Moderate deviations for stochastic variational inequalities

Mingjie Gao
School of Economics and Statistics, Guangzhou University,
Guangzhou, China
mjgao@gzhu.edu.cn

Ka-Fai Cedric Yiu
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hunghom, Kowloon, Hong Kong, China
cedric.yiu@polyu.edu.hk

Abstract

Stochastic variational inequalities (SVIs) have been used widely in modeling various optimization and equilibrium problems subject to data uncertainty. The sample average approximation (SAA) solution is an asymptotically consistent point estimator for the true solution to a stochastic variational inequality. Some central limit results and large deviation estimates for the SAA solution have been obtained. The purpose of this paper is to study the convergences in regimes of moderate deviations for the SAA solution. Using the delta method and the exponential approximation, we establish some results on moderate deviations. We apply the results to the hypotheses testing for the SVIs, and prove that the rejection region constructed by the central limit theorem has the probability of the type II error with exponential decay speed. We also give some simulations and numerical results for the tail probabilities.

Keywords: Moderate deviation; sample average approximation; stochastic optimization; stochastic variational inequality

1 Introduction

Stochastic variational inequality (SVI) has been used widely in engineering and economics as a model for a large class of equilibrium problems subject to data uncertainty, and is closely related to stochastic optimization problems. Let us first recall the model.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let ξ be a random vector on the probability space $(\Omega, \mathcal{F}, P)$ taking its values in a closed subset Ξ of $\mathbb{R}^d$. Let O be an open subset of $\mathbb{R}^q$, and let F be a measurable function from $O \times \Xi$ to $\mathbb{R}^q$, such that for each $x \in O$ the expectation $f_0(x) = E[F(x, \xi)]$ is well defined. Let S be a polyhedral convex set in $\mathbb{R}^q$. The stochastic variational inequality (SVI) problem is to find a point $x \in S \cap O$ such that

$$0 \in f_0(x) + N_S(x), \tag{1.1}$$

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

$$N_S(x) = \{v = (v^1, \cdots, v^q)^T \in \mathbb{R}^q, \langle v, s - x\rangle \leq 0 \text{ for each } s \in S\},$$

and $\langle\cdot,\cdot\rangle$ denotes the scalar product of two vectors of the same dimension. The point $x \in S\cap O$ is called a solution to (1.1).

Let $\xi_1, \cdots, \xi_n$ be independent and identically distributed (i.i.d.) random variables with distribution same as that of ξ. Define the sample average function $f_n : O \times \Omega \to \mathbb{R}^q$ by

$$f_n(x,\omega) = \frac{1}{n}\sum_{i=1}^{n} F(x, \xi_i(\omega)). \tag{1.2}$$

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

$$0 \in f_n(x,\omega) + N_S(x). \tag{1.3}$$

The point $x \in S \cap O$ is called a solution to (1.3). Let x_0 denote a solution to (1.1) and let x_n denote a solution to (1.3). We refer to a solution to (1.1) or its normal map formulation as a true solution, and a solution to (1.3) or its normal map formulation as a SAA solution.

It is known that under certain regularity conditions, the sample average approximation (SAA) solution almost surely converges to a true solution as the sample size n goes to infinity, see Gürkan, Özge and Robinson [11], King and Rockafellar [12], and Shapiro, Dentcheva and Ruszczyński [28, Section 5.2.1]. Xu [35] obtained an exponential rate in which the SAA solutions converge to the set of true solutions in probability under some assumptions on the moment generating functions of certain random variables, see [29] for related results on the exponential convergence rate. King and Rockafellar [12, Theorem 2.7] and Shapiro, Dentcheva and Ruszczyński [28, Section 5.2.2] provided the asymptotic distribution of SAA solutions. The asymptotic distribution result is not directly usable for confidence regions and hypothesis testings because it contains a function that depends discontinuously on the true solution. Lu and Budhiraja [16], Lu [17, 18], and Lamm, Lu and Budhiraja [13] proposed some new estimators and established their central limit theorems. They constructed asymptotically exact confidence regions and individual confidence intervals for the SVI by applying their asymptotic distribution results (see Demir [5] for related results on the confidence regions). In more detail, set $z = x_0 - f_0(x_0)$ and $z_n = x_n - f_n(x_n)$. Then $\sqrt{n}\Sigma_0^{-1/2} d(f)_S(z_0)(z_n - z_0)$ converges to a q-dimensional standard normal random variable (see Theorem 7 in [16]), i.e.,

$$\sqrt{n}\Sigma_0^{-1/2} d(f)_S(z_0)(z_n - z_0) \xrightarrow{d} N(0, I_q),$$

where Σ_0 is the covariance matrix of $F(x_0,\xi)$, and $d(f)_S(z_0)$ is the B-derivative of the normal map $(f_0)_S$ evaluated at z_0. For application to statistical inference, it is required to estimate Σ_0 and $d(f)_S(z_0)$ using the z_n and the sample data. It is natural to estimate Σ_0 by the sample covariance matrix of $\{F(x_n,\xi_i), i = 1, \cdots, n\}$. It is problematic to estimate $d(f)_S(z_0)$ since $d(f)_S(\cdot)$ is not continuous. In [18], Lu found that $d(f)_S(z_0)(z_n - z_0)$ and $d(f_n)_S(z_n)(z_n - z_0)$ have the same asymptotic distribution. In [16], the authors gave a function $\Phi_N : \Pi_S^{-1}(O) \times \mathbb{R}^q \times \Omega \to \mathbb{R}^q$ such that

$$\sqrt{n}\Sigma_n^{-1/2}\Phi_n(z_n)(z_n - z_0) \xrightarrow{d} N(0, I_q),$$

where $\Phi_n(\cdot) = \Phi_n(z_n(\omega), \cdot, \omega)$. These results were used to construct confidence regions and individual confidence intervals for the SVIs in [16, 17, 18, 13]. Liu and Zhang [14] builded the confidence regions based on error bound. Chen, Pong and Wets [2] proposed a two-stage stochastic variational inequality model. Chen, Shapiro and Sun [3] studied convergence of sample average approximation of the two-stage stochastic generalized equations. For a more detailed discussion and further references in stochastic variational inequalities, we refer to the recent book [10] and its references. Gwinner, Jadamba, Khan and Raciti in [10] present a comprehensive treatment of stochastic variational inequalities, including theory and applications.

In this paper, we study the convergences in regimes of moderate deviations for the SAA solution. We obtain some moderate deviation principles. Roughly, for any positive sequence $\lambda(n), n \geq 1$ in regime of moderate deviations, i.e.,

$$\lim_{n\to\infty} \lambda(n) = \infty, \text{ and } \lim_{n\to\infty} \frac{\sqrt{n}}{\lambda(n)} = \infty, \tag{1.4}$$

we have that for any $\delta > 0$, the probabilities $P\left(\left\|\frac{\sqrt{n}}{\lambda(n)}\Sigma_0^{-1/2} d(f)_S(z_0)(z_n - z_0)\right\| > \delta\right)$ and

$$P\left(\left\|\frac{\sqrt{n}}{\lambda(n)}\Sigma_n^{-1/2} d(f_n)_S(z_n)(z_n - z_0)\right\| > \delta\right), \quad P\left(\left\|\frac{\sqrt{n}}{\lambda(n)}\Sigma_n^{-1/2} \Phi_n(z_n)(z_n - z_0)\right\| > \delta\right),$$

are approximated by $\exp\left\{-\frac{\lambda^2(n)\delta^2}{2}\right\}$, for example,

$$P\left(\left\|\frac{\sqrt{n}}{\lambda(n)}\Sigma_n^{-1/2} d(f_n)_S(z_n)(z_n - z_0)\right\| > \delta\right) \approx \exp\left\{-\frac{\lambda^2(n)\delta^2}{2}\right\}. \tag{1.5}$$

These give a precise rate of deviation between the SAA solution and the true solution when $1 \ll \lambda(n) \ll \sqrt{n}$, where $a_n \ll b_n$ means $a_n/b_n \to 0$. The results are also used to the hypotheses testings. We show that the rejection region constructed by the central limit theorem has the probability of the type II error with exponential decay speed.

The paper is organized as follows. In Section 2, we recall conceptions and properties on the normal map formulation, piecewise linear functions, and large deviations. The main results and their proofs are presented in Section 3 and Section 4, respectively. In Section 5, we apply the moderate deviations to the hypotheses testing problems. In the last section, we give some simulations and numerical results.

2 Preliminaries

In this section, we first recall conceptions and properties on the normal map formulation, piecewise linear functions, B-differentiability and Hadamard differentiability (cf. [21, 22],[27],[32]). Then we introduce some assumptions and moment conditions, and give some explanations for the moment conditions. Finally, we recall differentiability and deviation inequalities for the solution of the SVIs. Some notations and some basic properties in large deviations are also included in this section (cf. Dembo and Zeitouni [4]).

2.1 Conceptions

A subset K of $\mathbb{R}^q$ is called a cone if $\mu x \in K$ whenever $x \in K$ and μ is a positive real number. For a set $C \subset \mathbb{R}^q$, $\text{int}\,(C)$ or C^o denotes its interior, and $\text{cone}\,(C)$ is the smallest cone that contains it. A subset C of $\mathbb{R}^q$ is said to be affine if $\lambda x + (1-\lambda)y \in C$ for every $x, y \in C$ and real number λ. The affine hull aff (C) of a convex set C is defined to be the intersection of all the affine sets which contains C. The related interior ri (C) of C is defined as the interior which results when C is regarded as a subset of aff (C).

Normal map. For any function g from a subset O of $\mathbb{R}^q$ to $\mathbb{R}^q$ and for any closed and convex set C in $\mathbb{R}^q$, the *normal map* induced by g and C is a function $g_C : \Pi_C^{-1}(O) \to \mathbb{R}^q$, defined as

$$g_C(z) = g(\Pi_C(z)) + (z - \Pi_C(z)) \text{ for each } z \in \Pi_C^{-1}(O). \tag{2.1}$$

where $\Pi_C(z)$ denotes the Euclidean projection of z on C, i.e., $\|\Pi_C(z) - z\| = \inf_{y \in C} \|y - z\|$, and $\Pi_C^{-1}(O)$ is the set of points $z \in \mathbb{R}^q$ such that $\Pi_C(z) \in O$.

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

$$(f_0)_S(z) = f_0(\Pi_S(z)) + (z - \Pi_S(z)) \text{ for each } z \in \Pi_S^{-1}(O), \tag{2.2}$$

If a point $x \in S \cap O$ satisfies (1.1), then the point $z = x - f_0(x)$ satisfies $\Pi_S(z) = x$ and

$$(f_0)_S(z) = 0. \tag{2.3}$$

Conversely, if z satisfies (2.3), then $x = \Pi_S(z)$ satisfies $x - f_0(x) = z$ and solves (1.1). Thus, equation (2.3) is an equivalent formulation for (1.1), and is referred to as the normal map formulation of (1.1).

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

$$(f_n)_S(z) = f_n(\Pi_S(z)) + (z - \Pi_S(z)) \text{ for each } z \in \Pi_S^{-1}(O). \tag{2.4}$$

The normal map formulation of the SAA problem (1.3) is

$$(f_n)_S(z) = 0. \tag{2.5}$$

where (2.5) is related to (1.3) in the same manner as (2.3), that is, if a point $x \in S \cap O$ satisfies (1.3), then the point $z = x - f_n(x)$ satisfies $\Pi_S(z) = x$ and (2.5). Conversely, if z satisfies (2.5), then $x = \Pi_S(z)$ satisfies $x - f_n(x) = z$ and solves (1.3).

Polyhedral subdivision. A polyhedral subdivision of $\mathbb{R}^q$ is defined to be a finite collection of convex polyhedra, $\Gamma = \{\gamma_1, \ldots, \gamma_m\} \subset \mathbb{R}^q$, satisfying the following three conditions:

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

If each of the γ_i is a polyhedra cone, then Γ is referred to as a conical subdivision.

Piecewise affine function. A continuous function $f : \mathbb{R}^q \to \mathbb{R}^m$ is piecewise affine if there exists a finite family of affine functions $f_j : \mathbb{R}^q \to \mathbb{R}^m$, $j = 1, \ldots, k$, such that for all $x \in \mathbb{R}^q$ $f(x) \in \{f_1(x), \ldots, f_k(x)\}$. The affine functions f_i, $i = 1, \ldots, k$, are referred to as the selection functions of f. When each selection function is linear, the function f is called piecewise linear.

Normal manifold. Π_S is a piecewise affine function. It coincides with an affine function on each of a family of finitely many q-dimensional polyhedral convex sets. This family is called the normal manifold of S, and each set in this family is called an q-cell. The relative interiors of all cells form a partition of $\mathbb{R}^q$.

B-differentiable. We denote a norm in a normed linear space by $\|\cdot\|$. Let $\mathcal{X}$, and $\mathcal{Y}$ be two Banach spaces, and let U be an open subset of $\mathcal{X}$. A function $g : U \to \mathcal{Y}$ is said to be *B-differentiable* at a point $x \in U$ if there is a positively homogeneous function $G : U \to \mathcal{Y}$, such that

$$g(x+v) = g(x) + G(v) + o(v). \tag{2.6}$$

Recall that $o(v)$ means $\lim_{\|v\|\to 0} \|o(v)\|/\|v\| = 0$ and a function $G : U \to \mathcal{Y}$ is positively homogeneous if $G(\lambda x) = \lambda G(x)$ for any $\lambda \geq 0$ and $x \in U$ with $\lambda x \in U$. Such a function G is called the B-derivative of g at x and is denoted as $dg(x)$. If $dg(x)$ is a bounded linear function, then g is Fréchet differentiable at x and $dg(x)$ is the Fréchet derivative of g at x.

Hadamard differentiability. Let us recall some conceptions of Hadamard differentiability (see Shapiro [27], van der Vaar and Wellner [32]). Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces. A map g defined on a subset $\mathcal{D}_g$ of $\mathcal{X}$ with values in $\mathcal{Y}$ is called Hadamard (directional) differentiable at x if there exists a continuous mapping $G : \mathcal{X} \mapsto \mathcal{Y}$ such that

$$\lim_{n\to\infty} \frac{g(x+t_n h_n) - g(x)}{t_n} = G(h) \tag{2.7}$$

holds for all sequences t_n converging to $0+$ and h_n converging to h in $\mathcal{X}$ such that $x + t_n h_n \in \mathcal{D}_g$ for every n. Such a function G is called the H-derivative of g at x and is also denoted as $dg(x)$.

When g is Lipschitz continuous in a neighborhood of x, Hadamard directional differentiability is equivalent to B-differentiability (Shapiro [26]).

Π_S is B-differentiable. For all points z in the relative interior of a cell, the B-derivative $d\Pi_S(z)$ is the same.

Let $f : \mathbb{R}^q \to \mathbb{R}^m$ be a piecewise affine function with the corresponding subdivision Γ. Set

$$\Gamma(x) = \{\gamma \in \Gamma; x \in \gamma\} \tag{2.8}$$

and

$$\Gamma'(x) = \{\text{cone}(\gamma - x); \gamma \in \Gamma(x)\}. \tag{2.9}$$

Then, the family $\Gamma'(x)$ is a conical subdivision of $\mathbb{R}^q$. The B-derivative $df(x)$ of f at x is a piecewise linear function from $\mathbb{R}^q$ to $\mathbb{R}^m$, whose corresponding subdivision is exactly $\Gamma'(x)$. From Theorem 2.2 in Lu [18], for any $x \in \mathbb{R}^q$ and $y \in \cup_{\gamma \in \Gamma(x)} \gamma$,

$$df(x)(y-x) = -df(y)(x-y). \tag{2.10}$$

2.2 Assumptions.

In the rest of the paper, let U be a nonempty compact subset of O, and let $C^1(U, \mathbb{R}^q)$ denote the Banach space of continuously differentiable mappings $f : U \to \mathbb{R}^q$, equipped with the norm

$$\|f\|_{1,U} = \sup_{x\in U} \|f(x)\| + \sup_{x\in U} \|df(x)\|. \tag{2.11}$$

Two basic assumptions in this paper are as follows.

(A1). The map $x \mapsto F(x, \xi(\omega))$ is continuously differentiable on O for a.s. $\omega \in \Omega$, and there exists a positive random variable κ such that

$$\|F(x, \xi(\omega)) - F(x', \xi(\omega))\| + \|d_x F(x, \xi(\omega)) - d_x F(x', \xi(\omega))\| \le \kappa(\omega)\|x - x'\|, \tag{2.12}$$

for all $x, x' \in O$ and a.s. $\omega \in \Omega$, where the notation $d_x F(x, \xi(\omega))$ stands for the partial derivative of F w.r.t. x, a $q \times q$ matrix. Furthermore, $E(\kappa^2) < \infty$, and

$$E\left(\|F(x, \xi(\omega))\|^2\right) < \infty, \quad E\left(\|d_x F(x, \xi(\omega))\|^2\right) < \infty, \quad \text{for all } x \in O. \tag{2.13}$$

(A2). Suppose that x_0 solves the variational inequality (1.1) and that x_0 belongs to the interior of U. Let $z_0 = x_0 - f_0(x_0)$, $L = df_0(x_0)$, $K = T_S(x_0) \cap \{z_0 - x_0\}^{\perp}$, and assume that the normal map L_K induced by L and K, defined as $L_K(h) = L(\Pi_K(h)) + h - \Pi_K(h)$ for each $h \in \mathbb{R}^q$, is a homeomorphism from $\mathbb{R}^q$ to $\mathbb{R}^q$, where the tangent cone $T_S(x_0)$ is defined by

$$T_S(x_0) = \{v \in \mathbb{R}^q, \text{ there exists } t > 0 \text{ such that } x_0 + tv \in S\}. \tag{2.14}$$

Let

$$M_x(t) = E\left(\exp\left\{\langle t, F(x, \xi) - f_0(x)\rangle\right\}\right), \quad t \in \mathbb{R}^q, \ x \in U$$

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

$$\mathcal{M}_x(T) = E\left(\exp\left\{\langle T, d_x F(x, \xi) - d_x f_0(x)\rangle\right\}\right), \quad T \in \mathbb{R}^{q \times q}, \ x \in U$$

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

The following assumptions will be required in this paper.

(A3). There exists a constant $\theta > 0$ such that

$$M_x(t) \le \exp\left\{\frac{\theta^2 \|t\|^2}{2}\right\} \text{ for any } x \in U, \ t \in \mathbb{R}^q. \tag{2.15}$$

(A4). There exists a constant $\vartheta > 0$ such that

$$\mathcal{M}_x(T) \le \exp\left\{\frac{\vartheta^2 \|T\|^2}{2}\right\} \text{ for any } x \in U, \ T \in \mathbb{R}^{q \times q}. \tag{2.16}$$

(A5). There exists a constant $\zeta > 0$ such that

$$E\left(\exp\{\zeta|\kappa|\}\right) < \infty, \tag{2.17}$$

(A6). There exists a constant $\zeta > 0$ such that

$$\sup_{x \in O} E\left(\exp\{\zeta \|F(x, \xi) F(x, \xi)^T\|\}\right) < \infty, \tag{2.18}$$

Remark 2.1. *The (A1) is a basic condition in the variational inequality (cf. [28, Theorem 5.15 in Section 5.2.1]). Some sufficient conditions on (A2) were discussed in [18, P.1465] (also see [21, 24]). If for some $\delta_0 \in (0, 1)$,*

$$\sup_{x \in U} E\left(\exp\{\delta_0(\|F(x, \xi)\|^2 + \|d_x F(x, \xi)\|^2)\}\right) < \infty, \tag{2.19}$$

then it is obvious that (A6) holds. By the Taylor expansion, we have

$$
\begin{aligned}
M_x(t) =& 1 + \sum_{k=2}^{\infty} \frac{1}{k!} E\left(\langle t, F(x,\xi) - f_0(x)\rangle^k\right) \\
\leq & 1 + \sum_{k=2}^{\infty} \frac{(2t)^k}{k!} E\left(\|F(x,\xi)\|^k\right) \\
\leq & 1 + \sum_{k=2}^{\infty} \frac{(4t^2)^k}{k!\delta_0^k} + \sum_{k=2}^{\infty} \frac{(2t)^k}{k!} E\left(\|F(x,\xi)\|^k I_{\{\|F(x,\xi)\|>2t/\delta_0\}}\right) \\
\leq & 1 + \sum_{k=2}^{\infty} \frac{(4t^2)^k}{k!\delta_0^k} + 4t^2 \sum_{k=2}^{\infty} \frac{\delta_0^{k-2}}{k!} E\left(\|F(x,\xi)\|^{2(k-1)}\right) \\
\leq & 1 + \sum_{k=2}^{\infty} \frac{(4t^2)^k}{k!\delta_0^k} + \frac{4t^2}{\delta_0} E\left(\exp\{\delta_0 \|F(x,\xi)\|^2\right) \\
\leq & \exp\left\{ \frac{4t^2}{\delta_0} E\left(\exp\{\delta_0 \|F(x,\xi)\|^2\right) \right\}.
\end{aligned}
$$

Therefore, (A3) is valid. Similarly, (A4) is valid. In particular, if $E(e^{\delta_0\|\xi\|^2}) < \infty$ for some $\delta_0 \in (0,1)$, and $F(x,\xi)$ satisfies the linear growth condition in ξ and κ in (A1) is also linear growth in $\|\xi\|$, i.e., for some constant $L \in (0,\infty)$, for any $x, x' \in U$,

$$\|F(x,\xi)\| + \|d_x F(x,\xi)\| \leq L(1 + \|\xi\|),$$

and

$$\|F(x,\xi) - F(x',\xi)\| + \|d_x F(x,\xi) - d_x F(x',\xi)\| \leq L(1 + \|\xi\|)\|x - x'\|,$$

then (A3)–(A6) hold.

2.3 Differentiability and deviation inequalities for the solution of the SVI.

By Theorem 4 in [13], under the assumptions (A1)-(A5), there exist positive real numbers β_1, μ_1, M_1 and σ_1 such that for each $\epsilon > 0$ and each $n \geq 1$,

$$P\left(\|f_n - f_0\|_{1,U} \geq \epsilon\right) \leq \beta_1 e^{-n\mu_1} + \frac{M_1}{\epsilon^q} \exp\left\{-\frac{n\epsilon^2}{\sigma_1}\right\}. \tag{2.20}$$

Because f_0 is differentiable at x_0 and Π_S is B-differentiable, the normal map $(f_0)_S$ is B-differentiable at z_0 by the chain rule of B-differentiability, with

$$d(f_0)_S(z_0)(h) = df_0(x_0)(d\Pi_S(z_0)(h)) + h - d\Pi_S(z_0)(h). \tag{2.21}$$

It was shown in [22] that L_K is exactly $d(f_0)_S(z_0)$, the B-derivative of $(f_0)_S$ at z_0.

Since $d\Pi_S(z_0)$ is a piecewise affine function,

$$\|d\Pi_S(z_0)\| := \sup_{h\in\mathbb{R}^q} \frac{\|d\Pi_S(z_0)(h)\|}{\|h\|} < \infty, \quad \|d(f_0)_S(z_0)\| := \sup_{h\in\mathbb{R}^q} \frac{\|d(f_0)_S(z_0)(h)\|}{\|h\|} < \infty. \tag{2.22}$$

Lemma 2.1. *[Lemma 1 in [16]] Assume that the assumptions (A1) and (A2) hold. Then there exists $\delta > 0$ such that*

$$\inf \left\{ \frac{\|L_K(z_1) - L_K(z_2)\|}{\|z_1 - z_2\|}; \quad z_1 \neq z_2, z_1, z_2 \in \mathbb{R}^q \right\} \geq \delta, \tag{2.23}$$

and for each $\eta > \delta^{-1}$, one can choose neighborhoods U_0 of x_0 in U, V_0 of z_0 in $\mathbb{R}^q$, and Γ_0 of f_0 in $C^1(U, \mathbb{R}^q)$, and a function $z : \Gamma_0 \to \mathbb{R}^q$ satisfying

1. *$z(f_0) = z_0$;*
2. *for each $f \in \Gamma_0$, $z(f)$ is the unique point in V_0 with $(f(\cdot))_S(z(f)) = 0$, and $x(f) = \Pi_S(z(f))$ is the unique point in U_0 with $0 \in f(x(f)) + N_S(x(f))$;*
3. *z is Lipschitz on Γ_0 with Lipschitz constant η.*

Furthermore, the functions $z(f)$ and $x(f)$ are B-differentiable at f_0 with

$$dz(f_0)(g) = (L_K)^{-1} \circ (-g(x_0)), \text{ and } dx(f_0) = \Pi_K \circ dz(f_0). \tag{2.24}$$

For each $\omega \in \Omega$ and integer n with $f_n \in \Gamma_0$, set

$$z_n = z(f_n) \text{ and } x_n = \Pi_S(z_n).$$

Then z_n is the unique solution for (2.5) in V_0, and x_n is the unique solution for the variational inequality $0 \in f_n(\cdot) + N_S(\cdot)$ in U_0.

Take $\epsilon_0 > 0$ such that $\{f \in C^1(U, \mathbb{R}^q); \|f - f_0\|_{1,U} < \epsilon_0/\delta\} \subset \Gamma_0$. Then by Theorem 7 in [13], under the assumptions (A1)-(A5), there exist positive real numbers β_0, μ_0, M_0 and σ_0 such that for each $\epsilon > 0$, each $n \geq 1$ and each $\epsilon \in (0, \epsilon_0)$,

$$P(\|x_n - x_0\| \geq \epsilon) \leq P(\|z_n - z_0\| \geq \epsilon) \leq \beta_0 e^{-n\mu_0} + \frac{M_0}{\epsilon^q} \exp\left\{ -\frac{n\epsilon^2}{\sigma_0} \right\}. \tag{2.25}$$

Let Σ_0 denote the covariance matrix of $F(x_0, \xi)$ and let $0 \leq \mu_1 \leq \mu_2 \leq \cdots \leq \mu_q$ be the all eigenvalues of Σ_0. Set $\rho = \min\{\mu_i > 0; 1 \leq i \leq q\}$. Let l be the number of positive eigenvalues of Σ_0 counted with regard to their algebraic multiplicities. As the same in [18], we decompose Σ_0 as

$$\Sigma_0 = U_0^T \begin{bmatrix} D_0 & 0 \\ 0 & 0 \end{bmatrix} U_0 = \begin{bmatrix} (U_0)_1^T & (U_0)_2^T \end{bmatrix} \begin{bmatrix} D_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (U_0)_1 \\ (U_0)_2 \end{bmatrix}, \tag{2.26}$$

where U_0 is a $q \times q$ orthogonal matrix, D_0 is diagonal matrix where the diagonal elements are decreasing positive eigenvalues, and the elements in matrices $(U_0)_1$ and $(U_0)_2$ are the first l and the last $q - l$ rows of U_0 respectively.

Let Σ_n denote the sample covariance matrix of $\{F(x_n, \xi_i)\}_{i=1}^n$:

$$\Sigma_n = \frac{1}{n-1} \sum_{i=1}^n \left(F(x_n, \xi_i) - \frac{1}{n} \sum_{j=1}^n F(x_n, \xi_j) \right) \left(F(x_n, \xi_i) - \frac{1}{n} \sum_{j=1}^n F(x_n, \xi_j) \right)^T.$$

Take ρ_0 with $0 < \rho_0 < \rho$. Let $0 \le \mu_1^n \le \mu_2^n \le \cdots \le \mu_q^n$ be the all eigenvalues of Σ_n and let l_n be the number of $\{1 \le i \le q; \mu_i \ge \rho_0\}$. Next, we write Σ_n as

$$\Sigma_n = U_n^T \Delta_n U_n, \tag{2.27}$$

where U_n is a $q \times q$ orthogonal matrix, Δ_n is diagonal matrix where the diagonal elements are decreasing positive eigenvalues. Let D_n be the upper-left submatrix of Δ_n whose diagonal elements are at least ρ_0. Let $(U_n)_1$ be the submatrix of U_n that consists of its first l_n rows, and let $(U_n)_2$ be the submatrix that consists of the remaining rows of U_n.

2.4 Large deviations

We conclude the section by introducing some notations, some basic properties in large deviations (see Dembo and Zeitouni [4]). Let $(\mathcal{X}, \rho)$ be a separable metric space and let $\lambda(n), n \ge 1$ be a sequence of positive numbers satisfying $\lambda(n) \to \infty$. Let Z_n: $\Omega \to \mathcal{X}$, $n \ge 1$, be a sequence of measurable maps and let $I : \mathcal{X} \to [0, \infty]$ be a good rate function, i.e., $\{I \le l\}$ is compact for all $l < \infty$. $\{Z_n, n \ge 1\}$ is said to satisfy a large deviation principle (LDP) with speed $\lambda^2(n)$ and with good rate function I if for any open subset G of $\mathcal{X}$,

$$\liminf_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(Z_n \in G\right) \ge -\inf_{x\in G} I(x), \tag{2.28}$$

and for any closed subset F of $\mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(Z_n \in F\right) \le -\inf_{x\in F} I(x). \tag{2.29}$$

The exponential approximation ([4, Theorem 4.2.13]). Let Z_n and Z_n', $n \ge 1$, be two sequence of measurable maps taking their values in $\mathcal{X}$. If $\{Z_n, n \ge 1\}$ satisfies a large deviation principle with speed $\lambda^2(n)$ and with good rate function I and for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\rho(Z_n, Z_n') \ge \epsilon\right) = -\infty, \tag{2.30}$$

then $\{Z_n', n \ge 1\}$ also satisfies the large deviation principle with the speed $\lambda^2(n)$ and with the good rate function I.

The contraction principle ([4, Theorem 4.2.1]). Let $\mathcal{X}$ and $\mathcal{Y}$ be two separable metric spaces and $f : \mathcal{X} \to \mathcal{Y}$ is continuous. If $\{Z_n, n \ge 1\}$ satisfies a large deviation principle with speed $\lambda^2(n)$ and with good rate function I, then $\{Z_n' = f(Z_n), n \ge 1\}$ satisfies the large deviation principle with the speed $\lambda^2(n)$ and with the good rate function I_f, where the rate function I_f given by

$$I_f(y) := \inf\{I(x); x \in \mathcal{X}, f(x) = y\},\ y \in \mathcal{Y}. \tag{2.31}$$

The following delta method in large deviation theory is given by Gao and Zhao in [8].

The delta method (Theorem 3.1 in Gao and Zhao [8]). Let $\mathcal{X}$ and $\mathcal{Y}$ be two separable Banach spaces. Let Φ: $\mathcal{D}_\Phi \subset \mathcal{X} \to \mathcal{Y}$ be Hadamard (directional) differentiable at z. Let Z_n: $\Omega \to \mathcal{D}_\Phi$, $n \ge 1$, be a sequence of maps and let r_n, $n \ge 1$, be a sequence of positive real numbers satisfying $r_n \to \infty$. If $\{r_n(Z_n - z), n \geq 1\}$ satisfies a large deviation principle with speed $\lambda^2(n)$ and with good rate function I, then, for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\|r_n(\Phi(Z_n) - \Phi(z)) - d\Phi(z)(r_n(Z_n - z))\| \geq \delta\right) = -\infty. \tag{2.32}$$

In particular, $\{r_n(\Phi(Z_n) - \Phi(z)), n \geq 1\}$ and $\{d\Phi(z)(r_n(Z_n - z)), n \geq 1\}$ satisfy a same large deviation principle.

3 Main results

In this section, we state the main results which are four moderate deviation principles for the solution of the SVIs, i.e., large deviations associated with the central limit theorems of the solution of the SVIs. The results give some convergence rates of the solution of the SVIs. The first basic result is the following theorem.

Theorem 3.1. *Suppose that the assumptions (A1)–(A5) hold. Then*

$$\left\{\frac{\sqrt{n}}{\lambda(n)} L_K(z_n - z_0), n \geq 1\right\}$$

satisfies the LDP in $\mathbb{R}^q$ with speed $\lambda^2(n)$ and with rate function I_K defined by

$$\begin{aligned} I_K(y) &= \sup_{v\in\mathbb{R}^q} \left\{\langle v, y\rangle - \frac{1}{2} v^T \Sigma_0 v\right\} \\ &= \sup_{v\in\mathbb{R}^q} \left\{\langle v, y\rangle - \frac{1}{2} \left\|D_0^{1/2}(U_0)_1 v\right\|^2\right\}, \quad y \in \mathbb{R}^q. \end{aligned} \tag{3.1}$$

In particular, if Σ_0 is invertible, then $I_K(y) = \frac{1}{2} y^T \Sigma_0^{-1} y$, and $\left\{\frac{\sqrt{n}}{\lambda(n)} \Sigma_0^{-1} L_K(z_n - z_0), n \geq 1\right\}$ satisfies the LDP in $\mathbb{R}^q$ with speed $\lambda^2(n)$ and with rate function $I(y) = \|y\|^2/2$.

Lu [18] proved that $L_K(z_n - z_0)$ and $d(f_n)_S(z_n)(z_n - z_0)$ have the same asymptotic distribution. We will prove the following results.

Theorem 3.2. *Suppose that the assumptions (A1)–(A5) hold. Then*

$$\left\{\frac{\sqrt{n}}{\lambda(n)} d(f_n)_S(z_n)(z_n - z_0), n \geq 1\right\} \tag{3.2}$$

satisfies the LDP in $\mathbb{R}^q$ with speed $\lambda^2(n)$ and with rate function I_K defined by (3.1).

Theorem 3.3. *Suppose that the assumptions (A1)-(A6) hold.*

(1). Assume that Σ_0 is nonsingular. Then

$$\left\{\frac{\sqrt{n}}{\lambda(n)} \Sigma_n^{-1/2} d(f_n)_S(z_n)(z_n - z_0), n \geq 1\right\} \tag{3.3}$$

satisfies the LDP in $\mathbb{R}^q$ with speed $\lambda^2(n)$ and with rate function $I(y) = \|y\|^2/2$.

(2). Assume that Σ_0 is singular. Then

$$\left\{\frac{n}{\lambda^2(n)}\left\|D_n^{-1/2}(U_n)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2, n\geq 1\right\} \tag{3.4}$$

satisfies the LDP in $\mathbb{R}$ with speed $\lambda^2(n)$ and with rate function

$$J(y)=\begin{cases} y/2 & \text{if } y\geq 0\\ \\ +\infty & \text{otherwise.}\end{cases} \tag{3.5}$$

In order to state the last moderate deviation result, let us recall the definition of the estimator Φ_n of the $d(f_0)_S(z_0)$ (see [13][16]). For each cell C_i in the normal manifold of S, define a function $d_i:\mathbb{R}^q\to\mathbb{R}$ by

$$d_i(z)=d(z,C_i)=\min_{x\in C_i}\|x-z\|, \tag{3.6}$$

Note that $d\Pi_S(z)$ is the same function on the relative interior of a cell. We can define a function $\Psi_i:\mathbb{R}^q\to\mathbb{R}^q$ by

$$\Psi_i(\cdot)=d\Pi_S(z)(\cdot) \text{ for any } z\in \mathrm{ri}(C_i). \tag{3.7}$$

Let $g:\mathbb{N}\to\mathbb{R}$ be a positive function satisfying

(i). $\lim_{n\to\infty} g(n)=\infty, \quad \lim_{n\to\infty}\frac{n}{g^2(n)}=\infty,$

(ii). $\lim_{n\to\infty} g^q(n)\exp\left\{-\frac{\theta_0 n^2}{g^2(n)}\right\}=0$ for $\theta_0=\min\left\{\frac{1}{4\sigma_0},\frac{1}{4\sigma_1},\frac{1}{4\sigma_0(E(\kappa))^2}\right\}$,

(iii). $\lim_{n\to\infty}\frac{n^{q/2}}{g^q(n)}\exp\left\{-\theta g^2(n)\right\}=0$ for each positive real number θ,

where σ_0 and σ_1 are as in (2.25) and (2.20) respectively and κ as in (A1).

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

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

$$\Lambda_n(z)(h)=\Psi_{i_0}(h), \tag{3.8}$$

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

$$\Phi_n(z,h,\omega)=df_n(\Pi_S(z))(\Lambda_n(z)(h))+h-\Lambda_n(z)(h). \tag{3.9}$$

Set $\Phi_n(z_n)(h)=\Phi_n(z_n(\omega),h,\omega)$. Let z_n^* denote a point in the relative interior of the cell C_{i_0} associated with (n,z_n). Then $d\Pi_S(z_n^*)=\Psi_{i_0}$ and

$$\Phi_n(z_n)(h)=df_n(\Pi_S(z_n))(d\Pi_S(z_n^*)(h))+h-d\Pi_S(z_n^*)(h). \tag{3.10}$$

By Lemma 3 in [16], Λ_n is jointly continuous with respect to (z,h). By Theorem 9 in [16], there exist positive number τ and integer n_0 such that for each $n\geq n_0$,

$$\begin{aligned}&P\left(g(n)\sup_{h\in\mathbb{R}^q}\frac{\|\Lambda_n(z_n)(h)-d\Pi_S(z_0)(h)\|}{\|h\|}\geq\tau\right)\\ &\leq P(\|z_n-z_0\|\geq\gamma_0/2)+P(g(n)\|z_n-z_0\|\geq 1/2)\\ &\leq 2P(g(n)\|z_n-z_0\|\geq 1/2),\end{aligned} \tag{3.11}$$

where $\gamma_0>0$ is the minimum of $d_i(z_0)$ among all of the cells C_i such that $z_0\notin C_i$.

Theorem 3.4. *Let*

$$\lim_{n\to\infty} \frac{n}{\lambda^2(n)g^2(n)} = \infty \ \text{and} \ \limsup_{n\to\infty} \frac{\log n}{\lambda^2(n)} < \infty. \tag{3.12}$$

Suppose that the assumptions (A1)–(A6) hold.

(1). Assume that Σ_0 is nonsingular. Then

$$\left\{ \frac{\sqrt{n}}{\lambda(n)} \Sigma_n^{-1/2} \Phi_n(z_n)(z_n - z_0), n \geq 1 \right\} \tag{3.13}$$

satisfies the LDP in $\mathbb{R}^q$ with speed $\lambda^2(n)$ and with rate function $I(y) = \|y\|^2/2$.

(2). Assume that Σ_0 is singular. Then

$$\left\{ \frac{n}{\lambda^2(n)} \left\| D_n^{-1/2}(U_n)_1 \Phi_n(z_n)(z_n - z_0) \right\|^2, n \geq 1 \right\} \tag{3.14}$$

satisfies the LDP in $\mathbb{R}$ with speed $\lambda^2(n)$ and with rate function $J(y)$ defined by (3.5).

Remark 3.1. *If $g(n) = o(n^p)$ for some $0 < p < 1/2$, then we can take $\lambda(n) = n^{1/2-p}$.*

4 Proofs of Main Results

In this section, we show the main results. Our proofs are based on the moderate deviation principle for i.i.d. random variables, the exponential approximation ([4, Theorem 4.2.13]), Delta method in large deviations (Theorem 3.1 in Gao and Zhao [8])) and some properties of the SVIs. Note that two exponential approximation random sequences have the same large deviation principle.

4.1 Proof of Theorem 3.1

We first apply the moderate deviation principle for i.i.d. random variables to derive the moderate deviations for $\{f_n - f_0, n \geq 1\}$ (Lemma 4.2), and then use the exponential approximation and Delta method in large deviations to show Theorem 3.1.

Let us first introduce a moderate deviation principle (MDP) in a separable Banach space (cf. Theorem 1.1. in [1], Theorem 5 in [34]).

Lemma 4.1. *Let $\mathbb{B}$ be a separable Banach space and let $\mathbb{B}^*$ be its dual space. Let $\{Z_n, n \geq 1\}$ be an i.i.d $\mathbb{B}$-valued random sequence satisfying*

$$E(g(Z_1)) = 0, \quad \text{and} \quad E(g^2(Z_1)) < \infty \ \text{for any} \ g \in \mathbb{B}^*$$

and

$$\frac{1}{\sqrt{n}\lambda(n)} \sum_{k=1}^{n} Z_k \to 0 \ \text{in probability.}$$

Assume that

$$E\left(\exp\{\epsilon \|Z_1\|\}\right) < \infty \ \text{for some} \ \epsilon > 0.$$

Then $\left\{\frac{1}{\sqrt{n}\lambda(n)}\sum_{k=1}^n Z_k, n\geq 1\right\}$ *satisfies the LDP in* $\mathbb{B}$ *with speed* $\lambda^2(n)$ *and with good rate function*

$$I(z)=\sup_{g\in\mathbb{B}^*}\left\{g(z)-\frac{1}{2}E\left(g^2(Z_1)\right)\right\}. \tag{4.1}$$

In particular, for any $r>0$

$$\lim_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|Z_n\|\geq r\right)=-\inf_{\|z\|\geq r}I(z)=-r^2\inf_{\|z\|\geq 1}I(z), \tag{4.2}$$

since

$$I(rz)=r^2I(z) \text{ for any } r\in\mathbb{R} \text{ and } z\in\mathbb{B}. \tag{4.3}$$

It follows from the compactness of $\{I(z)\leq L\}$ *for* $L\in(0,\infty)$ *that* $\inf_{\|z\|\geq 1}I(z)>0$.

Lemma 4.2. *Let the assumptions (A1)–(A5) hold. Then*

(1). $\left\{\frac{\sqrt{n}}{\lambda(n)}(f_n-f_0), n\geq 1\right\}$ *satisfies the LDP in* $C^1(U,\mathbb{R}^q)$ *with speed* $\lambda^2(n)$ *and with good rate function* $I(f)$ *satisfying*

$$I(rf)=r^2I(f) \text{ for any } r\in\mathbb{R} \text{ and } f\in C^1(U,\mathbb{R}^q), \quad \text{and} \quad \inf_{\|f\|_{1,U}\geq 1}I(f)>0. \tag{4.4}$$

(2). $\left\{\frac{\sqrt{n}}{\lambda(n)}(f_n(x_0)-f_0(x_0)), n\geq 1\right\}$ *satisfies the LDP in* $\mathbb{R}^q$ *with speed* $\lambda^2(n)$ *and with good rate function* I_K *defined by (3.1).*

Proof. By Theorem 5 in [16], we have that

$$\frac{\sqrt{n}}{\lambda(n)}\|f_n-f_0\|_{1,U}\to 0 \text{ in probability.}$$

Thus, applying Lemma 4.1 to $\mathbb{B}=C^1(U,\mathbb{R}^q)$ and $Z_n=F(\cdot,\xi_n)-f_0(\cdot)$, we obtain (1). Taking $\mathbb{B}=\mathbb{R}^q$ and $\tilde{Z}_n=F(x_0,\xi_n)-f_0(x_0)$ in Lemma 4.1, we get (2). □

Lemma 4.3. *Suppose that the assumptions (A1)–(A5) hold. Then for any* $\epsilon>0$,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\|L_K(z_n-z_0)+(f_n(x_0)-f_0(x_0))\|\geq\epsilon\right)=-\infty, \tag{4.5}$$

and

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|z_n-z_0\|\geq\epsilon\right)=-\infty, \tag{4.6}$$

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|x_n-x_0\|\geq\epsilon\right)=-\infty. \tag{4.7}$$

Proof. By Lemma 2.1, the functions $z(f)$ and $x(f)$ are B-differentiable at f_0 with $dz(f_0)(g)=(L_K)^{-1}\circ(-g(x_0))$. Thus, by the delta method in large deviation theory (see (2.32), or Theorem 3.1 in Gao and Zhao [8]) and Lemma 4.2, we have that

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\left\|(z_n-z_0)+L_K^{-1}(f_n(x_0)-f_0(x_0))\right\|\geq\epsilon\right)=-\infty,$$

which implies (4.5) since L_K is homeomorphism from $\mathbb{R}^q$ to $\mathbb{R}^q$.

(4.6) and (4.7) are consequences of (2.25). Here, we give a proof using Lemma 4.2. Choose $\epsilon > 0$ such that $B(f_0, \epsilon) := \{f \in C^1(U, \mathbb{R}^q); \|f - f_0\|_{1,U} < \epsilon\} \subset \Gamma_0$. Thus, by Lemma 4.2,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(f_n \notin \Gamma_0\right) \le \limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\|f_n - f_0\| \ge \epsilon\right) = -\infty.$$

Note that on $\{f_n \in \Gamma_0\}$, $\|L_K(z_n - z_0)\| \ge \delta\|z_n - z_0\|$. By

$$P\left(\|z_n - z_0\| \ge \epsilon\right) \le P\left(f_n \notin \Gamma_0\right) + P\left(f_n \in \Gamma_0, \|z_n - z_0\| \ge \epsilon\right),$$

we have that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\|z_n - z_0\| \ge \epsilon\right) \\
\le &\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\|L_K(z_n - z_0)\| \ge \delta\epsilon\right) \\
= &\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \max\Big\{ \log P\left(\|f_n(x_0) - f_0(x_0)\| \ge \delta\epsilon/2\right), \\
&\qquad \log P\left(\|L_K(z_n - z_0) - (f_n(x_0) - f_0(x_0))\| \ge \delta\epsilon/2\right) \Big\} = -\infty.
\end{aligned}$$

Finally, For each $\omega \in \Omega$ and integer n with $f_n \in \Gamma_0$, $x_n = \Pi_S(z_n)$ and $x_0 = \Pi(z_0)$, thus (4.7) holds. □

Proof of Theorem 3.1. By Lemma 4.3 and the exponential approximation ([4, Theorem 4.2.13]), we obtain the first claim of Theorem 3.1 which implies the second one by the contraction principle. □

4.2 Proof of Theorem 3.2

Proof of Theorem 3.2 is completed by two exponential approximations, i.e., (4.8) and (4.9).

Lemma 4.4. *Suppose that the assumptions (A1)–(A5) hold. Then for any $\epsilon > 0$,*

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)} \|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \ge \epsilon\right) = -\infty. \tag{4.8}$$

Proof. Recall that z_0 belongs to the interior of $\cup_{\gamma\in\Gamma(z_0)}\gamma$. By (4.6),

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(z_n \notin \cup_{\gamma\in\Gamma(z_0)}\gamma\right) = -\infty.$$

From [18, Theorem 2.2], when $z_n \in \cup_{\gamma\in\Gamma(z_0)}\gamma$,

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

Thus

$$P\left(\frac{\sqrt{n}}{\lambda(n)} \|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \ge \epsilon\right) \le P\left(z_n \notin \cup_{\gamma\in\Gamma(z_0)}\gamma\right),$$

and so, (4.8) holds.

□

Proof of Theorem 3.2. By the exponential approximation ([4, Theorem 4.2.13]), it is sufficient that for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)}\|d(f_0)_S(z_0)(z_n - z_0) + d(f_n)_S(z_n)(z_0 - z_n)\| > \epsilon\right) = -\infty. \quad (4.9)$$

Let Γ_1 be a closed subset of Γ_0 such that $f_0 \in \Gamma_1^o$ (the interior of Γ_1). Recall that $x_0 = \Pi_S(z_0)$ is a solution to (1.1). On $\{f_n \in \Gamma_1\}$, $x_n = \Pi_S(z_n)$ is a solution to (1.3). From (2.21), we have

$$d(f_0)_S(z_0)(z_n - z_0) = df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) + z_n - z_0 - d\Pi_S(z_0)(z_n - z_0). \quad (4.10)$$

Similarly, on $\{f_n \in \Gamma_1\}$,

$$d(f_n)_S(z_n)(z_0 - z_n) = df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n)) + z_0 - z_n - d\Pi_S(z_n)(z_0 - z_n). \quad (4.11)$$

It follows that on $\{f_n \in \Gamma_1\}$,

$$\begin{aligned}
&\|d(f_0)_S(z_0)(z_n - z_0) + d(f_n)_S(z_n)(z_0 - z_n)\| \\
\le&\|df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) + df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n))\| \\
&+ \|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \\
\le&\|df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) - df_n(x_n)(d\Pi_S(z_0)(z_n - z_0))\| \\
&+ \|df_n(x_n)(d\Pi_S(z_0)(z_n - z_0)) + df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n))\| \\
&+ \|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \\
\le& \|df_0(x_0) - df_n(x_n)\| \, \|d\Pi_S(z_0)(z_n - z_0)\| \\
&+ \|df_n(x_n) - df_0(x_0)\|\|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \\
&+ (\|df_0(x_0)\| + 1)\|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\|.
\end{aligned}$$

Therefore, for any $\epsilon > 0, \delta > 0$,

$$\begin{aligned}
&\left\{\frac{\sqrt{n}}{\lambda(n)}\|df_0(x_0)(d\Pi_S(z_0)(z_n - z_0)) + df_n(x_n)(d\Pi_S(z_n)(z_0 - z_n))\| > \epsilon\right\} \\
\subset&\{f_n \notin \Gamma_1\} \cup \{\|df_0(x_0) - df_n(x_n)\| \ge \delta) \cup \left\{\frac{\sqrt{n}}{\lambda(n)}\|d\Pi_S(z_0)\|\|(z_n - z_0)\| \ge \frac{\epsilon}{3\delta}\right\} \\
&\cup \left\{\frac{\sqrt{n}}{\lambda(n)}\|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \ge \frac{\epsilon}{3\delta}\right\} \\
&\cup \left\{\frac{\sqrt{n}}{\lambda(n)}\|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \ge \frac{\epsilon}{3(\|df_0(x_0)\| + 1)}\right\}.
\end{aligned} \quad (4.12)$$

Since $f_0 \in \Gamma_1^o$, by Lemma 4.2,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(f_n \notin \Gamma_1\right) = -\infty.$$

By Lemma 4.4,

$$\lim_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)}\|d\Pi_S(z_0)(z_n - z_0) + d\Pi_S(z_n)(z_0 - z_n)\| \ge \frac{\epsilon}{3\delta}\right) = -\infty,$$

and

$$\lim_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\|d\Pi_S(z_0)(z_n-z_0)+d\Pi_S(z_n)(z_0-z_n)\|\geq\frac{\epsilon}{3(\|df_0(x_0)\|+1)}\right)=-\infty.$$

By (2.25) or Theorem 3.1,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\|d\Pi_S(z_0)\|\|(z_n-z_0)\|\geq\frac{\epsilon}{3\delta}\right)=-\infty.$$

Thus, it suffices for (4.9) to show that for any $\delta>0$,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|df_0(x_0)-df_n(x_n)\|>\delta\right)=-\infty. \tag{4.13}$$

Since f_0 is continuously differentiable, $\|df_0(x_0)-df_0(x_n)\|\leq\|f\|_{1,U}\|x_n-x_0\|$. Thus, by (4.7), we have that for any $\delta>0$,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|df_0(x_0)-df_0(x_n)\|\geq\delta\right)=-\infty. \tag{4.14}$$

By Lemma 4.2, we have that

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|df_0(x_n)-df_n(x_n)\|\geq\delta\right)=-\infty. \tag{4.15}$$

Thus, by $\|df_n(x_n)-df_0(x_0)\|\leq\|df_n(x_n)-df_0(x_n)\|+\|df_0(x_n)-df_0(x_0)\|$, (4.13) follows from (4.14) and (4.15). Therefore, (4.9) is valid. □

4.3 Proof of Theorem 3.3

We use the exponential approximation method and the contraction principle to prove Theorem 3.3.

The Frobenius norm of a matrix $A=(a_{ij})_{p\times q}$ is defined as

$$\|A\|:=\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{q}|a_{ij}|^2}.$$

Lemma 4.5. *Suppose that the assumption (A6) holds. Then for any $\epsilon>0$,*

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|\Sigma_n-\Sigma_0\|\geq\epsilon\right)=-\infty. \tag{4.16}$$

Proof. By the moderate deviation principle (MDP) in separable Banach spaces (Lemma 4.1), for any $\epsilon>0$,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\sup_{x\in U}\left\|\frac{1}{n}\sum_{i=1}^{n}F(x,\xi_i)F(x,\xi_i)^T-E(F(x,\xi)F(x,\xi)^T)\right\|\geq\epsilon\right)=-\infty, \tag{4.17}$$

and

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\sup_{x\in U} \|f_n(x) - f_0(x)\| \geq \epsilon\right) = -\infty. \tag{4.18}$$

Using the same proof as (4.15), we can obtain for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\|\frac{1}{n}\sum_{i=1}^{n} F(x_n,\xi_i)F(x_n,\xi_i)^T - E(F(x_0,\xi)F(x_0,\xi)^T)\right\| \geq \epsilon\right) = -\infty, \tag{4.19}$$

and

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\|\frac{1}{n}\sum_{i=1}^{n} F(x_n,\xi_i) - E(F(x_0,\xi))\right\| \geq \epsilon\right) = -\infty. \tag{4.20}$$

Note that

$$\begin{aligned}\Sigma_n =& \frac{1}{n-1}\sum_{i=1}^{n}\left(F(x_n,\xi_i) - \frac{1}{n}\sum_{j=1}^{n} F(x_n,\xi_j)\right)\left(F(x_n,\xi_i) - \frac{1}{n}\sum_{j=1}^{n} F(x_n,\xi_j)\right)^T \\ =& \frac{1}{n-1}\sum_{i=1}^{n} F(x_n,\xi_i)F(x_n,\xi_i)^T - \frac{n+1}{n(n-1)}\sum_{j=1}^{n} F(x_n,\xi_j)\sum_{i=1}^{n} F(x_n,\xi_i)^T.\end{aligned}$$

By (4.19) and (4.20), we obtain (4.16). □

Lemma 4.6. *Let $\mathcal{S}^q$ denote the set of $q\times q$ non-negative definite symmetric matrices. For each $A = (a_{ij}) \in \mathcal{S}^q$, let $0 \leq \lambda_1(A) \leq \cdots \leq \lambda_q(A)$ be the all eigenvalues of A. Set $\mathcal{S}^q(l) = \{A \in \mathcal{S}^q;\ \lambda_{q-l+1}(A) \geq \rho_0\}$. For each $A \in \mathcal{S}$, we decompose A as*

$$A = U^T \Delta U, \tag{4.21}$$

where U is a $q\times q$ orthogonal matrix, Δ is diagonal matrix where the diagonal elements are decreasing non-negative eigenvalues. Let D be the upper-left submatrix of Δ whose diagonal elements are at least ρ_0. Let $(U)_1$ be the submatrix of U that consists of its first l rows, and let $(U)_2$ be the submatrix that consists of the remaining rows of U. Define

$$\Psi_1(A) = (U)_1^T D(U)_1, \quad \Psi_2(A) = (U)_1^T D^{-1}(U)_1, \quad A \in \mathcal{S}^q(l).$$

Then the maps Ψ_1 and Ψ_2 are continuous at Σ_0 in Frobenius norm.

Proof. Let $A_n = \left(A_{ij}^{(n)}\right) \in \mathcal{S}^q(l)$ converge to $A_0 = \Sigma_0$. For each $n \geq 0$, let U_n and Δ_n be the decomposition of A_n in (4.21). If we write

$$\Delta_n = \begin{bmatrix}(D_n)_1 & 0 \\ 0 & (D_n)_2\end{bmatrix}, \quad (U_n)_1 = \begin{bmatrix}(U_n)_{11} & (U_n)_{12}\end{bmatrix}, \quad (U_n)_2 = \begin{bmatrix}(U_n)_{21} & (U_n)_{22}\end{bmatrix}.$$

Then

$$\begin{aligned}A_n &= \begin{bmatrix}(U_n)_{11}^T & (U_n)_{21}^T \\ (U_n)_{12}^T & (U_n)_{22}^T\end{bmatrix}\begin{bmatrix}(D_n)_1 & 0 \\ 0 & (D_n)_2\end{bmatrix}\begin{bmatrix}(U_n)_{11} & (U_n)_{12} \\ (U_n)_{21} & (U_n)_{22}\end{bmatrix} \\ &= \begin{bmatrix}(U_n)_{11}^T(D_n)_1 & (U_n)_{21}^T(D_n)_2 \\ (U_n)_{12}^T(D_n)_1 & (U_n)_{22}^T(D_n)_2\end{bmatrix}\begin{bmatrix}(U_n)_{11} & (U_n)_{12} \\ (U_n)_{21} & (U_n)_{22}\end{bmatrix} \\ &= \begin{bmatrix}(U_n)_{11}^T(D_n)_1(U_n)_{11} + (U_n)_{21}^T(D_n)_2(U_n)_{21} & (U_n)_{11}^T(D_n)_1(U_n)_{12} + (U_n)_{21}^T(D_n)_2(U_n)_{22} \\ (U_n)_{12}^T(D_n)_1(U_n)_{11} + (U_n)_{22}^T(D_n)_2(U_n)_{21} & (U_n)_{12}^T(D_n)_1(U_n)_{12} + (U_n)_{22}^T(D_n)_2(U_n)_{22}\end{bmatrix}.\end{aligned}$$

Since the eigenvalues of a matrix are continuous functions of entries of the matrix, Δ_n tends to Δ_0 as $n \to \infty$. Note that $(D_n)_2 \to 0$, we have $A_n - U_n^T D_n U_n \to 0$. This yields that $U_n^T D_n U_n$ converge to $(U_0)_1^T D_0 (U_0)_1$. Thus, Ψ_1 is continuous at Σ_0.

Let A^+ denote generalized inverse of a matrix A. Then when A is a non-negative definite matrix, then $B = A^+$ if and only if

$$ABA = A,\ BAB = B,\ (AB)^T = AB,\ (BA)^T = BA.$$

Using the characterization, it is easy to get that $A_n^+ \to A^+$ when $A_n \to A$. Noting that $\Psi_2(A) = (\Psi_1(A))^+$, by continuity of Ψ_1 at Σ_0, we obtain that Ψ_2 is continuous at Σ_0.

□

Proof of Theorem 3.3. Let us first show (1). For any $\epsilon > 0$ and any $\delta > 0$,

$$\begin{aligned}
&P\left(\frac{\sqrt{n}}{\lambda(n)}\left\|\Sigma_n^{-1/2} d(f_n)_S(z_n)(z_n - z_0) - \Sigma_0^{-1/2} d(f_n)_S(z_n)(z_n - z_0)\right\| \geq \epsilon\right) \\
\leq &P\left(\|\Sigma_n^{-1/2} - \Sigma_0^{-1/2}\| \geq \delta\right) + P\left(\frac{\sqrt{n}}{\lambda(n)}\|d(f_n)_S(z_n)(z_n - z_0) - d(f_0)_S(z_0)(z_n - z_0)\| \geq \epsilon/(2\delta)\right) \\
&+ P\left(\frac{\sqrt{n}}{\lambda(n)}\|d(f_0)_S(z_0)(z_n - z_0)\| \geq \epsilon/(2\delta)\right).
\end{aligned}$$

By (4.16) and the delta method,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\|\Sigma_n^{-1/2} - \Sigma_0^{-1/2}\| \geq \delta\right) = -\infty.$$

By (4.9),

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)}\|d(f_n)_S(z_n)(z_n - z_0) - d(f_0)_S(z_0)(z_n - z_0)\| \geq \epsilon/(2\delta)\right) = -\infty.$$

By (2.25) or Theorem 3.1,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)}\|d(f_0)_S(z_0)(z_n - z_0)\| \geq \epsilon/(2\delta)\right) = -\infty.$$

Therefore, for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)}\left\|(\Sigma_n^{-1/2} - \Sigma_0^{-1/2}) d(f_n)_S(z_n)(z_n - z_0)\right\| \geq \epsilon\right) = -\infty,$$

and so, by Theorem 3.2 and the exponential approximation ([4, Theorem 4.2.13]), we obtain (1).

Next, let us show (2). By the Weyl theorem (see [31]),

$$\max_{1\leq i\leq q} |\mu_i^n - \mu_i| \leq \mathrm{tr}((\Sigma_n - \Sigma_0)(\Sigma_n - \Sigma_0)^T) = \|\Sigma_n - \Sigma_0\|^2.$$

Thus, by Lemma 4.5 and the contraction principle,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\max_{1\le i\le q}\|\mu_i^n-\mu_i\|\ge\min\{(\rho-\rho_0),\rho_0\}/2\right)$$
$$\le\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\|\Sigma_n-\Sigma_0\|\ge(\min\{(\rho-\rho_0),\rho_0\}/2)^{1/2}\right)=-\infty,$$

which implies that

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(l_n\neq l\right)=-\infty. \tag{4.22}$$

Noting that on $\{l_n=l\}$,

$$(\Delta_n)_{ii}>\rho_0 \text{ for each } i=1,\cdots,l \text{ and } (\Delta_n)_{ii}<\rho_0 \text{ for each } i=l+1,\cdots,q. \tag{4.23}$$

We define

$$\hat\Sigma_n=U_n^T\begin{bmatrix}D_n & 0\\ 0 & 0\end{bmatrix}U_n \text{ and } \hat\Sigma_n^+=U_n^T\begin{bmatrix}D_n^{-1} & 0\\ 0 & 0\end{bmatrix}U_n=(U_n)_1^TD_n^{-1}(U_n)_1. \tag{4.24}$$

By Lemma 4.6, there exist two functions Ψ_1 and Ψ_2 which are continuous at Σ_0 in the Frobenius norm such that on $\{l_n=l\}$,

$$\hat\Sigma_n=\Psi_1(\Sigma_n),\hat\Sigma_n^+=\Psi_2(\Sigma_n),\ \text{and}\ \Psi_1(\Sigma_0)=\Sigma_0,\quad \Psi_2(\Sigma_0)=\Sigma_0^+.$$

Since Ψ_1 and Ψ_2 are continuous at Σ_0, for any $\epsilon>0$, there exists $\delta>0$ such that

$$\sup_{\|\Sigma_n-\Sigma_0\|\le\delta}\max\left\{\|\Psi_1(\Sigma_n)-\Psi_1(\Sigma_0)\|,\|\Psi_2(\Sigma_n)-\Psi_2(\Sigma_0)\|\right\}<\epsilon.$$

Thus,

$$\left\{\max\left\{\left\|\hat\Sigma_n-\Sigma_0\right\|,\left\|\hat\Sigma_n^+-\Sigma_0^+\right\|\right\}\ge\epsilon\right\}\subset\left\{\|\Sigma_n-\Sigma_0\|\ge\delta\right\},$$

and so, by Lemma 4.5 , for any $\epsilon>0$,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\max\left\{\left\|\hat\Sigma_n-\Sigma_0\right\|,\left\|\hat\Sigma_n^+-\Sigma_0^+\right\|\right\}\ge\epsilon\right)=-\infty, \tag{4.25}$$

which implies that

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\left\|(U_n)_1^TD_n^{-1}(U_n)_1-(U_0)_1^TD_0^{-1}(U_0)_1\right\|\ge\epsilon\right)=-\infty. \tag{4.26}$$

By Theorem 3.2 and the contraction proinciple, we have

$$\left\{\frac{\sqrt{n}}{\lambda(n)}D_0^{-1/2}(U_0)_1d(f_n)_S(z_n)(z_n-z_0),n\ge1\right\} \tag{4.27}$$

satisfies the LDP in $\mathbb{R}^l$ with speed $\lambda^2(n)$ and rate function

$$\begin{aligned}\tilde I(y)=&\inf\left\{\sup_{v\in\mathbb{R}^q}\left\{\langle v,z\rangle-\frac12\left\|D_0^{1/2}(U_0)_1v\right\|^2\right\};\ D_0^{-1/2}(U_0)_1z=y\right\}\\ =&\sup_{v\in\mathbb{R}^l}\left\{\langle v,(U_0)_1^TD_0^{1/2}y\rangle-\frac12\left\|D_0^{1/2}(U_0)_1v\right\|^2\right\}\\ =&\sup_{v\in\mathbb{R}^l}\left\{\langle D_0^{1/2}(U_0)_1^Tv,y\rangle-\frac12\left\|D_0^{1/2}(U_0)_1v\right\|^2\right\}=\frac{\|y\|^2}{2}.\end{aligned}$$

Applying the contraction principle,

$$\left\{\left\|\frac{\sqrt{n}}{\lambda(n)}D_0^{-1/2}(U_0)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2, n\geq 1\right\} \tag{4.28}$$

satisfies the LDP in $\mathbb{R}$ with speed $\lambda^2(n)$ and rate function

$$\tilde{I}(y)=\inf\left\{\frac{\|z\|^2}{2};\ \|z\|^2=y\right\}=J(y).$$

Now, noting that

$$\begin{aligned}&\left\|\frac{\sqrt{n}}{\lambda(n)}D_n^{-1/2}(U_n)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2\\ =&\frac{n}{\lambda^2(n)}(z_n-z_0)^T d(f_n)_S(z_n)^T(U_n)_1^T D_n^{-1}(U_n)_1 d(f_n)_S(z_n)(z_n-z_0),\end{aligned}$$

we have

$$\begin{aligned}&\left|\left\|\frac{\sqrt{n}}{\lambda(n)}D_n^{-1/2}(U_n)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2-\left\|\frac{\sqrt{n}}{\lambda(n)}D_0^{-1/2}(U_0)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2\right|\\ \leq&\left\|(U_n)_1^T D_n^{-1}(U_n)_1-(U_0)_1^T D_0^{-1}(U_0)_1\right\|\left\|\frac{\sqrt{n}}{\lambda(n)}d(f_n)_S(z_n)(z_n-z_0)\right\|^2.\end{aligned}$$

and for any $\epsilon>0$, $\delta>0$,

$$\begin{aligned}&\left\{\left|\left\|\frac{\sqrt{n}}{\lambda(n)}D_n^{-1/2}(U_n)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2-\left\|\frac{\sqrt{n}}{\lambda(n)}D_0^{-1/2}(U_0)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2\right|\geq\epsilon\right\}\\ \subset&\{l_n\neq l\}\cup\left\{\left\|(U_n)_1^T D_n^{-1}(U_n)_1-(U_0)_1^T D_0^{-1}(U_0)_1\right\|\geq\delta\right\}\\ &\cup\left\{\left\|\frac{\sqrt{n}}{\lambda(n)}d(f_n)_S(z_n)(z_n-z_0)\right\|\geq\sqrt{\epsilon/\delta}\right\}.\end{aligned}$$

Thus, by (4.22) , (4.26) and Theorem 3.2,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\Bigg(\Bigg|&\left\|\frac{\sqrt{n}}{\lambda(n)}D_n^{-1/2}(U_n)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2\\ &-\left\|\frac{\sqrt{n}}{\lambda(n)}D_0^{-1/2}(U_0)_1 d(f_n)_S(z_n)(z_n-z_0)\right\|^2\Bigg|\geq\epsilon\Bigg)=-\infty.\end{aligned} \tag{4.29}$$

Finally, by (4.28) and (4.29), by the exponential approximation, we complete the proof of (2). □

4.4 Proof of Theorem 3.4

The following exponential approximation is a main step of proof of Theorem 3.4.

Lemma 4.7. *Suppose that assumptions (A1)–(A5) hold. Then under the condition (3.12), for any* $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\frac{\sqrt{n}}{\lambda(n)} \|\Phi_n(z_n)(z_n - z_0) - d(f_0)_S(z_0)(z_n - z_0)\| \geq \epsilon\right) = -\infty. \tag{4.30}$$

Proof. Set

$$V_n = \sup_{h\in\mathbb{R}^q} \frac{\|\Phi_n(z_n)(h) - d(f_0)_S(z_0)(h)\|}{\|h\|}.$$

Then for any $\epsilon > 0$, and any $M > \|f_0\|_{1,U}$,

$$\begin{aligned}
&P\left(\frac{\sqrt{n}}{\lambda(n)} \|\Phi_n(z_n)(z_n - z_0) - d(f_0)_S(z_0)(z_n - z_0)\| \geq \epsilon\right) \\
\leq &P\left(\frac{\sqrt{n}}{\lambda(n)} V_n\|z_n - z_0\| \geq \epsilon\right) \\
\leq &P\left(g(n)V_n > 4\tau(M+1)\right) + P\left(\frac{\sqrt{n}}{\lambda(n)}\|z_n - z_0\| \geq \frac{g(n)\epsilon}{4\tau(M+1)}\right).
\end{aligned} \tag{4.31}$$

Noting that on $\{f_n \in \Gamma_0\}$, $x_n = \Pi_S(z_n)$ and $x_0 = \Pi_S(z_0)$, by (2.21), (3.8) and (3.10), we have that on $\{f_n \in \Gamma_0\}$, for any $h \in \mathbb{R}^q$,

$$\begin{aligned}
&\|\Phi_n(z_n)(h) - d(f_0)_S(z_0)(h)\| \\
=&\|\left(df_n(\Pi_S(z_n))(\Lambda_n(z_n)(h)) - \Lambda_n(z_n)(h)\right) - \left(df_0(\Pi_S(z_0))(d\Pi_S(z_0)(h)) - d\Pi_S(z_0)(h)\right)\| \\
\leq &\|df_n(x_n)(\Lambda_n(z_n)(h)) - df_0(x_0)(d\Pi_S(z_0)(h))\| + \|\Lambda_n(z_n)(h) - d\Pi_S(z_0)(h)\| \\
\leq &\|df_n(x_n)(\Lambda_n(z_n)(h) - d\Pi_S(z_0)(h))\| + \|df_n(x_n)(d\Pi_S(z_0)(h)) - df_0(x_0)(d\Pi_S(z_0)(h))\| \\
&+ \|\Lambda_n(z_n)(h) - d\Pi_S(z_0)(h)\| \\
\leq &\left(\|df_n(x_n)\| + 1\right) \|\Lambda_n(z_n)(h) - d\Pi_S(z_0)(h)\| + \|df_n(x_n) - df_0(x_n)\|\|d\Pi_S(z_0)\|\|h\| \\
&+ \|f_0\|_{1,U}\|x_n - x_0\|\|\|d\Pi_S(z_0)\|\|h\|.
\end{aligned}$$

Thus for any $M > \|f_0\|_{1,U}$,

$$\begin{aligned}
&P\left(g(n)V_n > 4\tau(M+1)\right) \\
\leq &P(f_n \notin \Gamma_0) + P\left(\|df_n(x_n)\| \geq M\right) + P\left(\|df_n(x_n) - df_0(x_n)\|\|d\Pi_S(z_0)\| \geq \frac{\tau(M+1)}{g(n)}\right) \\
&+ P\left(\|f_0\|_{1,U}\|x_n - x_0\|\|\|d\Pi_S(z_0)\| \geq \frac{\tau(M+1)}{g(n)}\right) \\
&+ P\left(g(n) \sup_{h\in\mathbb{R}^q} \frac{\|\Lambda_n(z_n)(h) - d\Pi_S(z_0)(h)\|}{\|h\|} \geq \tau\right).
\end{aligned}$$

By (2.20)

$$\begin{aligned}
&P\left(\|df_n(x_n) - df_0(x_n)\|\|d\Pi_S(z_0)\| \geq \frac{\tau(M+1)}{g(n)}\right) \\
\leq &\beta_1 e^{-n\mu_1} + \frac{M_1}{(\frac{\tau(M+1)}{g(n)\|d\Pi_S(z_0)\|})^q} \exp\left\{-\frac{n(\frac{\tau(M+1)}{g(n)\|d\Pi_S(z_0)\|})^2}{\sigma_1}\right\},
\end{aligned} \tag{4.32}$$

and

$$P\left(\|df_n(x_n)\|\geq M\right)\leq P\left(\|df_n(x_n)-df_0(x_n)\|\geq M-\|f_0\|_{1,U}\right)$$
$$\leq\beta_1 e^{-n\mu_1}+\frac{M_1}{(M-\|f_0\|_{1,U})^q}\exp\left\{-\frac{n(M-\|f_0\|_{1,U})^2}{\sigma_1}\right\}. \tag{4.33}$$

Under the condition (3.12), by (2.25) and (3.11), there exists $n_0\geq 1$ such that for all $n\geq n_0$,

$$P\left(\frac{\sqrt{n}}{\lambda(n)}\|z_n-z_0\|\geq\frac{g(n)\epsilon}{4\tau(M+1)}\right)\leq\beta_0 e^{-n\mu_0}+\frac{M_0}{(\frac{g(n)\lambda(n)\epsilon}{4\tau(M+1)\sqrt{n}})^q}\exp\left\{-\frac{n(\frac{g(n)\lambda(n)\epsilon}{4\tau(M+1)\sqrt{n}})^2}{\sigma_0}\right\}, \tag{4.34}$$

$$P\left(\|f_0\|_{1,U}\|x_n-x_0\|\|\|d\Pi_S(z_0)\|\geq\frac{\tau(M+1)}{g(n)}\right)$$
$$\leq\beta_0 e^{-n\mu_0}+\frac{M_0}{(\frac{\tau(M+1)}{g(n)\|f_0\|_{1,U}\|d\Pi_S(z_0)\|})^q}\exp\left\{-\frac{n(\frac{\tau(M+1)}{g(n)\|f_0\|_{1,U}\|d\Pi_S(z_0)\|})^2}{\sigma_0}\right\}, \tag{4.35}$$

and

$$P\left(g(n)\sup_{h\in\mathbb{R}^q}\frac{\|\Lambda_n(z_n)(h)-d\Pi_S(z_0)(h)\|}{\|h\|}\geq\tau\right)$$
$$\leq 2P(g(n)\|z_n-z_0\|\geq 1/2) \tag{4.36}$$
$$\leq 2\beta_0 e^{-n\mu_0}+\frac{2M_0}{(2g(n))^{-q}}\exp\left\{-\frac{n/(2g(n))^2}{\sigma_0}\right\}.$$

Therefore,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\|\Phi_n(z_n)(z_0-z_n)-d(f_0)_S(z_0)(z_n-z_0)\|\geq\epsilon\right)$$
$$\leq\max\left\{\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(g(n)V_n>4\tau(M+1)\right),\right. \tag{4.37}$$
$$\left.\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\|z_n-z_0\|\geq\frac{g(n)\epsilon}{4\tau(M+1)}\right)\right\}$$
$$=-\infty.$$

□

Proof of Theorem 3.4. We only show (1). By Lemma 4.5 and Lemma 4.7, we have that for any $\epsilon>0$,

$$\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log P\left(\frac{\sqrt{n}}{\lambda(n)}\left\|\Sigma_n^{-1/2}\Phi_n(z_n)(z_n-z_0)-\Sigma_0^{-1/2}d(f_0)_S(z_0)(z_n-z_0)\right\|\geq\epsilon\right)=-\infty. \tag{4.38}$$

Thus, Theorem 3.4 follows from Theorem 3.1 and the exponential approximation ([4, Theorem 4.2.13]).

□

5 Application to statistical inferences

In this section, we apply the moderate deviations to the hypotheses testing of z_0. We show that the rejection region constructed by the central limit theorems in [16] has the probability of the type II error with an exponential decay speed. This property shows that the test has very good behaviors in the sense of large sample.

Let $\alpha \in (0,1)$ be given. Assume that (A1)–(A6) hold . We consider the hypotheses testing problem

$$H_0 : z_0 = \hat{z}_0, H_1 : z_0 \neq \hat{z}_0.$$

Let χ^2_m denote a χ^2 random variable with m degrees of freedom, and let $\chi^2_m(\alpha)$ denote the number that satisfies $P(\chi^2_m > \chi^2_m(\alpha)) = \alpha$ for $\alpha \in (0,1)$.

Case 1: Σ_0 is nonsingular. For $\alpha \in (0,1)$ given, by Theorem 2 in [16], we can take a rejection region

$$\left\{ \left\| n^{1/2}\Sigma_n^{-1/2}\Phi_n(z_n)(z_n - \hat{z}_0) \right\|^2 > \chi^2_q(\alpha) \right\}.$$

By Theorem 3.4 (1), for each $\hat{z} \neq \hat{z}_0$, the probability of the type II error

$$\begin{aligned}
\beta_n(\hat{z}) =& P\left(\left\| n^{1/2}\Sigma_n^{-1/2}\Phi_n(z_n)(z_n - \hat{z}_0) \right\|^2 < \chi^2_q(\alpha) \middle| z_0 = \hat{z} \right) \\
\leq& P\Bigg(\frac{\sqrt{n}}{\lambda(n)} \left\| \Sigma_n^{-1/2}\Phi_n(z_n)(z_n - \hat{z}) \right\| \\
&\qquad \geq \frac{\sqrt{n}}{\lambda(n)} \left\| \Sigma_n^{-1/2}\Phi_n(z_n)(\hat{z}_0 - \hat{z}) \right\| - \frac{\sqrt{\chi^2_q(\alpha)}}{\lambda(n)} \Bigg| z_0 = \hat{z} \Bigg) \\
\leq& P\Bigg(\frac{\sqrt{n}}{\lambda(n)} \left\| \Sigma_n^{-1/2}\Phi_n(z_n)(z_n - \hat{z}) \right\| \\
&\qquad \geq \frac{\sqrt{n}}{2\lambda(n)} \|\eta\| - \frac{\sqrt{\chi^2_q(\alpha)}}{\lambda(n)} \Bigg| z_0 = \hat{z} \Bigg) \\
&+ P\Bigg(\left\| \Sigma_n^{-1/2}\Phi_n(z_n)(\hat{z}_0 - \hat{z}) - \eta \right\| \geq \|\eta\|/2 \Bigg| z_0 = \hat{z} \Bigg),
\end{aligned}$$

where $\eta = \Sigma_0^{-1/2} d(f_0)_S(\hat{z})(\hat{z}_0 - \hat{z})$. Since $d(f_0)_S(\hat{z}) = L_K$, by Lemma 2.1, $\eta \neq 0$.

Since for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\| \Sigma_n^{-1/2} - \Sigma_0^{-1/2} \right\| \geq \epsilon \middle| z_0 = \hat{z} \right) = -\infty,$$

and

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\|\Phi_n(z_n) - d(f_0)_S(\hat{z})\| \geq \epsilon \middle| z_0 = \hat{z} \right) = -\infty,$$

we have that for any $\epsilon > 0$

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\| \Sigma_n^{-1/2}\Phi_n(z_n) - \Sigma_0^{-1/2} d(f_0)_S(\hat{z}) \right\| \geq \epsilon \middle| z_0 = \hat{z} \right) = -\infty.$$

Thus, by $\frac{\sqrt{n}}{2\lambda(n)}\|\eta\| - \frac{\sqrt{\chi_q^2(\alpha)}}{\lambda(n)} \to \infty$, we have

$$\limsup_{n\to\infty} \frac{1}{\lambda^2(n)} \log \beta_n(\hat{z}) = -\infty.$$

Case 2: Σ_0 is singular. In this case, for $\epsilon > 0$, for $\alpha \in (0,1)$ given, we can take the rejection region

$$\left\{ \left\| \sqrt{n}(U_n)_1^T D_n^{-1/2}(U_n)_1 \Phi_n(z_n)(z_n - \hat{z}_0) \right\|^2 > \chi_l^2(\alpha) \right\}.$$

Then, by Theorem 3.4 (2), the probability of the type II error with an exponential decay speed.

6 An example and simulations

In this subsection, we apply the above results to the same example used in [17, 18, 16], and give some simulations and numerical results.

In this example, $q = 2$, $d = 6$, $O = \mathbb{R}^2$, $S = \mathbb{R}^2_+$, $F : \mathbb{R}^2 \times \mathbb{R}^6 \to \mathbb{R}^2$ is defined by

$$F(x,\xi) = \begin{bmatrix} \xi^{(1)} & \xi^{(2)} \\ \xi^{(3)} & \xi^{(4)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi^{(5)} \\ \xi^{(6)} \end{bmatrix}, \tag{6.1}$$

and the random vector $\xi = (\xi^{(1)}, \cdots, \xi^{(6)})^T$ follows the uniform distribution over the box $[0,2] \times [0,1] \times [0,2] \times [0,4] \times [-1,1] \times [-1,1]$. The true problem is

$$0 \in f_0(x) + N_{\mathbb{R}^2_+}(x), \quad f_0(x) = M_0\, x, \tag{6.2}$$

where

$$M_0 = \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix}.$$

Then, $\Pi_S^{-1}(O) = \mathbb{R}^2$ and $(f_0)_{\mathbb{R}^2_+} : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$(f_0)_{\mathbb{R}^2_+}(z) = \begin{cases} f_0(z) & \text{if } z \in \mathbb{R}^2_+ \\ f_0(z_1, 0) + \begin{bmatrix} 0 \\ z_2 \end{bmatrix} & \text{if } z \in \mathbb{R}_+ \times \mathbb{R}_- \\ f_0(0, z_2) + \begin{bmatrix} z_1 \\ 0 \end{bmatrix} & \text{if } z \in \mathbb{R}_- \times \mathbb{R}_+ \\ \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} & \text{if } z \in \mathbb{R}^2_-. \end{cases}$$

It is obvious that $(f_0)_{\mathbb{R}^2_+}(z) = 0$ if and only if $\Pi(z) = 0$. Thus, the solution to (6.2) is $x_0 = 0$, and the solution of the corresponding normal map formulation is $z_0 = x_0 - E[F(x_0,\xi)] = 0$. The covariance matrix of $F(x_0,\xi) = (\xi^{(5)}, \xi^{(6)})$ is given by

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

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

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

in orthants $\mathbb{R}^2_+$, $\mathbb{R}_+ \times \mathbb{R}_-$, $\mathbb{R}_- \times \mathbb{R}_+$ and $\mathbb{R}^2_-$ respectively, that is,

$$d(f_0)_{\mathbb{R}^2_+}(z)x = \begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} xI_{\mathbb{R}^2_+}(z) + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} xI_{\mathbb{R}_+\times\mathbb{R}_-}(z) + \begin{bmatrix} 1 & 1/2 \\ 0 & 2 \end{bmatrix} xI_{\mathbb{R}_-\times\mathbb{R}_+}(z) + xI_{\mathbb{R}^2_-}(z),$$

where I_A denotes the indicator function of set A.

The sample average function $f_n : \mathbb{R}^2 \times \Omega \to \mathbb{R}^2$ is given by

$$f_n(x, \omega) = M_n x + b_n,$$

where

$$M_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \xi_i^{(1)} & \frac{1}{n}\sum_{i=1}^n \xi_i^{(2)} \\ \frac{1}{n}\sum_{i=1}^n \xi_i^{(3)} & \frac{1}{n}\sum_{i=1}^n \xi_i^{(4)} \end{bmatrix}, \quad b_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \xi_i^{(5)} \\ \frac{1}{n}\sum_{i=1}^n \xi_i^{(6)} \end{bmatrix}.$$

The sample average approximation (SAA) problem is

$$0 \in f_n(x, \omega) + N_{\mathbb{R}^2_+}(x),$$

and $(f_n)_{\mathbb{R}^2_+} : \mathbb{R}^2 \to \mathbb{R}^2$ is

$$(f_n)_{\mathbb{R}^2_+}(z) = \begin{cases} M_n z + b_n & \text{if } z \in \mathbb{R}^2_+ \\ M_n \begin{bmatrix} z_1 \\ 0 \end{bmatrix} + b_n + \begin{bmatrix} 0 \\ z_2 \end{bmatrix} & \text{if } z \in \mathbb{R}_+ \times \mathbb{R}_- \\ M_n \begin{bmatrix} 0 \\ z_2 \end{bmatrix} + b_n + \begin{bmatrix} z_1 \\ 0 \end{bmatrix} & \text{if } z \in \mathbb{R}_- \times \mathbb{R}_+ \\ b_n + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} & \text{if } z \in \mathbb{R}^2_-. \end{cases}$$

Let z_n be a solution of $(f_n)_{\mathbb{R}^2_+}(z) = 0$ and set $x_n = \Pi_{\mathbb{R}^2_+}(z_n)$. Then

$$z_n = \begin{cases} -M_{n,1}^{-1} b_n & \text{if } z_n \in \mathbb{R}^2_+ \\ -M_{n,2}^{-1} b_n & \text{if } z_n \in \mathbb{R}_+ \times \mathbb{R}_- \\ -M_{n,3}^{-1} b_n & \text{if } z_n \in \mathbb{R}_- \times \mathbb{R}_+ \\ -M_{n,4}^{-1} b_n & \text{if } z_n \in \mathbb{R}^2_-, \end{cases}$$

where $M_{n,1} = M_n$, and

$$M_{n,2} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \xi_i^{(1)} & 0 \\ \frac{1}{n}\sum_{i=1}^n \xi_i^{(3)} & 1 \end{bmatrix}, \quad M_{n,3} = \begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^n \xi_i^{(2)} \\ 0 & \frac{1}{n}\sum_{i=1}^n \xi_i^{(4)} \end{bmatrix}, \quad M_{n,4} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The B-derivative $d(f_n)_{\mathbb{R}^2_+}(z_n)$ is a piecewise linear function represented by matrices $M_{n,1}$, $M_{n,2}$, $M_{n,3}$, $M_{n,4}$ in orthants $\mathbb{R}^2_+$, $\mathbb{R}_+ \times \mathbb{R}_-$, $\mathbb{R}_- \times \mathbb{R}_+$ and $\mathbb{R}^2_-$ respectively, that is,

$$d(f_n)_{\mathbb{R}^2_+}(z_n)x = M_{n,1} x I_{\mathbb{R}^2_+}(z_n) + M_{n,2} x I_{\mathbb{R}_+\times\mathbb{R}_-}(z_n) + M_{n,3} x I_{\mathbb{R}_-\times\mathbb{R}_+}(z_n) + M_{n,4} x I_{\mathbb{R}^2_-}(z_n).$$

Therefore

$$
\begin{aligned}
& d(f_0)_{\mathbb{R}^2_+}(z_0)(z_n - z_0) \\
= & -\begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} M_n^{-1} b_n \\
= & -\begin{bmatrix} 1 & 1/2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \xi_i^{(1)} & \frac{1}{n}\sum_{i=1}^n \xi_i^{(2)} \\ \frac{1}{n}\sum_{i=1}^n \xi_i^{(3)} & \frac{1}{n}\sum_{i=1}^n \xi_i^{(4)} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \xi_i^{(5)} \\ \frac{1}{n}\sum_{i=1}^n \xi_i^{(6)} \end{bmatrix},
\end{aligned} \tag{6.3}
$$

and

$$
d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) = -b_n = -\begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \xi_i^{(5)} \\ \frac{1}{n}\sum_{i=1}^n \xi_i^{(6)} \end{bmatrix}. \tag{6.4}
$$

It is obvious that the assumptions (A1)-(A6) hold. Now, applying Theorem 3.1, Theorem 3.2 and Theorem 3.3 to this example, we have that for any $r > 0$

$$
\lim_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\| \frac{\sqrt{n}}{\lambda(n)} \Sigma_0^{-1/2} d(f_0)_{\mathbb{R}^2_+}(z_0)(z_n - z_0) \right\| > r \right) = -\frac{r^2}{2}, \tag{6.5}
$$

$$
\lim_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\| \frac{\sqrt{n}}{\lambda(n)} \Sigma_0^{-1/2} d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) \right\| > r \right) = -\frac{r^2}{2}, \tag{6.6}
$$

and

$$
\lim_{n\to\infty} \frac{1}{\lambda^2(n)} \log P\left(\left\| \frac{\sqrt{n}}{\lambda(n)} \Sigma_n^{-1/2} d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) \right\| > r \right) = -\frac{r^2}{2}, \tag{6.7}
$$

where Σ_n is the sample covariance matrix of $\{F(x_n, \xi_i)\}_{i=1}^n$:

$$
\Sigma_n = \frac{1}{n-1} \sum_{i=1}^n \left(F(x_n, \xi_i) - \frac{1}{n}\sum_{j=1}^n F(x_n, \xi_j) \right) \left(F(x_n, \xi_i) - \frac{1}{n}\sum_{j=1}^n F(x_n, \xi_j) \right)^T.
$$

Roughly, (6.5), (6.6) and (6.7) can be written respectively by

$$
P\left(\left\| \frac{\sqrt{n}}{\lambda(n)} \Sigma_0^{-1/2} d(f_0)_{\mathbb{R}^2_+}(z_0)(z_n - z_0) \right\| > r \right) \approx \exp\left\{ -\frac{\lambda^2(n) r^2}{2} \right\}, \tag{6.8}
$$

$$
P\left(\left\| \frac{\sqrt{n}}{\lambda(n)} \Sigma_0^{-1/2} d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) \right\| > r \right) \approx \exp\left\{ -\frac{\lambda^2(n) r^2}{2} \right\}, \tag{6.9}
$$

and

$$
P\left(\left\| \frac{\sqrt{n}}{\lambda(n)} \Sigma_n^{-1/2} d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) \right\| > r \right) \approx \exp\left\{ -\frac{\lambda^2(n) r^2}{2} \right\}. \tag{6.10}
$$

Next, we give some simulations and numerical results for the tail probabilities

$$
p_1(\alpha, r) = P\left(\left\| \frac{\sqrt{n}}{n^\alpha} \Sigma_0^{-1/2} d(f_0)_{\mathbb{R}^2_+}(z_0)(z_n - z_0) \right\| > r \right),
$$

$$
p_2(\alpha, r) = P\left(\left\| \frac{\sqrt{n}}{n^\alpha} \Sigma_0^{-1/2} d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) \right\| > r \right)
$$

and

$$
p_3(\alpha, r) = P\left(\left\| \frac{\sqrt{n}}{n^\alpha} \Sigma_n^{-1/2} d(f_n)_{\mathbb{R}^2_+}(z_n)(z_n - z_0) \right\| > r \right),
$$

where $0 < \alpha < 1/2$ and $r > 0$. Set

$$p(\alpha, r) = \exp\left\{-\frac{n^{2\alpha}r^2}{2}\right\}.$$

Case I: n fixed. We generate 10000 SAA problems with $n = 1000$, and consider $\alpha = 1/3$. The numerical results are summarized in the following Table. We select different r to verify the effectiveness of the theorems.

r	$p_1(1/3, r)$	$p_2(1/3, r)$	$p_3(1/3, r)$	$p(1/3, r)$
0.05	0.8853	0.8858	0.8855	0.8825
0.1	0.6098	0.6083	0.6081	0.6065
0.15	0.3228	0.3247	0.3252	0.3247
0.2	0.1367	0.1360	0.1375	0.1353
0.25	0.0462	0.0460	0.0461	0.0439

Table 1: $n = 1000$

Case II: $n \to \infty$. In this case, we also generate 10000 SAA problems with $\alpha = 1/3$, but n is from 200 to 1000 and $r = 0.3$. The numerical results are summarized in the following Table. From Table 1 and Table 2, we can see that the simulation results are highly consistent

n	$p_1(1/3, 0.3)$	$p_2(1/3, 0.3)$	$p_3(1/3, 0.3)$	$p(1/3, 0.3)$
200	0.2107	0.2071	0.2084	0.2146
400	0.0865	0.0855	0.0867	0.0869
600	0.0437	0.0424	0.0430	0.0407
800	0.0217	0.0221	0.0219	0.0207
1000	0.0118	0.0112	0.0117	0.0111

Table 2: $r = 0.3$

with the theorem results (Theorem 3.1, Theorem 3.2 and Theorem 3.3). Numerical results demonstrate the validity of the estimates.

Acknowledgment

The authors thank the editor and two anonymous referees for helpful comments and suggestions, which greatly improves the quality of the paper. This work is supported in part by the National Natural Science Foundation of China under Grant 11801184, and Grant 12171168, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515010368, and Grant 2020A1515010489, and in part by PolyU Grant G-UAHF.

References

[1] X. Chen. Moderate deviations for m-dependent random variables with Banach space values, *Stat. Probab. Letters*, 35 (1997), 123-134.

[2] X. Chen, T. K. Pong and R. J.-B. Wets, Two-stage stochastic variational inequalities: an ERM-solution procedure, *Mathematical Programming*, 165 (2017), 71–111.

[3] X. Chen, A. Shapiro and H. Sun, Convergence analysis of sample average approximation of two-stage stochastic generalized equations, *SIAM Journal on Optimization*, 29 (2019), 135–161.

[4] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, *Stochastic Modeling and Applied Probability 38*, Springer-Verlag, Berlin, 2010.

[5] M. C. Demir, *Asymptotics and confidence regions for stochastic variational inequalities*, Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin–Madison, Madison, WI, 2000.

[6] D. Dentcheva and W. Römisch, Differential stability of two-stage stochastic programs, *SIAM Journal on Optimization*, 11 (2000), 87–112.

[7] J. Dupačová and R. Wets, Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems, *The Annals of Statistics*, 16 (1988), 1517–1549.

[8] F. Q. Gao and X. Q. Zhao, Delta Method in Large Deviations and Moderate Deviations for Estimators, *The Annals of Statistics*, 39 (2011), 1211–1240.

[9] G. H. Golub, A. Hoffman and G. W. Stewart, A generalization of the Eckart-Young-Mirsky matrix approximation theorem, *Linear Algebra and its Applications*, 88-89 (1987), 317–327.

[10] J. Gwinner, B. Jadamba, A. A. Khan and F. Raciti, *Uncertainty Quantification in Variational Inequalities: Theory, Numerics, and Applications*. Chapman and Hall/CRC, 2022.

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

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

[13] M. Lamm, S. Lu and A. Budhiraja, Individual confidence intervals for solutions to expected value formulations of stochastic variational inequalities, *Mathematical Programming*, 165 (2017), 151–196.

[14] Y. Liu and J. Zhang, Confidence regions of stochastic variational inequalities: error bound approach, *Optimization*, 71 (2022), 2157–2184.

[15] R. C. Li, Relative perturbation theory: (II) eigenspace and singular space variations, *SIAM Journal on Matrix Analysis and Applications*, 20 (1999), 471–492.

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

[17] S. Lu, A new method to build confidence regions for solutions of stochastic variational inequalities, *Optimization*, 63 (2014), 1431–1443.

[18] S. Lu, Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities, *SIAM Journal on Optimization*, 24 (2014), 1458–1484.

[19] S. Lu, Y. Liu, L. Yin and K. Zhang, Confidence intervals and regions for the lasso by using stochastic variational inequality techniques in optimization, *J. R. Statist. Soc. B*, 79 (2017), 589–611.

[20] G. C. Pflug, Stochastic optimization and statistical inference, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., vol. 10 of Handbooks in Operations Research and Management Science, Elsevier, 2003.

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

[22] S. M. Robinson, Sensitivity analysis of variational inequalities by normal-map techniques, in *Variational Inequalities and Network Equilibrium Problems*, F. Giannessi and A. Maugeri, eds., New York, 1995, Plenum Press, 257–269.

[23] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[24] R. T. Rockafellar and R. Wets, *Variational Analysis*, vol. 317 of A Series of Comprehensive Studies in Mathematics, Springer-Verlag, Berlin, 2009.

[25] W. Romisch, Stability of stochastic programming, in *Stochastic Programming*, A. Ruszczy'nski and A. Shapiro, eds., vol. 10 of Handbooks in Operations Research and Management Science, Elsevier, 2003.

[26] A. Shapiro, On concepts of directional differentiability, *Journal of Optimization Theory and Applications*, 66 (1990), 477–487.

[27] A. Shapiro, Asymptotic behavior of optimal solutions in stochastic programming, *Mathematics of Operations Research*, 18 (1993), 829–845.

[28] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory,* Second edition, *SIAM*, Philadelphia, 2014.

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

[30] G. W. Stewart, On the perturbation of pseudo-inverses, projections and linear least squares problems, *SIAM Review*, 19 (1977), 634–662.

[31] G. W.Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.

[32] A. W. Van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer-Verlag, New York, 1996.

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

[34] L. M. Wu, Large deviations, moderate deviations and LIL for empirical processes, *Ann. Probab.*, 22 (1994), 17–27.

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