A Supplementary Material for "Non-Gaussian Component Analysis with Log-Density Gradient Estimation"

Hiroaki Sasaki
Grad. School of Info. Sci.
Nara Institute of Sci. & Tech.
Nara, Japan
hsasaki@is.naist.jp

Gang Niu
Grad. School of Frontier Sci.
The University of Tokyo
Tokyo, Japan
gang@ms.k.u-tokyo.ac.jp

Masashi Sugiyama
Grad. School of Frontier Sci.
The University of Tokyo
Tokyo, Japan
sugi@k.u-tokyo.ac.jp

1 Proof of Theorem 1

Our proof can be divided into two parts as mentioned in the main manuscript.

1.1 Part One: Convergence of LSLDG

1.1.1 Step 1.1

First of all, we establish the growth condition of LSLDG (see *Definition 6.1* in Bonnans and Shapiro (1998) for the detailed definition of the growth condition). Denote the expected and empirical objective functions by

$$\begin{aligned} J_j^*(\boldsymbol{\theta}) &= \boldsymbol{\theta}^\top \mathbf{G}_j^* \boldsymbol{\theta} + 2\boldsymbol{\theta}^\top \boldsymbol{h}_j^* + \lambda_j^* \boldsymbol{\theta}^\top \boldsymbol{\theta}, \\ \widehat{J}_j(\boldsymbol{\theta}) &= \boldsymbol{\theta}^\top \widehat{\mathbf{G}}_j \boldsymbol{\theta} + 2\boldsymbol{\theta}^\top \widehat{\boldsymbol{h}}_j + \lambda_j \boldsymbol{\theta}^\top \boldsymbol{\theta}. \end{aligned}$$

Then $\boldsymbol{\theta}_j^* = \mathrm{argmin}_{\boldsymbol{\theta}} J_j^*(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\theta}}_j = \mathrm{argmin}_{\boldsymbol{\theta}} \widehat{J}_j(\boldsymbol{\theta})$, and we have

Lemma 1. *The following second-order growth condition holds*

$$J_j^*(\boldsymbol{\theta}) \geq J_j^*(\boldsymbol{\theta}_j^*) + \epsilon_\lambda \|\boldsymbol{\theta} - \boldsymbol{\theta}_j^*\|_2^2.$$

Proof. $J_j^*(\boldsymbol{\theta})$ is strongly convex with parameter at least $2\lambda_j^*$, since $\mathbf{G}_j^*$ is symmetric and positive-definite. Hence,

$$\begin{aligned} J_j^*(\boldsymbol{\theta}) &\geq J_j^*(\boldsymbol{\theta}_j^*) + (\nabla J_j^*(\boldsymbol{\theta}_j^*))^\top (\boldsymbol{\theta} - \boldsymbol{\theta}_j^*) + \lambda_j^* \|\boldsymbol{\theta} - \boldsymbol{\theta}_j^*\|_2^2 \\ &\geq J_j^*(\boldsymbol{\theta}_j^*) + \epsilon_\lambda \|\boldsymbol{\theta} - \boldsymbol{\theta}_j^*\|_2^2, \end{aligned}$$

where we used the optimality condition $\nabla J_j^*(\boldsymbol{\theta}_j^*) = \mathbf{0}$ and the first condition $\lambda_j^* \geq \epsilon_\lambda$ of the theorem. □

1.1.2 Step 1.2

Second, we study the stability (with respect to perturbation) of $J_j^*(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_j^*$. Let

$$\boldsymbol{u} = \{\boldsymbol{u}_G \in \mathcal{S}_+^b, \boldsymbol{u}_h \in \mathbb{R}^b, u_\lambda \in \mathbb{R}\}$$

be a set of perturbation parameters, where b is the number of centers in $\psi_{ij}(\boldsymbol{x})$ and $\mathcal{S}_+^b \subset \mathbb{R}^{b\times b}$ is the cone of b-by-b symmetric positive semi-definite matrices. Define our perturbed objective function by

$$J_j(\boldsymbol{\theta}, \boldsymbol{u}) = \boldsymbol{\theta}^\top(\mathbf{G}_j^* + \boldsymbol{u}_G)\boldsymbol{\theta} + 2\boldsymbol{\theta}^\top(\boldsymbol{h}_j^* + \boldsymbol{u}_h) + (\lambda_j^* + u_\lambda)\boldsymbol{\theta}^\top\boldsymbol{\theta}.$$

It is clear that $J_j^*(\boldsymbol{\theta}) = J_j(\boldsymbol{\theta}, \mathbf{0})$, and then the stability of $J_j^*(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_j^*$ can be characterized as follows.

Lemma 2. *The difference function $J_j(\cdot, \boldsymbol{u}) - J_j^*(\cdot)$ is Lipschitz continuous modulus*

$$\omega(\boldsymbol{u}) = \mathcal{O}(\|\boldsymbol{u}_G\|_{\mathrm{Fro}} + \|\boldsymbol{u}_h\|_2 + |u_\lambda|)$$

on a sufficiently small neighborhood of $\boldsymbol{\theta}_j^*$.

Proof. The difference function is

$$J_j(\boldsymbol{\theta}, \boldsymbol{u}) - J_j^*(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top\boldsymbol{u}_G\boldsymbol{\theta} + 2\boldsymbol{\theta}^\top\boldsymbol{u}_h + u_\lambda\boldsymbol{\theta}^\top\boldsymbol{\theta},$$

with a partial gradient

$$\frac{\partial}{\partial\boldsymbol{\theta}}(J_j(\boldsymbol{\theta}, \boldsymbol{u}) - J_j^*(\boldsymbol{\theta})) = 2\boldsymbol{u}_G\boldsymbol{\theta} + 2\boldsymbol{u}_h + 2u_\lambda\boldsymbol{\theta}.$$

Notice that due to the ℓ_2-regularization in $J_j^*(\boldsymbol{\theta})$, $\exists M > 0$ such that $\|\boldsymbol{\theta}_j^*\|_2 \le M$. Now given a δ-ball of $\boldsymbol{\theta}_j^*$, i.e., $B_\delta(\boldsymbol{\theta}_j^*) = \{\boldsymbol{\theta} \mid \|\boldsymbol{\theta} - \boldsymbol{\theta}_j^*\|_2 \le \delta\}$, it is easy to see that $\forall\boldsymbol{\theta} \in B_\delta(\boldsymbol{\theta}_j^*)$,

$$\|\boldsymbol{\theta}\|_2 \le \|\boldsymbol{\theta} - \boldsymbol{\theta}_j^*\|_2 + \|\boldsymbol{\theta}_j^*\|_2 \le \delta + M,$$

and consequently

$$\left\|\frac{\partial}{\partial\boldsymbol{\theta}}(J_j(\boldsymbol{\theta}, \boldsymbol{u}) - J_j^*(\boldsymbol{\theta}))\right\|_2 \le 2(\delta + M)(\|\boldsymbol{u}_G\|_{\mathrm{Fro}} + |u_\lambda|) + 2\|\boldsymbol{u}_h\|_2.$$

This says that the gradient $\frac{\partial}{\partial\boldsymbol{\theta}}(J_j(\boldsymbol{\theta}, \boldsymbol{u}) - J_j^*(\boldsymbol{\theta}))$ has a bounded norm of order $\mathcal{O}(\|\boldsymbol{u}_G\|_{\mathrm{Fro}} + \|\boldsymbol{u}_h\|_2 + |u_\lambda|)$, and proves that the difference function $J_j(\boldsymbol{\theta}, \boldsymbol{u}) - J_j^*(\boldsymbol{\theta})$ is Lipschitz continuous on the ball $B_\delta(\boldsymbol{\theta}_j^*)$, with a Lipschitz constant of the same order. □

1.1.3 Step 1.3

Intuitively, Lemma 1 guarantees that the unperturbed objective function $J_j^*(\boldsymbol{\theta})$ grows quickly when $\boldsymbol{\theta}$ leaves $\boldsymbol{\theta}_j^*$. Lemma 2 guarantees that the perturbed objective function $J_j(\boldsymbol{\theta}, \boldsymbol{u})$ changes slowly for $\boldsymbol{\theta}$ around $\boldsymbol{\theta}_j^*$, where the slowness is with respect to the perturbation $\boldsymbol{u}$ it suffers. Based on Lemma 1, Lemma 2, and *Proposition 6.1* in Bonnans and Shapiro (1998),

$$\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^*\|_2 \le \frac{\omega(\boldsymbol{u})}{\epsilon_\lambda} = \mathcal{O}(\|\boldsymbol{u}_G\|_{\mathrm{Fro}} + \|\boldsymbol{u}_h\|_2 + |u_\lambda|),$$

since $\widehat{\boldsymbol{\theta}}_j$ is the exact solution to $\widehat{J}_j(\boldsymbol{\theta}) = J_j(\boldsymbol{\theta}, \boldsymbol{u})$ given $\boldsymbol{u}_G = \widehat{\mathbf{G}}_j - \mathbf{G}_j^*$, $\boldsymbol{u}_h = \widehat{\mathbf{h}}_j - \mathbf{h}_j^*$, and $u_\lambda = \lambda_j - \lambda_j^*$.

According to the *central limit theorem* (CLT), $\|\boldsymbol{u}_G\|_{\mathrm{Fro}} = \mathcal{O}_p(n^{-1/2})$ and $\|\boldsymbol{u}_h\|_2 = \mathcal{O}_p(n^{-1/2})$. Furthermore, we have already assumed that $|u_\lambda| = \mathcal{O}(n^{-1/2})$ in the first condition of the theorem. Hence, as $n \to \infty$,

$$\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^*\|_2 = \mathcal{O}_p\left(n^{-1/2}\right). \tag{1}$$

1.1.4 Step 1.4

Considering the empirical estimate of the log-density gradient $\widehat{g}^{(j)}(\boldsymbol{x})$ and the optimal estimate of the log-density gradient $g^{*(j)}(\boldsymbol{x})$, their gap in terms of the infinity norm is bounded below:

$$\begin{aligned}\|\widehat{g}^{(j)} - g^{*(j)}\|_\infty &= \sup_{\boldsymbol{x}} |\widehat{g}^{(j)}(\boldsymbol{x}) - g^{*(j)}(\boldsymbol{x})| \\ &= \sup_{\boldsymbol{x}} |(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^*)^\top \boldsymbol{\psi}_j(\boldsymbol{x})| \\ &\le \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^*\|_2 \cdot \sup_{\boldsymbol{x}} \|\boldsymbol{\psi}_j(\boldsymbol{x})\|_2,\end{aligned}$$

where the *Cauchy-Schwarz inequality* is used. Recall that $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_b$ are the centers, and for any i,

$$|\psi_{ij}(\boldsymbol{x})| = \frac{|[\boldsymbol{c}_i - \boldsymbol{x}]^{(j)}|}{\sigma_j^2} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{c}_i\|_2^2}{2\sigma_j^2}\right) \le \frac{|[\boldsymbol{c}_i - \boldsymbol{x}]^{(j)}|}{\sigma_j^2} \left(-\frac{([\boldsymbol{x} - \boldsymbol{c}_i]^{(j)})^2}{2\sigma_j^2}\right).$$

It is obvious that $|\psi_{ij}(\boldsymbol{x})|$ is bounded, since $\exp(-z^2)$ converges to zero much faster than $|z|$ diverges to infinity. Therefore, $\sup_{\boldsymbol{x}} \|\boldsymbol{\psi}_j(\boldsymbol{x})\|_2$ is a finite number, and we could know from Eq. (1) that

$$\|\widehat{g}^{(j)} - g^{*(j)}\|_\infty \le \mathcal{O}(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^*\|_2) = \mathcal{O}_p\left(n^{-1/2}\right). \tag{2}$$

1.2 Part Two: Convergence of LSNGCA

1.2.1 Step 2.1

To begin with, we focus on the convergence of $\widehat{\boldsymbol{\Gamma}}$. Given any $\boldsymbol{y}$, let $\widehat{\boldsymbol{z}} = \widehat{\boldsymbol{g}}(\boldsymbol{y})$ and $\boldsymbol{z}^* = \boldsymbol{g}^*(\boldsymbol{y})$, then

$$\begin{aligned}(\widehat{\boldsymbol{z}} + \boldsymbol{y})(\widehat{\boldsymbol{z}} + \boldsymbol{y})^\top - (\boldsymbol{z}^* + \boldsymbol{y})(\boldsymbol{z}^* + \boldsymbol{y})^\top &= \widehat{\boldsymbol{z}}\widehat{\boldsymbol{z}}^\top - \boldsymbol{z}^*\boldsymbol{z}^{*\top} + (\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)\boldsymbol{y}^\top + \boldsymbol{y}(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)^\top \\ &= (\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)\widehat{\boldsymbol{z}}^\top + \boldsymbol{z}^*(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)^\top + (\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)\boldsymbol{y}^\top + \boldsymbol{y}(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)^\top.\end{aligned}$$

As a result, based on Eq. (2),

$$\begin{aligned}\|(\widehat{\boldsymbol{z}} + \boldsymbol{y})(\widehat{\boldsymbol{z}} + \boldsymbol{y})^\top - (\boldsymbol{z}^* + \boldsymbol{y})(\boldsymbol{z}^* + \boldsymbol{y})^\top\|_{\mathrm{Fro}} &\le \|(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)\widehat{\boldsymbol{z}}^\top\|_{\mathrm{Fro}} + \|\boldsymbol{z}^*(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)^\top\|_{\mathrm{Fro}} \\ &\quad + \|(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)\boldsymbol{y}^\top\|_{\mathrm{Fro}} + \|\boldsymbol{y}(\widehat{\boldsymbol{z}} - \boldsymbol{z}^*)^\top\|_{\mathrm{Fro}} \\ &\le (\|\widehat{\boldsymbol{z}}\|_2 + \|\boldsymbol{z}^*\|_2 + 2\|\boldsymbol{y}\|) \cdot \|\widehat{\boldsymbol{z}} - \boldsymbol{z}^*\|_2 \\ &= \mathcal{O}(\|\widehat{\boldsymbol{z}} - \boldsymbol{z}^*\|_2) \\ &= \mathcal{O}_p\left(n^{-1/2}\right).\end{aligned}$$

This has proved the point-wise convergence from $(\widehat{\boldsymbol{g}}(\boldsymbol{y}) + \boldsymbol{y})(\widehat{\boldsymbol{g}}(\boldsymbol{y}) + \boldsymbol{y})^\top$ to $(\boldsymbol{g}^*(\boldsymbol{y}) + \boldsymbol{y})(\boldsymbol{g}^*(\boldsymbol{y}) + \boldsymbol{y})^\top$.

Define an intermediate matrix based on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ as

$$\widetilde{\boldsymbol{\Gamma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{g}^*(\boldsymbol{y}_i) + \boldsymbol{y}_i)(\boldsymbol{g}^*(\boldsymbol{y}_i) + \boldsymbol{y}_i)^\top.$$

Subsequently, $\widehat{\boldsymbol{\Gamma}}$ converges to $\widetilde{\boldsymbol{\Gamma}}$ in $\mathcal{O}_p(n^{-1/2})$ due to the point-wise convergence from $(\widehat{\boldsymbol{g}}(\boldsymbol{y}) + \boldsymbol{y})(\widehat{\boldsymbol{g}}(\boldsymbol{y}) + \boldsymbol{y})^\top$ to $(\boldsymbol{g}^*(\boldsymbol{y}) + \boldsymbol{y})(\boldsymbol{g}^*(\boldsymbol{y}) + \boldsymbol{y})^\top$ that was just proved, and $\widetilde{\boldsymbol{\Gamma}}$ converges to $\boldsymbol{\Gamma}^*$ in $\mathcal{O}_p(n^{-1/2})$ due to CLT. A combination of these two results gives us

$$\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}^*\|_{\mathrm{Fro}} \le \|\widehat{\boldsymbol{\Gamma}} - \widetilde{\boldsymbol{\Gamma}}\|_{\mathrm{Fro}} + \|\widetilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}^*\|_{\mathrm{Fro}} = \mathcal{O}_p\left(n^{-1/2}\right). \tag{3}$$

1.2.2 Step 2.2

Now let us consider the eigenvalues of $\boldsymbol{\Gamma}^*$. Let $\mu_1 > \cdots > \mu_r > \mu_{r+1}$ be the first $r+1$ eigenvalues of $\boldsymbol{\Gamma}^*$ *counted without multiplicity*, such that μ_r is the $d_{\mathbf{s}}$-th largest eigenvalue of $\boldsymbol{\Gamma}^*$ if counted with multiplicity. Define the eigen-gap by

$$\epsilon_\mu = \min_{i=1,\ldots,r}\{\mu_i - \mu_{i+1}\}.$$

We have assumed that $\mu_1 < +\infty$ and $\mu_r > 0$ in the second condition of the theorem, and thus it must hold that $0 < \epsilon_\mu < +\infty$. In the case that $\mathbf{\Gamma}^*$ has only one eigenvalue, we can simply assign $\epsilon_\mu = 1$.

According to *Lemma 5.2* of Koltchinskii and Giné (2000) as well as the appendix of Koltchinskii (1998), we can derive the stability of the eigen-decomposition of $\mathbf{\Gamma}^*$ with respect to some perturbation $\boldsymbol{u}_\Sigma = \widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}^*$. Whenever $\|\boldsymbol{u}_\Sigma\|_{\mathrm{Fro}} < \epsilon_\mu/4$:

- The first $r+1$ eigenvalues $\mu'_1 > \cdots > \mu'_r > \mu'_{r+1}$ of $\widehat{\mathbf{\Gamma}} = \mathbf{\Gamma}^* + \boldsymbol{u}_\Sigma$, counted without multiplicity, satisfy that $|\mu'_i - \mu_i| \le \|\boldsymbol{u}_\Sigma\|_{\mathrm{Fro}}$ for $1 \le i \le r$, and $\mu_r - \mu'_{r+1} \ge \epsilon_\mu - \|\boldsymbol{u}_\Sigma\|_{\mathrm{Fro}}$;
- Denote by $\Pi_i(\mathbf{\Gamma}^*)$ the orthogonal projection onto the eigen-spaces of $\mathbf{\Gamma}^*$ associated with μ_i, and by $\Pi_i(\widehat{\mathbf{\Gamma}})$ that of $\widehat{\mathbf{\Gamma}} = \mathbf{\Gamma}^* + \boldsymbol{u}_\Sigma$ associated with μ'_i, then for $1 \le i \le r$,

$$\|\Pi_i(\widehat{\mathbf{\Gamma}}) - \Pi_i(\mathbf{\Gamma}^*)\|_{\mathrm{Fro}} \le \frac{4}{\epsilon_\mu}\|\boldsymbol{u}_\Sigma\|_{\mathrm{Fro}}.$$

We have employed simplified notations above to avoid sophisticated names in operator theory. Intuitively, the first item guarantees that the eigenvalues of the perturbed matrix $\widehat{\mathbf{\Gamma}}$ are close to that of $\mathbf{\Gamma}^*$, and the second item guarantees that the eigen-spaces of $\widehat{\mathbf{\Gamma}}$ are also close to that of $\mathbf{\Gamma}^*$.

By noting that $\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}^*\|_{\mathrm{Fro}}$ was shown to have an order of $\mathcal{O}_p(n^{-1/2})$ in (3), whereas the eigen-gap ϵ_μ for fixed $\mathbf{\Gamma}^*$ is a constant value, we could obtain that as $n \to \infty$ for all i,

$$\|\Pi_i(\widehat{\mathbf{\Gamma}}) - \Pi_i(\mathbf{\Gamma}^*)\|_{\mathrm{Fro}} = \mathcal{O}_p\left(n^{-1/2}\right). \tag{4}$$

1.2.3 Step 2.3

Finally, we can bound $\mathcal{D}(\widehat{\mathcal{L}}, \mathcal{L}^*)$. The eigenvalues of $\mathbf{\Gamma}^*$ and $\widehat{\mathbf{\Gamma}}$ were counted without multiplicity, and hence the bases of $\Pi_i(\widehat{\mathbf{\Gamma}})$ and $\Pi_i(\mathbf{\Gamma}^*)$ may not be unique. Nevertheless, let $\mathbf{E}_{\mathcal{I}^*}$ be the matrix form of a fixed orthonormal basis of $\mathcal{I}^*$, then there exists a sequence of matrices $\{\mathbf{E}_{\widehat{\mathcal{I}},1}, \dots, \mathbf{E}_{\widehat{\mathcal{I}},n}, \dots\}$ such that

- $\mathbf{E}_{\widehat{\mathcal{I}},n}$ is the matrix form of a certain orthonormal basis of $\widehat{\mathcal{I}}$ based on a set of data samples of size n;
- The sequence converges to $\mathbf{E}_{\mathcal{I}^*}$ in $\mathcal{O}_p(n^{-1/2})$, i.e.,

$$\|\mathbf{E}_{\widehat{\mathcal{I}},n} - \mathbf{E}_{\mathcal{I}^*}\|_{\mathrm{Fro}} = \mathcal{O}_p\left(n^{-1/2}\right), \tag{5}$$

based on Eq. (4). Denote by $\mathbf{E}_{\mathcal{L}^*} = \mathbf{\Sigma}^{-1/2}\mathbf{E}_{\mathcal{I}^*}$ and $\mathbf{E}_{\widehat{\mathcal{L}},n} = \widehat{\mathbf{\Sigma}}^{-1/2}\mathbf{E}_{\widehat{\mathcal{I}},n}$, and then

$$\mathbf{E}_{\widehat{\mathcal{L}},n} - \mathbf{E}_{\mathcal{L}^*} = \widehat{\mathbf{\Sigma}}^{-1/2}\mathbf{E}_{\widehat{\mathcal{I}},n} - \mathbf{\Sigma}^{-1/2}\mathbf{E}_{\mathcal{I}^*} = (\widehat{\mathbf{\Sigma}}^{-1/2} - \mathbf{\Sigma}^{-1/2})\mathbf{E}_{\widehat{\mathcal{I}},n} + \mathbf{\Sigma}^{-1/2}(\mathbf{E}_{\widehat{\mathcal{I}},n} - \mathbf{E}_{\mathcal{I}^*}).$$

Therefore, we can prove that

$$\begin{aligned}
\mathcal{D}(\widehat{\mathcal{L}}, \mathcal{L}^*) &= \inf_{\widehat{\mathbf{E}},\mathbf{E}^*} \|\widehat{\mathbf{E}} - \mathbf{E}^*\|_{\mathrm{Fro}} \\
&\le \|\mathbf{E}_{\widehat{\mathcal{L}},n} - \mathbf{E}_{\mathcal{L}^*}\|_{\mathrm{Fro}} \\
&\le \|\mathbf{E}_{\widehat{\mathcal{I}},n}\|_{\mathrm{Fro}} \cdot \|\widehat{\mathbf{\Sigma}}^{-1/2} - \mathbf{\Sigma}^{-1/2}\|_{\mathrm{Fro}} + \|\mathbf{\Sigma}^{-1/2}\|_{\mathrm{Fro}} \cdot \|\mathbf{E}_{\widehat{\mathcal{I}},n} - \mathbf{E}_{\mathcal{I}^*}\|_{\mathrm{Fro}} \\
&= \mathcal{O}(\|\widehat{\mathbf{\Sigma}}^{-1/2} - \mathbf{\Sigma}^{-1/2}\|_{\mathrm{Fro}}) + \mathcal{O}(\|\mathbf{E}_{\widehat{\mathcal{I}},n} - \mathbf{E}_{\mathcal{I}^*}\|_{\mathrm{Fro}}) \\
&= \mathcal{O}_p\left(n^{-1/2}\right),
\end{aligned}$$

according to CLT and Eq. (5). □

References

F. Bonnans and A. Shapiro. Optimization problems with perturbations, a guided tour. *SIAM Review*, 40(2): 228–264, 1998.

V. Koltchinskii. Asymptotics of spectral projections of some random matrices approximating integral operators. *Progress in Probabilty*, 43:191–227, 1998.

V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6: 113–167, 2000.