Asymptotic Behaviors of Semidefinite Programming with a Covariance Perturbation

Mingjie Gao *
School of Mathematical Sciences, South China Normal University
Guangzhou, 510631, China
mjgao@m.scnu.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

In this paper, we study asymptotic behaviors of semidefinite programming with a covariance perturbation. We obtain some moderate deviations, Cramér-type moderate deviations and a law of the iterated logarithm of estimates of the respective optimal value and optimal solutions when the covariance matrix is estimated by its sample covariance. As an example, we also apply the main results to the Minimum Trace factor Analysis.

Keywords: Asymptotic, law of the iterated logarithm, minimum trace factor analysis, moderate deviations, perturbation, semidefinite programming.

1 Introduction

Let $\mathbb{S}^p$ denote the linear space of $p\times p$ symmetric matrices. For $A\in\mathbb{S}^p$, $A\succcurlyeq 0$ means that the matrix A is positive semidefinite. Let Σ_0 and $A_i, i=1,\cdots,n$ be $p\times p$ symmetric matrices. Shapiro ([26]) considered the following semidefinite programming (SDP) problem

$$\min_{x\in\mathbb{R}^n} c^T x \quad \text{subject to } \Sigma_0+\mathcal{A}(x)\succcurlyeq 0, \tag{1.1}$$

where $\mathcal{A}:\mathbb{R}^n\to\mathbb{S}^p$ is the linear mapping $\mathcal{A}(x):=\sum_{i=1}^n x_iA_i$, Σ_0 is viewed as a covariance matrix of a $p\times 1$ random vector Y. The Factor Analysis model is a classical example of the models (see Bentler ([1]), Shapiro ([26])).

A stochastic average approximation (SAA) of the original problem(1.1) is defined by

$$\min_{x\in\mathbb{R}^n} c^T x \quad \text{subject to } S+\mathcal{A}(x)\succcurlyeq 0, \tag{1.2}$$

*Supported by the National Natural Science Foundation of China (NSFC) Grant 11801184 and 11771157.
†Supported by GRF Grant PolyU. 152200/14E and PolyU grant G-YBVQ.

where

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$$

is the sample covariance matrix based on a sample $Y_1, \cdots, Y_N$ of the random vector Y, and $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ is the sample mean.

Let x^* and ϑ^* denote an optimal solution and the optimal value of problem (1.1), respectively, and let $\hat{x}_N$ and $\hat{\vartheta}_N$ be their counterparts of the problem (1.2). Shapiro ([26]) studied the central limit theorems of $\hat{x}_N$ and $\hat{\vartheta}_N$ based on the modern theory of sensitivity analysis of parameterized SDP problems.

The asymptotic distributions and the consistency of general SAA estimators were derived in [15], [22], [23]. Asymptotic confidence intervals on the optimal value of stochastic programs are given in e.g., [3], [5], [12] for risk-neutral problems and in [9] for risk-averse programs. Nonasymptotic confidence intervals on the optimal value and large deviation estimates are obtained in [7], [8], [12], [13], [16], [29]. Pflug [17] constructed universal confidence sets for the true solution of a stochastic optimization problem based on properties about boundedness in probability with known tail behavior. For further references, we refer to see [28].

In this paper, we are interesting in convergence rates on $\hat{x}_N$ and $\hat{\vartheta}_N$. For the weak convergence rate, we present some moderate deviation results. Under some conditions, we prove that there exists $v > 0$ such that for any $x > 0$,

$$\lim_{N\to\infty}\frac{1}{b_N^2}\log P\left(\frac{\sqrt{N}}{b_N}\left|\hat{x}_N - x^*\right| \geq x\right) = -\frac{x^2}{2v^2} \tag{1.3}$$

where

$$b_N \to \infty \text{ and } b_N^2/N \to 0 \text{ as } N \to \infty; \tag{1.4}$$

and that under some conditions, there exists $\nu > 0$ such that as $N \to \infty$, for $0 \leq x = o(N^{1/6})$ uniformly,

$$\frac{P\left(\sqrt{N}(\hat{\vartheta}_N - \vartheta^*) \leq -\nu x\right)}{\Phi(-x)} \to 1 \text{ and } \frac{P\left(\sqrt{N}(\hat{\vartheta}_N - \vartheta^*) \geq \nu x\right)}{1-\Phi(x)} \to 1, \tag{1.5}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. For the strong convergence rate, we will show that under some conditions, the following law of the iterated logarithm holds, i.e., there exists $\nu > 0$ such that

$$P\left(\limsup_{N\to\infty}\sqrt{\frac{N}{2\nu^2\log\log N}}(\hat{\vartheta}_N - \vartheta^*) = 1\right) = 1. \tag{1.6}$$

The results on moderate deviations can be applied to asymptotic confidence intervals and hypotheses testing (e.g., see [6]).

Our main results are stated and proved in Section 2. We also apply the main results to the Minimum Trace factor Analysis in Section 3. For convenience, we introduce differentiability properties of the optimal value and optimal solutions in Appendix A.

We conclude the section by introducing some notations and terminology. Let $\mathbb{S}_+^p$ and $\mathbb{S}_{++}^p$ denote cones of symmetric positive semidefinite and positive definite $p \times p$ matrices, respectively. As usual, $\mathrm{tr}\,(A) = \sum_{i=1}^{p} a_{ii}$ denotes the trace of $p \times p$ square matrix $A = [a_{ij}]$ and $A \bullet B = \mathrm{tr}\,(AB)$ for $A, B \in \mathbb{S}^p$. $A \otimes B$ denotes Kronecker product of $p \times q$ matrix $A = [a_{ij}]$ and $r \times s$ matrix $B = [b_{ij}]$. For a $p \times q$ matrix A, $\mathrm{vec}\,(A)$ denotes the pq-dimensional vector obtained by stacking columns of matrix A. By $A^\dagger$ we denote the Moore-Penrose pseudoinverse of matrix A. For $A \in \mathbb{S}^p$ with rank r, $A = NDN^T$ is its singular value decomposition, i.e., $N^T N = I_r$ and D is $r \times r$ diagonal matrix with diagonal entries given by nonzero eigenvalues of A, then $A^\dagger = ND^{-1}N^T$.

2 Asymptotic behaviors

Let Y be a $\mathbb{R}^p$-valued random variable with the covariance matrix Σ_0.

(A0). Assume that there exists $\epsilon_0 > 0$ such that

$$E\left(\exp\left\{\epsilon_0 |Y|^2\right\}\right) < \infty.$$

Let $Y_1, \cdots, Y_N$ be a sample of size N of the random vector Y, $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$, and

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$$

denote the sample mean and the sample covariance matrix, respectively.

For given matrix $\Sigma \in \mathbb{S}^p$, let $\vartheta(\Sigma)$ and $\bar{x}(\Sigma)$ be the optimal value and an optimal solution of the problem (A.1) in Appendix A (see [26]). Define

$$\hat{\vartheta}_N = \vartheta(S), \quad \hat{x}_N = \bar{x}(S); \quad \vartheta^* = \vartheta(\Sigma_0), \quad x^* = \bar{x}(\Sigma_0). \tag{2.1}$$

Shapiro([26]) studied the central limit theorems of $\hat{\vartheta}_N$ and $\hat{x}_N$. He obtained the following results.

- Under the condition of Proposition A.1,

  $$N^{1/2}(\hat{\vartheta}_N - \vartheta^*) \to \mathcal{N}(0, \lambda_0^T \Gamma \lambda) \text{ in law},$$

  where $\lambda_0 = \mathrm{vec}\,(\Lambda_0)$, and Γ is $p^2 \times p^2$ covariance matrix. In particular, if $Y \sim N(\mu, \Sigma_0)$, then

  $$\Gamma = 2M_p(\Sigma_0 \otimes \Sigma_0), \tag{2.2}$$

  and $M_p = \left[\frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\right]_{ij,kl}$ is $p^2 \times p^2$ symmetric matrix, $\delta_{ik} = 1_{\{i\}}(k)$.

- Under the condition of Proposition A.2,

  $$N^{1/2}(\hat{x}_N - x^*) \to \mathcal{N}(0, J^T \Gamma J) \text{ in law}.$$

In this section, we establish moderate deviations, Cramér-type moderate deviations and a law of the iterated logarithm for $\hat{\vartheta}_N$ and $\hat{x}_N$. Our proofs are based on Bernstein's inequality, Cramér-type moderate deviations and the law of the iterated logarithm for independent and identically distributed random variables.

2.1 Moderate deviations

Lemma 2.1 (Bernstein's inequality (cf. Proposition 3.3.1 in [31])). *Let $\xi_i, i \geq 1$ be a sequence of independent real random variables with mean $E(\xi_i) = 0$. Assume that there exist constants $v_i > 0$ and $H > 0$ such that for all $k \geq 2$*

$$E(|\xi_i|^k) \leq \frac{1}{2} k! v_i^2 H^{k-2}.$$

Set $B_N = \sqrt{v_1^2 + \cdots + v_N^2}$. Then for all $x > 0$,

$$P\left(\sum_{i=1}^N \xi_i \geq x B_N\right) \leq \exp\left\{-\frac{1}{2} x^2 (1 + xH/B_N)^{-1}\right\}.$$

Lemma 2.2 (cf. Theorem 1 in Chapter VIII, [14]). *Let $\xi, \xi_i, i \geq 1$ be a sequence of independent and identically distributed real random variables with mean $E(\xi) = 0$ and variance $E(\xi^2) = 1$. Assume that there exists $\epsilon_0 > 0$ such that*

$$E(e^{\epsilon_0 |\xi|}) < \infty.$$

Then for all $0 \leq x = o(\sqrt{N})$,

$$\frac{P\left(\sum_{i=1}^N \xi_i \geq x\sqrt{N}\right)}{1 - \Phi(x)} = \exp\left\{\frac{x^3}{\sqrt{N}} \lambda\left(\frac{x}{\sqrt{N}}\right)\right\} \left(1 + O\left(\frac{1+x}{\sqrt{N}}\right)\right), \tag{2.3}$$

and

$$\frac{P\left(\sum_{i=1}^N \xi_i \leq -x\sqrt{N}\right)}{\Phi(-x)} = \exp\left\{-\frac{x^3}{\sqrt{N}} \lambda\left(-\frac{x}{\sqrt{N}}\right)\right\} \left(1 + O\left(\frac{1+x}{\sqrt{N}}\right)\right), \tag{2.4}$$

where $\Phi(x)$ is the standard normal distribution function and $\lambda(t) = \sum_{k=0}^\infty a_k t^k$ is a power series with coefficients depending on the cumulants of the random variable ξ which converges for sufficiently small values of $|t|$.

In particular, for all $0 \leq x = O(N^{1/6})$, uniformly

$$\frac{P\left(\sum_{i=1}^N \xi_i \geq x\sqrt{N}\right)}{1 - \Phi(x)} = 1 + O\left(\frac{1+x^3}{\sqrt{N}}\right) \tag{2.5}$$

and

$$\frac{P\left(\sum_{i=1}^N \xi_i \leq -x\sqrt{N}\right)}{\Phi(-x)} = 1 + O\left(\frac{1+x^3}{\sqrt{N}}\right). \tag{2.6}$$

Theorem 2.1. *Let the condition (A0) hold. Suppose that the optimal value ϑ^* is finite and Slater condition for the true problem holds. Then*

(1). For any $\epsilon > 0$,

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left|\hat{\vartheta}_N - \vartheta^* - \sup_{\Lambda \in \mathrm{Sol}\,(D)} \Lambda \bullet (S - \Sigma_0)\right| \geq \epsilon\right) = -\infty. \tag{2.7}$$

Moreover, if $\mathrm{Sol}\,(D) = \{\Lambda_0\}$ *is a singleton, then for any* $\epsilon > 0$,

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N} \left| \hat{\vartheta}_N - \vartheta^* - \frac{1}{N}\sum_{i=1}^N Z_i \right| \geq \epsilon \right) = -\infty. \tag{2.8}$$

where

$$\begin{aligned} Z_i =& \mathrm{tr}\,(\Lambda_0(Y_iY_i^T - E(Y)Y_i^T - Y_iE(Y^T))) - E\left(\mathrm{tr}\,(\Lambda_0(Y_iY_i^T - E(Y)Y_i^T - Y_iE(Y^T)))\right) \\ =& \lambda_0^T \mathrm{vec}\,\left((Y_iY_i^T - E(Y)Y_i^T - Y_iE(Y^T))) - E\left(Y_iY_i^T - E(Y)Y_i^T - Y_iE(Y^T)\right)\right) \end{aligned} \tag{2.9}$$

(2). If $\mathrm{Sol}\,(D) = \{\Lambda_0\}$ *is a singleton, then for any* $x > 0$,

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N} \frac{1}{\sqrt{\lambda_0^T \Gamma \lambda_0}} \left(\hat{\vartheta}_N - \vartheta^*\right) \geq x \right) = -\frac{x^2}{2}, \tag{2.10}$$

and

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N} \frac{1}{\sqrt{\lambda_0^T \Gamma \lambda_0}} \left(\hat{\vartheta}_N - \vartheta^*\right) \leq -x \right) = -\frac{x^2}{2}. \tag{2.11}$$

Proof. (1). Firstly, we show that there exist constants $M > 0, L > 0$ such that for any $x > 0$,

$$P\left(|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq x\right) \leq M \exp\left\{-LN \min\{x, x^2\}\right\}. \tag{2.12}$$

We can write

$$S - \Sigma_0 = \frac{1}{N-1}\sum_{i=1}^N (Y_iY_i^T - E(YY^T)) - \frac{N}{N-1}\left(\bar{Y}\bar{Y}^T - E(Y)E(Y^T)\right),$$

and

$$\bar{Y}\bar{Y}^T - E(Y)E(Y^T) = (\bar{Y} - E(Y))(\bar{Y}^T - E(Y^T)) + E(Y)(\bar{Y}^T - E(Y^T)) + (\bar{Y} - E(Y))E(Y^T)$$

Then under the assumption (A0), for any $k \geq 2$,

$$E\left(\left|\mathrm{vec}\,(YY^T) - E(\mathrm{vec}\,(YY^T))\right|^k\right) \leq 2^k E((|Y|^2)^k) \leq k!\left(\frac{2}{\epsilon_0}\right)^k E\left(\exp\{\epsilon_0|Y|^2\}\right)$$

and

$$E\left(|Y - E(Y)|^k\right) \leq 2^k E(|Y|^k) \leq 2^k (E((|Y|^2)^k))^{1/2} \leq k!\left(\frac{2}{\epsilon_0}\right)^k E\left(\exp\{\epsilon_0|Y|^2\}\right)$$

Set $u = \sqrt{8E\left(\exp\{\epsilon_0|Y|^2\}\right)/\epsilon_0^2}$, $H = \frac{2}{\epsilon_0}$. Note that for a m-dimensional random vector $Z = (Z_1, \cdots, Z_m)^T$, for any $x > 0$,

$$\{|Z| \geq x\} \subset \cup_{k=1}^m \{|Z_k| \geq x/m\}.$$

Then by Bernstein's inequality for any $x > 0$,

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N}(\mathrm{vec}\,(Y_iY_i^T) - E(\mathrm{vec}\,(YY^T)))\right| \geq xu\right) \leq 2p^2 \exp\left\{-\frac{1}{2p^4}Nx^2(1+xH/(p^2u))^{-1}\right\}, \tag{2.13}$$

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Y_i - E(Y))\right| \geq xu\right) \leq 2p\exp\left\{-\frac{1}{2p^2}Nx^2(1+xH/(pu))^{-1}\right\}, \tag{2.14}$$

and

$$\begin{aligned} P\left(\left|\mathrm{vec}\,((\bar{Y}-E(Y))(\bar{Y}^T - E(Y^T)))\right| \geq xu\right) \leq & P\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Y_i - E(Y))\right| \geq \sqrt{xu}\right) \\ \leq & 2p\exp\left\{-\frac{1}{2p^2u}Nx(1+\sqrt{x}H/(p\sqrt{u}))^{-1}\right\}. \end{aligned} \tag{2.15}$$

Therefore, (2.12) holds.

Next, let us prove (2.7). By Proposition 3 in [26] (see Proposition A.1 in Appendix A), for any $\zeta > 0$, there exists $\eta > 0$, such that

$$\begin{aligned} & \{|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \leq \eta\} \\ \subset & \left\{\left|\hat{\vartheta}_N - \vartheta^* - \sup_{\Lambda\in \mathrm{Sol}\,(D)} \Lambda \bullet (S-\Sigma_0)\right| \leq \zeta\,|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)|\right\}. \end{aligned}$$

Thus, for any $\epsilon > 0$, $\zeta > 0$,

$$\begin{aligned} & \left\{\frac{\sqrt{N}}{b_N}\left|\hat{\vartheta}_N - \vartheta^* - \sup_{\Lambda\in \mathrm{Sol}\,(D)} \Lambda \bullet (S-\Sigma_0)\right| \geq \epsilon\right\} \\ \subset & \{|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \eta\} \cup \left\{\frac{\sqrt{N}}{b_N}\,|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \epsilon/\zeta\right\}. \end{aligned}$$

Now, (2.7) follows from

$$\lim_{\zeta\to 0}\limsup_{N\to\infty}\frac{1}{b_N^2}\log P\left(\frac{\sqrt{N}}{b_N}\,|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \epsilon/\zeta\right) = -\infty,$$

and

$$\limsup_{N\to\infty}\frac{1}{b_N^2}\log P\left(|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \eta\right) = -\infty.$$

If $\mathrm{Sol}\,(D) = \{\Lambda_0\}$ is a singleton, then

$$\sup_{\Lambda\in \mathrm{Sol}\,(D)} \Lambda \bullet (S-\Sigma_0) - \frac{1}{N-1}\sum_{i=1}^{N} Z_i = \frac{N}{N-1}\Lambda_0(\bar{Y}-E(Y))(\bar{Y}^T - E(Y^T))\Lambda_0^T.$$

Thus, by (2.15), we obtain (2.8) .

(2). Note that

$$E\left(Z_1^2\right) = \lambda_0^T \Gamma \lambda_0, \tag{2.16}$$

and that for any $x > 0$, $\lim_{N\to\infty} \frac{1}{b_N^2} \log(1-\Phi(b_N x)) = -\frac{1}{2}x^2$, and

$$\lim_{N\to\infty} \frac{1}{b_N^2} \left\{ \frac{(b_N x)^3}{\sqrt{N}} \lambda\left(\frac{b_N x}{\sqrt{N}}\right) \right\} = 0, \quad \lim_{N\to\infty} \frac{1}{b_N^2} \log\left(1 + O\left(\frac{1+b_N x}{\sqrt{N}}\right)\right) = 0.$$

Thus, by (2.3) and (2.4), we obtain that

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{1}{\sqrt{N} b_N} \frac{1}{\sqrt{\lambda_0^T \Gamma \lambda_0}} \sum_{i=1}^N Z_i \geq x \right) = -\frac{x^2}{2}, \tag{2.17}$$

and

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{1}{\sqrt{N} b_N} \frac{1}{\sqrt{\lambda_0^T \Gamma \lambda_0}} \sum_{i=1}^N Z_i \leq -x \right) = -\frac{x^2}{2}. \tag{2.18}$$

Finally, since for all $x > 0$, $0 < \epsilon < x$,

$$\begin{aligned}
& P\left(\frac{\sqrt{N}}{b_N \sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \geq x \right) \\
\leq & P\left(\frac{\sqrt{N}}{b_N \sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \geq x, \frac{\sqrt{N}}{b_N} \left| \hat{\vartheta}_N - \vartheta^* - \frac{1}{N} \sum_{i=1}^N Z_i \right| \leq \epsilon \right) \\
& + P\left(\frac{\sqrt{N}}{b_N} \left| \hat{\vartheta}_N - \vartheta^* - \frac{1}{N} \sum_{i=1}^N Z_i \right| \geq \epsilon \right) \\
\leq & P\left(\frac{1}{\sqrt{N} b_N} \frac{1}{\sqrt{\lambda_0^T \Gamma \lambda_0}} \sum_{i=1}^N Z_i \geq x - \epsilon \right) + P\left(\frac{\sqrt{N}}{b_N} \left| \hat{\vartheta}_N - \vartheta^* - \frac{1}{N} \sum_{i=1}^N Z_i \right| \geq \epsilon \right).
\end{aligned}$$

and

$$\begin{aligned}
& P\left(\frac{1}{\sqrt{N} b_N} \frac{1}{\sqrt{\lambda_0^T \Gamma \lambda_0}} \sum_{i=1}^N Z_i \geq x + \epsilon \right) \\
\leq & P\left(\frac{\sqrt{N}}{b_N \sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \geq x \right) + P\left(\frac{\sqrt{N}}{b_N} \left| \hat{\vartheta}_N - \vartheta^* - \frac{1}{N} \sum_{i=1}^N Z_i \right| \geq \epsilon \right).
\end{aligned}$$

By (2.8) and (2.17), we have that

$$\begin{aligned}
-\frac{(x+\epsilon)^2}{2} & \leq \liminf_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N \sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \geq x \right) \\
& \leq \limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N \sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \geq x \right) \leq -\frac{(x-\epsilon)^2}{2}.
\end{aligned}$$

Letting $\epsilon \to 0$, we obtain (2.10). The proof of (2.11) is similar to (2.10).

$\square$

Remark 2.1. *(1). Applying the moderate deviation principle for independent and identically distributed real random variables (cf. Theorem 3.7.1 in [4]) and the exponential approximation (cf. Theorem 4.2.13 in [4])), by* (2.8)*, we obtain a moderate deviation principle for $\hat{\vartheta}_N$, that is, for any open subset G of $\mathbb{R}$,*

$$\liminf_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N\sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \in G \right) \geq - \inf_{x\in G} \frac{x^2}{2}, \tag{2.19}$$

and for any closed subset F of $\mathbb{R}$,

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N\sqrt{\lambda_0^T \Gamma \lambda_0}} (\hat{\vartheta}_N - \vartheta^*) \in F \right) \leq - \inf_{x\in F} \frac{x^2}{2}. \tag{2.20}$$

(2). Theorem 2.1 can be obtained by using the delta methed in large deviations (see [6]) .

By Theorem 1 in [26] (see Proposition A.2 in Appendix), using the same proof in Theorem 2.1, we have the following result.

Theorem 2.2. *Let the condition (A0) hold. Suppose that* $\mathrm{Sol}\,(P) = \{x^*\}$ *is a singleton, and that x^* is a nondegenerate point of $\Sigma_0 + \mathcal{A}(\cdot)$ and the strict complementarity condition holds. Then*

(1). For any $\epsilon > 0$,

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N} \left| \hat{x}_N - x^* - \frac{1}{N}\sum_{i=1}^N X_i \right| \geq \epsilon \right) = -\infty, \tag{2.21}$$

where

$$X_i = J^T \left(\mathrm{vec}\,(Y_iY_i^T - E(Y)Y_i^T - Y_iE(Y^T)) - E\left(\mathrm{vec}\,(Y_iY_i^T - E(Y)Y_i^T - Y_iE(Y^T)) \right) \right). \tag{2.22}$$

(2). For any $r > 0$,

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N} |\hat{x}_N - x^*| \geq r \right) = - \inf_{|x|\geq r} I(x), \tag{2.23}$$

where

$$I(x) = \sum_{\alpha\in\mathbb{R}^n} \left\{ \alpha^T x - \frac{1}{2}\alpha^T J^T \Gamma J \alpha \right\}, \quad x \in \mathbb{R}^n, \tag{2.24}$$

is called the rate function.

Proof. By the moderate deviation principle for independent and identically distributed random variables (cf. Theorem 3.7.1 in [4]), for any open subset G of $\mathbb{R}^n$,

$$\liminf_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{1}{\sqrt{N} b_N} \sum_{i=1}^N X_i \in G \right) \geq - \inf_{x\in G} I(x), \tag{2.25}$$

and for any closed subset F of $\mathbb{R}^n$,

$$\limsup_{N\to\infty}\frac{1}{b_N^2}\log P\left(\frac{1}{\sqrt{N}b_N}\sum_{i=1}^N X_i\in F\right)\le -\inf_{x\in F}I(x). \tag{2.26}$$

In particular, for any $r>0$,

$$\lim_{N\to\infty}\frac{1}{b_N^2}\log P\left(\frac{1}{\sqrt{N}b_N}\left|\sum_{i=1}^N X_i\right|\ge r\right)=-\inf_{|x|\ge r}I(x). \tag{2.27}$$

Thus, by (2.21), (2.27) can be deduced by the same proof in Theorem 2.1 (2). □

2.2 Cramér-type moderate deviations

Theorem 2.3. *Let the condition (A0) hold. Suppose that* $\mathrm{Sol}\,(P)=\{x^*\}$ *is a singleton, and that* x^* *is a nondegenerate point of* $\Sigma_0+\mathcal{A}(\cdot)$ *and the strict complementarity condition holds. Then*

(1). For any $\epsilon>0$*,*

$$\limsup_{N\to\infty}\frac{1}{N^{1/3}}\log P\left(\sqrt{N}\left|\hat{\vartheta}_N-\vartheta^*-\frac{1}{N}\sum_{i=1}^N Z_i\right|\ge\epsilon/N^{1/6}\right)<0. \tag{2.28}$$

(2). For any $0\le x=o(N^{1/6})$*, uniformly*

$$\frac{P\left(\sqrt{N}(\hat{\vartheta}_N-\vartheta^*)\ge x\sqrt{\lambda_0^T\Gamma\lambda_0}\right)}{1-\Phi(x)}=1+O\left(\frac{1+x^3}{\sqrt{N}}\right) \tag{2.29}$$

and

$$\frac{P\left(\sqrt{N}(\hat{\vartheta}_N-\vartheta^*)\le -x\sqrt{\lambda_0^T\Gamma\lambda_0}\right)}{\Phi(-x)}=1+O\left(\frac{1+x^3}{\sqrt{N}}\right). \tag{2.30}$$

Proof. (1). By Proposition A.2, there exist constants $L>0$ and $\eta>0$, such that

$$\begin{aligned}&\{|\mathrm{vec}\,(S)-\mathrm{vec}\,(\Sigma_0)|\le\eta\}\\ \subset&\left\{\left|\hat{\vartheta}_N-\vartheta^*-\Lambda\bullet(S-\Sigma_0)\right|\le L\,|\mathrm{vec}\,(S)-\mathrm{vec}\,(\Sigma_0)|^2\right\}.\end{aligned}$$

Thus, for any $\epsilon>0$,

$$\begin{aligned}&\left\{\sqrt{N}\left|\hat{\vartheta}_N-\vartheta^*-\Lambda\bullet(S-\Sigma_0)\right|\ge\epsilon/N^{1/6}\right\}\\ \subset&\{|\mathrm{vec}\,(S)-\mathrm{vec}\,(\Sigma_0)|\ge\eta\}\cup\left\{|\mathrm{vec}\,(S)-\mathrm{vec}\,(\Sigma_0)|\ge\frac{\sqrt{\epsilon}}{\sqrt{L}N^{1/3}}\right\}.\end{aligned}$$

By

$$\limsup_{N\to\infty}\frac{1}{N^{1/3}}\log P\left(|\mathrm{vec}\,(S)-\mathrm{vec}\,(\Sigma_0)|\ge\frac{\sqrt{\epsilon}}{\sqrt{L}N^{1/3}}\right)<0,$$

and

$$\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(|\text{vec}\,(S) - \text{vec}\,(\Sigma_0)| \geq \eta\right) = -\infty,$$

we obtain

$$\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\sqrt{N}\left|\hat{\vartheta}_N - \vartheta^* - \Lambda \bullet (S - \Sigma_0)\right| \geq \epsilon/N^{1/6}\right) < 0. \tag{2.31}$$

This implies (2.28) from

$$\Lambda \bullet (S - \Sigma_0) - \frac{1}{N-1}\sum_{i=1}^{N} Z_i = \frac{N}{N-1}\Lambda(\bar{Y} - E(Y))(\bar{Y}^T - E(Y^T))\Lambda^T$$

and for any $\epsilon > 0$,

$$\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\sqrt{N}\left|\text{vec}\left((\bar{Y} - E(Y))(\bar{Y}^T - E(Y^T))\right)\right| \geq \epsilon/N^{1/6}\right) < 0.$$

(2). We only show (2.29). By condition (2.28), for any $\epsilon > 0$, there exists $C > 0$ such that for N large enough, and for all $0 \leq x \leq o(N^{1/6})$, uniformly

$$\frac{P\left(\sqrt{N}\left|\hat{\vartheta}_N - \vartheta^* - \frac{1}{N}\sum_{i=1}^N Z_i\right| \geq \epsilon/N^{1/6}\right)}{1 - \Phi(x)} = O\left(e^{-CN^{1/3}}\right).$$

Thus, for all $0 \leq x \leq o(N^{1/6})$, uniformly

$$\begin{aligned}
&\frac{P\left(\sqrt{N}(\hat{\vartheta}_N - \vartheta^*) \geq x\sqrt{\lambda_0^T\Gamma\lambda_0}\right)}{1 - \Phi(x)} \\
\leq&\frac{P\left(\sqrt{N}(\hat{\vartheta}_N - \vartheta^*) \geq x\sqrt{\lambda_0^T\Gamma\lambda_0}, \sqrt{N}\left|\hat{\vartheta}_N - \vartheta^* - \frac{1}{N}\sum_{i=1}^N Z_i\right| \leq \epsilon/N^{1/6}\right)}{1 - \Phi(x)} \\
&+ \frac{P\left(\sqrt{N}\left|\hat{\vartheta}_N - \vartheta^* - \frac{1}{N}\sum_{i=1}^N Z_i\right| \geq \epsilon/N^{1/6}\right)}{1 - \Phi(x)} \\
\leq&\frac{P\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i \geq x\sqrt{\lambda_0^T\Gamma\lambda_0} - \epsilon/N^{1/6}\right)}{1 - \Phi(x)} + O\left(e^{-CN^{1/3}}\right).
\end{aligned}$$

Similarly, Then for all $0 \leq x \leq o(N^{1/6})$, uniformly

$$\begin{aligned}
&\frac{P\left(\sqrt{N}(\hat{\vartheta}_N - \vartheta^*) \geq x\sqrt{\lambda_0^T\Gamma\lambda_0}\right)}{1 - \Phi(x)} \\
\geq&\frac{P\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i \geq x\sqrt{\lambda_0^T\Gamma\lambda_0} + \frac{\epsilon}{N^{1/6}}\right)}{1 - \Phi(x)} + O\left(e^{-CN^{1/3}}\right).
\end{aligned}$$

By lemma 2.2,

$$
\begin{aligned}
&\frac{P\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i \geq x\sqrt{\lambda_0^T\Gamma\lambda_0} \pm \frac{\epsilon}{N^{1/6}}\right)}{1-\Phi(x)} \\
=&\frac{P\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i \geq x\sqrt{\lambda_0^T\Gamma\lambda_0} \pm \frac{\epsilon}{N^{1/6}}\right)}{1-\Phi\left(x\sqrt{\lambda_0^T\Gamma\lambda_0} \pm \frac{\epsilon}{N^{1/6}}\right)} \frac{1-\Phi\left(x\sqrt{\lambda_0^T\Gamma\lambda_0} \pm \frac{\epsilon}{N^{1/6}}\right)}{1-\Phi(x)} \\
=&\left(1+O\left(\frac{1+x^3}{N^{1/3}}\right)\right)\left(1+O\left(\frac{1+x^2}{N^{1/3}}\right)\right) = 1+O\left(\frac{1+x^3}{N^{1/3}}\right).
\end{aligned}
$$

Thus, the first one of (2.29) holds.

□

2.3 A law of the iterated logarithm

In this subsection, we apply the moderate deviations to establish a law of the iterated logarithm

Theorem 2.4. *Let the condition (A0) hold. Suppose that* $\mathrm{Sol}\,(P) = \{x^*\}$ *is a singleton, and that* x^* *is a nondegenerate point of* $\Sigma_0 + \mathcal{A}(\cdot)$ *and the strict complementarity condition holds. Then*

$$
\limsup_{N\to\infty} \sqrt{\frac{N}{2\lambda_0^T\Gamma\lambda_0 \log\log N}}(\hat{\vartheta}_N - \vartheta^*) = 1, \quad a.s. \tag{2.32}
$$

and

$$
\liminf_{N\to\infty} \sqrt{\frac{N}{2\lambda_0^T\Gamma\lambda_0 \log\log N}}(\hat{\vartheta}_N - \vartheta^*) = -1, \quad a.s. \tag{2.33}
$$

Proof. By Proposition A.2, there exist constants $L > 0$ and $\eta > 0$, such that

$$
\begin{aligned}
&\{|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \leq \eta\} \\
\subset&\left\{\left|\hat{\vartheta}_N - \vartheta^* - \frac{1}{N}\sum_{i=1}^N Z_i\right| \leq L\,|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)|^2\right\}.
\end{aligned}
$$

Thus, for any $\epsilon > 0$,

$$
\begin{aligned}
&\left\{\sqrt{N}\left|\hat{\vartheta}_N - \vartheta^* - \frac{1}{N}\sum_{i=1}^N Z_i\right| \geq \epsilon\right\} \\
\subset&\{|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \eta\} \cup \left\{\sqrt{N}\,|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \frac{N^{1/4}\sqrt{\epsilon}}{\sqrt{L}}\right\}.
\end{aligned}
$$

By (2.12), there exist constants $C_1, C_2 \in (0,\infty)$ such that

$$
P\left(|\mathrm{vec}\,(S) - \mathrm{vec}\,(\Sigma_0)| \geq \eta\right) \leq C_1 \exp\left\{-C_2 N \min\{\eta,\eta^2\}\right\}
$$

and

$$P\left(\sqrt{N}\,|\mathrm{vec}\,(S)-\mathrm{vec}\,(\Sigma_0)|\geq \frac{N^{1/4}\sqrt{\epsilon}}{\sqrt{L}}\right)\leq C_1\exp\left\{-C_2N\min\left\{\frac{\sqrt{\epsilon}}{N^{1/4}\sqrt{L}},\frac{\epsilon}{N^{1/2}L}\right\}\right\}.$$

Thus,

$$\sum_{N=1}^{\infty}P\left(\sqrt{N}\left|\hat{\vartheta}_N-\vartheta^*-\frac{1}{N}\sum_{i=1}^{N}Z_i\right|\geq\epsilon\right)<\infty,$$

and so, by the Borel-Cantelli lemma,

$$\lim_{N\to\infty}\sqrt{N}\left|\hat{\vartheta}_N-\vartheta^*-\frac{1}{N}\sum_{i=1}^{N}Z_i\right|=0,\quad a.s. \tag{2.34}$$

Finally, (2.32) and (2.33) follow from (2.34) and the law of the iterated logarithm for independent and identically distributed random variables (see Theorem 8 in Chapter X of [14]):

$$\limsup_{N\to\infty}\frac{1}{\sqrt{2\lambda_0^T\Gamma\lambda_0N\log\log N}}\sum_{i=1}^{N}Z_i=1,\quad a.s.$$

and

$$\liminf_{N\to\infty}\frac{1}{\sqrt{2\lambda_0^T\Gamma\lambda_0N\log\log N}}\sum_{i=1}^{N}Z_i=-1,\quad a.s.$$

□

3 Factor analysis

In this section we apply general results of Section 2 to the so-called Minimum Trace Factor Analysis (MTFA) problem

$$\min_{x\in\mathbb{R}^p}\mathbf{1}^Tx\quad\text{ subject to }\Sigma_0+X\succcurlyeq 0, \tag{3.1}$$

where the (population) covariance matrix Σ_0 is decomposed into $(\Sigma_0-\Psi)+\Psi$ with Ψ being a diagonal matrix and matrix $\Sigma_0-\Psi\succcurlyeq 0$ having rank $r<p$, $\mathbf{1}=(1,...,1)^T$ and $X=\mathrm{diag}\,(x_1,...,x_p)$, i.e., in (1.1), $A_i=\mathrm{diag}\,(0,...,0,1,0,...,0)$, $i=1,...,p$, $c=\mathbf{1}$.

Consider now the estimates $\hat{\vartheta}_N$ and $\hat{x}_N$ of the optimal value and optimal solution of the MTFA problem (3.1).

The stochastic Minimum Trace Factor Analysis problem associated with SDP problem (3.1) is defined by

$$\min_{x\in\mathbb{R}^p}\mathbf{1}^Tx\quad\text{ subject to }S+X\succcurlyeq 0, \tag{3.2}$$

Now let $\hat{\vartheta}_N$ and $\hat{x}_N$ be the estimates of the optimal value and optimal solution of the problem (3.2). Then the moderate deviations of $\hat{\vartheta}_N$ follows from Theorem 2.1. Furthermore, if the nondegeneracy and strict complementarity conditions hold, then the set of optimal solutions of the primal problem (3.1) is a singleton. Let x^* be the optimal solution of the

MTFA problem (3.1) and let Λ be an optimal solution of the dual problem. Set $X^* = \operatorname{diag}(x_1^*, \cdots, x_p^*)$ and $\Upsilon = \Sigma_0 + X^*$. Then the problem (A.5) takes here the following form

$$\begin{cases} \min\limits_{h \in \mathbb{R}^p} & h^T \left(\Lambda \circ \Upsilon^\dagger\right) h + 2\operatorname{tr}\left(H\Lambda\Delta\Upsilon^\dagger\right) \\ \text{s.t.} & E^T HE + E^T \Delta E = 0, \end{cases} \tag{3.3}$$

where $H = \operatorname{diag}(h_1, \cdots, h_p)$ and the symbol $\circ$ denotes Hadamard product.

The moderate deviations of $\hat{x}_N$ and the Cramér-type moderate deviations and the law of iterated logarithm of $\hat{\vartheta}_N$ can be obtained from Theorem 2.2, Theorem 2.3 and Theorem 2.4.

Theorem 3.1. *Let x^* be the optimal solution of the MTFA problem (3.1), $\Upsilon = \Sigma_0 + X^*$ and $r = \operatorname{rank}(\Upsilon)$. Suppose that the point x^* is nondegenerate and the strict complementarity condition holds. Then*

(1). For any $r > 0$, (2.23) holds for $\hat{x}_N$ with rate function

$$I(x) = \sum_{\alpha \in \mathbb{R}^n} \left\{\alpha^T x - \frac{1}{2}\alpha^T J^T \Gamma J \alpha\right\}, \quad x \in \mathbb{R}^n. \tag{3.4}$$

where J is a $p^2 \times n$ matrix such that $J^T\delta$ is the optimal solution of problem (3.1).

(2). (2.29) and (2.30) hold uniformly for any $0 \le x = o(N^{1/6})$.

(3). (2.32) and (2.33) hold for $\hat{\vartheta}_N$.

A Differentiability properties of the optimal value and an optimal solution

For convenience, in this Appendix, we recall some results on differentiability properties of the optimal value $\vartheta(\Sigma)$ and an optimal solution $\bar{x}(\Sigma)$ of the following problem (A.1) considered as functions of matrix $\Sigma \in \mathbb{S}^p$ (see [26]):

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{subject to } \Sigma + \mathcal{A}(x) \succcurlyeq 0, \tag{A.1}$$

which can be viewed as an SDP problem parameterized by matrix $\Sigma \in \mathbb{S}^p$.

The (Lagrangian) dual of problem (A.1) can be written as

$$\max_{\Lambda \in \mathbb{S}^p_+} \Lambda \bullet \Sigma \quad \text{subject to } \Lambda \bullet A_i = c_i,\ i = 1, \cdots, n \tag{A.2}$$

The problems (A.1) and (A.2) are refered as the primal (P) and dual (D) problems, respectively. We also use notation $\sigma := \operatorname{vec}(\Sigma)$, $\bar{x}(\sigma) := \bar{x}(\Sigma)$ and $\vartheta(\sigma) := \vartheta(\Sigma)$.

Slater condition. It is said that Slater condition holds for the primal problem (P) if there exists $x^* \in \mathbb{R}^n$ such that $\Sigma + \mathcal{A}(x^*) \in \mathbb{S}^p_{++}$. If Slater condition holds, then optimal values of problems (P) and (D) are equal to each other.

Let $\mathcal{W}_r$ denote the space of matrices $A \in \mathbb{S}^p$ of $\operatorname{rank}(A) = r \le p$. Then by Proposition 1.1, Chapter 5 in [10], $\mathcal{W}_r$ is a smooth manifold of dimension

$$\dim(\mathcal{W}_r) = p(p+1)/2 - (p-r)(p-r+1)/2 = pr - r(r-1)/2,$$

and the tangent space of the manifold $\mathcal{W}_r$ at $A \in \mathcal{W}_r$ is

$$T_{\mathcal{W}_r}(A) = \left\{ \Delta A + A\Delta^T;\ \Delta \text{ is } p \times p \text{ matrix } \right\}.$$

Nondegenerate point. It is said that $x^* \in \mathbb{R}^n$ is a nondegenerate point of mapping $x \to \Sigma + \mathcal{A}(x)$ if for $\Upsilon := \Sigma + \mathcal{A}(x^*)$ and $r := \text{rank}\,(\Upsilon)$ it follows that

$$\mathcal{A}(\mathbb{R}^n) + T_{\mathcal{W}_r}(\Upsilon) = \mathbb{S}^p,$$

otherwise point x^* is said to be degenerate.

A.1 Differentiability of the optimal value $\vartheta(\Sigma)$

Let $\text{Sol}\,(P)$ denote the set of optimal solutions of the reference (true) problem (1.1), and let $\text{Sol}\,(D)$ be the set of optimal solutions of its dual problem (A.1) for $\Sigma = \Sigma_0$. By the classical convex analysis and Theorem 4.1.9 in [25], the following result holds.

Proposition A.1. *(Proposition 3 in [26]) Suppose that Slater condition holds for the reference problem (1.1) and its optimal value $\vartheta(\Sigma_0)$ is finite. Then the set $\text{Sol}\,(D)$ is nonempty, convex and compact and the optimal value function $\vartheta(\cdot)$ is continuous convex function and Fréchet directionally differentiable at Σ_0 with*

$$\vartheta'(\Sigma_0, H) = \sup_{\Lambda \in \text{Sol}\,(D)} \Lambda \bullet H. \tag{A.3}$$

That is,

$$\vartheta(\Sigma) - \vartheta(\Sigma_0) = \sup_{\Lambda \in \text{Sol}\,(D)} \Lambda \bullet (\Sigma - \Sigma_0) + o(\|\sigma - \sigma_0\|) \tag{A.4}$$

A.2 The second order differentiability of the optimal value $\vartheta(\Sigma)$

Suppose that $\text{Sol}\,(P) = \{x^*\}$ and that x^* is a nondegenerate point of $\Sigma_0 + \mathcal{A}(\cdot)$, and so $\text{Sol}\,(D) = \{\Lambda\}$ is a singleton.

Complementarity condition. Assume that Slater condition holds for the reference problem (1.1). Then by the first order optimality conditions we have that for $x^* \in \text{Sol}\,(P)$ and $\Lambda \in \text{Sol}\,(D)$ the following **complementarity condition** follows

$$(\Sigma_0 + \mathcal{A}(x^*)) \bullet \Lambda = 0.$$

Note that since $(\Sigma_0 + \mathcal{A}(x^*)) \succcurlyeq 0$ and $\Lambda \succcurlyeq 0$, this complementarity condition is equivalent to $(\Sigma_0 + \mathcal{A}(x^*))\Lambda = 0$ and hence $\text{rank}\,(\Lambda) \le p - r$, where

$$r := \text{rank}\,(\Sigma_0 + \mathcal{A}(x^*)).$$

It is said that **the strict complementarity condition** holds at $\Lambda \in \text{Sol}\,(D)$ if $\text{rank}\,(\Lambda) = p - r$.

Suppose also that the strict complementarity condition holds. Let $\Upsilon = NDN^T$ be the spectral decomposition of matrix $\Upsilon = \Sigma_0 + \mathcal{A}(x^*)$, and $\Lambda = EE^T$ for some $p \times (p - r)$ matrix E of rank $p - r$ such that $N^T E = 0$. It is known (see [26]) that the following optimization problem (A.5) depending on $\Delta \in \mathbb{S}^p$ has a unique optimal solution $J^T\delta$ and the optimal value is a quadratic function $\delta^T Q\delta$ where $\delta := \text{vec}\,(\Delta)$, J is a $p^2 \times n$ matrix and Q is a $p^2 \times p^2$ matrix.

$$\begin{cases} \min\limits_{h\in\mathbb{R}^n} \ \text{tr}\,\Big(\Lambda(\mathcal{A}(h)+\Delta)\Upsilon^{\dagger}(\mathcal{A}(h)+\Delta)\Big) \\ \quad s.t.\ \ E^T\mathcal{A}(h)E + E^T\Delta E = 0. \end{cases} \tag{A.5}$$

The following result is Theorem 1 in [26] which can be obtained from Section 5.3.6 in [2].

Proposition A.2 (Theorem 1 in [26])**.** *Suppose that* $\text{Sol}\,(P) = \{x^*\}$ *is a singleton, and that* x^* *is a nondegenerate point of* $\Sigma_0 + \mathcal{A}(\cdot)$ *and the strict complementarity condition holds. Then* $\bar{x}(\cdot)$ *is differentiable at* $\sigma_0 = \text{vec}\,(\Sigma_0)$ *and*

$$\bar{x}(\sigma) = \bar{x}(\sigma_0) + J^T(\sigma - \sigma_0) + o(\|\sigma - \sigma_0\|), \tag{A.6}$$

where $J^T\delta$ *is the optimal solution of problem (A.5). Moreover*

$$\vartheta(\sigma) = \vartheta(\sigma_0) + \Lambda \bullet (\Sigma - \Sigma_0) + (\sigma - \sigma_0)^T Q(\sigma - \sigma_0) + o(\|\sigma - \sigma_0\|^2), \tag{A.7}$$

where Λ *is the optimal solution of the dual problem and* $\delta^T Q\delta$ *is the optimal value of problem (A.5).*

Acknowledgements

The authors are very grateful to two anonymous referees for their helpful comments and suggestions.

References

[1] P.M. Bentler. Lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*, 1:343–357, 1972.

[2] J.F. Bonnans and A. Shapiro. Perturbation Analysis of Optimization Problems. *Springer*, New York, 2000.

[3] A. Chiralaksanakul and D.P. Morton. Assessing policy quality in multi-stage stochastic programming. *Stochastic Programming E-Print Series*, 12, 2004.

[4] A. Dembo and O. Zeitouni. Large Deviation Technique and Applicatons. *Springer*, 1998.

[5] A. Eichorn and W. Römisch. Stochastic integer programming: Limit theorems and confidence intervals. *Math. Oper. Res.*, 32:118–135, 2007.

[6] F.Q. Gao and X.Q. Zhao, Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.*, 39 (2011), 1211–1240.

[7] V. Guigues. Multistep stochastic mirror descent for risk-averse convex stochastic programs based on extended polyhedral risk measures. *Mathematical Programming*, 163:169–212, 2017.

[8] V. Guigues, A. Juditsky, and A. Nemirovski. Non-asymptotic confidence bounds for the optimal value of a stochastic program. *Optimization Methods and Software*, 32:1033–1058, 2017.

[9] V. Guigues, V. Krätschmer, and A. Shapiro. Statistical inference and hypotheses testing of risk averse stochastic programs. *SIAM Journal on Optimization*, 28:1337–1366, 2018.

[10] U. Helmke and J.B. Moore. Optimization and Dynamical Systems. *Springer*, London, 2nd Edition, 1996.

[11] A.J. King and R.T. Rockafellar. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.*, 18:148–162, 1993.

[12] A.J. Kleywegt, A. Shapiro, and T. Homem de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.*, 12:479–502, 2001.

[13] G. Lan, A. Nemirovski, and A. Shapiro. Validation analysis of mirror descent stochastic approximation method. *Math. Program.*, 134:425–458, 2012.

[14] V.V. Petrov. Sums of Independent Random Variables. *Springer-Verlag*, New York, 1975.

[15] G. Pflug. Asymptotic stochastic programs. *Math. Oper. Res.*, 20:769–789, 1995.

[16] G. Pflug. Stochastic programs and statistical data. *Ann. Oper. Res.*, 85:59–78, 1999.

[17] G. 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.

[18] W. Römisch. Delta method infinite dimensional. In S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic (eds.). *Encyclopedia of Statistical Sciences* 16, Wiley, New York (2006) (2nd ed.).

[19] K. Scheinberg. Parametric linear semidefinite programming. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming*, chapter 4, pages 93–110. Kluwer Academic Publishers, Boston, 2000.

[20] A. Shapiro. Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika*, 47:187–199, 1982.

[21] A. Shapiro. Weighted Minimum Trace Factor Analysis. *Psychometrika*, 47:243–264, 1982.

[22] A. Shapiro. Asymptotic properties of statistical estimators in stochastic programming. *Ann. Statist.*, 17:841–858, 1989.

[23] A. Shapiro. Asymptotic analysis of stochastic programs. *Ann. Oper. Res.*, 30:169–186, 1991.

[24] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming*, 77:301–320, 1997.

[25] A. Shapiro. Duality, optimality conditions, and perturbation analysis. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming*, chapter 4, pages 67–92. Kluwer Academic Publishers, Boston, 2000.

[26] A. Shapiro. Statistical inference of semidefinite programming. *Mathematical Programming, Ser B.* First Online, 2018.

[27] A. Shapiro and J.M.F. Ten Berge. Statistical inference of minimum rank factor analysis. *Psychometrika*, 67:79–94, 2002.

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

[29] A. Shapiro and T. Homem de Mello. On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM J. Optim.*, 11:70–86, 2000.

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

[31] V. Yurinsky. Sum and Gaussian Vectors. Lecture Notes in Mathematics 1617. *Springer*, New York, 1995.