1/2}\Phi_N(z_N)$. From (1.9) and (1.29) it follows that the probability of $d\Pi_S(z_N^*)$ equalling Π_{K_0} goes to one as the sample size goes to infinity. By adapting this setting to deterministic functions, we prove the following lemma.

Lemma 6. *Let $f, g : \mathbb{R}^n \to \mathbb{R}^n$ be piecewise linear functions with g being a homeomorphism. Suppose that $\{f_N\}_{N=1}^\infty$ and $\{g_N\}_{N=1}^\infty$ are two sequences of piecewise linear functions satisfying the following conditions.*

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

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

Proof. From Proposition 2 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}. Without loss of generality we can assume that for all N the functions g_N are invertible, $f_N = f$ and f, g and g_N share

a common conical subdivision $\{K_1, \ldots, K_l\}$. To finish the proof, it suffices to show that $h_j^\alpha(f, g_N, x) \to h_j^\alpha(f, g, x)$ for any $x \in \text{int}K_i$, $i = 1, \ldots, l$.

When $x \in \text{int}K_i$ and $h_j^\alpha(f, g, x) = 0$, it follows from Lemma 5 that $h_j^\alpha(f, g_N, x) = 0$. In the case of $x \in \text{int}K_i$ and $h_j^\alpha(f, g, x) > 0$, the convergence can be shown using an argument analogous to the proofs of Lemma 1 and Lemma 3 and is therefore omitted.

□

The main result of this section, Theorem 7, can now be proven. We will use the same notation used in Theorem 6. The conical subdivision associated with $df_{0,S}^{\text{nor}}(z_0)$ is denoted by $\Gamma'(z_0) = \{K_1, \ldots, K_l\}$, with $df_{0,S}^{\text{nor}}(z_0)|_{K_i} = M_i$. Each K_i is given by $K_i = \text{cone}(P_i - z_0)$, where $P_1, \ldots, P_l$ are all n-cells in the normal manifold of S that contain z_0. As before, we define the following random variables:

$$Y^i = M_i^{-1}\Sigma_0^{1/2}Z, \quad Y_0 = \Sigma_0^{1/2}Z \quad \text{and} \quad Y^* = df_{0,S}^{\text{nor}}(z_0)^{-1}\Sigma_0^{1/2}Z.$$

Additionally, we use $\Pi_{K_0}|_{K_i} = Q_i$ to denote the selection function of Π_{K_0} on K_i.

Theorem 7. *Let Assumptions 2, 3 and 4 hold. Let $\Phi_N(z_N)(\cdot)$ and z_N^* be as defined in (1.28). For all $j = 1, \ldots, n$ and $\alpha \in (0, 1)$,*

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(x_N - x_0)_j| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right) \ge 1 - \alpha. \tag{4.5}$$

Proof. As in the proof of Theorem 6, let C_i, $i = 1, \ldots, m$ denote the cells in the normal manifold of S, and for each N let the event A_N be as defined in (3.12). For $\omega \in A_N$ the equality $\Pi_{K_0} = d\Pi_S(z_N^*)$ holds, and $\{K_1, \ldots, K_l\}$ provides a common conical subdivision for Π_{K_0}, $df_{0,S}^{\text{nor}}(z_0)$ and $\Phi_N(z_N(\omega))$. From (3.13) and Lemma 6 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 $h_j^\alpha(\Pi_{K_0}, \Sigma_0^{-1/2}df_{0,S}^{\text{nor}}(z_0), u)$.

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

$$
\begin{aligned}
& \lim_{N\to\infty} \Pr\left(\sqrt{N}|(x_N - x_0)_j| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right) \\
= \quad & \lim_{N\to\infty} \Pr\left(\sqrt{N}\left|\left(\Pi_S(z_N) - \Pi_S(z_0)\right)_j\right| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*);\ A_N\right) \\
= \quad & \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(\sqrt{N}\left|\left(\Pi_{K_0}(z_N - z_0)\right)_j\right| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*);\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right) \\
= \quad & \sum_{i=1}^{l} \lim_{N\to\infty} \Pr\left(\sqrt{N}|(Q_i)_j(z_N - z_0)| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i);\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right) \qquad (4.6)
\end{aligned}
$$

where u_i in the last expression is any point in $\mathrm{int}K_i$. The first equality uses the relation between the solution to a variational inequality and that of its normal map formulation, while the second equality combines the almost sure convergence of z_N to z_0 with (1.8). The final equality holds, because both $z_N - z_0$ and $z_N - z_N^*$ will be contained in $\mathrm{int}K_i$ whenever $\omega \in A_N$ and $z_N \in \mathrm{int}P_i$, in which case $z_N - z_N^*$ may be replaced with u_i and $\Pi_{K_0}(z_N - z_0) = Q_i(z_N - z_0)$.

Evaluation of each term in (4.6) depends on whether $(Q_i)_j$ is zero or not. If $(Q_i)_j$ is the zero vector for some i, then

$$
\begin{aligned}
& \lim_{N\to\infty} \Pr\left(\sqrt{N}|(Q_i)_j(z_N - z_0)| \le h_j^\alpha(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i);\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right) \\
& = \lim_{N\to\infty} \Pr\left(\sqrt{N}(z_N - z_0) \in \mathrm{int}K_i\right) = \Pr\left(Y^* \in \mathrm{int}K_i\right) \\
& = \Pr\left(d(f_0)_S^{-1}(z_0)\Sigma_0^{1/2}Z \in K_i\right). \qquad (4.7)
\end{aligned}
$$

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 $\bar{v}^{i,j}$ to be such that $\bar{v}^{i,j} \notin K_i$ and $|(Q_i)_j\bar{v}^{i,j}| > h_j^\alpha(\Pi_{K_0}, \Sigma_0^{-1/2}df_{0,S}^{\mathrm{nor}}(z_0), u_i)$. With these we define random vectors

$$
\begin{aligned}
v_N^{i,j} \quad & = \sqrt{N}(z_N - z_0)\mathbb{1}_{z_N \in B \cap \mathrm{int}P_i} + \bar{v}^{i,j}\mathbb{1}_{z_N \notin B \cap \mathrm{int}P_i}, \\
\hat{Y}^{i,j} \quad & = Y^i\mathbb{1}_{Y^i \in \mathrm{int}K_i} + \bar{v}^{i,j}\mathbb{1}_{Y^i \notin \mathrm{int}K_i}, \\
\hat{h}_N^{i,j} \quad & = h_j^\alpha\left(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i\right)\mathbb{1}_{z_N \in B \cap \mathrm{int}P_i} + h_j^\alpha\left(\Pi_{K_0}, \Sigma_0^{-1/2}df_{0,S}^{\mathrm{nor}}(z_0), u_i\right)\mathbb{1}_{z_N \notin B \cap \mathrm{int}P_i}.
\end{aligned}
$$

Using the same arguments as in Theorem 6 it follows that

$$\frac{v_N^{i,j}}{\hat{h}_N^{i,j}} \Rightarrow \frac{\hat{Y}^{i,j}}{h_j^{\alpha}\left(\Pi_{K_0}, \Sigma_0^{-1/2} df_{0,S}^{\mathrm{nor}}(z_0), u_i\right)}$$

and

$$\begin{aligned}
&\lim_{N\to\infty} \Pr\left(\sqrt{N}\frac{|(Q_i)_j(z_N - z_0)|}{h_j^{\alpha}(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i)} \leq 1;\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right) \\
&= (1-\alpha)\Pr\left(df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2} Z \in K_i\right). \qquad (4.8)
\end{aligned}$$

Combining (4.7) and (4.8), with the fact that $z_N - z_N^* \in \mathrm{int}K_i$ on A_N, we have

$$\begin{aligned}
&\lim_{N\to\infty} \Pr\left(\sqrt{N}|(x_N - x_0)_j| \leq h_j^{\alpha}(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), z_N - z_N^*)\right) \\
&= \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(\sqrt{N}|(Q_i)_j(z_N - z_0)| \leq h_j^{\alpha}(d\Pi_S(z_N^*), \Sigma_N^{-1/2}\Phi_N(z_N), u_i);\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right) \\
&\geq (1-\alpha)\sum_{i=1}^{l} \Pr\left(df_{0,S}^{\mathrm{nor}}(z_0)^{-1}\Sigma_0^{1/2} Z \in K_i\right) = 1 - \alpha.
\end{aligned}$$

□

Again it is possible to relax Assumption 3 in the above theorem. In addition to the conditions specified after Theorem 6, a strict inequality needs to hold in (4.4) for each selection function when Z is replaced by Y_0. With these conditions, the limit in equation (4.5) will remain well defined and converge to the same quantity.

An important fact that can be seen in the proof of Theorem 7 is that

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

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. Let P_i be the n-cell in the normal manifold of S that contains z_0 and satisfies $K_i = \mathrm{cone}\,(P_i - z_0)$. If $z_N \in \mathrm{int}P_i$, then

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

and by Lemma 5

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

This means that the method of Theorem 7 returns the correct point estimate $(x_N(\omega))_j = (x_0)_j$ whenever $z_N \in \text{int}P_i$. Therefore, equation (4.5) holds as a strict inequality if and only if there is nonzero probability for the event $(x_N)_j = (x_0)_j$ to happen. While such point estimates are asymptotically correct, an incorrect estimate may be returned at small sample sizes if z_N is contained in an n-cell P_k for which the jth component of $\Pi_S|_{P_k}$ is zero but $z_0 \notin P_k$. By Lemma 5, the third method may be made robust against returning incorrect point estimates, by replacing $(Q_i)_j = 0$ with any nonzero vector. This modification does not change the third method's asymptotic level of confidence, but ensures that an interval with nonzero length is always returned.

We also observe from the proof of Theorem 7 that this approach for building confidence intervals, like the method in §3.3, will with high probability require working with only a single selection function. Moreover h_j^{α} is computed in largely the same manner as η_j^{α} using the approach of Section 3.4.

4.3 Numerical examples

In this section we apply the direct method of computing individual confidence intervals to two numerical examples. The first example is the same nonlinear complementarity problem considered in §3.5.3. This example will again allow us to examine the proposed method when the estimate z_N^* deviates from its asymptotic properties at small sample sizes. For the second example we consider an SVI where the set S that defines the problem is not a box. In this case the indirect methods based on the approaches of Chapter 3 cannot be applied.

For each of the examples, we will use the more conservative implementation of the direct approach that does not return point estimates. To do so, whenever $d\Pi_S(z_N^*)$ has a selection function with jth component equal to zero, we replace that component with the unit vector e_j whose jth element is one. With this modification a value which is equal to η_j^{α} will be returned instead of zero. While the resulting interval is more conservative than the point

estimate, this modification does not change the intervals' asymptotic level of confidence and can be more robust at small sample sizes, as discussed after Theorem 7.

4.3.1 Example 1

Recall the invariant capital stock problem in §3.5.3, which can be formulated as the nonlinear complementarity problem,

$$\begin{aligned} 0 \leq \nabla v(q) + (A - \gamma B)^T + C^T u \quad &\perp \quad q \geq 0 \\ 0 \leq (B - A)q \quad &\perp \quad y \geq 0 \\ 0 \leq -Cq + w \quad &\perp \quad u \geq 0. \end{aligned}$$

Assuming the same distribution used in §3.5.3 for the elements of A, B, C and w, we find that all components of z_0 are nonzero and between -0.6575 and 0.6833. Moreover we observe that the true solution x_0 has components $(x_0)_j = 0$ for $j = 1, 4, 5, 8, 9$, and 13.

In addition to the direct approach proposed in this chapter, we can also apply the indirect approach of projecting individual confidence intervals for $(z_0)_j$ onto S. Since the set $S = \mathbb{R}^{14}_+$, each selection function of $d\Pi_S(z)$ is represented by a diagonal matrix with values of zero and one along the diagonal. With the conservative implementation of the direct approach, we replace a selection function's jth row by the basis vector e_j, if that row has all zero entries. In view of (3.10) and (4.3), the value of h_j^α from such a replacement will be equal to η_j^α. If additionally $(z_N)_j \geq 0$, then the direct approach of §4.2 will produce the same interval for $(x_0)_j$ as the indirect approach of §3.3. If $(z_N)_j < 0$, then the indirect approach of §3.3 returns the interval $\left[0, \max\left\{0,\, (z_N)_j + N^{-1/2}\eta_j^\alpha\right\}\right]$, whereas the conservative implementation of the direct approach returns the interval $\left[0, N^{-1/2}h_j^\alpha\right] = \left[0, N^{-1/2}\eta_j^\alpha\right]$. Thus, if the jth row of a selection function contains all zeros, then the above two approaches will provide the same coverage rate for $(x_0)_j = 0$, with the approach of §4.2 returning a slightly longer interval when $(z_N)_j$ is negative.

Solving the same 1,000 replications of the SAA problem for N =100 and 3,000, used in §3.5.3, we see the benefit of not returning point estimates for $(x_0)_j$ when using the direct approach of §4.2. For eighty two replications at $N = 100$, the SAA solution z_N lies in

a different n-cell from z_0. For each of these replications $(z_N)_{10} < 0 < (x_0)_{10}$, and the original implementation will return an incorrect point estimate for $(x_0)_{10}$. Similar, but less frequent, errors occur for $(x_0)_j$ for $j = 2, 3, 6, 7$ and 14. The conservative implementation fixes those errors effectively. The remaining components of $(x_0)_j$ are equal to zero, and the conservative implementation does not affect their coverage rates. At the larger sample size $N = 3,000$, z_N and z_0 are always contained in the same n-cell and both implementations of the direct approach result in the same coverage rates.

Table 4.1: Coverage rates for $(x_0)_j$, $N = 100$ and $N = 3,000$, $\alpha = .05$

	$N = 100$			$N = 3,000$			
Component	υ_j^α	η_j^α	h_j^α	υ_j^α	η_j^α	h_j^α	a^0
$(x_0)_1$	99.7%	99.7%	99.7%	100%	100%	100%	91.5%
$(x_0)_2$	88.8%	93.4%	93.4%	93.9%	93.9%	93.9%	83.7%
$(x_0)_3$	89.3%	92.6%	92.6%	94.2%	94.3%	94.3%	83.7%
$(x_0)_4$	99.7%	99.7%	99.7%	100%	100%	100%	91.5%
$(x_0)_5$	99.7%	99.7%	99.7%	100%	100%	100%	91.5%
$(x_0)_6$	88.3%	91.4%	91.4%	95.4%	95.4%	95.4%	83.7%
$(x_0)_7$	89.5%	91.9%	91.9%	96.1%	96.1%	96.1%	83.7%
$(x_0)_8$	99.7%	99.7%	99.7%	100%	100%	100%	91.5%
$(x_0)_9$	97.3%	97.3%	97.3%	100%	100%	100%	91.5%
$(x_0)_{10}$	95.1%	96.4%	96.4%	95.1%	95.2%	95.2%	83.7%
$(x_0)_{11}$	90.5%	92%	92%	95.3%	95.3%	95.3%	83.7%
$(x_0)_{12}$	90.7%	93.7%	93.7%	94.7%	95.1%	95.1%	83.7%
$(x_0)_{13}$	99.7%	99.7%	99.7%	100%	100%	100%	91.5%
$(x_0)_{14}$	92%	93.3%	98.8%	93.8%	93.8%	93.8%	83.7%

The coverage rates for the direct approach and indirect approach using η_j^α are largely the same. The one component for which their performance differs is $(x_N)_{14}$ when $N = 100$. This is also due to the small sample size deviations in the location of z_N. The slightly longer interval obtained by the direct approach when $(z_N)_j < 0$ results in the higher coverage rate due to the samples where $(z_N)_{14} < 0 < (x_0)_{14}$. Overall, at $N = 100$ both conditioning based approaches perform largely in line with (or exceed) the desired level of confidence, and the coverage rates are further improved at $N = 3000$.

4.3.2 Example 2

For the second example, the SVI is defined by

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

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

$$f_0(x) = \begin{bmatrix} 4 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and the 2-cells in the normal manifold are of the form $C_i = \left\{ x \in \mathbb{R}^2 | A_i x \leq 0 \right\}$ with

$$A_1 = \begin{bmatrix} .5 & -1 \\ -2 & 1 \end{bmatrix},\ A_2 = \begin{bmatrix} 2 & -1 \\ -.5 & -1 \end{bmatrix},\ A_3 = \begin{bmatrix} .5 & 1 \\ -2 & 1 \end{bmatrix} \text{ and } A_4 = \begin{bmatrix} -2 & -1 \\ -.5 & 1 \end{bmatrix}.$$

These four cones provide the conical subdivision associated with $f_{0,S}^{\text{nor}}$. The corresponding family of selection functions for $f_{0,S}^{\text{nor}}$ are given by matrices

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

The SVI and its normal map formulation have true solutions $x_0 = z_0 = 0$. The function $df_{0,S}^{\text{nor}}(z_0)$ is equal to the $f_{0,S}^{\text{nor}}$, with its conical subdivision given by $K_i = \text{cone}\,(C_i - 0) = C_i$, $i = 1, \cdots, 4$.

Table 4.2: Coverage rates 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%	97.05%

For this example we consider 2,000 replications of the SAA problem at samples sizes of $N = 50$, 100, 200, and 2,000. Since the set S defining the SVI is not a box, for any real num-

bers $l \leq u$ neither $\Pi_S(\mathbb{R} \times [l, u])$ nor $\Pi_S([l, u]) \times \mathbb{R})$ results 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 is applicable. Combining (4.5) 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 $(x_0)_j$. Table 4.2 summarizes the coverage rates for $(x_0)_1$ and

Table 4.3: Intervals for $(x_0)_i$ by cone, $N = 2,000$, $\alpha = .05$

	Coverage rate		Average length	
Cone (samples in cone)	$(x_0)_1$	$(x_0)_2$	$(x_0)_1$	$(x_0)_2$
$K_1(80)$	88.75%	90%	.0104	.0132
K_2 (689)	95.36%	95.36%	.0089	.0177
K_3 (824)	100%	100%	.0246	.0526
K_4 (407)	95.33%	95.33%	.0073	.0036

$(x_0)_2$ obtained from 2,000 problems at each sample size with $\alpha = .05$. In Table 4.3 we examine the coverage rates and interval lengths from the SAA problems with $N = 2,000$, broken down by the location of $z_N - z_0$. Since the selection function corresponding to $d\Pi_S(z_0)|_{K_3}$ is represented by the zero matrix, the equality $(x_N)_j = (x_0)_j$ holds when $z_N - z_0 \in K_3$, leading to coverage rates of 100% for that case. The nonzero interval lengths we obtain for this case are due to the aforementioned conservative implementation of the direct method.

CHAPTER 5
Relaxed confidence intervals

5.1 Introduction

In this chapter we consider relaxing the conditions required for the methods developed in Chapters 3 and 4, and propose a procedure for computing intervals that meet a minimum level of confidence under the relaxed conditions. In particular, the method we develop does not require the use of the consistent estimate $\Phi_N(z_N)$. This allows us to omit Assumption 4 and the condition that z_N converge to z_0 in probability at an exponential rate. The exponential rate of convergence is required for the the appropriate choice of z_N^*. The conditioning based approaches of §3.3 and §4.2 rely on z_N^* to construct consistent estimates for the transformations in (1.12) and (1.14), and to identify a single selection function of these estimates. The approach developed in this chapter builds from these conditioning based approaches but uses only z_N to estimate the selection function used to compute an interval's width.

The compromise we make for omitting Assumption 4 is to allow for some error in the estimation of the selection function used to compute the width of an interval. Without a known rate for the convergence of z_N to z_0, we are unable to specify what the probability of an error occurring converges to as the sample size increases. We are able to provide an asymptotic upper bound for this probability and can then make conservative adjustments to the calculations that follow the estimation of the selection function. These adjustments allow us to specify a minimum level of confidence for the resulting intervals. The justification for this approach follows from the more general framework considered in Theorems 8 and 9, where we consider a set of estimates for the selection function that contains a consistent estimate with a sufficiently large probability. In §5.2 we formally define and justify the proposed approach, and in §5.3 we apply the proposed method to a numerical example.

5.2 Methodology

The conditioning based approaches to computing individual confidence intervals require estimating the selection function of $d\Pi_S(z_0)$ characterized by the matrix $M_{i(z_N)}$ and cone $K_{i(z_N)}$, where $K_{i(z_N)}$ is contained in the conical subdivision of $d\Pi_S(z_0)$, $z_N - z_0 \in K_{i(z_N)}$, and $d\Pi_S(z_0)|_{K_{i(z_N)}} = M_{i(z_N)}$. The approaches taken in §3.3 and §4.2 require estimates $M_N(\omega)$ and $K_N(\omega)$ that satisfy,

$$\lim_{N\to\infty} \Pr\left(K_N = K_{i(z_N)}\right) = 1 \tag{5.1}$$

and that for all $\epsilon > 0$,

$$\lim_{N\to\infty} \Pr\left(\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|M_N h - M_{i(z_N)} h\|}{\|h\|} \leq \epsilon\right) = 1. \tag{5.2}$$

In Chapters 3 and 4, the estimates $M_N(\omega)$ and $K_N(\omega)$ are selected using the location of $z_N - z_N^*$ in the conical subdivision of $d\Pi_S(z_N^*)$. The high probability of choosing a correct z_N^* and satisfying (5.1) and (5.2), followed from Assumption 4 and the resulting convergence of z_N to z_0 in probability at an exponential rate.

Our approach to computing intervals that maintain their desired asymptotic properties without requiring Assumption 4 is to relax (5.1) while still satisfying (5.2). To do so we make use of the following observation. Let P_N be a set in the polyhedral subdivision associated with Π_S with $z_N \in \text{int}P_N$. Then for any $z \in P_N$ we have $z_N - z \in \text{cone}(\text{int}P_N - z) = \text{int}K_z$, and

$$d\Pi_S(z)|_{K_z} = M_N = d\Pi_S(z_N).$$

The estimate $d\Pi_S(z_N)$ therefore satisfies (5.2) since under Assumptions 1 and 2 for a.e. N sufficiently large there exists a set P_N in the polyhedral subdivision of Π_S such that $z_N \in \text{int}P_N$ and $z_0 \in P_N$.

We next allow for (5.1) to be relaxed in two ways. First, we allow for the cone K_N to be replaced with a set of estimates. Second, we will only require that the limiting probability of this set containing $K_{i(z_N)}$ be greater than or equal to a target value. More formally, we

require a set $\mathcal{K}_N^{\alpha_1}$ comprised of polyhedral cones of dimension n with

$$\lim_{N\to\infty} \Pr\left(K_{i(z_N)} \in \mathcal{K}_N^{\alpha_1}\right) \geq 1 - \alpha_1. \tag{5.3}$$

To incorporate this relaxation into the computation of the confidence intervals we consider the follow generalization of the function η_j^α. Let M be an invertible $n \times n$ matrix and $\mathcal{K} = \left\{\tilde{K}_1, \ldots, \tilde{K}_m\right\}$ be a collection of polyhedral convex sets of dimension n. We then define the function $\tilde{\eta}_j^{\alpha_2}(\mathcal{K}, M)$ to be,

$$\inf\left\{\ell \geq 0 \middle| \frac{\Pr\left(|(M^{-1}Z)_j| \leq \ell, \text{ and } M^{-1}Z \in K_i\right)}{\Pr\left(M^{-1}Z \in K_i\ \right)} \geq 1 - \alpha_2 \text{ for all } K_i \in \mathcal{K}\right\}$$

Using the function $\tilde{\eta}_j^{\alpha_2}$ to determine an interval's width, we can now show the following convergence result analogous to (3.11) in Theorem 6. As before let $\Gamma'(z_0) = \{K_1, \ldots, K_l\}$ denote the conical subdivision associated with $df_{0,S}^{\text{nor}}(z_0)$ such that $df_{0,S}^{\text{nor}}(z_0)|_{K_i} = M_i$ and $K_i = \text{cone}(P_i - z_0)$ where $P_1, \ldots, P_l$ are all n-cells in the normal manifold of S that contain z_0.

Theorem 8. *Suppose that Assumptions 1, 2, and 3 hold, and that $\mathcal{K}_N^{\alpha_1}$ satisfies (5.3) for $\alpha_1 \in (0, 1]$. Then for every $j = 1, \ldots, n$ and $\alpha_2 \in (0, 1)$,*

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df_{N,S}^{nor}(z_N)\right)\right) \geq 1 - (\alpha_1 + \alpha_2). \tag{5.4}$$

Proof. For each N let,

$$\Phi_0(z_N)(h) = df\left(\Pi_S(z_N)\right) \circ d\Pi_S(z_0)(h) + h - d\Pi_S(z_0)(h)$$

and define the event $A_N = \left\{K_{i(z_N)} \in \mathcal{K}_N^{\alpha_1}\right\}$. Let B be a fixed neighborhood of z_0 such that $B \cap (z_0 + K_i) = B \cap P_i$ for $i = 1, \ldots, l$. Then

$$
\begin{aligned}
&\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df_{N,S}^{\mathrm{nor}}(z_N)\right)\right) \\
&\ge \lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df_{N,S}^{\mathrm{nor}}(z_N)\right);\ A_N\right) \\
&= \sum_{i=1}^{l} \lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df_{N,S}^{\mathrm{nor}}(z_N)\right);\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right)
\end{aligned}
$$

Note that when A_N holds and $z_N \in B \cap \mathrm{int}P_i$, it follows that, $K_{i(z_N)} = K_i$, $\Phi_0(z_N)|_{K_i} = df_{N,S}^{\mathrm{nor}}(z_N)$, and

$$
\begin{aligned}
&\frac{\Pr\left(|(df_{N,S}^{\mathrm{nor}}(z_N)^{-1}\Sigma_N^{1/2}Z)_j| \le \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df_{N,S}^{\mathrm{nor}}(z_N)\right)\ \text{ and }\ df_{N,S}^{\mathrm{nor}}(z_N)^{-1}\Sigma_N^{1/2}Z \in K_i\right)}{\Pr\left(df_{N,S}^{\mathrm{nor}}(z_N)^{-1}\Sigma_N^{1/2}Z \in K_i\ \right)} \\
&\ge \frac{\Pr\left(|(\Phi_0(z_N)^{-1}\Sigma_N^{1/2}Z)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right)\ \text{ and }\ \Phi_0(z_N)^{-1}\Sigma_N^{1/2}Z \in K_i\right)}{\Pr\left(\Phi_0(z_N)^{-1}\Sigma_N^{1/2}Z \in K_i\ \right)}
\end{aligned}
$$

where x_i is any point in $\mathrm{int}K_i$.

Next we observe that,

$$
\begin{aligned}
&\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df_{N,S}^{\mathrm{nor}}(z_N)\right)\right) \\
&\ge \sum_{i=1}^{l} \lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right);\ A_N;\ z_N \in B \cap \mathrm{int}P_i\right) \\
&= \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right);\ z_N \in B \cap \mathrm{int}P_i\right) \\
&\qquad - \sum_{i=1}^{l} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right); A_N^{\mathrm{c}};\ z_N \in B \cap \mathrm{int}P_i\right) \\
&\ge \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right);\ z_N \in B \cap \mathrm{int}P_i\right) \\
&\qquad - \sum_{i=1}^{l} \Pr\left(A_N^{\mathrm{c}}\ z_N \in B \cap \mathrm{int}P_i\right) \\
&= \lim_{N\to\infty} \sum_{i=1}^{l} \left[\Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right);\ z_N \in B \cap \mathrm{int}P_i\right)\right] - \Pr\left(A_N^{\mathrm{c}}\right) \\
&\ge \lim_{N\to\infty} \sum_{i=1}^{l} \left[\Pr\left(\sqrt{N}|(z_N - z_0)_j| \le \eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right);\ z_N \in B \cap \mathrm{int}P_i\right)\right] - \alpha_1
\end{aligned}
$$

where the final inequality follows from the definition of A_N and (5.3).

Under Assumptions 1 and 2, for all $\epsilon > 0$,

$$\lim_{N\to\infty} \Pr\left(\sup_{h\in\mathbb{R}^n, h\neq 0} \frac{\|\Phi_0(z_N)(h) - df^{\mathrm{nor}}_{0,S}(z_0)\|}{\|h\|} \leq \epsilon \right) = 1$$

which by Lemma 3 and Assumptions 3 implies that

$$\eta_j^{\alpha_2}\left(\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right) \Rightarrow \eta_j^{\alpha_2}\left(\Sigma_0^{-1/2} df^{\mathrm{nor}}_{0,S}(z_0), x_i\right).$$

Therefore by the same argument used to prove Theorem 6 it follows that

$$\lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq \eta_j^{\alpha_2}\left((\Sigma_N^{-1/2}\Phi_0(z_N), x_i\right);\ z_N \in B \cap \mathrm{int} P_i\right) = 1 - \alpha_2$$

and thus

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq \tilde{\eta}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, \Sigma_N^{-1/2} df^{\mathrm{nor}}_{N,S}(z_N)\right)\right) \geq 1 - (\alpha_1 + \alpha_2).$$

□

Before discussing approaches for choosing the collection of cones $\mathcal{K}_N^{\alpha_1}$ and the computation of intervals using $\tilde{\eta}_j^{\alpha_2}$, we present the analogous extension of the direct approach for computing confidence intervals. To do so, let M and $\mathcal{K} = \left\{\tilde{K}_1, \ldots, \tilde{K}_m\right\}$ be as above, and let Q be a $n \times n$ matrix. We then define the function $\tilde{h}_j^{\alpha_2}(\mathcal{K}, Q, M)$ to be,

$$\inf\left\{ \ell \geq 0 \middle| \frac{\Pr\left(|(Q)_j M^{-1} Z| \leq \ell, \text{ and } M^{-1}Z \in K_i\right)}{\Pr\left(M^{-1}Z \in K_i\ \right)} \geq 1 - \alpha_2 \text{ for all } K_i \in \mathcal{K} \right\}.$$

Theorem 9. *Suppose that Assumptions 1, 2, and 3 hold, and that $\mathcal{K}_N^{\alpha_1}$ satisfies (5.3) for $\alpha_1 \in (0, 1]$. Then for every $j = 1, \ldots, n$ and $\alpha_2 \in (0, 1)$,*

$$\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq \tilde{h}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, d\Pi_S(z_N), \Sigma_N^{-1/2} df^{nor}_{N,S}(z_N)\right)\right) \geq 1 - (\alpha_1 + \alpha_2). \quad (5.5)$$

Proof. For each N define let,

$$\Phi_0(z_N)(h) = df\left(\Pi_S(z_N)\right) \circ d\Pi_S(z_0)(h) + h - d\Pi_S(z_0)(h)$$

and define the event $A_N = \left\{K_{i(z_N)} \in \mathcal{K}_N^{\alpha_1}\right\}$. Let B be a fixed neighborhood of z_0 such that $B \cap (z_0 + K_i) = B \cap P_i$ for $i = 1, \ldots, l$. Then using the same arguments in Theorem 8 it follows that for $u_i \in \text{int}K_i$,

$$\begin{aligned}
&\lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq \tilde{h}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, d\Pi_S(z_N), \Sigma_N^{-1/2} df_{N,S}^{\text{nor}}(z_N)\right)\right) \\
&\geq \sum_{i=1}^{l} \lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq \tilde{h}_j^{\alpha_2}\left(\mathcal{K}_N^{\alpha_1}, d\Pi_S(z_N), \Sigma_N^{-1/2} df_{N,S}^{\text{nor}}(z_N)\right);\ A_N;\ z_N \in B \cap \text{int}P_i\right) \\
&\geq \sum_{i=1}^{l} \lim_{N\to\infty} \Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq h_j^{\alpha_2}\left(d\Pi_S(z_0), \Sigma_N^{-1/2}\Phi_0(z_N), u_i\right);\ A_N;\ z_N \in B \cap \text{int}P_i\right) \\
&\geq \lim_{N\to\infty} \sum_{i=1}^{l} \left[\Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq h_j^{\alpha_2}\left(d\Pi_S(z_0), \Sigma_N^{-1/2}\Phi_0(z_N), u_i\right);\ z_N \in B \cap \text{int}P_i\right)\right] - \Pr\left(A_N^c\right) \\
&\geq \lim_{N\to\infty} \sum_{i=1}^{l} \left[\Pr\left(\sqrt{N}|(z_N - z_0)_j| \leq h_j^{\alpha_2}\left(d\Pi_S(z_0), \Sigma_N^{-1/2}\Phi_0(z_N), u_i\right);\ z_N \in B \cap \text{int}P_i\right)\right] - \alpha_1 \\
&\geq 1 - (\alpha_1 + \alpha_2).
\end{aligned}$$

□

The results (5.4) and (5.5) will still hold with the appropriate relaxations of Assumption 3 discussed after Theorems 6 and 7.

While we have a good deal of latitude in how to choose $\mathcal{K}_N^{\alpha_1}$, for the methods of Theorems 8 and 9 to be computationally tractable we would like to limit the number of cones $K_i \in \mathcal{K}_N^{\alpha_1}$. In the following Lemma we show that from the sample data we can identify a single cone that will satisfy (5.3). To do so we use the asymptotically exact confidence regions in equations (1.20) to identify a subset of k-cells in the normal manifold of S, and select a cell with the lowest dimension from this subset. The proof of Lemma 7 does not require Σ_N to be invertible. To limit the notation involved we will use $R_{N,\epsilon}$ to denote the confidence region, though the same argument will hold when the confidence regions R_N are used. In the proof of Lemma 7 we will use the same notation as in Theorem 6. With this notation, the conical subdivision of $d\Pi_S(z_0)$ is comprised of sets $K_i = \text{cone}(P_i - z_0)$ where $P_1, \ldots, P_l$ are all n-cells in the normal manifold of S that contain z_0. The element of the conical

subdivision of $d\Pi_S(z_0)$ that contains $z_N - z_0$ for a particular sample shall be denoted by $K_{i(z_N)}$.

Lemma 7. *Suppose that Assumptions 1 and 2 hold. Let $\alpha_1 \in (0,1)$ and $R_{N,\epsilon}$ be a $(1-\alpha_1)*100\%$ confidence region for z_0 as given in equation (1.20). Let P_N be the n-cell in the normal manifold of S with $z_N \in P_N$ and let C_{i_N} be the k-cell that has the smallest dimension of all cells that intersect $R_{N,\epsilon}$ and P_N. Then for $\tilde{z}_{i_N} \in ri\, C_{i_N}$, $\mathcal{K}_N^{\alpha_1} = \{cone(P_N - \tilde{z}_{i_N})\}$ satisfies (5.3)*

Proof. Let $C_1, \ldots, C_m$ denote all of the k-cells in the normal manifold of S, $d_i(z)$ be as defined in equation (1.24) and $\mathcal{I} = \{i \mid z_0 \notin C_i\}$ be the collection of indices for the k-cells that do not contain z_0. Then $\min_{i\in\mathcal{I}} d_i(z_0) = \delta > 0$ since there are finitely many k-cells, each of which is a closed set. Let C_{i_0} denote the unique k-cell that contains z_0 in its relative interior. As shown in (Lu and Budhiraja, 2013, Proposition 5.1) the cell C_{i_0} is the cell of lowest dimension to contain z_0. Then for any cell C_i with $C_i \neq C_{i_0}$ and dimension less than or equal to that of C_{i_0}, it follows that $i \in \mathcal{I}$.

For any $i \in \mathcal{I}$ and $z \in C_i$,

$$\begin{aligned} \|z_N - z\| &= \|z_0 - z + z_N - z_0\| \\ &\geq \|z_0 - z\| - \|z_N - z_0\| \\ &\geq \delta - \|z_N - z_0\|. \end{aligned} \tag{5.6}$$

Let G_N denote the event that $\delta/2 \geq \min_{i\in\mathcal{I}} d_i(z_N)$, then by (5.6) and the almost sure convergence of z_N to z_0, that $\Pr(G_N) \to 1$.

Next, recall the simultaneous confidence intervals for z_0 computed by finding the minimum axis aligned bounding box that contains $R_{N,\epsilon}$. Let $\Sigma_N = U_N^T \Delta_N U_N$, where U_N is an orthogonal matrix with rows $u_{N,1}, \ldots, u_{N,n}$ and Δ_N is a diagonal matrix with elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then by Proposition 1, each component interval, $j = 1, \ldots, n$, has half-width

$$w_{N,j}^{\epsilon} = \sqrt{\frac{\chi^2_{l_N}(\alpha) \sum_{i=1}^{l_N} (c_{N,j} u_{N,i}^T)^2 \lambda_i}{N}} + \frac{\epsilon}{\sqrt{N}} \sum_{i=l_N+1}^{n} |c_{N,j} u_{N,i}^T|.$$

where $c_{N,j}$ is the jth row of $df^{\text{nor}}_{N,S}(z_N)^{-1}$. Let $w_N = (w^{\epsilon}_{N,1}, \ldots, w^{\epsilon}_{N,j})$. From the almost sure convergence of Σ_N to Σ_0 and $df_N(z_N)$ to $df_0(z_0)$, it follows that $\|w_N\| \to 0$ almost surely.

Define the event A_N to be $\{\|w_N\| < \delta/2 \quad \text{and } z_0 \in R_{N,\epsilon}\}$, and let B be a fixed neighborhood of z_0 such that $B \cap (z_0 + K_i) = B \cap P_i$ for $i = 1, \ldots, l$. Then,

$$\begin{aligned}
&\lim_{N\to\infty} \Pr\left(K_{i(z_N)} = \text{cone}(P_N - \tilde{z}_{i_N})\right) \\
&= \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(K_i = \text{cone}(P_i - \tilde{z}_{i_N});\ z_N \in B \cap \text{int}P_i\right) \\
&\geq \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(K_i = \text{cone}(P_i - \tilde{z}_{i_N});\ A_N;\ G_N;\ z_N \in B \cap \text{int}P_i\right) \\
&= \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(A_N;\ G_N;\ z_N \in B \cap \text{int}P_i\right).
\end{aligned}$$

The final equality follows from the fact that when A_N and G_N both occur, for any $z \in R_{N,\epsilon}$,

$$\|z_N - z\| \leq \|w_N\| < \delta/2 \leq \min_{i\in\mathcal{I}} d_i(z_N),$$

and thus no k-cell with index $i \in \mathcal{I}$ intersects with $R_{N,\epsilon}$. Since A_N requires $z_0 \in R_{N,\epsilon}$, it follows that C_{i_0} is the cell of lowest dimension to intersect with $R_{N,\epsilon}$. Therefore $C_{i_0} = C_{i_N}$ and

$$\begin{aligned}
&\lim_{N\to\infty} \Pr\left(K_{i(z_N)} = \text{cone}(P_N - \tilde{z}_{i_N})\right) \\
&\geq \lim_{N\to\infty} \sum_{i=1}^{l} \Pr\left(A_N;\ G_N;\ z_N \in B \cap \text{int}P_i\right) \\
&= \lim_{N\to\infty} \Pr\left(A_N;\ G_N\right) \\
&\geq \lim_{N\to\infty} \Pr\left(z_0 \in R_{N,\epsilon}\right) = 1 - \alpha_1
\end{aligned}$$

and $\mathcal{K}^{\alpha_1}_N = \{\text{cone}(P_N - \tilde{z}_{i_N})\}$ satisfies (5.3) □

From the proof of Lemma 7 we see that the same result will hold if we use the simultaneous confidence intervals computed from $R_{N,\epsilon}$ to identify the set of k-cells from which C_{i_N} is chosen. When the set S is a box, working with the simultaneous confidence intervals has the computational benefit of allowing us to identify the cell C_{i_N} by making n

component wise comparisons. For more general sets S the search for this face can still be simplified by using the fact that it is required to have the lowest possible dimension. We can begin by identifying the facets of P_N that intersect $R_{N,\epsilon}$ as this is typically an easier problem, and from these facets we can then restrict the search to only their intersection. An algorithm for finding the desired cell can be initialized with a cell of the lowest possible dimension, designed to search over faces of increasing dimension, and terminate as soon as a cell that intersects $R_{N,\epsilon}$ is found. Checking if a cell intersects with $R_{N,\epsilon}$ reduces to solving a quadratic programming problem.

Examining the combination of Lemma 7 and Theorems 8 and 9, we see how the proposed methods generalize the conditioning based approaches of §3.3 and §4.2. In Lemma 7 we use the confidence regions to estimate which k-cell in the normal manifold of S contains z_0 in its relative interior, and therefore the choice of cone K_N. Since the limiting probability of making a correct choice is bounded below by the regions' level of confidence, we are able to adjust for this error in the calculation as $\tilde{\eta}_j^{\alpha_2}$. In the equations (5.4) and (5.5) we see that this approach allows us to construct intervals that meet at least a $(1-\alpha)*100\%$ level of confidence by balancing between the error in estimating the cone, bounded above by α_1, and the probability of not covering the true component using the correct cone, which equals α_2. This is similar to the approaches of §3.3 and §4.2, but in those settings the exponential rate of convergence allows us to remove the error in estimating the cone and set $\alpha_1 = 0$ and $\alpha_2 = \alpha$.

Additionally, the framework of Theorems 8 and 9 considers more conservative approaches were we include multiple cones in the set $\mathcal{K}_N^{\alpha_1}$. This provides us with an opportunity to balance between increasing how conservative an interval may be and the amount of additional computation required. This may range from the choice of a single cone as in Lemma 7, to the choice of a potentially large set of cones that will always contain the correct choice, in which case $\tilde{\eta}_j^{\alpha_2}$ would provide a bound on the width of an asymptotically exact confidence interval.

5.3 Numerical example

For this example we consider the linear complementarity problem

$$0 \leq Mx + \xi \perp x \geq 0$$

where $x \in \mathbb{R}^{20}$. The matrix M is of the form

$$M = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$$

where each M_i is a 10×10 row diagonally dominant matrix. We generate each component of ξ from a uniform distribution over an interval of length 1.5, with with $E[\xi_i] > 0$, $i = 1, 2, 4, 5, 8$, $E[\xi_i] < 0$, $i = 3, 6, 7, 9, 10$, and $E[\xi_i] = 0$, $i = 11, \ldots, 20$. With this distribution of ξ and the block diagonal form of the matrix M, $df^{\text{nor}}_{0,S}(z_0)$ has 2^{10} selection functions. Similar to the invariant capital stock example in §3.5.3 and §4.3.1, this example poses a challenge to the proposed method as all nonzero components of z_0 are less than 1 in absolute value, with one component less than 0.01.

We consider 1,000 SAA problems at each sample size of $N = 100$ and 2,000. When implementing the approach of §5.2, we select $\tilde{z}_{i_N}$ from a k-cell of that has the lowest dimension of cells which intersect a set of 97.5% simultaneous confidence intervals for z_0. For the samples with $N = 100$ this resulted in eleven to thirteen components of $\tilde{z}_{i_N}$ equaling zero, and between nine to eleven components equaling zero at the samples of size $N = 2,000$. For this example, the method of §3.3 would also be appropriate. Using this approach with component wise comparisons of z_N to $N^{-1/3}$, z^*_N had either fifteen or sixteen components equal to zero for the samples with $N = 100$ and twelve components equal to zero for the samples of size $N = 2,000$. For comparison, we also consider the intervals computed using the method analyzed in Theorem 3.

In Table 5.1 we summarize the performance of the intervals computed using all three approaches. We see that the large number of selection functions for $df^{\text{nor}}_{0,S}(z_0)$ results in poor performance for the method of Theorem 3 which does not account for the effects of

Table 5.1: Coverage rates for $(z_0)_j$, $N = 100$ and $N = 2,000$, $\alpha_1 = \alpha_2 = .025$

	$N = 100$			$N = 2,000$		
Component	$\upsilon_j^{.05}$	$\eta_j^{.05}$	$\tilde{\eta}_j^{.025}$	$\upsilon_j^{.05}$	$\eta_j^{.05}$	$\tilde{\eta}_j^{.025}$
$(z_0)_1$	81.5%	94.2%	96.9%	81.9%	95.7%	97.4%
$(z_0)_2$	80.7%	94.1%	96.4%	81.4%	95%	96.9%
$(z_0)_3$	82.3%	95.2%	98.1%	81.9%	96%	98.2%
$(z_0)_4$	81.5%	95.8%	97.7%	81.2%	94.7%	97.2%
$(z_0)_5$	81.9%	95.1%	97.2%	81.7%	95.5%	98.3%
$(z_0)_6$	82.8%	94.3%	97.5%	81.2%	95%	96.9%
$(z_0)_7$	81.1%	95.3%	97.7%	80.7%	94.5%	98%
$(z_0)_8$	81%	94.5%	96.6%	80.4%	95.5%	97.4%
$(z_0)_9$	87.8%	97.3%	98.5%	82.6%	95.9%	98%
$(z_0)_{10}$	79.7%	94.5%	97.8%	80.4%	93.7%	96.9%
$(z_0)_{11}$	81.1%	94.2%	97%	82.2%	95.9%	98.5%
$(z_0)_{12}$	80.8%	95.5%	97.7%	79.5%	94.5%	96%
$(z_0)_{13}$	79.8%	93.8%	96.9%	80.5%	93.2%	97.6%
$(z_0)_{14}$	82.1%	95.6%	98.1%	80.2%	95%	97.2%
$(z_0)_{15}$	80.8%	95.5%	97.4%	81.3%	95.3%	97.2%
$(z_0)_{16}$	82.1%	94.9%	97%	80.6%	94.7%	97.5%
$(z_0)_{17}$	79.3%	93.5%	96.8%	80.1%	95.4%	97.7%
$(z_0)_{18}$	79.3%	94.5%	97.1%	78.9%	94.7%	97.4%
$(z_0)_{19}$	82.6%	96.2%	97.9%	81.6%	94.6%	97.2%
$(z_0)_{20}$	81.6%	95.3%	97.4%	79.9%	95.4%	98%

the piecewise structure of $df_{0,S}^{\text{nor}}(z_0)$. At the sample size $N = 2,000$, this approach does not cover any component of $(z_0)_i$ at a rate exceeding 83%. In this example we also see the potentially conservative performance of the intervals with width $\tilde{\eta}^{\alpha_2}$. The potential error in estimating the cone $K_{i(z_N)}$ is conservatively accounted for by setting $\alpha_1 = \alpha_2 = 0.025$, but for this example the error in estimating $K_{i(z_N)}$ does not have a large impact on the intervals' performance. As a result, the intervals perform largely in line with a $(1 - \alpha_2) * 100\%$ level of confidence.

When computing confidence intervals for $(x_0)_j$ we use the robust approach that does not return a point estimate when $(z_0)_j < 0$. This approach is once again seen to be beneficial, since at the sample size of $N = 100$ for 214 samples, $(z_N)_9 < 0 < (z_0)_9 = (x_0)_9$, and an incorrect point estimate would be returned if no adjustment is made. In Table 5.2 we summarize the coverage rates of $(x_0)_i$ for the intervals computed using $h_j^{0.05}$, $\tilde{h}_j^{0.025}$ and by projecting the intervals for $(z_0)_j$ with width $\upsilon_j^{0.05}$ onto S. Comparing the coverage rates of $(z_0)_j$ and $(x_0)_j$, we see an increase for the components with $(z_0)_j = 0$ for all three

Table 5.2: Coverage rates for $(x_0)_j$, $N = 100$ and $N = 2,000$, $\alpha_1 = \alpha_2 = .025$

	$N = 100$			$N = 2,000$		
Component	$v_j^{.05}$	$h_j^{.05}$	$\tilde{h}_j^{.025}$	$v_j^{.05}$	$h_j^{.05}$	$\tilde{h}_j^{.025}$
$(x_0)_1$	100%	100%	100%	100%	100%	100%
$(x_0)_2$	100%	100%	100%	100%	100%	100%
$(x_0)_3$	82.3%	95.2%	98.1%	81.9%	96%	98.2%
$(x_0)_4$	100%	100%	100%	100%	100%	100%
$(x_0)_5$	100%	100%	100%	100%	100%	100%
$(x_0)_6$	82.8%	94.3%	97.5%	81.2%	95%	96.9%
$(x_0)_7$	81.2%	95.3%	97.7%	80.7%	94.5%	98%
$(x_0)_8$	100%	100%	100%	100%	100%	100%
$(x_0)_9$	90.1%	97.6%	98.6%	82.6%	95.9%	98%
$(x_0)_{10}$	79.7%	94.5%	97.8%	80.4%	93.7%	96.9%
$(x_0)_{11}$	91.2%	97.5%	98.9%	92.3%	98.6%	99.7%
$(x_0)_{12}$	93.3%	98.6%	99.5%	92%	97.6%	98.7%
$(x_0)_{13}$	84.8%	94.7%	97.3%	85.4%	94.5%	96.4%
$(x_0)_{14}$	92%	98.2%	99.2%	90.7%	98.1%	99.1%
$(x_0)_{15}$	91.8%	99%	99.3%	92.9%	98.4%	98.9%
$(x_0)_{16}$	89.1%	96.4%	97.8%	87.8%	96.3%	98.2%
$(x_0)_{17}$	92%	97.8%	99.1%	93%	98.6%	99.5%
$(x_0)_{18}$	90.6%	97.5%	98.5%	90.2%	97.5%	98.9%
$(x_0)_{19}$	92.6%	98.2%	99.1%	91.8%	97.7%	98.8%
$(x_0)_{20}$	89.9%	97.3%	98.4%	89.8%	98.4%	99.1%

methods, and 100% coverage of $(x_0)_j$ for those components with $(z_0)_j < 0$. At the sample size of $N = 100$ we also observe an increase in the coverage of $(x_0)_9$ due to the samples for which $(z_N)_9 < 0 < (z_0)_9 = (x_0)_9$. Comparing the three methods for the components with $(x_0)_j > 0$, we see similar performance as in the case of the normal map solutions. In particular, the intervals with width v_j^{α} are not close to the desired level of confidence for such components, while the intervals with width h_j^{α} and $\tilde{h}_j^{\alpha_2}$ perform largely in line with the values of $\alpha = 0.05$ and $\alpha_2 = 0.025$ respectively.

BIBLIOGRAPHY

Anitescu, M. and Petra, C. (2011). Higher-order confidence intervals for stochastic programming using bootstrapping. Technical Report ANL/MCS-P1964-1011, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL.

Attouch, H., Cominetti, R., and Teboulle, M., editors (2009). *Special Issue on Nonlinear Convex Optimization and Variational Inequalities*. Springer. Mathematical Programming 116, Numbers 1-2.

Chen, Y., Lan, G., and Yuyuan, O. (2014). Accelerated schemes for a class of variational inequalities. *Submitted*.

Demir, M. C. (2000). *Asymptotics and confidence regions for stochastic variational inequalities*. PhD thesis, University of Wisconsin, Madison.

Dirkse, S. P. and Ferris, M. C. (1995a). Mcplib: A collection of nonlinear mixed complementarity problems. *Optimization Methods and Software*, 5(4):319–345.

Dirkse, S. P. and Ferris, M. C. (1995b). The PATH solver: a nonmonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5(2):123–156.

Facchinei, F. and Pang, J.-S. (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume I. Springer, New York.

Ferris, M. C. (2005). MATLAB and GAMS : Interfacing optimization and visualization software. Technical Report TR98-19, Computer Sciences Department, University of Wisconsin, Madison.

Ferris, M. C. and Pang, J.-S. (1997a). *Complementarity and Variational Problems: State of the Art*. SIAM.

Ferris, M. C. and Pang, J.-S. (1997b). Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713.

Floudas, C. A., Pardalos, P. M., Adjiman, C. S., Esposito, W. R., Gümüs, Z. H., Harding, S. T., Klepeis, J. L., Meyer, C. A., and Schweiger, C. A. (1999). *Handbook of Test Problems in Local and Global Optimization*, volume 33. Springer, New York.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2013). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9996.

Giannessi, F. and Maugeri, A., editors (1995). *Variational Inequalities and Network Equilibrium Problems*. Plenum Press, New York.

Giannessi, F., Maugeri, A., and Pardalos, P. M., editors (2001). *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, volume 58 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers.

Gürkan, G., Yonca Özge, A., and Robinson, S. M. (1999). Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 84:313–333.

Hansen, T. and Koopmans, T. C. (1972). On the definition and computation of a capital stock invariant under optimization. *Journal of Economic Theory*, 5(3):487–523.

Harker, P. T. and Pang, J.-S. (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms, and applications. *Mathematical Programming*, 48:161–220.

Jiang, H. and Huifu, X. (2008). Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Transactions on Automatic Control*, 53(6):1462–1475.

Juditsky, A., Nemirovski, A., and Tauvel, C. (2011). Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58.

King, A. J. and Rockafellar, R. T. (1993). Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, 18:148–162.

Koshal, J., Nedić, A., and Shanbhag, U. V. (2013). Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, 58(3):594–609.

Lu, S. (2012). A new method to build confidence regions for solutions of stochastic variational inequalities. *Optimization*. Published online before print at http://www.tandfonline.com/doi/abs/10.1080/02331934.2012.727556.

Lu, S. (2014). Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities. *SIAM Journal on Optimization*. Forthcoming.

Lu, S. and Budhiraja, A. (2013). Confidence regions for stochastic variational inequalities. *Mathematics of Operations Research*, 38:545–568.

MATLAB (2010). *Version 7.11 (R2010b)*. The MathWorks Inc., Natick, Massachusetts.

Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. (2009a). Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609.

Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. (2009b). Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609.

Pang, J. (1990). Newton's method for b-differentiable equations. *Mathematics of Operations Research*, 15:311–341.

Pang, J.-S. and Ralph, D., editors (2009). *Special Issue on Nonlinear Programming, Variational Inequalities, and Stochastic Programming*. Springer. Mathematical Programming 117, Numbers 1-2.

Polyak, B. T. (1990). New stochastic approximation type procedures. *Avtomatica i Telemekhanika*, 7:98–107.

Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855.

Ralph, D. (1994). On branching numbers of normal manifolds. *Nonlinear Analysis: Theory, Methods & Applications*, 22(8):1041–1050.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22:400–407.

Robinson, S. M. (1991). An implicit-function theorem for a class of nonsmooth functions. *Mathematics of Operations Research*, 16(2):292–309.

Robinson, S. M. (1992). Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17(3):691–714.

Robinson, S. M. (1995). Sensitivity analysis of variational inequalities by normal-map techniques. In Giannessi, F. and Maugeri, A., editors, *Variational Inequalities and Network Equilibrium Problems*, pages 257–269, New York. Plenum Press.

Rosenthal, R. (2012). GAMS–A User's Guide. GAMS Development Corporation, Washington, D.C. *Available online http://www.gams.com.*

Scholtes, S. (1996). A proof of the branching number bound for normal manifolds. *Linear Algebra and its Applications*, 246:83–95.

Scholtes, S. (2012). *Introduction to Piecewise Differentiable Equations.* Springer.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. P. (2009). *Lectures on Stochastic Programming: Modeling and Theory.* Society for Industrial and Applied Mathematics and Mathematical Programming Society.

Shapiro, A. and Xu, H. (2008). Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. *Optimization*, 57:395–418.

Vogel, S. (2008). Universal confidence sets for solutions of optimization problems. *SIAM Journal on Optimization*, 19(3):1467–1488.

Xu, H. (2010). Sample average approximation methods for a class of stochastic variational inequality problems. *Asia-Pacific Journal of Operational Research*, 27(1):103–119.