J. Eur. Math. Soc. 23, 765–843 (2021) DOI 10.4171/JEMS/1023

Vladimir Koltchinskii

Asymptotically efficient estimation of smooth functionals of covariance operators

Received January 15, 2018

Abstract. Let X be a centered Gaussian random variable in a separable Hilbert space $\mathbb{H}$ with covariance operator Σ. We study the problem of estimation of a smooth functional of Σ based on a sample $X_1, \ldots, X_n$ of n independent observations of X. More specifically, we are interested in functionals of the form $\langle f(\Sigma), B\rangle$, where $f : \mathbb{R} \to \mathbb{R}$ is a smooth function and B is a nuclear operator in $\mathbb{H}$. We prove concentration and normal approximation bounds for the plug-in estimator $\langle f(\hat{\Sigma}), B\rangle$, $\hat{\Sigma} := n^{-1} \sum_{j=1}^n X_j \otimes X_j$ being the sample covariance based on $X_1, \ldots, X_n$. These bounds show that $\langle f(\hat{\Sigma}), B\rangle$ is an asymptotically normal estimator of its expectation $\mathbb{E}_\Sigma \langle f(\hat{\Sigma}), B\rangle$ (rather than of the parameter of interest $\langle f(\Sigma), B\rangle$) with parametric convergence rate $O(n^{-1/2})$ provided that the effective rank $\mathbf{r}(\Sigma) := \mathrm{tr}(\Sigma)/\|\Sigma\|$ ($\mathrm{tr}(\Sigma)$ the trace and $\|\Sigma\|$ the operator norm of Σ) satisfies the assumption $\mathbf{r}(\Sigma) = o(n)$. At the same time, we show that the bias of this estimator is typically as large as $\mathbf{r}(\Sigma)/n$ (which is larger than $n^{-1/2}$ if $\mathbf{r}(\Sigma) \geq n^{1/2}$). When $\mathbb{H}$ is a finite-dimensional space of dimension $d = o(n)$, we develop a method of bias reduction and construct an estimator $\langle h(\hat{\Sigma}), B\rangle$ of $\langle f(\Sigma), B\rangle$ that is asymptotically normal with convergence rate $O(n^{-1/2})$. Moreover, we study the asymptotic properties of the risk of this estimator and prove asymptotic minimax lower bounds for arbitrary estimators showing the asymptotic efficiency of $\langle h(\hat{\Sigma}), B\rangle$ in a semiparametric sense.

Keywords. Asymptotic efficiency, sample covariance, bootstrap, effective rank, concentration inequalities, normal approximation, perturbation theory

1. Introduction

Let X be a random variable in a separable Hilbert space $\mathbb{H}$ sampled from a Gaussian distribution with mean 0 and covariance operator $\Sigma := \mathbb{E}(X \otimes X)$ (denoted $N(0; \Sigma)$). The purpose of this paper is to study the problem of estimation of smooth functionals of unknown covariance Σ based on a sample $X_1, \ldots, X_n$ of i.i.d. observations of X. Specifically, we deal with the functionals of the form $\langle f(\Sigma), B\rangle$, where $f : \mathbb{R} \to \mathbb{R}$ is a smooth function[1] and B is a nuclear operator. The estimation of bilinear forms of spectral projection operators of Σ, which is of importance in the principal component analysis,

V. Koltchinskii: School of Mathematics, Georgia Institute of Technology,
Atlanta, GA 30332-0160, USA; e-mail: vlad@math.gatech.edu

Mathematics Subject Classification (2020): Primary 62H12; Secondary 62G20, 62H25, 60B20

[1] More precisely, the "smoothness" in this paper means that the function belongs to the Besov space $B^s_{\infty,1}(\mathbb{R})$ for a certain value of $s > 0$ (see Subsection 1.2).

can be easily reduced to this basic problem. Moreover, the estimation of $\langle f(\Sigma), B\rangle$ is a major building block in the development of methods of statistical estimation of more general functionals of covariance like $\langle f_1(\Sigma), B_1\rangle \dots \langle f_k(\Sigma), B_k\rangle$ and their linear combinations.

Throughout the paper, we use the following notations. Given $A, B \geq 0$, $A \lesssim B$ means that $A \leq CB$ for a numerical (most often, unspecified) constant $C > 0$; $A \gtrsim B$ is equivalent to $B \lesssim A$; $A \asymp B$ is equivalent to $A \lesssim B$ and $B \lesssim A$. Sometimes, constants in the above relationships might depend on some parameter(s). In such cases, the signs $\lesssim$, $\gtrsim$ and $\asymp$ are provided with subscripts: say, $A \lesssim_\gamma B$ means that $A \leq C_\gamma B$ for a constant $C_\gamma > 0$ that depends on γ.

Let $\mathcal{B}(\mathbb{H})$ denote the space of all bounded linear operators in a separable Hilbert space $\mathbb{H}$ equipped with the operator norm and let $\mathcal{B}_{sa}(\mathbb{H})$ denote the subspace of all self-adjoint operators.[2] In what follows, A^* denotes the adjoint of $A \in \mathcal{B}(\mathbb{H})$, $\mathrm{tr}(A)$ denotes its trace (provided that A is trace class) and $\|A\|$ denotes its operator norm. We use the notation $\|A\|_p$ for the Schatten p-norm of A: $\|A\|_p^p := \mathrm{tr}(|A|^p)$, $|A| = (A^*A)^{1/2}$, $p \in [1, \infty]$. In particular, $\|A\|_1$ is the nuclear norm of A, $\|A\|_2$ is its Hilbert–Schmidt norm and $\|A\|_\infty = \|A\|$ is its operator norm. We denote the space of self-adjoint operators A with $\|A\|_p < \infty$ (p-th Schatten class operators) by $\mathcal{S}_p = \mathcal{S}_p(\mathbb{H})$, $1 \leq p \leq \infty$. The space of compact self-adjoint operators in $\mathbb{H}$ is denoted by $\mathcal{C}_{sa}(\mathbb{H})$. The inner product notation $\langle \cdot, \cdot \rangle$ is used both for inner products in the underlying Hilbert space $\mathbb{H}$ and for the Hilbert–Schmidt inner product between operators. Moreover, it is also used to denote bounded linear functionals on spaces of operators (for instance, $\langle A, B\rangle$, where A is a bounded operator and B is a nuclear operator, is a value of such a linear functional on the space of bounded operators). For $u, v \in \mathbb{H}$, $u \otimes v$ denotes the tensor product of vectors u and v: $(u \otimes v)x := u\langle v, x\rangle$ for $x \in \mathbb{H}$. The operator $u \otimes v$ is of rank 1 and finite linear combinations of rank 1 operators are operators of finite rank. The rank of A is denoted by $\mathrm{rank}(A)$. Finally, $\mathcal{C}_+(\mathbb{H})$ denotes the cone of self-adjoint positive semidefinite nuclear operators in $\mathbb{H}$ (covariance operators).

In what follows, we often use exponential bounds for random variables of the following form: for all $t \geq 1$, with probability at least $1 - e^{-t}$ we have $\xi \leq Ct$. Sometimes, our derivation would yield a slightly different probability bound, for instance: for all $t \geq 1$ with probability at least $1 - 3e^{-t}$ we have $\xi \leq Ct$. Such bounds could be easily rewritten again as $1 - e^{-t}$ by adjusting the value of C: for $t \geq 1$ with probability at least $1 - e^{-t} = 1 - 3e^{-t-\log(3)}$ we have $\xi \leq C(t + \log(3)) \leq 2\log(3)Ct$. Such an adjustment of constants will be used in many proofs without further notice.

[2] The main results of the paper are proved in the case when $\mathbb{H}$ is a real Hilbert space. However, on a couple of occasions, especially in auxiliary statements, its complexification $\mathbb{H}^{\mathbb{C}} = \{u + iv : u, v \in \mathbb{H}\}$ with a standard extension of the inner product and complexification of the operators acting in $\mathbb{H}$ is needed. With some abuse of notation, we keep in such cases the notation $\mathbb{H}$ for the complex Hilbert space.

1.1. Sample covariance and effective rank

Let $\hat{\Sigma}$ denote the sample covariance based on the data $X_1, \dots, X_n$:

$$\hat{\Sigma} := n^{-1} \sum_{j=1}^{n} X_j \otimes X_j.$$

It is well known that $\hat{\Sigma}$ is a complete sufficient statistic and equals the maximum likelihood estimator in the problem of estimation of the unknown covariance of i.i.d. observations $X_1, \dots, X_n$ sampled from $N(0; \Sigma)$.

In what follows, we often use the so called *effective rank* of the covariance Σ as a complexity parameter of the covariance estimation problem. It is defined as

$$\mathbf{r}(\Sigma) := \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|}.$$

Note that $\mathbf{r}(\Sigma) \le \mathrm{rank}(\Sigma) \le \dim(\mathbb{H})$. The following result of Koltchinskii and Lounici [KL2] shows that, in the Gaussian case, the size of the random variable $\|\hat{\Sigma} - \Sigma\|/\|\Sigma\|$ (which is a relative operator norm error of the estimator $\hat{\Sigma}$ of Σ) is completely characterized by the ratio $\mathbf{r}(\Sigma)/n$.

Theorem 1. *The following bound holds:*

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \asymp \|\Sigma\| \biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n}\biggr). \tag{1.1}$$

Moreover, for all $t \ge 1$, with probability at least $1 - e^{-t}$,

$$\bigl|\|\hat{\Sigma} - \Sigma\| - \mathbb{E}\|\hat{\Sigma} - \Sigma\|\bigr| \lesssim \|\Sigma\| \biggl(\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee 1\biggr)\sqrt{\frac{t}{n}} \vee \frac{t}{n}\biggr). \tag{1.2}$$

It follows from the expectation bound (1.1) and the concentration inequality (1.2) that for all $t \ge 1$, with probability at least $1 - e^{-t}$,

$$\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\biggr) \tag{1.3}$$

and, for all $p \ge 1$,

$$\mathbb{E}^{1/p}\|\hat{\Sigma} - \Sigma\|^p \lesssim_p \|\Sigma\| \biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n}\biggr). \tag{1.4}$$

To avoid the dependence of the constant on p, the following modification of the above bound will be used on a couple of occasions:

$$\mathbb{E}^{1/p}\|\hat{\Sigma} - \Sigma\|^p \lesssim \|\Sigma\| \biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{p}{n}} \vee \frac{p}{n}\biggr). \tag{1.5}$$

Since $\mathbf{r}(\Sigma) \leq d := \dim(\mathbb{H})$, the bounds in terms of effective rank imply well known bounds in terms of dimension. For instance, for all $t \geq 1$, with probability at least $1-e^{-t}$,

$$\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \left(\sqrt{\frac{d}{n}} \vee \frac{d}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n} \right) \tag{1.6}$$

(see, e.g., [Ver]). Of course, the bound (1.6) is meaningless in the infinite-dimensional case. In the finite-dimensional case, it is sharp if Σ is isotropic ($\Sigma = cI_d$ for a constant c), or if it is of *isotropic type*, that is, the spectrum of Σ is bounded above and bounded away from zero (by constants). In this case, $\mathbf{r}(\Sigma) \asymp d$, which makes (1.6) sharp. This is the case, for instance, for popular *spiked covariance models* introduced by Johnstone [Jo] (see also [JoLu, Paul, BJNP]). However, in the case of fast decay of eigenvalues of Σ, the effective rank $\mathbf{r}(\Sigma)$ could be significantly smaller than d and it becomes the right complexity parameter in covariance estimation.

In what follows, we are interested in problems in which $\mathbf{r}(\Sigma)$ is allowed to be large, but $\mathbf{r}(\Sigma) = o(n)$ as $n \to \infty$. This is a necessary and sufficient condition for $\hat{\Sigma}$ to be an operator norm consistent estimator of Σ, which also means that $\hat{\Sigma}$ is a small perturbation of Σ when n is large and methods of perturbation theory can be used to analyze the behavior of $f(\hat{\Sigma})$ for smooth functions f.

1.2. Overview of main results

In this subsection, we state and discuss the main results of the paper concerning asymptotically efficient estimation of the functionals $\langle f(\Sigma), B\rangle$ for a smooth function $f : \mathbb{R} \to \mathbb{R}$ and a nuclear operator B. It turns out that the proper notion of smoothness of f in these problems is in terms of Besov spaces and Besov norms. The relevant definitions (of the spaces $B^s_{\infty,1}(\mathbb{R})$ and the corresponding norms), notations and references are provided in Section 2.

A standard approach to asymptotic analysis of plug-in estimators (in particular, such as $\langle f(\hat{\Sigma}), B\rangle$) in statistics is the Delta Method based on the first order Taylor expansion of $f(\hat{\Sigma})$. Due to a result by Peller (see Section 2), for any $f \in B^1_{\infty,1}(\mathbb{R})$, the mapping $A \mapsto f(A)$ is Fréchet differentiable with respect to the operator norm on the space of bounded self-adjoint operators in $\mathbb{H}$. Let Σ be a covariance operator with spectral decomposition $\Sigma := \sum_{\lambda \in \sigma(\Sigma)} \lambda P_\lambda$, $\sigma(\Sigma)$ the spectrum of Σ, λ an eigenvalue of Σ and P_λ the corresponding spectral projection (the orthogonal projection onto the eigenspace of Σ). Then the derivative $Df(\Sigma)(H) = Df(\Sigma; H)$ of the operator function $f(A)$ at $A = \Sigma$ in the direction H is given by

$$Df(\Sigma; H) = \sum_{\lambda,\mu \in \sigma(\Sigma)} f^{[1]}(\lambda,\mu) P_\lambda H P_\mu,$$

where $f^{[1]}(\lambda,\mu) = \frac{f(\lambda)-f(\mu)}{\lambda-\mu}$ for $\lambda \neq \mu$ and $f^{[1]}(\lambda,\lambda) = f'(\lambda)$ (see Section 2). Moreover, if $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$, then the following first order Taylor expansion holds:

$$f(\hat{\Sigma}) - f(\Sigma) = Df(\Sigma; \hat{\Sigma} - \Sigma) + S_f(\Sigma; \hat{\Sigma} - \Sigma)$$

with the linear term $Df(\Sigma; \hat{\Sigma}-\Sigma) = n^{-1}\sum_{j=1}^{n} Df(\Sigma; X_j \otimes X_j - \Sigma)$ and the remainder $S_f(\Sigma; \hat{\Sigma} - \Sigma)$ satisfying the bound

$$\|S_f(\Sigma; \hat{\Sigma} - \Sigma)\| \lesssim_s \|f\|_{B^s_{\infty,1}} \|\hat{\Sigma} - \Sigma\|^s$$

(see (2.15)). Since the linear term $Df(\Sigma; \hat{\Sigma}-\Sigma)$ is the sum of i.i.d. random variables, it is easy to check (for instance, using the Berry–Esseen bound) that $\sqrt{n}\,\langle Df(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$ is asymptotically normal with limit mean zero and limit variance

$$\sigma_f^2(\Sigma; B) := 2\|\Sigma^{1/2} Df(\Sigma; B)\Sigma^{1/2}\|_2^2.$$

Using the exponential bound (1.3) on $\|\hat{\Sigma} - \Sigma\|$, one can easily conclude that the remainder $\langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$ is asymptotically negligible (that is, of order $o(n^{-1/2})$) if $(\mathbf{r}(\Sigma)/n)^{s/2} = o(n^{-1/2})$, or equivalently $\mathbf{r}(\Sigma) = o(n^{1-1/s})$. In the case when $s = 2$, this means that $\mathbf{r}(\Sigma) = o(n^{1/2})$. This implies that $\langle f(\hat{\Sigma}), B\rangle$ is an asymptotically normal estimator of $\langle f(\Sigma), B\rangle$ with convergence rate $n^{-1/2}$ and limit normal distribution $N(0; \sigma_f^2(\Sigma; B))$ (under the assumption that $\mathbf{r}(\Sigma) = o(n^{1-1/s})$). The above perturbation analysis is essentially the same as for spectral projections of $\hat{\Sigma}$ in the case of fixed finite dimension (see Anderson [A]), or in the infinite-dimensional case when the "complexity" of the problem (characterized by $\operatorname{tr}(\Sigma)$ or $\mathbf{r}(\Sigma)$) is fixed (see Dauxois, Pousse and Romain [DPR]). Note also that the bias of the estimator $\langle f(\hat{\Sigma}), B\rangle$,

$$\langle \mathbb{E}_\Sigma f(\hat{\Sigma}) - f(\Sigma), B\rangle = \langle \mathbb{E}_\Sigma S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle,$$

is $\lesssim \|f\|_{B^s_{\infty,1}} \|B\|_1 (\mathbf{r}(\Sigma)/n)^{s/2}$, so it is of order $o(n^{-1/2})$ (asymptotically negligible) under the same condition $\mathbf{r}(\Sigma) = o(n^{1-1/s})$. Moreover, it is easy to see that this bound on the bias is sharp for generic smooth functions f. For instance, if $f(x) = x^2$ and $B = u \otimes u$, then one can check by a straightforward computation that

$$\sup_{\|u\|\le 1} |\langle \mathbb{E}_\Sigma f(\hat{\Sigma}) - f(\Sigma), u \otimes u\rangle| = \frac{\|\operatorname{tr}(\Sigma)\Sigma + \Sigma^2\|}{n} \asymp \|\Sigma\|^2 \frac{\mathbf{r}(\Sigma)}{n}.$$

This means that, as long as $\mathbf{r}(\Sigma) \ge n^{1/2}$, one can choose a vector u from the unit ball (for which the supremum is "nearly attained") such that both the bias and the remainder are not asymptotically negligible, and moreover it turns out that if $\mathbf{r}(\Sigma)/n^{1/2} \to \infty$, then $\langle f(\hat{\Sigma}), B\rangle$ is not even a $\sqrt{n}$-consistent estimator of $\langle f(\Sigma), B\rangle$. If in addition the operator norm $\|\Sigma\|$ is bounded by a constant $R > 0$, one can find a function in the space $B^2_{\infty,1}(\mathbb{R})$ that coincides with $f(x) = x^2$ in a neighborhood of the interval $[0, R]$, and the above claims hold for this function, too (see also Remark 2 below).

Our first goal is to show that $\langle f(\hat{\Sigma}), B\rangle$ is an asymptotically normal estimator of its own expectation $\langle \mathbb{E}_\Sigma f(\hat{\Sigma}), B\rangle$ with convergence rate $n^{-1/2}$ and limit variance $\sigma_f^2(\Sigma; B)$ in the class of covariances with effective rank of order $o(n)$. Given $r > 1$ and $a > 0$, define $\mathcal{G}(r; a) := \{\Sigma : \mathbf{r}(\Sigma) \le r,\ \|\Sigma\| \le a\}$.

Theorem 2. *Suppose $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$. Let $a, \sigma_0 > 0$. Suppose that $r_n > 1$ and $r_n = o(n)$ as $n \to \infty$. Then*

$$\sup_{\Sigma \in \mathcal{G}(r_n;a),\, \|B\|_1 \le 1,\, \sigma_f(\Sigma;B) \ge \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{n^{1/2}\langle f(\hat{\Sigma}) - \mathbb{E}_\Sigma f(\hat{\Sigma}), B\rangle}{\sigma_f(\Sigma;B)} \le x \right\} - \Phi(x) \right| \to 0 \tag{1.7}$$

as $n \to \infty$, where $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2}\, dt$, $x \in \mathbb{R}$.

This result is a consequence of Corollary 4 proved in Section 4 that provides an explicit bound on the accuracy of normal approximation. Its proof is based on a concentration bound for the remainder $\langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$ of the first order Taylor expansion developed in Section 3. This bound essentially shows that the centered remainder

$$\langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle - \mathbb{E}\langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$$

is of order $(\mathbf{r}(\Sigma)/n)^{(s-1)/2}\sqrt{1/n}$, which is $o(n^{-1/2})$ as long as $\mathbf{r}(\Sigma) = o(n)$.

Theorem 2 shows that the naive plug-in estimator $\langle f(\hat{\Sigma}), B\rangle$ "concentrates" around its expectation with approximately standard normal distribution of the random variables

$$\frac{n^{1/2}\langle f(\hat{\Sigma}) - \mathbb{E}_\Sigma f(\hat{\Sigma}), B\rangle}{\sigma_f(\Sigma;B)}.$$

At the same time, as discussed above, the plug-in estimator has a large bias when the effective rank of Σ is sufficiently large (say, $\mathbf{r}(\Sigma) \ge n^{1/2}$ for functions f of smoothness $s = 2$). In the case when $\Sigma \in \mathcal{G}(r_n; a)$ with $r_n = o(n^{1/2})$ and $\sigma_f(\Sigma; B) \ge \sigma_0$, the bias is negligible and $\langle f(\hat{\Sigma}), B\rangle$ becomes an asymptotically normal estimator of $\langle f(\Sigma), B\rangle$. Moreover, we will also derive the asymptotics of the risk of the plug-in estimator for loss functions satisfying the following assumption:

Assumption 1. Let $\ell : \mathbb{R} \to \mathbb{R}_+$ be a loss function such that $\ell(0) = 0$, $\ell(u) = \ell(-u)$ for $u \in \mathbb{R}$, ℓ is nondecreasing and convex on $\mathbb{R}_+$ and, for some constants $c_1, c_2 > 0$, $\ell(u) \le c_1 e^{c_2 u}$ for $u \ge 0$.

Corollary 1. *Suppose $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$. Let $a, \sigma_0 > 0$. Suppose that $r_n > 1$ and $r_n = o(n^{1-1/s})$ as $n \to \infty$. Then*

$$\sup_{\Sigma \in \mathcal{G}(r_n;a),\, \|B\|_1 \le 1,\, \sigma_f(\Sigma;B) \ge \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{n^{1/2}(\langle f(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma;B)} \le x \right\} - \Phi(x) \right| \to 0 \tag{1.8}$$

as $n \to \infty$. Moreover, under the same assumptions on f and r_n, and for any loss function ℓ satisfying Assumption 1,

$$\sup_{\Sigma \in \mathcal{G}(r_n;a),\, \|B\|_1 \le 1,\, \sigma_f(\Sigma;B) \ge \sigma_0} \left| \mathbb{E}_\Sigma \ell\left(\frac{n^{1/2}(\langle f(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma;B)} \right) - \mathbb{E}\ell(Z) \right| \to 0 \tag{1.9}$$

as $n \to \infty$, where Z is a standard normal random variable.

The main difficulty in asymptotically efficient estimation of $\langle f(\Sigma), B\rangle$ is related to the development of bias reduction methods. We will now discuss an approach to this problem in the case when $\mathbb{H}$ is finite-dimensional of dimension $d = d_n = o(n)$ and the covariance operator Σ is of isotropic type (the spectrum of Σ is bounded from above and bounded away from zero by constants that do not depend on n). In this case, the effective rank $\mathbf{r}(\Sigma)$ is of the same order as the dimension d, so d will be used as a complexity parameter. Developing a similar approach in a more general setting (when the effective rank $\mathbf{r}(\Sigma)$ is a relevant complexity parameter) remains an open problem.

Consider the integral operator

$$\mathcal{T}g(\Sigma) := \mathbb{E}_\Sigma g(\hat{\Sigma}) = \int_{\mathcal{C}_+(\mathbb{H})} g(S)P(\Sigma; dS), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}),$$

where $\mathcal{C}_+(\mathbb{H})$ is the cone of positive semidefinite self-adjoint operators in $\mathbb{H}$ (covariance operators) and $P(\Sigma; \cdot)$ is the distribution of the sample covariance $\hat{\Sigma}$ based on n i.i.d. observations sampled from $N(0; \Sigma)$ (which is a rescaled Wishart distribution). In what follows, $\mathcal{T}$ will be called the *Wishart operator*. We will view it as an operator acting on bounded measurable functions on $\mathcal{C}_+(\mathbb{H})$ taking values either in the real line, or in the space of self-adjoint operators. Such operators play an important role in the theory of Wishart matrices (see, e.g., James [James, James1, James2], Graczyk, Letac and Massam [GLM, GLM1], Letac and Massam [LetMas]). Their properties will be discussed in detail in Section 5. To find an unbiased estimator $g(\hat{\Sigma})$ of $f(\Sigma)$, one has to solve the integral equation $\mathcal{T}g(\Sigma) = f(\Sigma)$, $\Sigma \in \mathcal{C}_+(\mathbb{H})$ (the *Wishart equation*). Let $\mathcal{B} := \mathcal{T} - \mathcal{I}$, $\mathcal{I}$ being the identity operator. Then the solution of the Wishart equation can be formally written as the Neumann series

$$g(\Sigma) = (\mathcal{I}+\mathcal{B})^{-1} f(\Sigma) = (\mathcal{I} - \mathcal{B} + \mathcal{B}^2 - \cdots) f(\Sigma) = \sum_{j=0}^{\infty} (-1)^j \mathcal{B}^j f(\Sigma).$$

We do not use this representation in what follows and do not need any facts about the convergence of the series. Instead, we will define an approximate solution of the Wishart equation in terms of a partial sum of the Neumann series,

$$f_k(\Sigma) := \sum_{j=0}^{k} (-1)^j \mathcal{B}^j f(\Sigma), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}).$$

With this definition, we have

$$\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma) = (-1)^k \mathcal{B}^{k+1} f(\Sigma), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}).$$

It remains to show that $\langle \mathcal{B}^{k+1} f(\Sigma), B\rangle$ is small for smooth enough functions f, which would imply that the bias $\langle \mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma), B\rangle$ of the estimator $\langle f_k(\hat{\Sigma}), B\rangle$ of $\langle f(\Sigma), B\rangle$ is also small. [Very recently, a similar approach was considered by Jiao, Han and Weissman [JHW] in the case of estimation of a function $f(\theta)$ of the parameter θ of binomial model $B(n; \theta)$, $\theta \in [0, 1]$. In this case, $\mathcal{T} f$ is the Bernstein polynomial of degree n approximating f, and some results of classical approximation theory ([GonZ], [Tot]) were used in [JHW] to control $\mathcal{B}^k f$.]

Note that $P(\cdot;\cdot)$ is a Markov kernel and it could be viewed as the transition kernel of a Markov chain $\hat{\Sigma}^{(t)}, t = 0, 1, \dots,$ in the cone $\mathcal{C}_+(\mathbb{H})$, where $\hat{\Sigma}^{(0)} = \Sigma$, $\hat{\Sigma}^{(1)} = \hat{\Sigma}$, and, in general, for any $t \geq 1$, $\hat{\Sigma}^{(t)}$ is the sample covariance based on n i.i.d. observations sampled from the distribution $N(0; \hat{\Sigma}^{(t-1)})$ (conditionally on $\hat{\Sigma}^{(t-1)}$). In other words, the Markov chain $\{\hat{\Sigma}^{(t)}\}$ is based on iterative applications of bootstrap, and it will be called the *bootstrap chain*. As a consequence of (1.6), with a high probability (conditionally on $\hat{\Sigma}^{(t-1)}$), $\|\hat{\Sigma}^{(t)} - \hat{\Sigma}^{(t-1)}\| \lesssim \|\hat{\Sigma}^{(t-1)}\|\sqrt{d/n}$, so when $d = o(n)$, the Markov chain $\{\hat{\Sigma}^{(t)}\}$ moves in "small steps" of order $\asymp \sqrt{d/n}$. Clearly, with the above definitions,

$$\mathcal{T}^k f(\Sigma) = \mathbb{E}_\Sigma f(\hat{\Sigma}^{(k)}).$$

Note that, by Newton's binomial formula,

$$\mathcal{B}^k f(\Sigma) = (\mathcal{T} - \mathcal{I})^k f(\Sigma) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mathcal{T}^j f(\Sigma) = \mathbb{E}_\Sigma \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\Sigma}^{(j)}).$$

The expression $\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\Sigma}^{(j)})$ can be viewed as the k-th order difference of f along the Markov chain $\{\hat{\Sigma}^{(t)}\}$. It is well known that, for a k times continuously differentiable function f on the real line, the k-th order difference $\Delta_h^k f(x)$ (where $\Delta_h f(x) := f(x+h) - f(x)$) is of order $O(h^k)$ for a small increment h. Thus, at least heuristically, one can expect that $\mathcal{B}^k f(\Sigma)$ would be of order $O((d/n)^{k/2})$ (since $\sqrt{d/n}$ is the size of the "steps" of the Markov chain $\{\hat{\Sigma}^{(t)}\}$). This means that, for d much smaller than n, one can achieve a significant bias reduction in a relatively small number of steps k. The justification of this heuristic is rather involved. It is based on a representation of the operator function $f(\Sigma)$ in the form $\mathcal{D}g(\Sigma) := \Sigma^{1/2} Dg(\Sigma)\Sigma^{1/2}$, where g is a real valued function on $\mathcal{C}_+(\mathbb{H})$ invariant with respect to the orthogonal group. The properties of orthogonally invariant functions are then used to derive an integral representation for the function $\mathcal{B}^k f(\Sigma) = \mathcal{B}^k \mathcal{D}g(\Sigma) = \mathcal{D}\mathcal{B}^k g(\Sigma)$, which implies, for a sufficiently smooth f, bounds on $\mathcal{B}^k f(\Sigma)$ of order $O((d/n)^{k/2})$ and, as a consequence, bounds on the bias of the estimator $\langle f_k(\hat{\Sigma}), B\rangle$ of $\langle f(\Sigma), B\rangle$ of order $o(n^{-1/2})$, provided that $d = o(n)$ and k is sufficiently large (see (5.15), and Theorem 8 and Corollary 5 in Section 6).

The next step in the analysis of the estimator $\langle f_k(\hat{\Sigma}), B\rangle$ is to derive normal approximation bounds for $\langle f_k(\hat{\Sigma}), B\rangle - \mathbb{E}_\Sigma\langle f_k(\hat{\Sigma}), B\rangle$. To this end, in Section 7 we study smoothness properties of the functions $\mathcal{D}\mathcal{B}^k g(\Sigma)$ for a smooth orthogonally invariant function g that are later used to prove proper smoothness of $\langle f_k(\Sigma), B\rangle$ and derive concentration bounds on the remainder $\langle S_{f_k}(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$ of the first order Taylor expansion of $\langle f_k(\hat{\Sigma}), B\rangle$, which is the main step in showing that the centered remainder is asymptotically negligible and proving the normal approximation. In addition, we show that the limit variance in the normal approximation of $\langle f_k(\hat{\Sigma}), B\rangle - \mathbb{E}_\Sigma\langle f_k(\hat{\Sigma}), B\rangle$ coincides with $\sigma_f^2(\Sigma; B)$ (which is exactly the same as the limit variance in the normal approximation of $\langle f(\hat{\Sigma}), B\rangle - \mathbb{E}_\Sigma\langle f(\hat{\Sigma}), B\rangle$). This finally yields normal approximation bounds of Theorems 10 and 11 in Section 8.

Given $d > 1$ and $a \geq 1$, denote by $\mathcal{S}(d; a)$ the set of all covariance operators in a d-dimensional space $\mathbb{H}$ such that $\|\Sigma\|, \|\Sigma^{-1}\| \leq a$. The following result on uniform normal approximation of the estimator $\langle f_k(\hat{\Sigma}), B\rangle$ of $\langle f(\Sigma), B\rangle$ is an immediate consequence of Theorem 11.

Theorem 3. *Let $a \geq 1$ and $\sigma_0 > 0$. Suppose that, for some $\alpha \in (0,1)$, $1 < d_n \leq n^{\alpha}$ for $n \geq 1$. Suppose also that $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s > \frac{1}{1-\alpha}$. Let k be an integer such that $\frac{1}{1-\alpha} < k+1+\beta \leq s$ for some $\beta \in (0,1]$. Then*

$$\sup_{\Sigma\in\mathcal{S}(d_n;a),\, \|B\|_1\leq 1,\, \sigma_f(\Sigma;B)\geq\sigma_0}\ \sup_{x\in\mathbb{R}} \left|\mathbb{P}_{\Sigma}\left\{\frac{n^{1/2}(\langle f_k(\hat{\Sigma}),B\rangle - \langle f(\Sigma),B\rangle)}{\sigma_f(\Sigma;B)} \leq x\right\} - \Phi(x)\right| \to 0 \tag{1.10}$$

as $n \to \infty$. Moreover, if ℓ is a loss function satisfying Assumption 1, then

$$\sup_{\Sigma\in\mathcal{S}(d_n;a),\, \|B\|_1\leq 1,\, \sigma_f(\Sigma;B)\geq\sigma_0} \left|\mathbb{E}_{\Sigma}\ell\left(\frac{n^{1/2}(\langle f_k(\hat{\Sigma}),B\rangle - \langle f(\Sigma),B\rangle)}{\sigma_f(\Sigma;B)}\right) - \mathbb{E}\ell(Z)\right| \to 0 \tag{1.11}$$

as $n \to \infty$.

Remark 1. Note that for $\alpha \in (0,1/2)$ and $s > \frac{1}{1-\alpha}$, one can choose $k=0$, implying that $f_k(\hat{\Sigma}) = f(\hat{\Sigma})$ in Theorem 3 is a usual plug-in estimator (cf. Corollary 1). However, for $\alpha = 1/2$, we have to assume that $s > 2$ and choose $k=1$ to satisfy the condition $k+1+\beta > \frac{1}{1-\alpha} = 2$. Thus, in this case, the bias correction is already nontrivial. For larger values of α, even more smoothness of f is required and more iterations k in our bias reduction method are needed.

Remark 2. It easily follows from well known embedding theorems for Besov spaces (see, e.g., [Tr, Section 2.3.2]) that, for $s' > s > 0$, the Hölder space $C^{s'}(\mathbb{R})$ is contained in $B^s_{\infty,1}(\mathbb{R})$. Moreover, it is easy to see that any $C^{s'}$ function defined locally in a neighborhood of the spectrum of Σ could be extended to a function from $C^{s'}(\mathbb{R})$. These observations show that Theorem 3 could be applied to all C^s functions defined in a neighborhood of the spectrum of Σ for all $s > \frac{1}{1-\alpha}$.

To show the asymptotic efficiency of $\langle f_k(\hat{\Sigma}), B\rangle$, it remains to prove a minimax lower bound on the risk of an arbitrary estimator $T_n(X_1,\dots,X_n)$ of $\langle f(\Sigma), B\rangle$ that would imply the optimality of the variance $\sigma_f^2(\Sigma;B)$ in normal approximation (1.10), (1.11). Let $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$. Given $a > 1$, let $\mathring{\mathcal{S}}(d;a)$ be the set of all covariance operators in a Hilbert space $\mathbb{H}$ of dimension d such that $\|\Sigma\|, \|\Sigma^{-1}\| < a$. Given $\sigma_0 > 0$, denote

$$\mathring{\mathcal{S}}_{f,B}(d;a;\sigma_0) := \mathring{\mathcal{S}}(d;a) \cap \{\Sigma : \sigma_f(\Sigma;B) > \sigma_0\}.$$

Note that the set $\mathring{\mathcal{S}}_{f,B}(d;a;\sigma_0)$ is open in the operator norm topology, which easily follows from the continuity of the functions $\Sigma \mapsto \|\Sigma\|$, $\Sigma \mapsto \|\Sigma^{-1}\|$ (on the set of nonsingular operators) and $\Sigma \mapsto \sigma_f^2(\Sigma;B)$ (see Lemma 26 in Section 9) with respect to the operator norm. This set could be empty. For instance, since $\sigma_f^2(\Sigma; B) \leq 2\|f'\|_{L_\infty}^2 \|\Sigma\|^2 \|B\|_2^2$, we have $\mathring{\mathcal{S}}_{f,B}(d; a; \sigma_0) = \emptyset$ if $\sigma_0^2 > 2\|f'\|_{L_\infty}^2 \|\Sigma\|^2 \|B\|_2^2$. Denote

$$\mathfrak{B}_f(d; a; \sigma_0) := \{B : \|B\|_1 \leq 1, \ \mathring{\mathcal{S}}_{f,B}(d; a; \sigma_0) \neq \emptyset\}.$$

The following theorem provides an asymptotic minimax lower bound on the mean squared error of estimation of the functionals $\langle f(\Sigma)B\rangle$ for $\|B\|_1 \leq 1$. By convention, it will be assumed that $\inf \emptyset = +\infty$.

Theorem 4. *Let $a > 1$, $\sigma_0^2 > 0$ and let $\{d_n\}$ be an arbitrary sequence of integers $d_n \geq 2$. Then, for all $a' \in (1, a)$ and $\sigma_0' > \sigma_0$,*

$$\liminf_{n\to\infty} \inf_{T_n} \inf_{B\in\mathfrak{B}_f(d_n;a';\sigma_0')} \sup_{\Sigma\in\mathring{\mathcal{S}}(d_n;a),\, \sigma_f(\Sigma;B)>\sigma_0} \frac{n\mathbb{E}_\Sigma(T_n - \langle f(\Sigma), B\rangle)^2}{\sigma_f^2(\Sigma; B)} \geq 1, \tag{1.12}$$

where the first infimum is taken over all statistics $T_n = T_n(X_1, \dots, X_n)$ based on i.i.d. observations $X_1, \dots, X_n$ sampled from $N(0; \Sigma)$.

The proof is given in Section 9.

Remark 3. If $C \subset \sigma(\Sigma)$ is a "component" of the spectrum of Σ such that the distance $\mathrm{dist}(C, \sigma(\Sigma) \setminus C)$ from C to the rest of the spectrum is bounded away from zero by a sufficiently large gap and P_C is the orthogonal projection on the direct sum of the eigenspaces of Σ corresponding to the eigenvalues from C, then it is easy to represent P_C as $f(\Sigma)$ for a smooth function f that is equal to 1 on C and vanishes outside of a neighborhood of C that does not contain other eigenvalues. The problem of efficient estimation of linear functionals of P_C (such as its matrix entries in a given basis or general bilinear forms) is of importance in principal component analysis. A related problem of estimation of linear functionals of principal components was recently studied in [KLN] in the case of one-dimensional spectral projections. The methods of efficient estimation developed in [KLN] are rather specialized and they could not be easily extended even to spectral projections of rank higher than 1. This, in part, was our motivation to study the problem for more general smooth functionals and to develop a more general approach. Similarly, one can represent the operator $P_C \Sigma P_C$ as a smooth function of Σ and use the approach of the current paper to develop efficient estimators of bilinear forms or matrix entries of such operators. This could be of interest in the case of covariance matrices of the form $\Sigma = \Sigma_0 + \sigma^2 I_d$, where Σ_0 is a low rank covariance matrix (say, the covariance matrix whose eigenvectors are "spikes" of a spiked covariance model). If C is the set of top eigenvalues of Σ that correspond to its "spikes", then estimation of Σ_0 could be reduced to estimation of $P_C \Sigma P_C$.

Remark 4. The results of this paper could not be directly applied to estimation of functionals of the form $\mathrm{tr}(f(\Sigma))$ since in this case B is the identity operator and its nuclear norm is not bounded by a constant. In such cases, $\sqrt{n}$-consistent estimators do not always exist in high-dimensional problems, minimax optimal convergence rates are lower than $n^{-1/2}$ and they do depend on the dimension (see, for instance, [CLZ] for an example of estimation of the log-determinant $\log\det(\Sigma) = \mathrm{tr}(\log(\Sigma))$). Although some elements of our approach (in particular, the bias reduction method) could be useful in this case, a comprehensive theory of estimation of the functionals $\langle f(\Sigma), B\rangle$ in the case of B with unbounded nuclear norm remains an open problem and it is beyond the scope of this paper.

Remark 5. In this paper, the problem was studied only in the case of Gaussian models with known mean (without loss of generality, set to be zero) and unknown covariance operators. In [KZh], a similar problem of efficient estimation of smooth functionals of unknown mean in Gaussian shift models with known covariance was studied. The problem becomes more complicated when both mean and covariance are unknown (in particular, it would require a more difficult analysis of the operators $\mathcal{T}$ and $\mathcal{B}$ involved in the bias reduction method).

Remark 6. The computation of estimators $f_k(\hat{\Sigma})$ could be based on Monte Carlo simulation of the bootstrap chain. To this end, one has to simulate a segment of this chain of length $k+1$ starting at the sample covariance $\hat{\Sigma}$. This would allow us to compute the sum $\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j} f(\hat{\Sigma}^{(j+1)})$. Averaging such sums over a sufficiently large number N of independent copies of the bootstrap chain provides a Monte Carlo approximation of $\mathcal{B}^k f(\hat{\Sigma})$, which allows us to approximate $f_k(\hat{\Sigma})$. A total of $(k+1)N$ computations of the function f of covariance operators (each of them based on a singular value decomposition) would be required to implement this procedure.

1.3. Related results

To the best of our knowledge, the problem of efficient estimation for general classes of smooth functionals of covariance operators in the setting of the current paper has not been studied before. However, many results in the literature on nonparametric, semiparametric and high-dimensional statistics as well as some results in random matrix theory are relevant in our context. Below we provide a brief discussion of some of these results.

Asymptotically efficient estimation of smooth functionals of infinite-dimensional parameters has been an important topic in nonparametric statistics for a number of years; it also has deep connections to efficiency in semiparametric estimation (see, e.g., [BKRW], [GN] and references therein). The early references include Levit [Lev1, Lev2] and the book of Ibragimov and Khasminskii [IKh]. In the paper by Ibragimov, Nemirovski and Khasminskii [IKhN] and later in the paper [Nem1] and in the Saint-Flour lectures [Nem2] by Nemirovski, sharp results on efficient estimation of general smooth functionals of parameters of Gaussian white noise models were obtained, precisely describing the dependence between the rate of decay of Kolmogorov's diameters of parameter space (used as a measure of its complexity) and the degree of smoothness of functionals for which efficient estimation is possible. A general approach to construction of efficient estimators of smooth functionals in Gaussian white noise models was also developed in those papers. The result of Theorem 3 is in the same spirit, with the growth rate α of the dimension of the space being the complexity parameter instead of the rate of decay of Kolmogorov's diameters. At this point, we do not know whether the smoothness threshold $s > \frac{1}{1-\alpha}$ for efficient estimation obtained in this theorem is sharp (although the sharpness of the same smoothness threshold was proved in [KZh] in the case of the Gaussian shift model).

More recently, there has been a lot of interest in semiparametric efficiency properties of regularization-based estimators (such as LASSO) in various models of high-dimensional statistics (see, e.g., [GBRD], [JaMont], [ZZ], [JG]) as well as in minimax optimal rates of estimation of the special functionals (in particular, linear and quadratic) in such models [CL1], [CL2], [CCT].

In a series of pioneering papers in the 80s–90s, Girko obtained a number of results on asymptotically normal estimation of many special functionals of covariance matrices in high-dimensional setting, in particular, on estimation of the Stieltjes transform of the spectral function $\mathrm{tr}((I + t\Sigma)^{-1})$ (see [Gir] and also [Gir1] and references therein). His estimators were typically functions of a sample covariance $\hat{\Sigma}$ defined in terms of certain equations (so called G-estimators) and the proofs of their asymptotic normality were largely based on martingale CLT. The centering and normalizing parameters in the limit theorems in those papers are often hard to interpret and the estimators were not proved to be asymptotically efficient.

Asymptotic normality of so called linear spectral statistics $\mathrm{tr}(f(\hat{\Sigma}))$ centered either by their own expectations, or by the integral of f with respect to a Marchenko–Pastur type law has been an active subject of research in random matrix theory both in the case of high-dimensional sample covariance (or Wishart matrices) and in other random matrix models such as Wigner matrices (see, e.g., Bai and Silverstein [BaiS], Lytova and Pastur [LP], Sosoe and Wong [SW]). Although these results do not have direct statistical implications since $\mathrm{tr}(f(\hat{\Sigma}))$ does not "concentrate" around the corresponding population parameter, probabilistic and analytic techniques developed in those papers are highly relevant.

There are many results in the literature on special cases of the above problem, such as asymptotic normality of the statistic $\log\det(\hat{\Sigma}) = \mathrm{tr}(\log(\hat{\Sigma}))$ (the log-determinant). If $d = d_n \leq n$, then it was shown that the sequence

$$\frac{\log\det(\hat{\Sigma}) - a_{n,d} - \log\det(\Sigma)}{b_{n,d}}$$

converges in distribution to a standard normal random variable for explicitly given sequences $a_{n,d}, b_{n,d}$ that depend only on the sample size n and on the dimension d. This means that $\log\det(\hat{\Sigma})$ is an asymptotically normal estimator of $\log\det(\Sigma) = \mathrm{tr}(\log(\Sigma))$ subject to a simple bias correction (see, e.g., Girko [Gir] and more recent paper by Cai, Liang and Zhou [CLZ]). The convergence rate of this estimator is typically lower than $n^{-1/2}$: for instance, if $d = n^{\alpha}$ for $\alpha \in (0,1)$, then the convergence rate is $\asymp n^{-(1-\alpha)/2}$ (and, for $\alpha = 1$, the estimator is not consistent). In this case, the problem is relatively simple since $\log\det(\hat{\Sigma}) - \log\det(\Sigma) = \log\det(W)$, where W is the sample covariance based on a sample of n i.i.d. standard normal random vectors.

In a recent paper by Koltchinskii and Lounici [KL1] (see also [KL3, KL4]), the problem of estimation of bilinear forms of spectral projections of covariance operators was studied in the setting when $\mathbf{r}(\Sigma) = o(n)$ as $n \to \infty$.[3] Normal approximation and concentration results for bilinear forms centered by their expectations were proved using first order perturbation expansions for empirical spectral projections and concentration inequalities for their remainder terms (which is similar to the approach of the current paper). Special properties of the bias of these estimators were studied that, in the case of one-dimensional spectral projections, led to the development of a bias reduction method based on sample splitting that resulted in a construction of $\sqrt{n}$-consistent and asymptotically normal estimators of linear forms of eigenvectors of the true covariance (principal components) in the case when $\mathbf{r}(\Sigma) = o(n)$ as $n \to \infty$. This approach has been further developed in a very recent paper by Koltchinskii, Loeffler and Nickl [KLN] in which asymptotically efficient estimators of linear forms of eigenvectors of Σ were studied.

Other recent references on estimation of functionals of covariance include Fan, Rigollet and Wang [FRW] (optimal rates of estimation of special functionals of covariance under sparsity assumptions), Gao and Zhou [GaoZ] (Bernstein–von Mises theorems for functionals of covariance), Kong and Valiant [KoVa] (estimation of "spectral moments" $\mathrm{tr}(\Sigma^k)$).

2. Analysis and operator theory preliminaries

In this section, we discuss several results in operator theory concerning perturbations of smooth functions of self-adjoint operators in Hilbert spaces. They are simple modifications of known results due to several authors (see recent survey by Aleksandrov and Peller [AP2]).

2.1. Entire functions of exponential type and Besov spaces

Let $f : \mathbb{C} \to \mathbb{C}$ be an entire function and let $\sigma > 0$. We say that f is *of exponential type* σ (more precisely, $\le \sigma$) if for any $\varepsilon > 0$ there exists $C = C(\varepsilon, \sigma, f) > 0$ such that

$$|f(z)| \le Ce^{(\sigma+\varepsilon)|z|}, \qquad z \in \mathbb{C}.$$

In what follows, $\mathcal{E}_\sigma = \mathcal{E}_\sigma(\mathbb{C})$ denotes the space of all entire functions of exponential type σ. It is straightforward to see (and well known) that $f \in \mathcal{E}_\sigma$ if and only if

$$\limsup_{R\to\infty} \frac{\log \sup_{\varphi\in[0,2\pi]} |f(Re^{i\varphi})|}{R} =: \sigma(f) \le \sigma.$$

With a little abuse of notation, the restriction $f_{\restriction\mathbb{R}}$ of f to $\mathbb{R}$ will also be denoted by f; $\mathcal{F}f$ will denote the Fourier transform of f: $\mathcal{F}f(t) = \int_{\mathbb{R}} e^{-itx} f(x)\,dx$ (if f is not square integrable, its Fourier transform is understood in the sense of tempered distributions). According to the *Paley–Wiener theorem*,

$$\mathcal{E}_\sigma \cap L_\infty(\mathbb{R}) = \{f \in L_\infty(\mathbb{R}) : \mathrm{supp}(\mathcal{F}f) \subset [-\sigma, \sigma]\}.$$

[3] For other recent results on covariance estimation under assumptions on its effective rank see [NSU, RW].

It is also well known that $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$ if and only if $|f(z)| \le \|f\|_{L_\infty(\mathbb{R})} e^{\sigma|\mathrm{Im}(z)|}$ for $z \in \mathbb{C}$.

We will use the *Bernstein inequality* $\|f'\|_{L_\infty(\mathbb{R})} \le \sigma\|f\|_{L_\infty(\mathbb{R})}$ that holds for all $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$. Moreover, since $f \in \mathcal{E}_\sigma$ implies $f' \in \mathcal{E}_\sigma$, we also have $\|f''\|_{L_\infty(\mathbb{R})} \le \sigma^2\|f\|_{L_\infty(\mathbb{R})}$, and similar bounds hold for all the derivatives of f.

The next elementary lemma is a corollary of the Bernstein inequality. It provides bounds on the remainder of the first order Taylor expansion of $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$.

Lemma 1. *Let $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$. Denote*

$$S_f(x; h) := f(x+h) - f(x) - f'(x)h, \quad x, h \in \mathbb{R}.$$

Then

$$|S_f(x;h)| \le \frac{\sigma^2}{2}\|f\|_{L_\infty(\mathbb{R})} h^2, \quad x, h \in \mathbb{R},$$

and

$$|S_f(x;h') - S_f(x;h)| \le \sigma^2\|f\|_{L_\infty(\mathbb{R})}\delta(h,h')|h'-h|, \quad x, h, h' \in \mathbb{R}.$$

where $\delta(h,h') := (|h| \wedge |h'|) + |h'-h|/2$.

We also need an extension of the Bernstein inequality to functions of several complex variables. Let $f : \mathbb{C}^k \to \mathbb{C}$ be an entire function and let $\sigma := (\sigma_1, \dots, \sigma_k)$, $\sigma_j > 0$. The function f is *of exponential type* $\sigma = (\sigma_1, \dots, \sigma_k)$ if for any $\varepsilon > 0$ there exists $C = C(\varepsilon, \sigma, f) > 0$ such that

$$|f(z_1, \dots, z_k)| \le C e^{\sum_{j=1}^k (\sigma_j + \varepsilon)|z_j|}, \quad z_1, \dots, z_k \in \mathbb{C}.$$

Let $\mathcal{E}_{\sigma_1,\dots,\sigma_k}$ be the set of all such functions. The following extension of the Bernstein inequality can be found in the paper by Nikol'skii [Nik], who actually proved it for an arbitrary L_p-norm, $1 \le p \le \infty$. If $f \in \mathcal{E}_{\sigma_1,\dots,\sigma_k} \cap L_\infty(\mathbb{R})$, then for any $m \ge 0$ and any $m_1, \dots, m_k \ge 0$ such that $\sum_{j=1}^k m_j = m$,

$$\left\| \frac{\partial^m f}{\partial x_1^{m_1} \dots \partial x_k^{m_k}} \right\|_{L_\infty(\mathbb{R}^k)} \le \sigma_1^{m_1} \dots \sigma_k^{m_k} \|f\|_{L_\infty(\mathbb{R}^k)}. \tag{2.1}$$

Let $w \ge 0$ be a C^∞ function on the real line with $\mathrm{supp}(w) \subset [-2,2]$ such that $w(t) = 1$ for $t \in [-1,1]$ and $w(-t) = w(t)$ for $t \in \mathbb{R}$. Define $w_0(t) := w(t/2) - w(t)$ for $t \in \mathbb{R}$, which implies that $\mathrm{supp}(w_0) \subset \{t : 1 \le |t| \le 4\}$. Let $w_j(t) := w_0(2^{-j}t)$ for $t \in \mathbb{R}$ with $\mathrm{supp}(w_j) \subset \{t : 2^j \le |t| \le 2^{j+2}\}$, $j = 0, 1, \dots$. These definitions immediately imply that

$$w(t) + \sum_{j \ge 0} w_j(t) = 1, \quad t \in \mathbb{R}.$$

Finally, define functions $W, W_j \in \mathcal{S}(\mathbb{R})$ (the Schwartz space of functions in $\mathbb{R}$) by their Fourier transforms as follows:

$$w(t) = (\mathcal{F}W)(t), \quad w_j(t) = (\mathcal{F}W_j)(t), \quad t \in \mathbb{R},\ j \ge 0.$$

For a tempered distribution $f \in \mathcal{S}'(\mathbb{R})$, one can define its Littlewood–Paley dyadic decomposition as the family of functions $f_0 := f * W$, $f_n := f * W_{n-1}$, $n \geq 1$ with compactly supported Fourier transforms. Note that, by the Paley–Wiener theorem, $f_n \in \mathcal{E}_{2^{n+1}} \cap L_\infty(\mathbb{R})$. It is well known that $\sum_{n\geq 0} f_n = f$ with convergence in $\mathcal{S}'(\mathbb{R})$.

We use the Besov norms

$$\|f\|_{B^s_{\infty,1}} := \sum_{n\geq 0} 2^{ns} \|f_n\|_{L_\infty(\mathbb{R})}, \qquad s \in \mathbb{R},$$

and the corresponding Besov spaces

$$B^s_{\infty,1}(\mathbb{R}) := \{f \in \mathcal{S}'(\mathbb{R}) : \|f\|_{B^s_{\infty,1}} < \infty\}.$$

We do not use the whole scale of Besov spaces. Note that Besov norms are equivalent for different choices of the function w and the corresponding Besov spaces coincide. If $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \geq 0$, then the series $\sum_{n\geq 0} f_n$ converges uniformly to f in $\mathbb{R}$, which easily implies that $f \in C_u(\mathbb{R})$, where $C_u(\mathbb{R})$ is the space of all bounded uniformly continuous functions in $\mathbb{R}$ and $\|f\|_{L_\infty} \leq \|f\|_{B^s_{\infty,1}}$. Thus, for $s \geq 0$, the space $B^s_{\infty,1}(\mathbb{R})$ is continuously embedded in $C_u(\mathbb{R})$. Moreover, if $C^s(\mathbb{R})$ denotes the Hölder space of smoothness $s > 0$, then, for all $s' > s > 0$, $C^{s'}(\mathbb{R}) \subset B^s_{\infty,1}(\mathbb{R}) \subset C^s(\mathbb{R})$ (see [Tr, Sections 2.3.2, 2.5.7]). Further details on Besov spaces can also be found in [Tr].

2.2. *Taylor expansions for operator functions*

For a continuous (and even for a Borel measurable) function f in $\mathbb{R}$ and $A \in \mathcal{B}_{\text{sa}}(\mathbb{H})$, the operator $f(A)$ is well defined and self-adjoint (for instance, by the spectral theorem). By standard holomorphic functional calculus, the operator $f(A)$ is well defined for $A \in \mathcal{B}(\mathbb{H})$ and for any function $f : \mathbb{C} \supset G \to \mathbb{C}$ holomorphic in a neighborhood G of the spectrum $\sigma(A)$ of A. It is given by the Cauchy formula

$$f(A) := -\frac{1}{2\pi i} \oint_\gamma f(z) R_A(z)\, dz,$$

where $R_A(z) := (A - zI)^{-1}$ for $z \notin \sigma(A)$ is the resolvent of A and $\gamma \subset G$ is a contour surrounding $\sigma(A)$ with counterclockwise orientation. In particular, this holds for all entire functions f and the mapping $\mathcal{B}(\mathbb{H}) \ni A \mapsto f(A) \in \mathcal{B}(\mathbb{H})$ is Fréchet differentiable with derivative

$$Df(A; H) = \frac{1}{2\pi i} \oint_\gamma f(z) R_A(z) H R_A(z)\, dz, \qquad H \in \mathcal{B}(\mathbb{H}). \tag{2.2}$$

The last formula easily follows from the perturbation series for the resolvent

$$R_{A+H}(z) = \sum_{k=0}^{\infty} (-1)^k (R_A(z)H)^k R_A(z), \qquad z \in \mathbb{C} \setminus \sigma(A),$$

which converges in the operator norm as long as $\|H\| < \frac{1}{\|R_A(z)\|} = \frac{1}{\text{dist}(z,\sigma(A))}$.

We need to extend the bounds of Lemma 1 to functions of operators, establishing similar properties for the remainder of the first order Taylor expansion

$$S_f(A;H) := f(A+H) - f(A) - Df(A;H), \qquad A, H \in \mathcal{B}_{sa}(\mathbb{H}),$$

where f is an entire function of exponential type σ. This is related to a circle of problems studied in operator theory literature concerning operator Lipschitz and operator differentiable functions (see, in particular, the survey by Aleksandrov and Peller [AP2]).

We will need the following lemma.

Lemma 2. *Let $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$. Then, for all $A, H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\|f(A+H) - f(A)\| \le \sigma \|f\|_{L_\infty(\mathbb{R})} \|H\|, \tag{2.3}$$

$$\|Df(A;H)\| \le \sigma \|f\|_{L_\infty(\mathbb{R})} \|H\|, \tag{2.4}$$

$$\|S_f(A;H)\| \le \frac{\sigma^2}{2} \|f\|_{L_\infty(\mathbb{R})} \|H\|^2, \tag{2.5}$$

$$\|S_f(A;H') - S_f(A;H)\| \le \sigma^2 \|f\|_{L_\infty(\mathbb{R})} \delta(H,H') \|H'-H\|, \tag{2.6}$$

where $\delta(H,H') := (\|H\| \wedge \|H'\|) + \|H'-H\|/2$.

The bounds (2.3) and (2.4) are well known [AP2] (in fact, (2.3) means that, for $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$, $\mathcal{B}_{sa}(\mathbb{H}) \ni A \mapsto f(A) \in \mathcal{B}_{sa}(\mathbb{H})$ is operator Lipschitz with respect to the operator norm). The proof of (2.5) and (2.6) is also based on a nice approach by Aleksandrov and Peller [AP1, AP2] developed to prove the operator Lipschitz property. We give the proof of the lemma for completeness.

Proof of Lemma 2. Let E be a complex Banach space and let $\mathcal{E}_\sigma(E)$ be the space of entire functions $F : \mathbb{C} \to E$ of exponential type σ, that is, entire functions F such that for any $\varepsilon > 0$ there exists a constant $C = C(\varepsilon, \sigma, F) > 0$ for which $\|F(z)\| \le Ce^{(\sigma+\varepsilon)|z|}$, $z \in \mathbb{C}$. If $F \in \mathcal{E}_\sigma(E)$ and $\sup_{x\in\mathbb{R}} \|F(x)\| < \infty$, then the Bernstein inequality holds for F:

$$\sup_{x\in\mathbb{R}} \|F'(x)\| \le \sigma \sup_{x\in\mathbb{R}} \|F(x)\|. \tag{2.7}$$

Indeed, for any $l \in E^*$, $l(F(\cdot)) \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$, which implies that

$$\sup_{x\in\mathbb{R}} \|F'(x)\| = \sup_{\|l\|\le 1} \sup_{x\in\mathbb{R}} |l(F'(x))| \le \sigma \sup_{\|l\|\le 1} \sup_{x\in\mathbb{R}} |l(F(x))| = \sigma \sup_{x\in\mathbb{R}} \|F(x)\|$$

and

$$\|F(x+h) - F(x)\| \le \sigma \sup_{x\in\mathbb{R}} \|F(x)\| \, |h|. \tag{2.8}$$

A similar simple argument (now based on Lemma 1) shows that for $S_F(x;h) := F(x+h) - F(x) - F'(x)h$, we have

$$\|S_F(x;h)\| \le \frac{\sigma^2}{2} \sup_{x\in\mathbb{R}} \|F(x)\| h^2, \qquad x, h \in \mathbb{R}, \tag{2.9}$$

and

$$\|S_F(x;h') - S_F(x;h)\| \le \sigma^2 \sup_{x\in\mathbb{R}} \|F(x)\| \left[|h|\,|h'-h| + \frac{|h'-h|^2}{2} \right], \qquad x, h, h' \in \mathbb{R}. \tag{2.10}$$

Next, for $A, H \in \mathcal{B}_{sa}(\mathbb{H})$ and $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$, define $F(z) := f(A+zH)$ for $z \in \mathbb{C}$. Then $F \in \mathcal{E}_{\sigma\|H\|}(\mathcal{B}(\mathbb{H}))$. Indeed, F is complex-differentiable at any point $z \in \mathbb{C}$ with derivative $F'(z) = Df(A+zH; H)$, so it is an entire function with values in $E = \mathcal{B}(\mathbb{H})$. In addition, by the von Neumann theorem (see, e.g., [Dav, Theorem 9.5.3]),

$$\|F(z)\| = \|f(A+zH)\| \le \sup_{|\zeta|\le\|A\|+|z|\,\|H\|} |f(\zeta)| \le \|f\|_{L_\infty(\mathbb{R})} e^{\sigma\|A\|} e^{\sigma\|H\||z|}, \qquad z \in \mathbb{C},$$

implying that F is of exponential type $\sigma\|H\|$. Note also that

$$\sup_{x\in\mathbb{R}} \|F(x)\| = \sup_{x\in\mathbb{R}} \|f(A+xH)\| \le \sup_{x\in\mathbb{R}} |f(x)| = \|f\|_{L_\infty(\mathbb{R})}.$$

Hence, (2.7) and (2.8) imply that

$$\begin{aligned}\|f(A+H) - f(A)\| = \|F(1) - F(0)\| &\le \sup_{x\in\mathbb{R}} \|F'(x)\| \\ &\le \sigma\|H\| \sup_{x\in\mathbb{R}} \|F(x)\| \le \sigma\|f\|_{L_\infty(\mathbb{R})}\|H\|\end{aligned}$$

and

$$\|Df(A; H)\| = \|F'(0)\| \le \sigma\|f\|_{L_\infty(\mathbb{R})}\|H\|,$$

which proves (2.3) and (2.4). Similarly, using (2.9), we get

$$\begin{aligned}\|S_f(A; H)\| &= \|f(A+H) - f(A) - Df(A; H)\| = \|F(1) - F(0) - F'(0)(1-0)\| \\ &= \|S_F(0, 1)\| \le \frac{\sigma^2\|H\|^2}{2} \sup_{x\in\mathbb{R}} \|F(x)\| \le \frac{\sigma^2}{2}\|f\|_{L_\infty(\mathbb{R})}\|H\|^2,\end{aligned}$$

proving (2.5).

To prove (2.6), define

$$F(z) := f(A+H+z(H'-H)) - f(A+z(H'-H)), \qquad z \in \mathbb{C}.$$

As in the previous case, F is an entire function with values in $\mathcal{B}(\mathbb{H})$. The bound

$$\|F(z)\| \le \|f\|_{L_\infty(\mathbb{R})}(e^{\sigma\|A+H\|} + e^{\sigma\|A\|})e^{\sigma\|H'-H\||z|}$$

implies that $F \in \mathcal{E}_{\sigma\|H'-H\|}(\mathcal{B}(\mathbb{H}))$. Clearly, also $\sup_{x\in\mathbb{R}} \|F(x)\| \le 2\|f\|_{L_\infty(\mathbb{R})}$.

Note that

$$\begin{aligned}&S_f(A; H') - S_f(A; H) \\ &\quad = Df(A+H; H'-H) - Df(A; H'-H) + S_f(A+H; H'-H)\end{aligned} \tag{2.11}$$

and (2.5) implies

$$\|S_f(A+H; H'-H)\| \le \frac{\sigma^2}{2}\|f\|_{L_\infty(\mathbb{R})}\|H'-H\|^2.$$

On the other hand, we have (by the Bernstein inequality)

$$\|Df(A+H;H'-H)-Df(A;H'-H)\| = \|F'(0)\| \le \sigma\|H'-H\| \sup_{x\in\mathbb{R}}\|F(x)\|$$

and (2.3) implies that

$$\sup_{x\in\mathbb{R}}\|F(x)\| = \sup_{x\in\mathbb{R}}\|f(A+H+x(H'-H))-f(A+x(H'-H))\| \le \sigma\|f\|_{L_\infty(\mathbb{R})}\|H\|.$$

Now, it follows from (2.11) that

$$\|S_f(A;H')-S_f(A;H)\| \le \sigma^2\|f\|_{L_\infty(\mathbb{R})}\left(\|H\|+\frac{\|H'-H\|}{2}\right)\|H'-H\|,$$

which implies (2.6). □

Remark 7. In addition to (2.6), the following bound follows from (2.3) and (2.4):

$$\|S_f(A;H')-S_f(A;H)\| \le 2\sigma\|f\|_{L_\infty(\mathbb{R})}\|H'-H\|. \tag{2.12}$$

Note also that $\delta(H,H') \le \|H\|+\|H'\|$ for $H,H'\in\mathcal{B}_{sa}(\mathbb{H})$.

Following Aleksandrov and Peller [AP2], we use Littlewood–Paley dyadic decomposition and the corresponding family of Besov norms to extend the bounds of Lemma 2 to functions in Besov classes. It would be more convenient for our purposes to use inhomogeneous Besov norms instead of homogeneous norms used in [AP2]. Peller [Pel] proved that any function $f\in B^1_{\infty,1}(\mathbb{R})$ [4] is operator Lipschitz and operator differentiable on the space of self-adjoint operators with respect to the operator norm (in [AP2], these facts were proved using Littlewood–Paley theory and extensions of the Bernstein inequality for operator functions; see also the earlier paper [AP1]). We will state Peller's results in the next lemma in a convenient form along with some additional bounds on the remainder of the first order Taylor expansion $S_f(A;H)=f(A+H)-f(A)-Df(A;H)$ for f in suitable Besov spaces.

Lemma 3. *If $f\in B^1_{\infty,1}(\mathbb{R})$, then for all $A,H\in\mathcal{B}_{sa}(\mathbb{H})$,*

$$\|f(A+H)-f(A)\| \le 2\|f\|_{B^1_{\infty,1}(\mathbb{R})}\|H\|. \tag{2.13}$$

Moreover, the function $\mathcal{B}_{sa}(\mathbb{H})\ni A\mapsto f(A)\in\mathcal{B}_{sa}(\mathbb{H})$ is Fréchet differentiable with respect to the operator norm with derivative given by the following series (that converges in the operator norm):

$$Df(A;H)=\sum_{n\ge 0}Df_n(A;H). \tag{2.14}$$

If $f\in B^s_{\infty,1}(\mathbb{R})$ for some $s\in[1,2]$, then, for all $A,H,H'\in\mathcal{B}_{sa}(\mathbb{H})$,

$$\|S_f(A;H)\| \le 2^{3-s}\|f\|_{B^s_{\infty,1}}\|H\|^s \tag{2.15}$$

and

$$\|S_f(A;H')-S_f(A;H)\| \le 4\|f\|_{B^s_{\infty,1}}(\delta(H,H'))^{s-1}\|H'-H\|. \tag{2.16}$$

[4] In fact, Peller used modified homogeneous Besov classes instead of inhomogeneous Besov spaces we use in this paper.

Proof. Recall that, for $f \in B^1_{\infty,1}(\mathbb{R})$, the series $\sum_{n\geq 0} f_n$ converges to f uniformly in $\mathbb{R}$. Since A, $A+H$, $A+H'$ are bounded self-adjoint operators, we also get

$$\sum_{n\geq 0} f_n(A) = f(A), \quad \sum_{n\geq 0} f_n(A+H) = f(A+H), \quad \sum_{n\geq 0} f_n(A+H') = f(A+H'), \tag{2.17}$$

with all series converging in the operator norm.

To prove (2.13), observe that

$$\begin{aligned}\|f(A+H)-f(A)\| &= \Big\|\sum_{n\geq 0}[f_n(A+H)-f_n(A)]\Big\| \\ &\leq \sum_{n\geq 0}\|f_n(A+H)-f_n(A)\| \leq \sum_{n\geq 0} 2^{n+1}\|f_n\|_{L_\infty(\mathbb{R})}\|H\| = 2\|f\|_{B^1_{\infty,1}}\|H\|,\end{aligned}$$

where we use (2.3).

By (2.4),

$$\sum_{n\geq 0}\|Df_n(A;H)\| \leq \sum_{n\geq 0} 2^{n+1}\|f_n\|_{L_\infty(\mathbb{R})}\|H\| = 2\|f\|_{B^1_{\infty,1}}\|H\| < \infty,$$

implying the convergence of $\sum_{n\geq 0} Df_n(A;H)$ in the operator norm. We define

$$Df(A;H) := \sum_{n\geq 0} Df_n(A;H). \tag{2.18}$$

We will prove that this yields the Fréchet derivative of $f(A)$. To this end, note that (2.17) and (2.18) imply that

$$S_f(A;H) = \sum_{n\geq 0}[f_n(A+H)-f_n(A)-Df_n(A;H)] = \sum_{n\geq 0} S_{f_n}(A;H). \tag{2.19}$$

As a consequence,

$$\begin{aligned}\|S_f(A;H)\| &\leq \sum_{n\leq N}\|S_{f_n}(A;H)\| + \sum_{n>N}\|f_n(A+H)-f_n(A)\| + \sum_{n>N}\|Df_n(A;H)\| \\ &\leq \sum_{n\leq N} 2^{2(n+1)}\|f_n\|_{L_\infty(\mathbb{R})}\|H\|^2 + 2\sum_{n>N} 2^{n+1}\|f_n\|_{L_\infty(\mathbb{R})}\|H\|,\end{aligned}$$

where we use (2.3)–(2.5). Given $\varepsilon > 0$, take N such that $\sum_{n>N} 2^{n+1}\|f_n\|_{L_\infty} \leq \varepsilon/4$ and suppose $\|H\| \leq \frac{\varepsilon}{2\sum_{n\leq N} 2^{2(n+1)}\|f_n\|_{L_\infty(\mathbb{R})}}$. This implies that $\|S_f(A;H)\| \leq \varepsilon\|H\|$ and Fréchet differentiability of $f(A)$ with derivative $Df(A;H)$ follows.

To prove (2.16), use (2.6) and (2.12) to get

$$\begin{aligned}&\|S_{f_n}(A;H') - S_{f_n}(A;H)\| \\ &\quad\leq 2^{2(n+1)}\|f_n\|_{L_\infty(\mathbb{R})}\delta(H,H')\|H'-H\| \wedge 2^{n+2}\|f_n\|_{L_\infty(\mathbb{R})}\|H'-H\| \\ &\quad= 2^{n+2}\|f_n\|_{L_\infty(\mathbb{R})}(2^n\delta(H,H')\wedge 1)\|H'-H\|.\end{aligned}$$

It follows that

$$\|S_f(A;H') - S_f(A;H)\| \le \sum_{n\ge 0} \|S_{f_n}(A;H') - S_{f_n}(A;H)\|$$
$$\le \sum_{n\ge 0} 2^{n+2}\|f_n\|_{L_\infty(\mathbb{R})}(2^n\delta(H,H')\wedge 1)\|H'-H\|$$
$$= 4\Big(\sum_{2^n\le 1/\delta(H,H')} 2^{2n}\|f_n\|_{L_\infty(\mathbb{R})}\delta(H,H') + \sum_{2^n>1/\delta(H,H')} 2^n\|f_n\|_{L_\infty(\mathbb{R})}\Big)\|H'-H\|$$
$$\le 4\Big(\sum_{2^n\le 1/\delta(H,H')} 2^{sn}\|f_n\|_{L_\infty(\mathbb{R})}\Big(\frac{1}{\delta(H,H')}\Big)^{2-s}\delta(H,H')$$
$$+\sum_{2^{-n}<\delta(H,H')} 2^{sn}\|f_n\|_{L_\infty(\mathbb{R})}(\delta(H,H'))^{s-1}\Big)\|H'-H\|$$
$$\le 4\Big(\sum_{2^n\le 1/\delta(H,H')} 2^{sn}\|f_n\|_{L_\infty(\mathbb{R})} + \sum_{2^n>1/\delta(H,H')} 2^{sn}\|f_n\|_{L_\infty(\mathbb{R})}\Big)(\delta(H,H'))^{s-1}\|H'-H\|$$
$$= 4\|f\|_{B^s_{\infty,1}}(\delta(H,H'))^{s-1}\|H'-H\|,$$

which yields (2.16). The bound (2.15) follows from (2.16) when $H'=0$. □

Suppose $A\in\mathcal{B}_{sa}(\mathbb{H})$ is a compact operator with spectral representation $A=\sum_{\lambda\in\sigma(A)}\lambda P_\lambda$, where P_λ denotes the spectral projection corresponding to the eigenvalue λ. The following formula for the derivative $Df(A;H)$ with $f\in B^1_{\infty,1}(\mathbb{R})$ is well known (see [Bh, Theorem V.3.3] for a finite-dimensional version):

$$Df(A;H) = \sum_{\lambda,\mu\in\sigma(A)} f^{[1]}(\lambda,\mu)P_\lambda H P_\mu, \tag{2.20}$$

where $f^{[1]}(\lambda,\mu) := \frac{f(\lambda)-f(\mu)}{\lambda-\mu}$ for $\lambda\neq\mu$ and $f^{[1]}(\lambda,\lambda) := f'(\lambda)$. In other words, the operator $Df(A;H)$ can be represented in the basis of eigenvectors of A as a Schur product of the Loewner matrix $(f^{[1]}(\lambda,\mu))_{\lambda,\mu\in\sigma(A)}$ and the matrix of the operator H in this basis. We will need this formula only in the case of discrete spectrum, but there are also extensions to more general operators A with continuous spectrum (with the sums replaced by double operator integrals): see Aleksandrov and Peller [AP2, Theorems 3.5.11 and 1.6.4].

Finally, we need some extensions of the results stated above to higher order derivatives (see [Skr], [ACDS], [KS] and references therein for a number of subtle results in this direction). If $g:\mathcal{B}_{sa}(\mathbb{H})\to\mathcal{B}_{sa}(\mathbb{H})$ is a k times Fréchet differentiable function, its k-th derivative $D^kg(A)$ for $A\in\mathcal{B}_{sa}(\mathbb{H})$ can be viewed as a symmetric multilinear operator valued form

$$D^kg(A)(H_1,\dots,H_k) = D^kg(A;H_1,\dots,H_k),\qquad H_1,\dots,H_k\in\mathcal{B}_{sa}(\mathbb{H}).$$

Given such a form $M:\mathcal{B}_{sa}(\mathbb{H})\times\cdots\times\mathcal{B}_{sa}(\mathbb{H})\to\mathcal{B}_{sa}(\mathbb{H})$, define its operator norm as

$$\|M\| := \sup_{\|H_1\|,\dots,\|H_k\|\le 1}\|M(H_1,\dots,H_k)\|.$$

The derivatives $D^k g(A)$ are defined iteratively:

$$D^k g(A)(H_1, \dots, H_{k-1}, H_k) = D(D^{k-1} g(A)(H_1, \dots, H_{k-1}))(H_k).$$

For $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$, the k-th derivative $D^k f(A)$ is given by the formula

$$D^k f(A; H_1, \dots, H_k) \\ = \frac{(-1)^{k+1}}{2\pi i} \sum_{\pi \in S_k} \oint_\gamma f(z) R_A(z) H_{\pi(1)} R_A(z) H_{\pi(2)} \dots R_A(z) H_{\pi(k)} R_A(z)\, dz$$

for $H_1, \dots, H_k \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$, where $\gamma \subset \mathbb{C}$ is a contour surrounding $\sigma(A)$ with counter-clockwise orientation.

The following lemmas hold.

Lemma 4. *Let $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$. Then, for all $k \geq 1$,*

$$\|D^k f(A)\| \leq \sigma^k \|f\|_{L_\infty(\mathbb{R})}, \qquad A \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H}). \tag{2.21}$$

Proof. Given $A, H_1, \dots H_k \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$, denote

$$F(z_1, \dots, z_k) = f(A + z_1 H_1 + \dots + z_k H_k), \qquad (z_1, \dots, z_k) \in \mathbb{C}^k.$$

Then f is an entire operator valued function of exponential type $(\sigma\|H_1\|, \dots, \sigma\|H_k\|)$:

$$\|F(z_1, \dots, z_k)\| \leq \sup_{|\zeta| \leq \|A + z_1 H_1 + \dots + z_k H_k\|} |f(\zeta)| \\ \leq e^{\|A\|} \exp\{\sigma\|H_1\|\,|z_1| + \dots + \sigma\|H_k\|\,|z_k|\}.$$

By the Bernstein inequality (2.1) (extended to Banach space valued functions as at the beginning of the proof of Lemma 2), we get

$$\left\| \frac{\partial^k F(x_1, \dots, x_k)}{\partial x_1 \dots \partial x_k} \right\| \leq \sigma^k \|H_1\| \dots \|H_k\| \sup_{x_1, \dots, x_k \in \mathbb{R}} \|F(x_1, \dots, x_k)\|.$$

Therefore,

$$\|D^k f(A + x_1 H_1 + \dots + x_k H_k)(H_1, \dots, H_k)\| \leq \sigma^k \|H_1\| \dots \|H_k\|\, \|f\|_{L_\infty(\mathbb{R})}.$$

For $x_1 = \dots = x_k = 0$, this yields

$$\|D^k f(A)(H_1, \dots, H_k)\| \leq \sigma^k \|H_1\| \dots \|H_k\|\, \|f\|_{L_\infty(\mathbb{R})},$$

implying the claim of the lemma. □

Lemma 5. *Let $f \in \mathcal{E}_\sigma \cap L_\infty(\mathbb{R})$. Then, for all $k \geq 1$ and all $A, H_1, \dots, H_k, H \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\|D^k f(A+H; H_1, \dots, H_k) - D^k f(A; H_1, \dots, H_k)\| \leq \sigma^{k+1} \|f\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_k\| \, \|H\| \quad (2.22)$$

and

$$\|S_{D^k f(\cdot; H_1,\dots,H_k)}(A; H)\| \leq \frac{\sigma^{k+2}}{2} \|f\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_k\| \, \|H\|^2. \quad (2.23)$$

Proof. The bound (2.22) easily follows from (2.21) (applied to the derivative $D^{k+1} f$). The proof of (2.23) relies on the Bernstein inequality (2.1) and on a slight modification of the proof of (2.5). □

Lemma 6. *Suppose $f \in B^k_{\infty,1}(\mathbb{R})$. Then the function $\mathcal{B}_{sa}(\mathbb{H}) \ni A \mapsto f(A) \in \mathcal{B}_{sa}(\mathbb{H})$ is k times Fréchet differentiable and*

$$\|D^j f(A)\| \leq 2^j \|f\|_{B^j_{\infty,1}}, \quad A \in \mathcal{B}_{sa}(\mathbb{H}),\ j = 1, \dots, k. \quad (2.24)$$

Moreover, if $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (k, k+1]$, then

$$\|D^k f(A+H) - D^k f(A)\| \leq 2^{k+1} \|f\|_{B^s_{\infty,1}} \|H\|^{s-k}, \quad A, H \in \mathcal{B}_{sa}(\mathbb{H}). \quad (2.25)$$

Proof. As in the proof of Lemma 3, we use Littlewood–Paley decomposition of f. Since, by (2.21), for all $j = 1, \dots, k$,

$$\sum_{n\geq 0} \|D^j f_n(A; H_1, \dots, H_j)\| \leq \sum_{n\geq 0} 2^{(n+1)j} \|f_n\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_j\| \leq 2^j \|f\|_{B^j_{\infty,1}} \|H_1\| \dots \|H_j\| < \infty, \quad (2.26)$$

the series $\sum_{n\geq 0} D^j f_n(A; H_1, \dots, H_j)$ converges in operator norm and we can define symmetric j-linear forms

$$D^j f(A; H_1, \dots, H_j) := \sum_{n\geq 0} D^j f_n(A; H_1, \dots, H_j), \quad j = 1, \dots, k.$$

By the same argument as in the proof of claim (2.4) of Lemma 3 and using the bounds (2.22) and (2.23), we can now prove by induction that $D^j f(A; H_1, \dots, H_j)$, $j = 1, \dots, k$, are the consecutive derivatives of $f(A)$. Indeed, for $j = 1$, this was already proved in Lemma 3. Assume that it is true for some $j < k$; to prove that it is also true for $j + 1$ note that

$$\begin{aligned}
&\|D^j f(A+H; H_1, \dots, H_j) - D^j f(A; H_1, \dots, H_j) - D^{j+1} f(A; H_1, \dots, H_j, H)\| \\
&\leq \sum_{n\leq N} \|S_{D^j f_n(\cdot; H_1,\dots,H_j)}(A; H)\| \\
&\quad + \sum_{n>N} \|D^j f_n(A+H; H_1, \dots, H_j) - D^j f_n(A; H_1, \dots, H_j)\| \\
&\quad + \sum_{n>N} \|D^{j+1} f_n(A; H_1, \dots, H_j, H)\|
\end{aligned}$$

$$\leq \sum_{n\leq N} \frac{2^{(j+2)(n+1)}}{2} \|f_n\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_j\| \, \|H\|^2 + 2 \sum_{n>N} 2^{(j+1)(n+1)} \|f_n\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_j\| \, \|H\|.$$

Given $\varepsilon > 0$, take N such that $\sum_{n>N} 2^{(j+1)(n+1)} \|f_n\|_{L_\infty} \leq \varepsilon/4$, which is possible for $f \in B^{j+1}_{\infty,1}(\mathbb{R})$, and suppose $\|H\| \leq \frac{\varepsilon}{\sum_{n\leq N} 2^{(j+2)(n+1)} \|f_n\|_{L_\infty(\mathbb{R})}}$. Then

$$\|D^j f(A+H; H_1, \dots, H_j) - D^j f(A; H_1, \dots, H_j) - D^{j+1} f(A; H_1, \dots, H_j, H)\| \leq \varepsilon \|H_1\| \dots \|H_k\| \, \|H\|.$$

Therefore, the function $A \mapsto D^j f(A; H_1, \dots, H_j)$ is Fréchet differentiable with derivative $D^{j+1} f(A; H_1, \dots, H_j, H)$.

The bounds (2.24) now follow from (2.26).

To prove (2.25), note that

$$\|D^k f(A+H)(H_1, \dots, H_k) - D^k f(A)(H_1, \dots, H_k)\| \leq \sum_{n\geq 0} \|D^k f_n(A+H)(H_1, \dots, H_k) - D^k f_n(A)(H_1, \dots, H_k)\|.$$

Using (2.21) and (2.22), we get

$$\begin{aligned}
&\|D^k f(A+H)(H_1, \dots, H_k) - D^k f(A)(H_1, \dots, H_k)\| \\
&\quad \leq \sum_{2^n \leq 1/\|H\|} 2^{(n+1)(k+1)} \|f_n\|_{L_\infty(\mathbb{R})} \|H\| \, \|H_1\| \dots \|H_k\| \\
&\qquad + 2 \sum_{2^n > 1/\|H\|} 2^{(n+1)k} \|f_n\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_k\| \\
&\quad \leq 2^{k+1} \|H_1\| \dots \|H_k\| \\
&\qquad \times \Big[\sum_{2^n \leq 1/\|H\|} 2^{ns} \|f_n\|_{L_\infty(\mathbb{R})} 2^{n(k+1-s)} \|H\| + \sum_{2^n > 1/\|H\|} 2^{ns} \|f_n\|_{L_\infty(\mathbb{R})} 2^{n(k-s)} \Big] \\
&\quad \leq 2^{k+1} \|H_1\| \dots \|H_k\| \, \|H\|^{s-k} \Big[\sum_{2^n \leq 1/\|H\|} 2^{ns} \|f_n\|_{L_\infty(\mathbb{R})} + \sum_{2^n > 1/\|H\|} 2^{ns} \|f_n\|_{L_\infty(\mathbb{R})} \Big] \\
&\quad = 2^{k+1} \|f\|_{B^s_{\infty,1}} \|H\|^{s-k} \|H_1\| \dots \|H_k\|,
\end{aligned}$$

which implies (2.25). □

In what follows, we use the definition of Hölder space norms of functions of bounded self-adjoint operators. For an open set $G \subset \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$, a k times Fréchet differentiable function $g : G \to \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$ and, for $s = k + \beta$, $\beta \in (0, 1]$, define

$$\|g\|_{C^s(G)} := \max_{0\leq j\leq k} \sup_{A\in G} \|D^j g(A)\| \vee \sup_{A, A+H \in G,\, H\neq 0} \frac{\|D^k g(A+H) - D^k g(A)\|}{\|H\|^\beta}. \tag{2.27}$$

A similar definition applies to k times Fréchet differentiable functions $g : G \to \mathbb{R}$ (with $\|D^j g(A)\|$ being the operator norm of a j-linear form). In both cases, $C^s(G)$ denotes the space of functions g on G (operator valued or real valued) with $\|g\|_{C^s(G)} < \infty$. In particular, these norms apply to the operator functions $\mathcal{B}_{sa}(\mathbb{H}) \ni A \mapsto f(A) \in \mathcal{B}_{sa}(\mathbb{H})$, where f is a function on the real line. With a little abuse of notation, we write the norm of such an operator function as $\|f\|_{C^s(\mathcal{B}_{sa}(\mathbb{H}))}$. The next result immediately follows from Lemma 6.

Corollary 2. *Suppose that $f \in B^s_{\infty,1}(\mathbb{R})$ for some $k \geq 0$ and $s \in (k, k+1]$. Then $\|f\|_{C^s(\mathcal{B}_{sa}(\mathbb{H}))} \leq 2^{k+1}\|f\|_{B^s_{\infty,1}}$.*

3. Concentration bounds for the remainder of the first order Taylor expansion

Let $g : \mathcal{B}_{sa}(\mathbb{H}) \to \mathbb{R}$ be a Fréchet differentiable function with respect to the operator norm with derivative $Dg(A; H)$, $H \in \mathcal{B}_{sa}(\mathbb{H})$. Note that $Dg(A; \cdot)$, is a bounded linear functional on $\mathcal{B}_{sa}(\mathbb{H})$ and its restriction to the subspace $\mathcal{C}_{sa}(\mathbb{H}) \subset \mathcal{B}_{sa}(\mathbb{H})$ of compact self-adjoint operators in $\mathbb{H}$ can be represented as $Dg(A, H) = \langle Dg(A), H\rangle$, where $Dg(A) \in \mathcal{S}_1$ is a trace class operator in $\mathbb{H}$. Let $S_g(A; H)$ be the remainder of the first order Taylor expansion of g:

$$S_g(A; H) := g(A + H) - g(A) - Dg(A; H), \quad A, H \in \mathcal{B}_{sa}(\mathbb{H}).$$

Our goal is to obtain concentration inequalities for the random variable $S_g(\Sigma; \hat{\Sigma} - \Sigma)$ around its expectation. It will be done under the following assumption on the remainder $S_g(A; H)$:

Assumption 2. Let $s \in [1, 2]$. Assume there exists a constant $L_{g,s} > 0$ such that, for all $\Sigma \in \mathcal{C}_+(\mathbb{H})$ and $H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,

$$|S_g(\Sigma; H') - S_g(\Sigma; H)| \leq L_{g,s}(\|H\| \vee \|H'\|)^{s-1}\|H' - H\|.$$

Note that Assumption 2 implies (for $H' = 0$) that $|S_g(\Sigma; H)| \leq L_{g,s}\|H\|^s$ for $\Sigma \in \mathcal{C}_+(\mathbb{H})$ and $H \in \mathcal{B}_{sa}(\mathbb{H})$.

Theorem 5. *Suppose Assumption 2 holds for some $s \in (1, 2]$. Then there exists a constant $K_s > 0$ such that for all $t \geq 1$, with probability at least $1 - e^{-t}$,*

$$\begin{aligned}&|S_g(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_g(\Sigma; \hat{\Sigma} - \Sigma)| \\ &\leq K_s L_{g,s}\|\Sigma\|^s\biggl(\Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{s-1/2} \vee \Bigl(\frac{t}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{t}{n}\Bigr)^{s-1/2}\biggr)\sqrt{\frac{t}{n}}.\end{aligned} \tag{3.1}$$

Proof. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be such that $\varphi(u) = 1$ for $u \leq 1$, $\varphi(u) = 0$ for $u \geq 2$ and $\varphi(u) = 2 - u$ for $u \in (1, 2)$. Denote $E := \hat{\Sigma} - \Sigma$ and, given $\delta > 0$, define

$$h(X_1, \ldots, X_n) := S_g(\Sigma; E)\varphi(\|E\|/\delta). \tag{3.2}$$

We start by deriving a concentration bound for the function $h(X_1, \ldots, X_n)$ of Gaussian random variables $X_1, \ldots, X_n$. To this end, we will show that $h(X_1, \ldots, X_n)$ satisfies a Lipschitz condition. With a minor abuse of notation, we will assume for a while that $X_1, \ldots, X_n$ are nonrandom points of $\mathbb{H}$ and let $X'_1, \ldots, X'_n$ be another set of such points. Denote by $\hat{\Sigma}' := n^{-1} \sum_{j=1}^n X'_j \otimes X'_j$ the sample covariance based on $X'_1, \ldots, X'_n$ and let $E' := \hat{\Sigma}' - \Sigma$.

The following lemma establishes a Lipschitz condition for h.

Lemma 7. *Suppose Assumption 2 holds with some $s \in (1, 2]$. Then, for all $\delta > 0$ and h defined by (3.2), the following bound holds with some constant $C_s > 0$ for all $X_1, \ldots, X_n, X'_1, \ldots, X'_n \in \mathbb{H}$:*

$$|h(X_1, \ldots, X_n) - h(X'_1, \ldots, X'_n)| \leq \frac{C_s L_{g,s}(\|\Sigma\|^{1/2} + \sqrt{\delta})\delta^{s-1}}{\sqrt{n}} \Big(\sum_{j=1}^n \|X_j - X'_j\|^2 \Big)^{1/2}. \tag{3.3}$$

Proof. Using the fact that φ takes values in $[0, 1]$ and it is a Lipschitz function with constant 1, and taking into account Assumption 2, we get

$$\begin{aligned} |h(X_1, \ldots, X_n)| &\leq |S_g(\Sigma; E)| I(\|E\| \leq 2\delta) \leq L_{g,s}\|E\|^s I(\|E\| \leq 2\delta) \\ &\leq 2^s L_{g,s}\delta^s \end{aligned} \tag{3.4}$$

and similarly

$$|h(X'_1, \ldots, X'_n)| \leq 2^s L_{g,s}\delta^s. \tag{3.5}$$

We also have

$$\begin{aligned} |h(X_1, \ldots, X_n) - h(X'_1, \ldots, X'_n)| &\leq |S_g(\Sigma, E) - S_g(\Sigma, E')| + \frac{1}{\delta}|S_g(\Sigma, E')| \|E - E'\| \\ &\leq L_{g,s}(\|E\| \vee \|E'\|)^{s-1}\|E' - E\| + L_{g,s}\frac{1}{\delta}\|E'\|^s \|E' - E\|. \end{aligned} \tag{3.6}$$

If both $\|E\| \leq 2\delta$ and $\|E'\| \leq 2\delta$, then (3.6) implies

$$|h(X_1, \ldots, X_n) - h(X'_1, \ldots, X'_n)| \leq (2^{s-1} + 2^s)L_{g,s}\delta^{s-1}\|E' - E\|. \tag{3.7}$$

If both $\|E\| > 2\delta$ and $\|E'\| > 2\delta$, then $\varphi(\|E\|/\delta) = \varphi(\|E'\|/\delta) = 0$, implying that $h(X_1, \ldots, X_n) = h(X'_1, \ldots, X'_n) = 0$. If $\|E\| \leq 2\delta$, $\|E'\| > 2\delta$ and $\|E' - E\| > \delta$, then

$$\begin{aligned} |h(X_1, \ldots, X_n) - h(X'_1, \ldots, X'_n)| = |h(X_1, \ldots, X_n)| &\leq 2^s L_{g,s}\delta^s \\ &\leq 2^s L_{g,s}\delta^{s-1}\|E' - E\|. \end{aligned}$$

If $\|E\| \leq 2\delta$, $\|E'\| > 2\delta$ and $\|E' - E\| \leq \delta$, then $\|E'\| \leq 3\delta$ and, similarly to (3.7), we get

$$|h(X_1, \ldots, X_n) - h(X'_1, \ldots, X'_n)| \leq (3^{s-1} + 3^s)L_{g,s}\delta^{s-1}\|E' - E\|. \tag{3.8}$$

By these simple considerations, the bound (3.8) holds in all possible cases. This fact along with (3.4) and (3.5) yields

$$|h(X_1,\dots,X_n)-h(X_1',\dots,X_n')| \le (3^{s-1}+3^s)L_{g,s}\delta^{s-1}(\|E'-E\|\wedge\delta). \tag{3.9}$$

We now obtain an upper bound on $\|E'-E\|$. We have

$$\begin{aligned}
\|E'-E\| &= \Big\|n^{-1}\sum_{j=1}^n X_j\otimes X_j - n^{-1}\sum_{j=1}^n X_j'\otimes X_j'\Big\| \\
&\le \Big\|n^{-1}\sum_{j=1}^n (X_j-X_j')\otimes X_j\Big\| + \Big\|n^{-1}\sum_{j=1}^n X_j'\otimes (X_j-X_j')\Big\| \\
&= \sup_{\|u\|,\|v\|\le 1}\Big|n^{-1}\sum_{j=1}^n \langle X_j-X_j',u\rangle\langle X_j,v\rangle\Big| + \sup_{\|u\|,\|v\|\le 1}\Big|n^{-1}\sum_{j=1}^n \langle X_j',u\rangle\langle X_j-X_j',v\rangle\Big| \\
&\le \sup_{\|u\|\le 1}\Big(n^{-1}\sum_{j=1}^n\langle X_j-X_j',u\rangle^2\Big)^{1/2}\sup_{\|v\|\le 1}\Big(n^{-1}\sum_{j=1}^n\langle X_j,v\rangle^2\Big)^{1/2} \\
&\quad + \sup_{\|u\|\le 1}\Big(n^{-1}\sum_{j=1}^n\langle X_j',u\rangle^2\Big)^{1/2}\sup_{\|v\|\le 1}\Big(n^{-1}\sum_{j=1}^n\langle X_j-X_j',v\rangle^2\Big)^{1/2} \\
&\le \frac{\|\hat\Sigma\|^{1/2}+\|\hat\Sigma'\|^{1/2}}{\sqrt{n}}\Big(\sum_{j=1}^n\|X_j-X_j'\|^2\Big)^{1/2} \le (2\|\Sigma\|^{1/2}+\|E\|^{1/2}+\|E'\|^{1/2})\Delta,
\end{aligned}$$

where $\Delta := n^{-1/2}(\sum_{j=1}^n \|X_j-X_j'\|^2)^{1/2}$. Without loss of generality, assume that $\|E\|\le 2\delta$ (again, if both $\|E\|>2\delta$ and $\|E'\|>2\delta$, then $h(X_1,\dots,X_n)=h(X_1',\dots,X_n')=0$ and inequality (3.3) holds trivially). Then

$$\|E'-E\|\le (2\|\Sigma\|^{1/2}+2\sqrt{2\delta}+\|E'-E\|^{1/2})\Delta.$$

If $\|E'-E\|\le\delta$, the last bound implies that

$$\|E'-E\|\le (2\|\Sigma\|^{1/2}+(2\sqrt{2}+1)\sqrt{\delta})\Delta \le 4\|\Sigma\|^{1/2}\Delta\vee(4\sqrt{2}+2)\sqrt{\delta}\,\Delta.$$

Otherwise, if $\|E'-E\|>\delta$, we get

$$\|E'-E\|\le 4\|\Sigma\|^{1/2}\Delta\vee(4\sqrt{2}+2)\Delta\|E'-E\|^{1/2},$$

which yields

$$\|E'-E\|\le 4\|\Sigma\|^{1/2}\Delta\vee(4\sqrt{2}+2)^2\Delta^2.$$

Thus, either $\|E'-E\|\le 4\|\Sigma\|^{1/2}\Delta$, or $\|E'-E\|\le(4\sqrt{2}+2)^2\Delta^2$. In the last case, we also have (since $\delta<\|E'-E\|$)

$$\delta<\sqrt{\delta}\,\|E'-E\|^{1/2}\le(4\sqrt{2}+2)\sqrt{\delta}\,\Delta.$$

This shows that

$$\|E'-E\|\wedge\delta\le 4\|\Sigma\|^{1/2}\Delta\vee(4\sqrt{2}+2)\sqrt{\delta}\,\Delta \tag{3.10}$$

both when $\|E'-E\|\le\delta$ and when $\|E'-E\|>\delta$.

Substituting (3.10) in (3.9) yields

$$|h(X_1,\dots,X_n)-h(X'_1,\dots,X'_n)| \\ \le \frac{(3^{s-1}+3^s)L_{g,s}(4\|\Sigma\|^{1/2}+(4\sqrt{2}+2)\sqrt{\delta})\delta^{s-1}}{\sqrt{n}}\Big(\sum_{j=1}^n \|X_j-X'_j\|^2\Big)^{1/2},$$

which implies (3.3). □

In what follows, we set, for a given $t>0$,

$$\delta=\delta_n(t):=\mathbb{E}\|\hat{\Sigma}-\Sigma\|+C\|\Sigma\|\bigg[\bigg(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee 1\bigg)\sqrt{\frac{t}{n}}\vee\frac{t}{n}\bigg].$$

It follows from (1.3) that there exists an absolute constant $C>0$ such that

$$\mathbb{P}\{\|\hat{\Sigma}-\Sigma\|\ge\delta_n(t)\}\le e^{-t},\qquad t\ge 1. \tag{3.11}$$

Assuming that $t\ge\log(4)$, we get $\mathbb{P}\{\|E\|\ge\delta\}\le 1/4$. Let $M:=\mathrm{Med}(S_g(\Sigma;E))$ be a median of random variable $S_g(\Sigma;E)$. Then

$$\mathbb{P}\{h(X_1,\dots,X_n)\ge M\}\ge\mathbb{P}\{h(X_1,\dots,X_n)\ge M,\ \|E\|<\delta\} \\ \ge\mathbb{P}\{S_g(\Sigma;E)\ge M,\ \|E\|<\delta\}\ge 1/2-\mathbb{P}\{\|E\|\ge\delta\}\ge 1/4.$$

Similarly, $\mathbb{P}\{h(X_1,\dots,X_n)\le M\}\ge 1/4$. In view of the Lipschitz property of h (Lemma 7), we now use a relatively standard argument (see [KL2, Lemma 2 and its applications in Section 3]) based on a Gaussian isoperimetric inequality (see Ledoux [Led, Theorem 2.5 and (2.9)]) to conclude that with probability at least $1-e^{-t}$,

$$|h(X_1,\dots,X_n)-M|\lesssim_s L_{g,s}\delta^{s-1}(\|\Sigma\|^{1/2}+\delta^{1/2})\|\Sigma\|^{1/2}\sqrt{t/n}.$$

Moreover, since $S_g(\Sigma;E)=h(X_1,\dots,X_n)$ on the event $\{\|E\|<\delta\}$ of probability at least $1-e^{-t}$, we find that with probability $1-2e^{-t}$,

$$|S_g(\Sigma;E)-M|\lesssim_s L_{g,s}\delta^{s-1}(\|\Sigma\|^{1/2}+\delta^{1/2})\|\Sigma\|^{1/2}\sqrt{t/n}. \tag{3.12}$$

It follows from (1.1) that

$$\delta=\delta_n(t)\lesssim\|\Sigma\|\bigg(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\frac{\mathbf{r}(\Sigma)}{n}\vee\sqrt{\frac{t}{n}}\vee\frac{t}{n}\bigg). \tag{3.13}$$

Substituting (3.13) into (3.12) easily yields, with probability at least $1-2e^{-t}$,

$$|S_g(\Sigma;E)-M| \\ \lesssim_s L_{g,s}\|\Sigma\|^s\bigg(\bigg(\frac{\mathbf{r}(\Sigma)}{n}\bigg)^{(s-1)/2}\vee\bigg(\frac{\mathbf{r}(\Sigma)}{n}\bigg)^{s-1/2}\vee\bigg(\frac{t}{n}\bigg)^{(s-1)/2}\vee\bigg(\frac{t}{n}\bigg)^{s-1/2}\bigg)\sqrt{\frac{t}{n}}, \tag{3.14}$$

and moreover by adjusting the value of the constant in (3.14) the probability bound can be replaced by $1-e^{-t}$. By integrating out the tails of (3.14) one can get

$$|\mathbb{E}S_g(\Sigma; E) - M| \le \mathbb{E}|S_g(\Sigma; E) - M| \lesssim_s L_{g,s}\|\Sigma\|^s\biggl(\Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{s-1/2} \vee \Bigl(\frac{1}{n}\Bigr)^{(s-1)/2}\biggr)\sqrt{\frac{1}{n}}. \tag{3.15}$$

Combining (3.14) and (3.15) implies that for all $t \ge 1$, with probability at least $1-e^{-t}$,

$$|S_g(\Sigma; E) - \mathbb{E}S_g(\Sigma; E)| \lesssim_s L_{g,s}\|\Sigma\|^s\biggl(\Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{s-1/2} \vee \Bigl(\frac{t}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{t}{n}\Bigr)^{s-1/2}\biggr)\sqrt{\frac{t}{n}}, \tag{3.16}$$

which completes the proof. □

Example 1. Consider the functional

$$g(A) := \langle f(A), B\rangle = \mathrm{tr}(f(A)B^*), \qquad A \in \mathcal{B}_{sa}(\mathbb{H}),$$

where f is a given smooth function and $B \in \mathcal{S}_1$ is a given nuclear operator.

Corollary 3. *If $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$, then with probability at least $1-e^{-t}$ the following concentration inequality holds for the functional g:*

$$|S_g(\Sigma; \hat{\Sigma}-\Sigma) - \mathbb{E}S_g(\Sigma; \hat{\Sigma}-\Sigma)| \lesssim_s \|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s\biggl(\Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{s-1/2} \vee \Bigl(\frac{t}{n}\Bigr)^{(s-1)/2} \vee \Bigl(\frac{t}{n}\Bigr)^{s-1/2}\biggr)\sqrt{\frac{t}{n}}. \tag{3.17}$$

Proof. It easily follows from Lemma 3 that Assumption 2 is satisfied for $s \in [1,2]$ with $L_{g,s} = 2^{s+1}\|f\|_{B^s_{\infty,1}}\|B\|_1$. Therefore, Theorem 5 implies (3.17). □

In what follows, we need a more general version of the bound of Theorem 5 (under somewhat more general conditions than in Assumption 2).

Assumption 3. Assume that, for all $\Sigma \in \mathcal{C}_+(\mathbb{H})$ and $H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,

$$|S_g(\Sigma; H') - S_g(\Sigma; H)| \le \eta(\Sigma; \|H\| \vee \|H'\|)\|H'-H\|,$$

where $0 < \delta \mapsto \eta(\Sigma; \delta)$ is a nondecreasing function of the following form:

$$\eta(\Sigma; \delta) := \eta_1(\Sigma)\delta^{\alpha_1} \vee \cdots \vee \eta_m(\Sigma)\delta^{\alpha_m}$$

for given nonnegative functions $\eta_1, \dots, \eta_m$ on $\mathcal{C}_+(\mathbb{H})$ and positive numbers $\alpha_1, \dots, \alpha_m$.

The proof of the following result is a simple modification of the proof of Theorem 5.

Theorem 6. *Suppose Assumption 3 holds. Then for all $t \geq 1$, with probability at least $1 - e^{-t}$,*

$$|S_g(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_g(\Sigma; \hat{\Sigma} - \Sigma)| \\ \lesssim_\eta \eta(\Sigma; \delta_n(\Sigma; t))(\sqrt{\|\Sigma\|} + \sqrt{\delta_n(\Sigma; t)})\sqrt{\|\Sigma\|}\sqrt{t/n}, \qquad (3.18)$$

where

$$\delta_n(\Sigma; t) := \|\Sigma\|\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right). \qquad (3.19)$$

4. Normal approximation bounds for plug-in estimators

Let $g : \mathcal{B}_{sa}(\mathbb{H}) \to \mathbb{R}$ be a Fréchet differentiable function with respect to the operator norm with derivative $Dg(A; H)$, $H \in \mathcal{B}_{sa}(\mathbb{H})$. Recall that for $H \in \mathcal{C}_{sa}(\mathbb{H})$ we have $Dg(A; H) = \langle Dg(A), H\rangle$, where $Dg(A) \in \mathcal{S}_1$. Denote

$$\mathcal{D}g(\Sigma) := \Sigma^{1/2} Dg(\Sigma)\Sigma^{1/2}.$$

The following theorem is the main result of this section.

Theorem 7. *Suppose Assumption 2 holds for some $s \in (1, 2]$ and also that $\mathbf{r}(\Sigma) \leq n$. Define*

$$\gamma_s(g; \Sigma) := \log\left(\frac{L_{g,s}\|\Sigma\|^s}{\|\mathcal{D}g(\Sigma)\|_2}\right), \qquad t_{n,s}(g; \Sigma) := \left[-\gamma_s(g; \Sigma) + \frac{s-1}{2}\log\left(\frac{n}{\mathbf{r}(\Sigma)}\right)\right] \vee 1.$$

Then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n^{1/2}(g(\hat{\Sigma}) - \mathbb{E}g(\hat{\Sigma}))}{\sqrt{2}\|\mathcal{D}g(\Sigma)\|_2} \leq x\right\} - \Phi(x)\right| \lesssim_s \left(\frac{\|\mathcal{D}g(\Sigma)\|_3}{\|\mathcal{D}g(\Sigma)\|_2}\right)^3 \frac{1}{\sqrt{n}} \\ + \frac{L_{g,s}\|\Sigma\|^s}{\|\mathcal{D}g(\Sigma)\|_2}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2} \vee \left(\frac{t_{n,s}(g; \Sigma)}{n}\right)^{(s-1)/2} \vee \left(\frac{t_{n,s}(g; \Sigma)}{n}\right)^{s-1/2}\right)\sqrt{t_{n,s}(g; \Sigma)}. \qquad (4.1)$$

Proof. Note that

$$g(\hat{\Sigma}) - g(\Sigma) = \langle Dg(\Sigma), \hat{\Sigma} - \Sigma\rangle + S_g(\Sigma; \hat{\Sigma} - \Sigma),$$

and since $\mathbb{E}\langle Dg(\Sigma), \hat{\Sigma} - \Sigma\rangle = 0$, we have

$$\mathbb{E}g(\hat{\Sigma}) - g(\Sigma) = S_g(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_g(\Sigma; \hat{\Sigma} - \Sigma),$$

implying that

$$g(\hat{\Sigma}) - \mathbb{E}g(\hat{\Sigma}) = \langle Dg(\Sigma), \hat{\Sigma} - \Sigma\rangle + S_g(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_g(\Sigma; \hat{\Sigma} - \Sigma). \qquad (4.2)$$

The linear term

$$\langle Dg(\Sigma), \hat{\Sigma} - \Sigma\rangle = n^{-1}\sum_{j=1}^{n}\langle Dg(\Sigma)X_j, X_j\rangle - \mathbb{E}\langle Dg(\Sigma)X, X\rangle \tag{4.3}$$

is the sum of i.i.d. random variables and it will be approximated by a normal distribution.

We will need the following simple lemma.

Lemma 8. *Let $A \in \mathcal{S}_1$ be a self-adjoint trace class operator. Denote by λ_j, $j \geq 1$, the eigenvalues of the operator $\Sigma^{1/2}A\Sigma^{1/2}$ (repeated with their multiplicities and, to be specific, such that their absolute values are arranged in nonincreasing order). Then*

$$\langle AX, X\rangle \overset{d}{=} \sum_{k\geq 1}\lambda_k Z_k^2,$$

where $Z_1, Z_2, \ldots$ are i.i.d. standard normal random variables.

Proof. First assume that Σ is a finite rank operator, or equivalently that X takes values in a finite-dimensional subspace L of $\mathbb{H}$. In this case, $X = \Sigma^{1/2}Z$, where Z is a standard normal vector in L. Therefore,

$$\langle AX, X\rangle = \langle A\Sigma^{1/2}Z, \Sigma^{1/2}Z\rangle = \langle \Sigma^{1/2}A\Sigma^{1/2}Z, Z\rangle = \sum_{k\geq 1}\lambda_k Z_k^2,$$

where $\{Z_k\}$ are the coordinates of Z in the basis of eigenvectors of $\Sigma^{1/2}A\Sigma^{1/2}$.

In the infinite-dimensional case the result follows by standard finite-dimensional approximation. □

Note that $\mathbb{E}\langle AX, X\rangle = \sum_{k\geq 1}\lambda_k = \mathrm{tr}(\Sigma^{1/2}A\Sigma^{1/2})$ and

$$\mathrm{Var}(\langle AX, X\rangle) = \sum_{k\geq 1}\lambda_k^2\mathbb{E}(Z_k^2 - 1)^2 = 2\sum_{k\geq 1}\lambda_k^2 = 2\|\Sigma^{1/2}A\Sigma^{1/2}\|_2^2.$$

The following result immediately follows from the Berry–Esseen bound (see [Pet, Chapter 5, Theorem 3]; an extension of the inequality to infinite sums of independent r.v. is based on a straightforward approximation argument).

Lemma 9. *The following bound holds:*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n^{1/2}\langle Dg(\Sigma), \hat{\Sigma} - \Sigma\rangle}{\sqrt{2}\|\mathcal{D}g(\Sigma)\|_2} \leq x\right\} - \Phi(x)\right| \lesssim \left(\frac{\|\mathcal{D}g(\Sigma)\|_3}{\|\mathcal{D}g(\Sigma)\|_2}\right)^3\frac{1}{\sqrt{n}}.$$

Proof. Indeed, by (4.3) and Lemma 8 with $A = Dg(\Sigma)$,

$$\frac{n^{1/2}\langle Dg(\Sigma), \hat{\Sigma} - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2} \overset{d}{=} \frac{\sum_{j=1}^{n}\sum_{k\geq 1}\lambda_k(Z_{k,j}^2 - 1)}{\mathrm{Var}^{1/2}(\sum_{j=1}^{n}\sum_{k\geq 1}\lambda_k(Z_{k,j}^2 - 1))}, \tag{4.4}$$

where $\{Z_{k,j}\}$ are i.i.d. standard normal random variables. By the Berry–Esseen bound,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sum_{j=1}^n\sum_{k\geq 1}\lambda_k(Z_{k,j}^2-1)}{\mathrm{Var}^{1/2}(\sum_{j=1}^n\sum_{k\geq 1}\lambda_k(Z_{k,j}^2-1))}\right\}-\Phi(x)\right|$$
$$\lesssim \frac{\sum_{j=1}^n\sum_{k\geq 1}|\lambda_k|^3\mathbb{E}|Z_{k,j}^2-1|^3}{(\sum_{j=1}^n\sum_{k\geq 1}\lambda_k^2\mathbb{E}(Z_{k,j}^2-1)^2)^{3/2}}\lesssim \frac{\sum_{k\geq 1}|\lambda_k|^3}{(\sum_{k\geq 1}\lambda_k^2)^{3/2}}\frac{1}{\sqrt{n}}\lesssim\left(\frac{\|\mathcal{D}g(\Sigma)\|_3}{\|\mathcal{D}g(\Sigma)\|_2}\right)^3\frac{1}{\sqrt{n}}. \qquad \square$$

Finally, the following lemma will be used.

Lemma 10. *For random variables ξ, η, denote*

$$\Delta(\xi,\eta) := \sup_{x\in\mathbb{R}}|\mathbb{P}\{\xi\leq x\}-\mathbb{P}\{\eta\leq x\}|, \qquad \delta(\xi,\eta) := \inf_{\delta>0}[\mathbb{P}\{|\xi-\eta|\geq\delta\}+\delta].$$

Then, for a standard normal random variable Z, $\Delta(\xi,Z)\leq\Delta(\eta,Z)+\delta(\xi,\eta)$.

Proof. For all $x\in\mathbb{R}$ and $\delta>0$,

$$\begin{aligned}\mathbb{P}\{\xi\leq x\} &\leq \mathbb{P}\{\xi\leq x,\ |\xi-\eta|<\delta\}+\mathbb{P}\{|\xi-\eta|\geq\delta\}\\ &\leq \mathbb{P}\{\eta\leq x+\delta\}+\mathbb{P}\{|\xi-\eta|\geq\delta\}\\ &\leq \mathbb{P}\{Z\leq x+\delta\}+\Delta(\eta,Z)+\mathbb{P}\{|\xi-\eta|\geq\delta\}\\ &\leq \mathbb{P}\{Z\leq x\}+\delta+\Delta(\eta,Z)+\mathbb{P}\{|\xi-\eta|\geq\delta\},\end{aligned}$$

where we use the trivial bound $\mathbb{P}\{Z\leq x+\delta\}-\mathbb{P}\{Z\leq x\}\leq\delta$. Thus,

$$\mathbb{P}\{\xi\leq x\}-\mathbb{P}\{Z\leq x\}\leq\Delta(\eta,Z)+\mathbb{P}\{|\xi-\eta|\geq\delta\}+\delta.$$

Similarly,

$$\mathbb{P}\{\xi\leq x\}-\mathbb{P}\{Z\leq x\}\geq-\Delta(\eta,Z)-\mathbb{P}\{|\xi-\eta|\geq\delta\}-\delta,$$

implying that $\Delta(\xi,Z)\leq\Delta(\eta,Z)+\mathbb{P}\{|\xi-\eta|\geq\delta\}+\delta$ for all $\delta>0$. Taking the infimum over $\delta>0$ yields the claim of the lemma. $\square$

We apply the last lemma to the random variables

$$\xi := \frac{n^{1/2}(g(\hat{\Sigma})-\mathbb{E}g(\hat{\Sigma}))}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2} \quad\text{and}\quad \eta := \frac{n^{1/2}\langle Dg(\Sigma),\hat{\Sigma}-\Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2}.$$

By (4.2),

$$\xi-\eta=\frac{n^{1/2}(S_g(\Sigma;\hat{\Sigma}-\Sigma)-\mathbb{E}S_g(\Sigma;\hat{\Sigma}-\Sigma))}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2}.$$

Recall that Assumption 2 holds and $\mathbf{r}(\Sigma)\leq n$, and denote

$$\delta_{n,s}(g;\Sigma;t) := K_s L_{g,s}\frac{\|\Sigma\|^s}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2}\vee\left(\frac{t}{n}\right)^{(s-1)/2}\vee\left(\frac{t}{n}\right)^{s-1/2}\right)\sqrt{t}.$$

It immediately follows from Theorem 5 that $\mathbb{P}\{|\xi-\eta|\geq\delta_{n,s}(g;\Sigma;t)\}\leq e^{-t}$ for $t\geq 1$, and

$$\delta(\xi,\eta)\leq\inf_{t\geq 1}[\delta_{n,s}(g;\Sigma;t)+e^{-t}].$$

It follows from Lemmas 9 and 10 that, for some $C > 0$,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n^{1/2}(g(\hat{\Sigma}) - \mathbb{E}g(\hat{\Sigma}))}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2} \le x\right\} - \Phi(x)\right| \le C\left(\frac{\|\mathcal{D}g(\Sigma)\|_3}{\|\mathcal{D}g(\Sigma)\|_2}\right)^3 \frac{1}{\sqrt{n}} + \inf_{t\ge 1}[\delta_{n,s}(g;\Sigma;t) + e^{-t}]. \tag{4.5}$$

Recall that $\gamma_s(g;\Sigma) = \log\left(\frac{L_{g,s}\|\Sigma\|^s}{\|\mathcal{D}g(\Sigma)\|_2}\right)$ and

$$t_{n,s}(g;\Sigma) = \left[-\gamma_s(g;\Sigma) + \frac{s-1}{2}\log\left(\frac{n}{\mathbf{r}(\Sigma)}\right)\right] \vee 1.$$

Let $\bar{t} := t_{n,s}(g;\Sigma)$. Then

$$e^{-\bar{t}} \le L_{g,s}\frac{\|\Sigma\|^s\sqrt{\bar{t}}}{\sqrt{2}\,\|\mathcal{D}g(\Sigma)\|_2}\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2} \lesssim_s \delta_{n,s}(g;\Sigma;\bar{t}).$$

Therefore,

$$\inf_{t\ge 1}[\delta_{n,s}(g;\Sigma;t) + e^{-t}] \lesssim_s \delta_{n,s}(g;\Sigma;\bar{t}) \lesssim_s \frac{L_{g,s}\|\Sigma\|^s}{\|\mathcal{D}g(\Sigma)\|_2}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2} \vee \left(\frac{t_{n,s}(g;\Sigma)}{n}\right)^{(s-1)/2} \vee \left(\frac{t_{n,s}(g;\Sigma)}{n}\right)^{s-1/2}\right)\sqrt{t_{n,s}(g;\Sigma)}.$$

Substituting this into (4.5) completes the proof of Theorem 7. □

Our main example of interest is the functional $g(A) := \langle f(A), B\rangle$ for $A \in \mathcal{B}_{sa}(\mathbb{H})$, where f is a smooth function and $B \in \mathcal{S}_1(\mathbb{H})$ is a nuclear operator. If $f \in B^1_{\infty,1}(\mathbb{R})$, then the function $A \mapsto f(A)$ is operator differentiable, implying the differentiability of the functional $A \mapsto g(A)$ with derivative $Dg(A;H) = \langle Df(A;H), B\rangle$ for $A, H \in \mathcal{B}_{sa}(\mathbb{H})$. Moreover, for $A = \Sigma$ with spectral decomposition $\Sigma = \sum_{\lambda\in\sigma(\Sigma)} \lambda P_\lambda$, formula (2.20) holds, implying that $\mathcal{C}_{sa}(\mathbb{H}) \ni H \mapsto Df(\Sigma;H) = Df(\Sigma)H \in \mathcal{B}_{sa}(\mathbb{H})$ is a symmetric operator: $\langle Df(\Sigma)H_1, H_2\rangle = \langle H_1, Df(\Sigma)H_2\rangle$ for $H_1 \in \mathcal{C}_{sa}(\mathbb{H})$ and $H_2 \in \mathcal{S}_1(\mathbb{H})$. Therefore,

$$Dg(\Sigma;H) = \langle Df(\Sigma;B), H\rangle, \qquad H \in \mathcal{C}_{sa}(\mathbb{H}),$$

or, in other words, $Dg(\Sigma) = Df(\Sigma;B)$. Denote

$$\sigma_f(\Sigma;B) := \sqrt{2}\,\|\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}\|_2, \qquad \mu_f^{(3)}(\Sigma;B) = \|\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}\|_3.$$

The following result is a simple consequence of Theorem 7 and Corollary 3.

Corollary 4. *Let $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$. Define*

$$\gamma_s(f;\Sigma) := \log\left(\frac{2^{s+3/2}\|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s}{\sigma_f(\Sigma;B)}\right),$$
$$t_{n,s}(f;\Sigma) := \left[-\gamma_s(f;\Sigma) + \frac{s-1}{2}\log\left(\frac{n}{\mathbf{r}(\Sigma)}\right)\right]\vee 1.$$

Then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n^{1/2}\langle f(\hat\Sigma)-\mathbb{E}f(\hat\Sigma)),B\rangle}{\sigma_f(\Sigma,B)}\le x\right\}-\Phi(x)\right|$$
$$\lesssim_s \Delta_n^{(s)}(f;\Sigma;B) := \left(\frac{\mu_f^{(3)}(\Sigma;B)}{\sigma_f(\Sigma;B)}\right)^3\frac{1}{\sqrt{n}}$$
$$+\frac{\|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s}{\sigma_f(\Sigma;B)}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2}\vee\left(\frac{t_{n,s}(f;\Sigma)}{n}\right)^{(s-1)/2}\vee\left(\frac{t_{n,s}(f;\Sigma)}{n}\right)^{s-1/2}\right)$$
$$\times\sqrt{t_{n,s}(f;\Sigma)}. \tag{4.6}$$

We will now prove Theorem 2 and Corollary 1 from Section 1.

Proof of Theorem 2 and Corollary 1. The proof of (1.7) immediately follows from (4.6). It is also easy to prove (1.8) using (1.7), the bound on the bias

$$\|\mathbb{E}_\Sigma f(\hat\Sigma)-f(\Sigma)\| = \|\mathbb{E}_\Sigma S_f(\Sigma;\hat\Sigma-\Sigma)\| \lesssim \|f\|_{B^s_{\infty,1}}\mathbb{E}\|\hat\Sigma-\Sigma\|^s$$
$$\lesssim \|f\|_{B^s_{\infty,1}}\|\Sigma\|^s(\mathbf{r}(\Sigma)/n)^{s/2}, \tag{4.7}$$

and Lemma 10.

The proof of (1.9) is a bit more involved and requires a few more lemmas. The following fact is well known (it follows, e.g., from [Ver, Proposition 5.16]).

Lemma 11. *Let $\{\xi_i\}$ be i.i.d. standard normal random variables and let $\{\gamma_i\}$ be real numbers. Then for all $t\ge 0$, with probability at least $1-e^{-t}$,*

$$\left|\sum_{i\ge1}\gamma_i(\xi_i^2-1)\right| \lesssim \left(\sum_{i\ge1}\gamma_i^2\right)^{1/2}\sqrt{t}\vee\sup_{i\ge1}|\gamma_i|t.$$

Lemma 12. *If $f\in B^s_{\infty,1}(\mathbb{R})$ for some $s\in(1,2]$ and $\mathbf{r}(\Sigma)\le n$, then for all $t\ge1$, with probability at least $1-e^{-t}$,*

$$\left|\frac{n^{1/2}\langle f(\hat\Sigma)-f(\Sigma),B\rangle}{\sigma_f(\Sigma;B)}\right|$$
$$\lesssim_s \left(\frac{\|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s}{\sigma_f(\Sigma;B)}\vee\frac{\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma;B)}\vee 1\right)\left(\sqrt{t}\vee\frac{(\mathbf{r}(\Sigma))^{s/2}}{n^{(s-1)/2}}\right). \tag{4.8}$$

Proof. Recall that

$$\langle f(\hat\Sigma)-\mathbb{E}f(\hat\Sigma),B\rangle = \langle Df(\Sigma;\hat\Sigma-\Sigma),B\rangle + \langle S_f(\Sigma;\hat\Sigma-\Sigma)-\mathbb{E}S_f(\Sigma;\hat\Sigma-\Sigma),B\rangle.$$

It follows from (4.4) that

$$\frac{n^{1/2}\langle Df(\Sigma;\hat\Sigma-\Sigma),B\rangle}{\sigma_f(\Sigma;B)} \overset{d}{=} \frac{\sum_{j=1}^n\sum_{k\geq 1}\lambda_k(Z_{k,j}^2-1)}{\mathrm{Var}^{1/2}(\sum_{j=1}^n\sum_{k\geq 1}\lambda_k(Z_{k,j}^2-1))}, \tag{4.9}$$

where the $Z_{k,j}$ are i.i.d. standard normal r.v.'s and the λ_k are the eigenvalues (repeated with their multiplicities) of $\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}$. Using Lemma 11, we easily see that for all $t\geq 1$, with probability at least $1-e^{-t}$,

$$\left|\frac{n^{1/2}\langle Df(\Sigma;\hat\Sigma-\Sigma),B\rangle}{\sigma_f(\Sigma;B)}\right| \lesssim \sqrt{t}\vee\frac{t}{\sqrt{n}}. \tag{4.10}$$

To control $\langle S_f(\Sigma;\hat\Sigma-\Sigma)-\mathbb{E}S_f(\Sigma;\hat\Sigma-\Sigma),B\rangle$, we use the bound (3.17) to deduce that for all $t\geq 1$, with probability at least $1-e^{-t}$,

$$\begin{aligned}&|\langle S_f(\Sigma;\hat\Sigma-\Sigma)-\mathbb{E}S_f(\Sigma;\hat\Sigma-\Sigma),B\rangle|\\ &\lesssim_s \|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2}\vee\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{s-1/2}\vee\left(\frac{t}{n}\right)^{(s-1)/2}\vee\left(\frac{t}{n}\right)^{s-1/2}\right)\sqrt{\frac{t}{n}}.\end{aligned} \tag{4.11}$$

If $\mathbf{r}(\Sigma)\leq n$ and $t\leq n$, the bounds (4.10), (4.11) and (4.7) easily imply that with probability at least $1-e^{-t}$,

$$\left|\frac{n^{1/2}\langle f(\hat\Sigma)-f(\Sigma),B\rangle}{\sigma_f(\Sigma;B)}\right| \lesssim_s \left(\frac{\|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s}{\sigma_f(\Sigma;B)}\vee 1\right)\left(\sqrt{t}\vee\frac{(\mathbf{r}(\Sigma))^{s/2}}{n^{(s-1)/2}}\right). \tag{4.12}$$

Note also that, for all $t>n$,

$$\left|\frac{n^{1/2}\langle f(\hat\Sigma)-f(\Sigma),B\rangle}{\sigma_f(\Sigma;B)}\right| \leq \frac{2\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma;B)}\sqrt{t}. \tag{4.13}$$

The result immediately follows from (4.12) and (4.13). □

Lemma 13. *Let ℓ be a loss function satisfying Assumption 1. For any random variables ξ,η and for all $A>0$,*

$$|\mathbb{E}\ell(\xi)-\mathbb{E}\ell(\eta)|\leq 4\ell(A)\Delta(\xi;\eta)+\mathbb{E}\ell(\xi)I(|\xi|\geq A)+\mathbb{E}\ell(\eta)I(|\eta|\geq A).$$

Proof. Clearly,

$$\begin{aligned}|\mathbb{E}\ell(\xi)-\mathbb{E}\ell(\eta)|\leq{}&|\mathbb{E}\ell(\xi)I(|\xi|<A)-\mathbb{E}\ell(\eta)I(|\eta|<A)|\\ &+\mathbb{E}\ell(\xi)I(|\xi|\geq A)+\mathbb{E}\ell(\eta)I(|\eta|\geq A).\end{aligned} \tag{4.14}$$

Denoting by F_ξ, F_η the distribution functions of ξ, η, assuming that A is a continuity point of both F_ξ and F_η and using integration by parts, we get

$$\begin{aligned}|\mathbb{E}\ell(\xi)I(|\xi|<A)-\mathbb{E}\ell(\eta)I(|\eta|<A)| &= \left|\int_{-A}^{A}\ell(x)\,d(F_\xi-F_\eta)(x)\right| \\ &= \left|\ell(A)(F_\xi-F_\eta)(A)-\ell(-A)(F_\xi-F_\eta)(-A)-\int_{-A}^{A}(F_\xi-F_\eta)(x)\ell'(x)\,dx\right|.\end{aligned}$$

Using the properties of ℓ (in particular, that ℓ is an even function and ℓ' is nonnegative and nondecreasing on $\mathbb{R}_+$), we get

$$\begin{aligned}|\mathbb{E}\ell(\xi)I(|\xi|<A)-\mathbb{E}\ell(\eta)I(|\eta|<A)| &\le 2\ell(A)\Delta(\xi;\eta)+2\int_0^A\ell'(u)\,du\,\Delta(\xi,\eta) \\ &= 4\ell(A)\Delta(\xi,\eta),\end{aligned}$$

which together with (4.14) implies the claim. If A is not a continuity point of F_ξ or F_η, one can easily obtain the result by a limiting argument. □

The following lemma is elementary.

Lemma 14. *Let ξ be a random variable such that for some $\tau > 0$ and for all $t \ge 1$, with probability at least $1-e^{-t}$,*

$$|\xi| \le \tau\sqrt{t}. \tag{4.15}$$

Let ℓ be a loss function satisfying Assumption 1. Then

$$\mathbb{E}\ell^2(\xi) \le 2e\sqrt{2\pi}\,c_1^2e^{2c_2^2\tau^2}. \tag{4.16}$$

We now apply Lemmas 13 and 14 to the r.v.'s

$$\xi := \xi(\Sigma) := \frac{\sqrt{n}(\langle f(\hat{\Sigma}),B\rangle-\langle f(\Sigma),B\rangle)}{\sigma_f(\Sigma;B)}$$

and $\eta := Z$. The bound (4.8) and Lemma 14 along with the fact that under the conditions of the theorem $\frac{(\mathbf{r}(\Sigma))^{s/2}}{n^{(s-1)/2}} \le \frac{r_n^{s/2}}{n^{(s-1)/2}} \le 1$ (for large enough n) imply that (4.15) and (4.16) hold with

$$\tau := \frac{\|f\|_{B^s_{\infty,1}}\|B\|_1\|\Sigma\|^s}{\sigma_f(\Sigma;B)} \vee \frac{2\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma;B)} \vee 1.$$

It follows from the bound of Lemma 13 that

$$\begin{aligned}|\mathbb{E}\ell(\xi)-\mathbb{E}\ell(Z)| &\le 4\ell(A)\Delta(\xi;Z)+\mathbb{E}^{1/2}\ell^2(\xi)\mathbb{P}^{1/2}\{|\xi|\ge A\} \\ &\quad+\mathbb{E}^{1/2}\ell^2(Z)\mathbb{P}^{1/2}\{|Z|\ge A\}.\end{aligned} \tag{4.17}$$

Using (4.16), standard bounds on $\mathbb{E}\ell^2(Z)$, $\mathbb{P}\{|Z| \geq A\}$ and the bound of Corollary 4, we get

$$|\mathbb{E}\ell(\xi) - \mathbb{E}\ell(Z)| \lesssim_s 4c_1^2 e^{2c_2A^2} \Delta_n^{(s)}(f; \Sigma; B) + \sqrt{2e}\,(2\pi)^{1/4} c_1 e^{c_2^2\tau^2} e^{-A^2/(2\tau^2)} + c_1 e^{c_2^2} e^{-A^2/4}.$$

To complete the proof of (1.9), it remains to take the supremum over the class of covariances $\mathcal{G}(r_n; a) \cap \{\Sigma : \sigma_f(\Sigma; B) \geq \sigma_0\}$ and over all the operators B with $\|B\|_1 \leq 1$, and to pass to the limit first as $n \to \infty$ and then as $A \to \infty$. □

5. Wishart operators, bootstrap chains, invariant functions and bias reduction

In what follows, we assume that $\mathbb{H}$ is a finite-dimensional inner product space of dimension d. Recall that $\mathcal{C}_+(\mathbb{H}) \subset \mathcal{B}_{sa}(\mathbb{H})$ denotes the cone of covariance operators in $\mathbb{H}$ and let $L_\infty(\mathcal{C}_+(\mathbb{H}))$ be the space of uniformly bounded Borel measurable functions on $\mathcal{C}_+(\mathbb{H})$ equipped with the uniform norm. Define an operator $\mathcal{T} : L_\infty(\mathcal{C}_+(\mathbb{H})) \to L_\infty(\mathcal{C}_+(\mathbb{H}))$:

$$\mathcal{T}g(\Sigma) = \mathbb{E}_\Sigma g(\hat{\Sigma}), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}), \tag{5.1}$$

where $\hat{\Sigma} = \hat{\Sigma}_n := n^{-1}\sum_{j=1}^n X_j \otimes X_j$ is the sample covariance operator based on i.i.d. observations $X_1, \ldots, X_n$ sampled from $N(0; \Sigma)$. Let $P(\Sigma; \cdot)$ denote the probability distribution of $\hat{\Sigma}$ in the space $\mathcal{C}_+(\mathbb{H})$ (equipped with its Borel σ-algebra $\mathfrak{B}(\mathcal{C}_+(\mathbb{H}))$). Note that $P(\Sigma; n^{-1}A)$ for $A \in \mathfrak{B}(\mathcal{C}_+(\mathbb{H}))$ is a Wishart distribution $\mathcal{W}_d(\Sigma; n)$. Clearly, P is a Markov kernel,

$$\mathcal{T}g(\Sigma) = \int_{\mathcal{C}_+(\mathbb{H})} g(V)P(\Sigma; dV), \qquad g \in L_\infty(\mathcal{C}_+(\mathbb{H})),$$

and $\mathcal{T}$ is a contraction: $\|\mathcal{T}g\|_{L_\infty} \leq \|g\|_{L_\infty}$.

Let $\hat{\Sigma}^0 := \Sigma$, $\hat{\Sigma}^{(1)} := \hat{\Sigma}$ and, more generally, given $\hat{\Sigma}^{(k)}$, define $\hat{\Sigma}^{(k+1)}$ as the sample covariance based on n i.i.d. observations $X_1^{(k)}, \ldots, X_n^{(k)}$ sampled from $N(0; \hat{\Sigma}^{(k)})$. Then $\hat{\Sigma}^{(k)}$, $k \geq 0$, is a homogeneous Markov chain with values in $\mathcal{C}_+(\mathbb{H})$, with $\hat{\Sigma}^{(0)} = \Sigma$ and with transition probability kernel P. The operator $\mathcal{T}^k$ can be represented as

$$\mathcal{T}^k g(\Sigma) = \mathbb{E}_\Sigma g(\hat{\Sigma}^{(k)}) = \int_{\mathcal{C}_+(\mathbb{H})} \cdots \int_{\mathcal{C}_+(\mathbb{H})} g(V_k)P(V_{k-1}; dV_k)P(V_{k-2}; dV_{k-1}) \ldots P(V_1; dV_2)P(\Sigma; dV_1)$$

for $\Sigma \in \mathcal{C}_+(\mathbb{H})$. We will be interested in the operator $\mathcal{B} = \mathcal{T} - \mathcal{I}$, which can be called the *bias operator* since $\mathcal{B}g(\Sigma)$ represents the bias of the plug-in estimator $g(\hat{\Sigma})$ of $g(\Sigma)$:

$$\mathcal{B}g(\Sigma) = \mathbb{E}_\Sigma g(\hat{\Sigma}) - g(\Sigma), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}).$$

Note that, by Newton's binomial formula, $\mathcal{B}^k g(\Sigma)$ can be represented as

$$\mathcal{B}^k g(\Sigma) = (\mathcal{T} - \mathcal{I})^k g(\Sigma) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mathcal{T}^j g(\Sigma)$$
$$= \mathbb{E}_\Sigma \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} g(\hat{\Sigma}^{(j)}), \tag{5.2}$$

which could be viewed as the expectation of the k-th order difference of g along the sample path of the Markov chain $\hat{\Sigma}^{(t)}$, $t = 0, 1, \ldots$.

Denote

$$g_k(\Sigma) := \sum_{j=0}^{k} (-1)^j \mathcal{B}^j g(\Sigma), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}). \tag{5.3}$$

Proposition 1. *The bias of the estimator $g_k(\hat{\Sigma})$ of $g(\Sigma)$ is*

$$\mathbb{E}_\Sigma g_k(\hat{\Sigma}) - g(\Sigma) = (-1)^k \mathcal{B}^{k+1} g(\Sigma).$$

Proof. Indeed,

$$\mathbb{E}_\Sigma g_k(\hat{\Sigma}) - g(\Sigma) = \mathcal{T} g_k(\Sigma) - g(\Sigma) = (\mathcal{I} + \mathcal{B}) g_k(\Sigma) - g(\Sigma)$$
$$= \sum_{j=0}^{k} (-1)^j \mathcal{B}^j g(\Sigma) - \sum_{j=1}^{k+1} (-1)^j \mathcal{B}^j g(\Sigma) - g(\Sigma) = (-1)^k \mathcal{B}^{k+1} g(\Sigma). \qquad \square$$

Let now $L_\infty(\mathcal{C}_+(\mathbb{H}); \mathcal{B}_{\rm sa}(\mathbb{H}))$ be the space of uniformly bounded Borel measurable functions $g : \mathcal{C}_+(\mathbb{H}) \to \mathcal{B}_{\rm sa}(\mathbb{H})$. We will need a version of the linear operator defined by (5.1) acting from $L_\infty(\mathcal{C}_+(\mathbb{H}); \mathcal{B}_{\rm sa}(\mathbb{H}))$ into itself. With a little abuse of notation, we still denote it by $\mathcal{T}$ and also set $\mathcal{B} := \mathcal{T} - \mathcal{I}$. These operators have all the properties stated above. This allows one to define by (5.3) an operator valued function g_k for which Proposition 1 still holds. In what follows, it should be clear from the context whether $\mathcal{T}$ and $\mathcal{B}$ act on real valued or on operator valued functions.

Given a smooth function f on the real line, we would like to find an estimator of $f(\Sigma)$ with a small bias. To this end, we consider an estimator $f_k(\hat{\Sigma})$ and, in view of Proposition 1, we need to show that, for a proper choice of k (depending on α such that $d = \dim(\mathbb{H}) \le n^\alpha$),

$$\|\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma)\| = \|\mathcal{B}^{k+1} f(\Sigma)\| = o(n^{-1/2}).$$

At the same time, we need to show that f_k satisfies certain smoothness properties such as Assumption 3. As a consequence, the (properly normalized) random variables $n^{1/2}(\langle f_k(\hat{\Sigma}), B\rangle - \mathbb{E}_\Sigma \langle f_k(\hat{\Sigma}), B\rangle)$ would be close in distribution to a standard normal r.v.. Since, in addition, the bias $\mathbb{E}_\Sigma \langle f_k(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle$ is of order $o(n^{-1/2})$, we would be able to conclude that $\langle f_k(\hat{\Sigma}), B\rangle$ is an asymptotically normal estimator of $\langle f(\Sigma), B\rangle$ with the classical convergence rate $n^{-1/2}$.

Our approach is based on representing the operator valued function $f_k(\Sigma)$ as $f_k(\Sigma) = \mathcal{D}g_k(\Sigma)$, where $g : \mathcal{C}_+(\mathbb{H}) \to \mathbb{R}$ is a real valued orthogonally invariant function and $\mathcal{D}$ is a differential operator defined below and called the *lifting operator*. This approach allows us to derive certain integral representations for functions $\mathcal{B}^k f(\Sigma) = \mathcal{D}\mathcal{B}^k g(\Sigma)$ that are then used to obtain proper bounds on $\mathcal{B}^k f(\Sigma)$ and to study smoothness properties of $\mathcal{B}^k f(\Sigma)$ and $f_k(\Sigma)$.

A function $g \in L_\infty(\mathcal{C}_+(\mathbb{H}))$ is *orthogonally invariant* iff for all orthogonal transformations U of $\mathbb{H}$, $g(U\Sigma U^{-1}) = g(\Sigma)$ for $\Sigma \in \mathcal{C}_+(\mathbb{H})$. Note that any such g could be represented as $g(\Sigma) = \varphi(\lambda_1(\Sigma), \dots, \lambda_d(\Sigma))$, where $\lambda_1(\Sigma) \geq \cdots \lambda_d(\Sigma)$ are the eigenvalues of Σ and φ is a symmetric function of d variables. A typical example is $g(\Sigma) = \mathrm{tr}(\psi(\Sigma))$ for a function ψ of a real variable. Let $L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ be the space of all orthogonally invariant functions from $L_\infty(\mathcal{C}_+(\mathbb{H}))$. Clearly, the orthogonally invariant functions form an algebra. We will need several facts concerning the properties of the operators $\mathcal{T}, \mathcal{B}$ as well as the lifting operator $\mathcal{D}$ on the space of orthogonally invariant functions. In the case of orthogonally invariant polynomials, similar properties can be found in the literature on Wishart distribution (see, e.g., [FK, LetMas]).

Proposition 2. *If $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$, then $\mathcal{T}g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ and $\mathcal{B}g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$.*

Proof. Indeed, the transformation $\Sigma \mapsto U\Sigma U^{-1}$ is a bijection of $\mathcal{C}_+(\mathbb{H})$,

$$\mathcal{T}g(U\Sigma U^{-1}) = \mathbb{E}_{U\Sigma U^{-1}} g(\hat{\Sigma}) = \mathbb{E}_\Sigma g(U\hat{\Sigma}U^{-1}) = \mathbb{E}_\Sigma g(\hat{\Sigma}) = \mathcal{T}g(\Sigma),$$

and the function Tg is uniformly bounded. □

An operator valued function $g : \mathcal{C}_+(\mathbb{H}) \to \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$ is called *orthogonally equivariant* if for all orthogonal transformations U, $g(U\Sigma U^{-1}) = Ug(\Sigma)U^{-1}$ for $\Sigma \in \mathcal{C}_+(\mathbb{H})$.

We say that $g : \mathcal{C}_+(\mathbb{H}) \to \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$ is *differentiable* (resp., *continuously differentiable*, *k times continuously differentiable*, etc.) on $\mathcal{C}_+(\mathbb{H})$ if there exists a uniformly bounded, Lipschitz (with respect to the operator norm) and differentiable (resp., continuously differentiable, k times continuously differentiable, etc.) extension of g to an open set G with $\mathcal{C}_+(\mathbb{H}) \subset G \subset \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$. Note that g could be further extended from G to a uniformly bounded Lipschitz (with respect to the operator norm) function on $\mathcal{B}_{\mathrm{sa}}(\mathbb{H})$, which will be still denoted by g.

Proposition 3. *If $g : \mathcal{C}_+(\mathbb{H}) \to \mathbb{R}$ is orthogonally invariant and continuously differentiable on $\mathcal{C}_+(\mathbb{H})$ with derivative Dg, then Dg is orthogonally equivariant.*

Proof. First suppose that Σ is positive definite. Then, given $H \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$, $\Sigma + tH$ is a covariance operator for all small enough t. Thus, for all $H \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$,

$$\begin{aligned}
&\langle Dg(U\Sigma U^{-1}), H\rangle = \lim_{t\to 0} \frac{g(U\Sigma U^{-1} + tH) - g(U\Sigma U^{-1})}{t} \\
&= \lim_{t\to 0} \frac{g(U(\Sigma + tU^{-1}HU)U^{-1}) - g(U\Sigma U^{-1})}{t} = \lim_{t\to 0} \frac{g(\Sigma + tU^{-1}HU) - g(\Sigma)}{t} \\
&= \langle Dg(\Sigma), U^{-1}HU\rangle = \langle UDg(\Sigma)U^{-1}, H\rangle,
\end{aligned}$$

implying

$$Dg(U\Sigma U^{-1}) = UDg(\Sigma)U^{-1}. \tag{5.4}$$

It remains to observe that the positive definite covariance operators are dense in $\mathcal{C}_+(\mathbb{H})$ and to extend (5.4) to $\mathcal{C}_+(\mathbb{H})$ by continuity. □

We now define the differential operator

$$\mathcal{D}g(\Sigma) := \Sigma^{1/2} Dg(\Sigma)\Sigma^{1/2}$$

acting on continuously differentiable functions on $\mathcal{C}_+(\mathbb{H})$. It will be called the *lifting operator*. We will show that the operators $\mathcal{T}$ and $\mathcal{D}$ commute (and, as a consequence, $\mathcal{B}$ and $\mathcal{D}$ also commute).

Proposition 4. *Suppose $d \lesssim n$. For all functions $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ that are continuously differentiable on $\mathcal{C}_+(\mathbb{H})$ with a uniformly bounded derivative Dg and for all $\Sigma \in \mathcal{C}_+(\mathbb{H})$,*

$$\mathcal{D}\mathcal{T}g(\Sigma) = \mathcal{T}\mathcal{D}g(\Sigma) \quad \text{and} \quad \mathcal{D}\mathcal{B}g(\Sigma) = \mathcal{B}\mathcal{D}g(\Sigma).$$

Proof. Note that $\hat{\Sigma} \stackrel{d}{=} \Sigma^{1/2} W \Sigma^{1/2}$, where W is the sample covariance based on i.i.d. standard normal random variables $Z_1, \dots, Z_n$ in $\mathbb{H}$ (which is a rescaled Wishart matrix). Let $\Sigma^{1/2}W^{1/2} = RU$ be the polar decomposition of $\Sigma^{1/2}W^{1/2}$ with positive semidefinite R and orthogonal U. Then

$$\hat{\Sigma} = \Sigma^{1/2} W \Sigma^{1/2} = \Sigma^{1/2} W^{1/2} W^{1/2} \Sigma^{1/2} = RUU^{-1}R = R^2$$

and

$$\begin{aligned} W^{1/2}\Sigma W^{1/2} &= W^{1/2}\Sigma^{1/2}\Sigma^{1/2}W^{1/2} = U^{-1}RRU = U^{-1}R^2U = U^{-1}\Sigma^{1/2}W\Sigma^{1/2}U \\ &= U^{-1}\hat{\Sigma}U. \end{aligned}$$

Since g is orthogonally invariant, we have

$$\mathcal{T}g(\Sigma) = \mathbb{E}_\Sigma g(\hat{\Sigma}) = \mathbb{E}g(\Sigma^{1/2}W\Sigma^{1/2}) = \mathbb{E}g(W^{1/2}\Sigma W^{1/2}), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}). \tag{5.5}$$

Since we extended g to a uniformly bounded function on $\mathcal{B}_{sa}(\mathbb{H})$, the right hand side of (5.5) is well defined for all $\Sigma \in \mathcal{B}_{sa}(\mathbb{H})$, and it will be used to extend $\mathcal{T}g(\Sigma)$ to $\mathcal{B}_{sa}(\mathbb{H})$. Moreover, since g is Lipschitz with respect to the operator norm, and for $d \lesssim n$, $\mathbb{E}\|W\| \le 1 + \mathbb{E}\|W - I\| \le 1 + C\sqrt{d/n} \lesssim 1$ (see (1.6)), it is easy to check that $\mathcal{T}g(\Sigma)$ is Lipschitz with respect to the operator norm on $\mathcal{B}_{sa}(\mathbb{H})$.

Let $H \in \mathcal{B}_{sa}(\mathbb{H})$ and $\Sigma_t := \Sigma + tH$ for $t > 0$. Note that

$$\begin{aligned} \frac{\mathcal{T}g(\Sigma_t) - \mathcal{T}g(\Sigma)}{t} &= \frac{\mathbb{E}g(W^{1/2}\Sigma_t W^{1/2}) - \mathbb{E}g(W^{1/2}\Sigma W^{1/2})}{t} \\ &= \mathbb{E}\frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t} I(\|W\| \le 1/\sqrt{t}) \\ &\quad + \mathbb{E}\frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t} I(\|W\| > 1/\sqrt{t}). \end{aligned} \tag{5.6}$$

Recall that g is continuously differentiable on the open set $G \supset \mathcal{C}_+(\mathbb{H})$. Also, $W^{1/2}\Sigma W^{1/2} \in \mathcal{C}_+(\mathbb{H}) \subset G$ and $W^{1/2}\Sigma_t W^{1/2} \in G$ for all small enough $t > 0$. The last fact follows from the bound $\|W^{1/2}(\Sigma_t - \Sigma)W^{1/2}\| \le \|W\| t\|H\| \le \sqrt{t}\,\|H\|$, which holds for all $t \le 1/\|W\|^2$ (or $\|W\| \le 1/\sqrt{t}$). Therefore, we easily get

$$\lim_{t\to 0} \frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t} I(\|W\| \le 1/\sqrt{t})$$
$$= \langle Dg(W^{1/2}\Sigma W^{1/2}), W^{1/2}HW^{1/2}\rangle = \langle W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}, H\rangle.$$

Also, since g is Lipschitz with respect to the operator norm,

$$\left|\frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t} I(\|W\| \le 1/\sqrt{t})\right|$$
$$\lesssim_g \frac{\|W^{1/2}(\Sigma_t - \Sigma)W^{1/2}\|}{t} \le \frac{\|W\|\,\|\Sigma_t - \Sigma\|}{t} \le \|W\|\,\|H\|.$$

Since $\mathbb{E}\|W\| \lesssim 1$, we can use Lebesgue's dominated convergence theorem to prove that

$$\lim_{t\to 0} \mathbb{E}\frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t} I(\|W\| \le 1/\sqrt{t})$$
$$= \mathbb{E}\langle W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}, H\rangle = \langle \mathbb{E}W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}, H\rangle. \tag{5.7}$$

On the other hand, since g is uniformly bounded, we can use the bound (1.6) to prove that for some constant $C > 0$ and for all $t \le 1/C^2$,

$$\mathbb{E}\left|\frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t} I(\|W\| > 1/\sqrt{t})\right|$$
$$\lesssim_g \frac{1}{t}\mathbb{P}\{\|W\| \ge 1/\sqrt{t}\} \le \frac{1}{t}\exp\left\{-\frac{n}{C\sqrt{t}}\right\} \to 0 \quad \text{as } t \to 0. \tag{5.8}$$

It follows from (5.6)–(5.8) that

$$\langle D\mathcal{T}g(\Sigma), H\rangle = \langle \mathbb{E}W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}, H\rangle.$$

It is also easy to check that $\mathbb{E}W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}$ is a continuous function on G, implying that $\mathcal{T}g$ is continuously differentiable on G with Fréchet derivative

$$D\mathcal{T}g(\Sigma) = \mathbb{E}W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}.$$

Since $W^{1/2}\Sigma W^{1/2} = U^{-1}\hat{\Sigma}U$ and Dg is an orthogonally equivariant function (see Proposition 3), we get $Dg(W^{1/2}\Sigma W^{1/2}) = U^{-1}Dg(\hat{\Sigma})U$. Therefore,

$$\begin{aligned}
&\mathcal{D}\mathcal{T}g(\Sigma)\\
&= \Sigma^{1/2}D\mathcal{T}g(\Sigma)\Sigma^{1/2} = \Sigma^{1/2}\mathbb{E}(W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2})\Sigma^{1/2}\\
&= \mathbb{E}(\Sigma^{1/2}W^{1/2}Dg(W^{1/2}\Sigma W^{1/2})W^{1/2}\Sigma^{1/2}) = \mathbb{E}(\Sigma^{1/2}W^{1/2}U^{-1}Dg(\hat{\Sigma})UW^{1/2}\Sigma^{1/2})\\
&= \mathbb{E}(RUU^{-1}Dg(\hat{\Sigma})UU^{-1}R) = \mathbb{E}(RDg(\hat{\Sigma})R) = \mathbb{E}_\Sigma(\hat{\Sigma}^{1/2}Dg(\hat{\Sigma})\hat{\Sigma}^{1/2})\\
&= \mathbb{E}_\Sigma\mathcal{D}g(\hat{\Sigma}) = \mathcal{T}\mathcal{D}g(\Sigma).
\end{aligned}$$

A similar relationship for $\mathcal{B}$ and $\mathcal{D}$ follows easily. □

We will now derive useful representations of the operators $\mathcal{T}^k$ and $\mathcal{B}^k$ and prove that they also commute with the differential operator $\mathcal{D}$.

Proposition 5. *Suppose $d \lesssim n$. Let $W_1, \dots, W_k, \dots$ be i.i.d. copies of W.*[5] *Then, for all $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ and for all $k \geq 1$,*

$$\mathcal{T}^k g(\Sigma) = \mathbb{E} g(W_k^{1/2} \dots W_1^{1/2} \Sigma W_1^{1/2} \dots W_k^{1/2}) \tag{5.9}$$

and

$$\mathcal{B}^k g(\Sigma) = \mathbb{E} \sum_{I \subset \{1,\dots,k\}} (-1)^{k-|I|} g(A_I^* \Sigma A_I), \tag{5.10}$$

where $A_I := \prod_{i \in I} W_i^{1/2}$. Suppose, in addition, that g is continuously differentiable on $\mathcal{C}_+(\mathbb{H})$ with a uniformly bounded derivative Dg. Then

$$D\mathcal{B}^k g(\Sigma) = \mathbb{E} \sum_{I \subset \{1,\dots,k\}} (-1)^{k-|I|} A_I Dg(A_I^* \Sigma A_I) A_I^*, \tag{5.11}$$

and, for all $\Sigma \in \mathcal{C}_+(\mathbb{H})$,

$$\mathcal{D}\mathcal{T}^k g(\Sigma) = \mathcal{T}^k \mathcal{D} g(\Sigma) \quad \text{and} \quad \mathcal{D}\mathcal{B}^k g(\Sigma) = \mathcal{B}^k \mathcal{D} g(\Sigma). \tag{5.12}$$

Finally,

$$\mathcal{B}^k \mathcal{D} g(\Sigma) = \mathcal{D}\mathcal{B}^k g(\Sigma) = \mathbb{E}\Big(\sum_{I \subset \{1,\dots,k\}} (-1)^{k-|I|} \Sigma^{1/2} A_I Dg(A_I^* \Sigma A_I) A_I^* \Sigma^{1/2} \Big). \tag{5.13}$$

Proof. Since $\hat{\Sigma} \overset{d}{=} \Sigma^{1/2} W \Sigma^{1/2}$, $W^{1/2} \Sigma W^{1/2} = U^{-1} \Sigma^{1/2} W \Sigma^{1/2} U$, where U is an orthogonal operator, and g is orthogonally invariant, we have

$$\mathcal{T} g(\Sigma) = \mathbb{E}_\Sigma g(\hat{\Sigma}) = \mathbb{E} g(W^{1/2} \Sigma W^{1/2}) \tag{5.14}$$

(which has already been used in the proof of Proposition 4).

By Proposition 2, orthogonal invariance of g implies the same property of $\mathcal{T} g$ and, by induction, of $\mathcal{T}^k g$ for all $k \geq 1$. Then, also by induction, it follows from (5.14) that

$$\mathcal{T}^k g(\Sigma) = \mathbb{E} g(W_k^{1/2} \dots W_1^{1/2} \Sigma W_1^{1/2} \dots W_k^{1/2}).$$

If $I \subset \{1, \dots, k\}$ with $|I| = \mathrm{card}(I) = j$ and $A_I = \prod_{i \in I} W_i^{1/2}$, this clearly implies that

$$\mathcal{T}^j g(\Sigma) = \mathbb{E} g(A_I^* \Sigma A_I).$$

In view of (5.2), we easily see that (5.10) holds. If g is continuously differentiable on $\mathcal{C}_+(\mathbb{H})$ with a uniformly bounded derivative Dg, it follows from (5.10) that $\mathcal{B}^k g(\Sigma)$ is continuously differentiable on $\mathcal{C}_+(\mathbb{H})$ with Fréchet derivative given by (5.11). To prove this, it is enough to justify differentiation under the expectation sign, which is done exactly

[5] Recall that W is the sample covariance based on i.i.d. standard normal random variables $Z_1, \dots, Z_n$ in $\mathbb{H}$.

as in the proof of Proposition 4. Finally, it follows from (5.11) that the derivatives $D\mathcal{B}^k g$, $k \geq 1$, are uniformly bounded in $\mathcal{C}_+(\mathbb{H})$. Similarly, as a consequence of (5.9) and the properties of g, $\mathcal{T}^k g(\Sigma)$ is continuously differentiable on $\mathcal{C}_+(\mathbb{H})$ with uniformly bounded derivative $D\mathcal{T}^k g$ for all $k \geq 1$. Therefore, (5.12) follows from Proposition 4 by induction. Formula (5.13) follows from (5.12) and (5.11). □

Define the following functions providing the linear interpolation between the identity operator I and the operators $W_1^{1/2}, \dots, W_k^{1/2}$:

$$V_j(t_j) := I + t_j(W_j^{1/2} - I), \quad t_j \in [0,1],\ 1 \leq j \leq k.$$

Clearly, $V_j(t_j) \in \mathcal{C}_+(\mathbb{H})$ for all $j = 1, \dots, k$ and $t_j \in [0,1]$. Let

$$R = R(t_1, \dots, t_k) = V_1(t_1) \dots V_k(t_k), \quad L = L(t_1, \dots, t_k) = V_k(t_k) \dots V_1(t_1) = R^*.$$

Define

$$S = S(t_1, \dots, t_k) = L(t_1, \dots, t_k)\Sigma R(t_1, \dots, t_k), \quad (t_1, \dots, t_k) \in [0,1]^k.$$

Finally, let

$$\varphi(t_1, \dots, t_k)$$
$$:= \Sigma^{1/2} R(t_1, \dots, t_k) Dg(S(t_1, \dots, t_k)) L(t_1, \dots, t_k)\Sigma^{1/2}, \quad (t_1, \dots, t_k) \in [0,1]^k.$$

The following representation will play a crucial role in our further analysis.

Proposition 6. *Suppose $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+1$ times continuously differentiable with uniformly bounded derivatives $D^j g$, $j = 1, \dots, k+1$. Then the function φ is k times continuously differentiable on $[0,1]^k$ and*

$$\mathcal{B}^k \mathcal{D} g(\Sigma) = \mathbb{E} \int_0^1 \cdots \int_0^1 \frac{\partial^k \varphi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k, \quad \Sigma \in \mathcal{C}_+(\mathbb{H}). \tag{5.15}$$

Proof. Given a function $\phi : [0,1]^k \to \mathbb{R}$, define for $1 \leq i \leq k$ finite difference operators

$$\mathfrak{D}_i \phi(t_1, \dots, t_k) := \phi(t_1, \dots, t_{i-1}, 1, t_{i+1}, \dots, t_k) - \phi(t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_k)$$

(with obvious modifications for $i = 1, k$). Then $\mathfrak{D}_1 \dots \mathfrak{D}_k \phi$ does not depend on $t_1, \dots, t_k$ and is given by the formula

$$\mathfrak{D}_1 \dots \mathfrak{D}_k \phi = \sum_{(t_1, \dots, t_k) \in \{0,1\}^k} (-1)^{k-(t_1+\cdots+t_k)} \phi(t_1, \dots, t_k). \tag{5.16}$$

It is well known and easy to check that if ϕ is k times continuously differentiable on $[0,1]^k$, then

$$\mathfrak{D}_1 \dots \mathfrak{D}_k \phi = \int_0^1 \cdots \int_0^1 \frac{\partial^k \phi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k. \tag{5.17}$$

Similar definitions and formula (5.17) also hold for vector valued and operator valued functions ϕ.

It immediately follows from (5.13) and (5.16) that

$$\mathcal{B}^k \mathcal{D}g(\Sigma) = \mathbb{E}\mathfrak{D}_1 \dots \mathfrak{D}_k \varphi. \tag{5.18}$$

Since Dg is k times continuously differentiable and the functions $S(t_1, \dots, t_k)$ and $R(t_1, \dots, t_k)$ are polynomials with respect to $t_1, \dots, t_k$, the function φ is k times continuously differentiable on $[0, 1]^k$. The representation (5.15) follows from (5.18) and (5.17). □

6. Bounds on the iterated bias operator

Our goal in this section is to prove the following bound on the iterated bias operator $\mathcal{B}^k \mathcal{D}g(\Sigma)$.

Theorem 8. *Suppose $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+1$ times continuously differentiable with uniformly bounded derivatives $D^j g$, $j = 1, \dots, k+1$. Suppose also that $d \le n$ and $k \le n$. Then for some constant $C > 0$ and for all $\Sigma \in \mathcal{C}_+(\mathbb{H})$,*

$$\|\mathcal{B}^k \mathcal{D}g(\Sigma)\| \le C^{k^2} \max_{1 \le j \le k+1} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+1} \vee \|\Sigma\|) \left(\frac{d}{n} \vee \frac{k}{n}\right)^{k/2}. \tag{6.1}$$

It follows from the commutation relation (5.12) that

$$\mathcal{D}g_k(\Sigma) = (\mathcal{D}g)_k(\Sigma), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}),$$

where g_k is defined by (5.3) and

$$(\mathcal{D}g)_k(\Sigma) := \sum_{j=0}^{k} (-1)^j \mathcal{B}^j \mathcal{D}g(\Sigma), \qquad \Sigma \in \mathcal{C}_+(\mathbb{H}).$$

Clearly, we have (see Proposition 1)

$$\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma) = (-1)^k \mathcal{B}^{k+1} \mathcal{D}g(\Sigma). \tag{6.2}$$

The bound (6.1) is needed, in particular, to control the bias of the estimator $\mathcal{D}g_k(\hat{\Sigma})$ of $\mathcal{D}g(\Sigma)$. Namely, we have the following corollary.

Corollary 5. *Suppose that $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+2$ times continuously differentiable with uniformly bounded derivatives $D^j g$, $j = 1, \dots, k+2$, and also $d \le n$, $k+1 \le n$. Then*

$$\begin{aligned} &\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\| \\ &\quad \le C^{(k+1)^2} \max_{1 \le j \le k+2} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+2} \vee \|\Sigma\|) \left(\frac{d}{n} \vee \frac{k+1}{n}\right)^{(k+1)/2}. \end{aligned} \tag{6.3}$$

If, in addition, $k+1 \le d \le n$ and, for some $\delta > 0$,

$$k \ge \frac{\log d}{\log(n/d)} + \delta\left(1 + \frac{\log d}{\log(n/d)}\right), \tag{6.4}$$

then

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\| \le C^{(k+1)^2} \max_{1 \le j \le k+2} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+2} \vee \|\Sigma\|) n^{-(1+\delta)/2}. \tag{6.5}$$

The proof of this corollary immediately follows from (6.2) and (6.1). If $d = n^{\alpha}$ for some $\alpha \in (0,1)$, condition (6.4) becomes $k \geq \frac{\alpha+\delta}{1-\alpha}$. Thus, if

$$k(\alpha,\delta) := \min\Big\{k \geq \frac{\alpha+\delta}{1-\alpha}\Big\},$$

then the bound (6.5) holds with $k = k(\alpha,\delta)$.

Remark 8. In Section 7, we will obtain a sharper bound on the bias of the estimator $\mathcal{D}g_k(\hat{\Sigma})$ (under stronger smoothness assumptions, see Corollary 7).

The first step towards the proof of Theorem 8 is to compute the partial derivative $\frac{\partial^k \varphi}{\partial t_1 \dots \partial t_k}$, which will allow us to use representation (5.15). To this end, we first derive formulas for partial derivatives of the operator valued function $h(S(t_1,\dots,t_k))$, where $h = Dg$. To simplify the notations, given $T = \{t_{i_1},\dots,t_{i_m}\} \subset \{t_1,\dots,t_k\}$, we will write $\partial_T S$ instead of $\frac{\partial^m S(t_1,\dots,t_k)}{\partial t_{i_1}\dots\partial t_{i_m}}$ (similarly, we use the notation $\partial_T h(S)$ for a partial derivative of a function $h(S)$).

Let $\mathcal{D}_{j,T}$ be the set of all partitions $(\Delta_1,\dots,\Delta_j)$ of $T \subset \{t_1,\dots,t_k\}$ with nonempty sets $\Delta_i, i = 1,\dots,j$ (partitions with different order of $\Delta_1,\dots,\Delta_j$ are considered identical). For $\Delta = (\Delta_1,\dots,\Delta_j) \in \mathcal{D}_{j,T}$, set $\partial_\Delta S = (\partial_{\Delta_1}S,\dots,\partial_{\Delta_j}S)$. Denote $\mathcal{D}_T := \bigcup_{j=1}^{|T|}\mathcal{D}_{j,T}$. For $\Delta = (\Delta_1,\dots,\Delta_j) \in \mathcal{D}_T$, set $j_\Delta := j$.

Lemma 15. *Suppose, for some $m \leq k$, $h = Dg \in L_\infty(\mathcal{C}_+(\mathbb{H});\mathcal{B}_{sa}(\mathbb{H}))$ is m times continuously differentiable with derivatives $D^j h$, $j \leq m$.*[6] *Then the function $[0,1]^k \ni (t_1,\dots,t_k) \mapsto h(S(t_1,\dots,t_k))$ is m times continuously differentiable and for any $T \subset \{t_1,\dots,t_k\}$ with $|T| = m$,*

$$\partial_T h(S) = \sum_{\Delta\in\mathcal{D}_T} D^{j_\Delta}h(S)(\partial_\Delta S) = \sum_{j=1}^{m}\sum_{\Delta\in\mathcal{D}_{j,T}} D^j h(S)(\partial_\Delta S). \tag{6.6}$$

Proof. Since $[0,1]^k \ni (t_1,\dots,t_k) \mapsto S(t_1,\dots,t_k)$ is an operator valued polynomial and h is m times continuously differentiable, the function $[0,1]^k \ni (t_1,\dots,t_k) \mapsto h(S(t_1,\dots,t_k))$ is also m times continuously differentiable. We will now prove (6.6) by induction on m. For $m = 1$, it reduces to

$$\partial_{\{t_i\}}h(S) = \frac{\partial h(S)}{\partial t_i} = Dh(S)\Big(\frac{\partial S}{\partial t_i}\Big),$$

which is true by the chain rule. Assume that (6.6) holds for some $m < k$ and for any $T \subset \{t_1,\dots,t_k\}$ with $|T| = m$. Let $T' = T \cup \{t_l\}$ for some $t_l \notin T$. Then

$$\partial_{T'}h(S) = \partial_{\{t_l\}}\partial_T h(S) = \sum_{j=1}^{m}\sum_{\Delta\in\mathcal{D}_{j,T}} \partial_{\{t_l\}}D^j h(S)(\partial_\Delta S). \tag{6.7}$$

[6] Recall that $D^j h$ is an operator valued symmetric j-linear form on the space $\mathcal{B}_{sa}(\mathbb{H})$.

Given $\Delta = (\Delta_1, \ldots, \Delta_j) \in \mathcal{D}_{j,T}$, define partitions $\Delta^{(i)} \in \mathcal{D}_{j,T'}$, $i = 1, \ldots, j$, as follows:

$$\Delta^{(1)} := (\Delta_1 \cup \{t_l\}, \Delta_2, \ldots, \Delta_j), \qquad \Delta^{(2)} := (\Delta_1, \Delta_2 \cup \{t_l\}, \ldots, \Delta_j), \ \ldots,$$
$$\Delta^{(j)} := (\Delta_1, \ldots, \Delta_{j-1}, \Delta_j \cup \{t_l\}).$$

Also define a partition $\tilde{\Delta} \in \mathcal{D}_{j+1,T'}$ by $\tilde{\Delta} := (\Delta_1, \ldots, \Delta_j, \{t_l\})$. It is easy to see that any partition $\Delta' \in \mathcal{D}_{T'}$ is the image of a unique partition $\Delta \in \mathcal{D}_T$ under one of the transformations $\Delta \mapsto \Delta^{(i)}$, $i = 1, \ldots, j_\Delta$, and $\Delta \mapsto \tilde{\Delta}$. This implies that

$$\mathcal{D}_{T'} = \bigcup_{\Delta \in \mathcal{D}_T} \{\Delta^{(1)}, \ldots, \Delta^{(j_\Delta)}, \tilde{\Delta}\}.$$

It easily follows from the chain rule and the product rule that

$$\partial_{\{t_l\}} D^j h(S)(\partial_\Delta S) = \sum_{i=1}^{j} D^j h(S)(\partial_{\Delta^{(i)}} S) + D^{j+1} h(S)(\partial_{\tilde{\Delta}} S).$$

Substituting this into (6.7) easily yields

$$\partial_{T'} h(S) = \sum_{j=1}^{m+1} \sum_{\Delta \in \mathcal{D}_{j,T'}} D^j h(S)(\partial_\Delta S). \qquad \square$$

Next we derive upper bounds on $\|\partial_T S\|$, $\|\partial_T R\|$ and $\|\partial_T L\|$ for $T \subset \{t_1, \ldots, t_k\}$. Denote $\delta_i := \|W_i - I\|$, $i = 1, \ldots, k$.

Lemma 16. *For all* $T \subset \{t_1, \ldots, t_k\}$,

$$\|\partial_T R\| \le \prod_{t_i \in T} \frac{\delta_i}{1 + \delta_i} \prod_{i=1}^{k} (1 + \delta_i), \tag{6.8}$$

$$\|\partial_T L\| \le \prod_{t_i \in T} \frac{\delta_i}{1 + \delta_i} \prod_{i=1}^{k} (1 + \delta_i), \tag{6.9}$$

$$\|\partial_T S\| \le 2^k \|\Sigma\| \prod_{t_i \in T} \frac{\delta_i}{1 + \delta_i} \prod_{i=1}^{k} (1 + \delta_i)^2. \tag{6.10}$$

Remark 9. The bounds of the lemma hold for $T = \emptyset$ with an obvious convention that in this case $\prod_{t_i \in T} a_i = 1$.

Proof of Lemma 16. Observe that $\frac{\partial}{\partial t_i} V_i(t_i) = W_i^{1/2} - I$. Let $B_i^0 := V_i(t_i)$ and $B_i^1 := W_i^{1/2} - I$. For $R = V_1(t_1) \ldots V_k(t_k)$, we have $\partial_T R = \prod_{i=1}^{k} B_i^{I_T(t_i)}$ and

$$\|\partial_T R\| \le \prod_{t_i \in T} \|W_i^{1/2} - I\| \prod_{t_i \notin T} \|V_i(t_i)\|.$$

Due to the elementary inequality $|\sqrt{x}-1| \le |x-1|, x \ge 0$, we have $\|W_i^{1/2} - I\| \le \|W_i - I\| = \delta_i$ and $\|V_i(t_i)\| \le 1 + \|W_i^{1/2} - I\| \le 1 + \|W_i - I\| = 1 + \delta_i$. Therefore,

$$\|\partial_T R\| \le \prod_{t_i \in T} \delta_i \prod_{t_i \notin T} (1+\delta_i) = \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i),$$

which proves (6.8). Similarly, we have (6.9).

Note that, by the product rule,

$$\partial_T S = \partial_T (L \Sigma R) = \sum_{T' \subset T} (\partial_{T'} L) \Sigma (\partial_{T \setminus T'} R).$$

Therefore,

$$\|\partial_T S\| \le \|\Sigma\| \sum_{T' \subset T} \|\partial_{T'} L\| \|\partial_{T\setminus T'} R\|$$
$$\le \|\Sigma\| \sum_{T' \subset T} \prod_{t_i \in T'} \frac{\delta_i}{1+\delta_i} \prod_{t_i \in T \setminus T'} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^2 = 2^k \|\Sigma\| \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^2,$$

proving (6.10). □

Lemma 17. *Suppose that, for some $0 \le m \le k$, $h = Dg \in L_\infty(\mathcal{C}_+(\mathbb{H}); \mathcal{B}_{sa}(\mathbb{H}))$ is m times differentiable with uniformly bounded continuous derivatives $D^j h$, $j = 1, \dots, m$. Then for all $T \subset \{t_1, \dots, t_k\}$ with $|T| = m$,*

$$\|\partial_T h(S)\| \le 2^{m(k+m+1)} \max_{0 \le j \le m} \|D^j h\|_{L_\infty} (\|\Sigma\|^m \vee 1) \prod_{i=1}^k (1+\delta_i)^{2m} \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i}. \quad (6.11)$$

Proof. Assume that $m \ge 1$ (for $m = 0$, the bound of the lemma is trivial). Let $\Delta = (\Delta_1, \dots, \Delta_j) \in \mathcal{D}_{j,T}$, $j \le m$. Note that

$$\|D^j h(S)(\partial_{\Delta_1} S, \dots, \partial_{\Delta_j} S)\| \le \|D^j h(S)\| \, \|\partial_{\Delta_1} S\| \dots \|\partial_{\Delta_j} S\|$$
$$\le \|D^j h(S)\| 2^{kj} \|\Sigma\|^j \prod_{l=1}^j \prod_{t_i \in \Delta_l} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^{2j}$$
$$= \|D^j h(S)\| 2^{kj} \|\Sigma\|^j \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^{2j}.$$

Using Lemma 15, we get

$$\|\partial_T h(S)\| \le \sum_{j=1}^m \sum_{\Delta \in \mathcal{D}_{j,T}} \|D^j h(S)(\partial_\Delta S)\|$$
$$\le \sum_{j=1}^m \mathrm{card}(\mathcal{D}_{j,T}) \|D^j h(S)\| 2^{kj} \|\Sigma\|^j \prod_{i=1}^k (1+\delta_i)^{2j} \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i}.$$

Note that the number of all functions on T with values in $\{1,\dots,j\}$ is equal to j^m, and clearly $\mathrm{card}(\mathcal{D}_{j,T}) \le j^m$. Therefore,

$$\begin{aligned}\|\partial_T h(S)\| &\le \sum_{j=1}^{m} j^m \|D^j h(S)\| 2^{kj}\|\Sigma\|^j \prod_{i=1}^{k}(1+\delta_i)^{2j}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\\ &\le m^{m+1}2^{km}\max_{1\le j\le m}\|D^j h\|_{L_\infty}(\|\Sigma\|\vee\|\Sigma\|^m)\prod_{i=1}^{k}(1+\delta_i)^{2m}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\\ &\le 2^{m(k+m+1)}\max_{1\le j\le m}\|D^j h\|_{L_\infty}(\|\Sigma\|\vee\|\Sigma\|^m)\prod_{i=1}^{k}(1+\delta_i)^{2m}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i},\end{aligned} \tag{6.12}$$

which easily implies (6.11). □

Next we bound partial derivatives of the function $\Sigma^{1/2}Lh(S)R\Sigma^{1/2}$ (with $S = S(t_1,\dots,t_k)$, $L = L(t_1,\dots,t_k)$, $R = R(t_1,\dots,t_k)$ and $h = Dg$).

Lemma 18. *Assume that $d \le n$ and $k \le n$. Suppose $h = Dg \in L_\infty(\mathcal{C}_+(\mathbb{H}); \mathcal{B}_{sa}(\mathbb{H}))$ is k times differentiable with uniformly bounded continuous derivatives $D^j h$, $j = 1,\dots,k$. Then*

$$\begin{aligned}&\|\partial_{\{t_1,\dots,t_k\}}\Sigma^{1/2}Rh(S)L\Sigma^{1/2}\|\\ &\qquad\le 3^k 2^{k(2k+1)}\max_{0\le j\le k}\|D^j h\|_{L_\infty}(\|\Sigma\|^{k+1}\vee\|\Sigma\|)\prod_{i=1}^{k}(1+\delta_i)^{2k+1}\delta_i.\end{aligned} \tag{6.13}$$

Proof. Note that

$$\partial_{\{t_1,\dots,t_k\}}\Sigma^{1/2}Rh(S)L\Sigma^{1/2} = \sum_{T_1,T_2,T_3}\big(\Sigma^{1/2}(\partial_{T_1}R)(\partial_{T_2}h(S))(\partial_{T_3}L)\Sigma^{1/2}\big), \tag{6.14}$$

where the sum is over all the partitions of the set $\{t_1,\dots,t_k\}$ into disjoint subsets T_1, T_2, T_3. The number of such partitions is equal to 3^k. We have

$$\|\Sigma^{1/2}(\partial_{T_1}R)(\partial_{T_2}h(S))(\partial_{T_3}L)\Sigma^{1/2}\| \le \|\Sigma\|\,\|\partial_{T_1}L\|\,\|\partial_{T_2}h(S)\|\,\|\partial_{T_3}R\|. \tag{6.15}$$

Assume $|T_1| = m_1$, $|T_2| = m_2$, $|T_3| = m_3$. It follows from Lemma 17 that

$$\|\partial_{T_2}h(S)\| \le 2^{m_2(k+m_2+1)}\max_{0\le j\le m_2}\|D^j h\|_{L_\infty}(\|\Sigma\|^{m_2}\vee 1)\prod_{i=1}^{k}(1+\delta_i)^{2m_2}\prod_{t_i\in T_2}\frac{\delta_i}{1+\delta_i}.$$

On the other hand, by (6.8) and (6.9), we have

$$\|\partial_{T_3}R\| \le \prod_{t_i\in T_3}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i), \qquad \|\partial_{T_1}L\| \le \prod_{t_i\in T_1}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i).$$

It follows from these bounds and (6.15) that

$$
\begin{aligned}
&\|\Sigma^{1/2}(\partial_{T_1}R)(\partial_{T_2}h(S))(\partial_{T_3}L)\Sigma^{1/2}\| \\
&\le \|\Sigma\|2^{m_2(k+m_2+1)}\max_{0\le j\le m_2}\|D^j h\|_{L_\infty}(\|\Sigma\|^{m_2}\vee 1)\prod_{i=1}^k(1+\delta_i)^{2m_2+2}\prod_{t_i\in T_1\cup T_2\cup T_3}\frac{\delta_i}{1+\delta_i} \\
&\le 2^{k(2k+1)}\max_{0\le j\le k}\|D^j h\|_{L_\infty}(\|\Sigma\|^{k+1}\vee\|\Sigma\|)\prod_{i=1}^k(1+\delta_i)^{2k+2}\prod_{i=1}^k\frac{\delta_i}{1+\delta_i} \\
&= 2^{k(2k+1)}\max_{0\le j\le k}\|D^j h\|_{L_\infty}(\|\Sigma\|^{k+1}\vee\|\Sigma\|)\prod_{i=1}^k(1+\delta_i)^{2k+1}\delta_i.
\end{aligned}
$$

Since the number of terms in the sum on the right hand side of (6.14) is equal to 3^k, we easily see that (6.13) holds. □

Proof of Theorem 8. We use the representation (5.15) to get

$$
\|\mathcal{B}^k\mathcal{D}g(\Sigma)\| \le \int_0^1\cdots\int_0^1 \mathbb{E}\|\partial_{\{t_1,\dots,t_k\}}\Sigma^{1/2}Rh(S)L\Sigma^{1/2}\|\,dt_1\dots dt_k. \tag{6.16}
$$

Using the bounds (6.13) yields

$$
\|\mathcal{B}^k\mathcal{D}g(\Sigma)\| \le 3^k 2^{k(2k+1)}\max_{0\le j\le k}\|D^j h\|_{L_\infty}(\|\Sigma\|^{k+1}\vee\|\Sigma\|)\mathbb{E}\prod_{i=1}^k(1+\delta_i)^{2k+1}\delta_i. \tag{6.17}
$$

Note that

$$
\mathbb{E}\prod_{i=1}^k(1+\delta_i)^{2k+1}\delta_i = \prod_{i=1}^k\mathbb{E}(1+\delta_i)^{2k+1}\delta_i = \bigl(\mathbb{E}(1+\|W-I\|)^{2k+1}\|W-I\|\bigr)^k
$$

and

$$
\begin{aligned}
\mathbb{E}(1+\|W-I\|)^{2k+1}\|W-I\| &= 2^{2k+1}\mathbb{E}\left(\frac{1+\|W-I\|}{2}\right)^{2k+1}\|W-I\| \\
&\le 2^{2k+1}\mathbb{E}\frac{1+\|W-I\|^{2k+1}}{2}\|W-I\| = 2^{2k}\bigl(\mathbb{E}\|W-I\|+\mathbb{E}\|W-I\|^{2k+2}\bigr).
\end{aligned}
$$

Using the bound (1.5), we find that with some constant $C_1\ge 1$,

$$
\mathbb{E}\|W-I\| \le \mathbb{E}^{1/(2k+2)}\|W-I\|^{2k+2} \le C_1\left(\sqrt{\frac{d}{n}}\vee\sqrt{\frac{k}{n}}\right),
$$

which implies that

$$
\begin{aligned}
\mathbb{E}(1+\|W-I\|)^{2k+1}\|W-I\| &\le 2^{2k}\left[C_1\left(\sqrt{\frac{d}{n}}\vee\sqrt{\frac{k}{n}}\right)+C_1^{2k+2}\left(\frac{d}{n}\vee\frac{k}{n}\right)^{k+1}\right] \\
&\le 2^{2k}C_1^{2k+2}\left(\sqrt{\frac{d}{n}}\vee\sqrt{\frac{k}{n}}\right)
\end{aligned}
$$

and

$$\mathbb{E}\prod_{i=1}^{k}(1+\delta_i)^{2k+1}\delta_i \le 2^{2k^2}C_1^{2k^2+2k}\left(\frac{d}{n}\vee\frac{k}{n}\right)^{k/2}. \tag{6.18}$$

We substitute this bound into (6.17) to get

$$\|\mathcal{B}^k\mathcal{D}g(\Sigma)\| \le 3^k 2^{4k^2+k}C_1^{2k^2+2k}\max_{0\le j\le k}\|D^j h\|_{L_\infty}(\|\Sigma\|^{k+1}\vee\|\Sigma\|)\left(\frac{d}{n}\vee\frac{k}{n}\right)^{k/2}, \tag{6.19}$$

which implies the result. □

7. Smoothness properties of $\mathcal{D}g_k(\Sigma)$

Our goal in this section is to show that, for a smooth orthogonally invariant function g, the function $\mathcal{D}g_k(\Sigma)$ satisfies Assumption 3. This result will be used in the next section to prove normal approximation bounds for $\mathcal{D}g_k(\hat{\Sigma})$. We will assume in what follows that g is defined and properly smooth *on the whole space* $\mathcal{B}_{sa}(\mathbb{H})$ of self-adjoint operators in $\mathbb{H}$.

Recall that, by (5.15),

$$\mathcal{B}^k\mathcal{D}g(\Sigma) = \mathbb{E}\int_0^1\cdots\int_0^1 \frac{\partial^k\varphi(t_1,\dots,t_k)}{\partial t_1\dots\partial t_k}\,dt_1\dots dt_k,$$

where

$$\varphi(t_1,\dots,t_k) := \Sigma^{1/2}R(t_1,\dots,t_k)Dg(S(t_1,\dots,t_k))L(t_1,\dots,t_k)\Sigma^{1/2}$$

for $(t_1,\dots,t_k)\in[0,1]^k$.

Let $\delta\in(0,1/2)$ and let $\gamma:\mathbb{R}\to\mathbb{R}$ be a C^∞ function such that

$$0\le\gamma(u)\le\sqrt{u},\ u\ge 0,\qquad \gamma(u)=\sqrt{u},\ u\in[\delta,1/\delta],$$
$$\operatorname{supp}(\gamma)\subset[\delta/2,2/\delta],\qquad \|\gamma\|_{B^1_{\infty,1}}\lesssim\frac{\log(2/\delta)}{\sqrt{\delta}}.$$

For instance, one can take $\gamma(u) := \lambda(u/\delta)\sqrt{u}(1-\lambda(\delta u/2))$, where λ is a C^∞ nondecreasing function with values in $[0,1]$ such that $\lambda(u)=0$ for $u\le 1/2$ and $\lambda(u)=1$ for $u\ge 1$. The bound on the norm $\|\gamma\|_{B^1_{\infty,1}}$ can be proved using the equivalent definition of Besov norms in terms of difference operators (see [Tr, Section 2.5.12]). Clearly, for all $\Sigma\in\mathcal{C}_+(\mathbb{H})$ we have $\|\gamma(\Sigma)\|\le\|\Sigma\|^{1/2}$, and for all $\Sigma\in\mathcal{C}_+(\mathbb{H})$ with $\sigma(\Sigma)\subset[\delta,1/\delta]$ we have $\gamma(\Sigma)=\Sigma^{1/2}$.

Since we need further differentiation of $\mathcal{B}^k\mathcal{D}g(\Sigma)$ with respect to Σ, it will be convenient to introduce (for given $H,H'\in\mathcal{B}_{sa}(\mathbb{H})$) the following function:

$$\begin{aligned}\phi(t_1,\dots,t_k;s_1,s_2) &:= \gamma(\bar{\Sigma}(s_1,s_2))R(t_1,\dots,t_k)\\ &\quad\times Dg\big(L(t_1,\dots,t_k)\bar{\Sigma}(s_1,s_2)R(t_1,\dots,t_k)\big)L(t_1,\dots,t_k)\gamma(\bar{\Sigma}(s_1,s_2)),\end{aligned}$$

where $\bar{\Sigma}(s_1,s_2)=\Sigma+s_1H+s_2(H'-H)$ for $s_1,s_2\in\mathbb{R}$. Note that $\varphi(t_1,\dots,t_k)=\phi(t_1,\dots,t_k,0,0)$. By the argument already used at the beginning of the proof of Lemma 15, if $h:=Dg$ is k times continuously differentiable, then so is ϕ.

For simplicity, we write $B_k(\Sigma) := \mathcal{B}^k \mathcal{D}g(\Sigma)$ and $D_k(\Sigma) := \mathcal{D}g_k(\Sigma)$. Clearly, $D_k(\Sigma) := \sum_{j=0}^{k}(-1)^j B_j(\Sigma)$

$$B_k(\Sigma) := \mathbb{E}\int_0^1 \cdots \int_0^1 \frac{\partial^k \phi(t_1, \ldots, t_k, 0, 0)}{\partial t_1 \ldots \partial t_k} dt_1 \ldots dt_k, \qquad k \geq 1.$$

For $k = 0$, we have $B_0(\Sigma) := \mathcal{D}g(\Sigma)$.

Denote

$$\gamma_{\beta,k}(\Sigma; u) := (\|\Sigma\| \vee u \vee 1)^{k+1/2}(u \vee u^{\beta}), \qquad u > 0,\ \beta \in [0, 1],\ k \geq 1.$$

Recall definition (2.27) of C^s-norms of smooth operator valued functions defined in an open set $G \subset \mathcal{B}_{sa}(\mathbb{H})$. It is assumed in this section that $G = \mathcal{B}_{sa}(\mathbb{H})$ and we will write $\|\cdot\|_{C^s}$ instead of $\|\cdot\|_{C^s(\mathcal{B}_{sa}(\mathbb{H}))}$.

Theorem 9. *Suppose that, for some $k \leq d$, g is $k + 2$ times continuously differentiable on $\mathcal{B}_{sa}(\mathbb{H})$ and, for some $\beta \in (0, 1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. In addition, suppose that $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ and $\sigma(\Sigma) \subset [\delta, 1/\delta]$. Then, for some constant $C \geq 1$ and for all $H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\|S_{B_k}(\Sigma; H') - S_{B_k}(\Sigma; H)\| \leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} \left(\frac{d}{n}\right)^{k/2} \gamma_{\beta,k}(\Sigma; \|H\| \vee \|H'\|)\|H' - H\|. \qquad (7.1)$$

Corollary 6. *Suppose that, for some $k \leq d$, g is $k + 2$ times continuously differentiable and, for some $\beta \in (0, 1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. Suppose also that $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$, $d \leq n/2$ and $\sigma(\Sigma) \subset [\delta, 1/\delta]$. Then, for some constant $C \geq 1$ and for all $H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\|S_{D_k}(\Sigma; H') - S_{D_k}(\Sigma; H)\| \leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} \gamma_{\beta,k}(\Sigma; \|H\| \vee \|H'\|)\|H' - H\|. \qquad (7.2)$$

Proof. Indeed,

$$\begin{aligned}\|S_{D_k}(\Sigma; H') - S_{D_k}(\Sigma; H)\| &\leq \sum_{j=0}^{k} \|S_{B_j}(\Sigma; H') - S_{B_j}(\Sigma; H)\| \\ &\leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} \sum_{j=0}^{k} \left(\frac{d}{n}\right)^{j/2} \gamma_{\beta,k}(\Sigma; \|H\| \vee \|H'\|)\|H' - H\| \\ &\leq 2C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} \gamma_{\beta,k}(\Sigma; \|H\| \vee \|H'\|)\|H' - H\|,\end{aligned}$$

implying the bound of the corollary (after proper adjustment of the value of C). □

Proof of Theorem 9. Note that

$$
\begin{aligned}
&S_{B_k}(\Sigma; H') - S_{B_k}(\Sigma; H) \\
&\qquad = DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H) + S_{B_k}(\Sigma + H; H' - H),
\end{aligned}
$$

so we need to bound

$$
\|DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H)\| \quad \text{and} \quad \|S_{B_k}(\Sigma + H; H' - H)\|
$$

separately. To this end, note that

$$
\begin{aligned}
B_k(\Sigma + s_1 H) &= \mathbb{E}\int_0^1 \cdots \int_0^1 \frac{\partial^k \phi(t_1, \dots, t_k, s_1, 0)}{\partial t_1 \dots \partial t_k}\, dt_1 \dots dt_k, \\
DB_k(\Sigma; H) &= \mathbb{E}\int_0^1 \cdots \int_0^1 \frac{\partial^{k+1} \phi(t_1, \dots, t_k, 0, 0)}{\partial t_1 \dots \partial t_k \partial s_1}\, dt_1 \dots dt_k, \\
DB_k(\Sigma + s_1 H; H' - H) &= \mathbb{E}\int_0^1 \cdots \int_0^1 \frac{\partial^{k+1} \phi(t_1, \dots, t_k, s_1, 0)}{\partial t_1 \dots \partial t_k \partial s_2}\, dt_1 \dots dt_k.
\end{aligned}
$$

The last two formulas hold provided that g is $k+2$ times continuously differentiable with uniformly bounded derivatives $D^j g$, $j = 0, \dots, k+2$, and, as a consequence, the function $\phi(t_1, \dots, t_k, s_1, s_2)$ is $k+1$ times continuously differentiable (the proof of this fact is similar to the proof of differentiability of $\phi(t_1, \dots, t_k)$, see the proofs of Proposition 4 and Lemma 15).

As a consequence,

$$
\begin{aligned}
&DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H) \\
&\quad = \mathbb{E}\int_0^1 \cdots \int_0^1 \left[\frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k \partial s_2} - \frac{\partial^{k+1} \phi(t_1, \dots, t_k, 0, 0)}{\partial t_1 \dots \partial t_k \partial s_2}\right] dt_1 \dots dt_k
\end{aligned}
\tag{7.3}
$$

and

$$
\begin{aligned}
&S_{B_k}(\Sigma + H; H' - H) \\
&= \mathbb{E}\int_0^1 \cdots \int_0^1 \left[\frac{\partial^k \phi(t_1, \dots, t_k, 1, 1)}{\partial t_1 \dots \partial t_k} - \frac{\partial^k \phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k} - \frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k \partial s_2}\right] \\
&\qquad \times dt_1 \dots dt_k \\
&= \mathbb{E}\int_0^1 \cdots \int_0^1 \int_0^1 \left[\frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, s_2)}{\partial t_1 \dots \partial t_k \partial s_2} - \frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k \partial s_2}\right] ds_2\, dt_1 \dots dt_k.
\end{aligned}
\tag{7.4}
$$

The next two lemmas provide upper bounds on

$$
\|DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H)\| \quad \text{and} \quad \|S_{B_k}(\Sigma + H; H' - H)\|.
$$

Lemma 19. *Suppose that, for some $k \leq d$, $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+2$ times continuously differentiable and, for some $\beta \in (0,1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. In addition, suppose that $\sigma(\Sigma) \subset [\delta, 1/\delta]$. Then, for some constant $C > 0$ and for all $H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\begin{aligned}&\|DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H)\| \\ &\qquad\leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} ((\|\Sigma\| + \|H\|)^{k+1/2} \vee 1) \Big(\frac{d}{n}\Big)^{k/2} \\ &\qquad\times (\|H\| \vee \|H\|^\beta) \|H' - H\|. \end{aligned} \tag{7.5}$$

Lemma 20. *Suppose that, for some $k \leq d$, $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+2$ times continuously differentiable and, for some $\beta \in (0,1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. In addition, suppose that $\sigma(\Sigma) \subset [\delta, 1/\delta]$. Then, for some constant $C > 0$ and for all $H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\begin{aligned}&\|S_{B_k}(\Sigma + H; H' - H)\| \\ &\quad\leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} ((\|\Sigma\| + \|H\| + \|H'\|)^{k+1/2} \vee 1) \Big(\frac{d}{n}\Big)^{k/2} \\ &\quad\times (\|H' - H\|^{1+\beta} \vee \|H' - H\|^2). \end{aligned} \tag{7.6}$$

In the next section, we will also need the following lemma.

Lemma 21. *Suppose that, for some $k \leq d$, $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+2$ times differentiable with uniformly bounded continuous derivatives $D^j g$, $j = 0, \dots, k+2$. In addition, suppose that $\sigma(\Sigma) \subset [\delta, 1/\delta]$. Then, for some constant $C > 0$ and for all $H \in \mathcal{B}_{sa}(\mathbb{H})$,*

$$\|DB_k(\Sigma; H)\| \leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1}} (\|\Sigma\|^{k+1/2} \vee 1) \Big(\frac{d}{n}\Big)^{k/2} \|H\|. \tag{7.7}$$

We give a proof of Lemma 19 below. The proofs of Lemmas 20 and 21 are based on a similar approach.

Proof of Lemma 19. First, we derive an upper bound on the difference

$$\frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k \partial s_2} - \frac{\partial^{k+1} \phi(t_1, \dots, t_k, 0, 0)}{\partial t_1 \dots \partial t_k \partial s_2}$$

in the right hand side of (7.3). To this end, note that by the product rule,

$$\frac{\partial^{k+1} \phi(t_1, \dots, t_k, s_1, s_2)}{\partial t_1 \dots \partial t_k \partial s_2} = \frac{\partial}{\partial s_2} \sum_{T_1, T_2, T_3} \gamma(\bar{\Sigma})(\partial_{T_1} R)(\partial_{T_2} h(L\bar{\Sigma}R))(\partial_{T_3} L)\gamma(\bar{\Sigma}),$$

where $h = Dg$ and the sum extends over all partitions T_1, T_2, T_3 of the set of variables $\{t_1, \dots, t_k\}$. We can further write

$$\begin{aligned}\frac{\partial^{k+1}\phi(t_1,\dots,t_k,s_1,s_2)}{\partial t_1\dots\partial t_k\partial s_2} &= \sum_{T_1,T_2,T_3}\big[(\partial_{\{s_2\}}\gamma(\bar\Sigma))(\partial_{T_1}R)(\partial_{T_2}h(L\bar\Sigma R))(\partial_{T_3}L)\gamma(\bar\Sigma)\\ &\quad+\gamma(\bar\Sigma)(\partial_{T_1}R)(\partial_{T_2\cup\{s_2\}}h(L\bar\Sigma R))(\partial_{T_3}L)\gamma(\bar\Sigma)\\ &\quad+\gamma(\bar\Sigma)(\partial_{T_1}R)(\partial_{T_2}h(L\bar\Sigma R))(\partial_{T_3}L)(\partial_{\{s_2\}}\gamma(\bar\Sigma))\big]. \end{aligned}\tag{7.8}$$

Observe that $\partial_{\{s_2\}}\gamma(\bar\Sigma) = D\gamma(\bar\Sigma; H' - H)$ and deduce from (7.8) that

$$\frac{\partial^{k+1}\phi(t_1,\dots,t_k,1,0)}{\partial t_1\dots\partial t_k\partial s_2}-\frac{\partial^{k+1}\phi(t_1,\dots,t_k,0,0)}{\partial t_1\dots\partial t_k\partial s_2}=\sum_{T_1,T_2,T_3}[A_1+\dots+A_9],\tag{7.9}$$

where

$$\begin{aligned}
A_1 &:= [D\gamma(\Sigma+H;H'-H)-D\gamma(\Sigma;H'-H)](\partial_{T_1}R)(\partial_{T_2}h(L\bar\Sigma_{1,0}R))(\partial_{T_3}L)\gamma(\bar\Sigma_{1,0}),\\
A_2 &:= D\gamma(\Sigma;H'-H)(\partial_{T_1}R)\big(\partial_{T_2}h(L(\Sigma+H)R)-\partial_{T_2}h(L\Sigma R)\big)(\partial_{T_3}L)\gamma(\bar\Sigma_{1,0}),\\
A_3 &:= D\gamma(\Sigma;H'-H)(\partial_{T_1}R)(\partial_{T_2}h(L\Sigma R))(\partial_{T_3}L)(\gamma(\Sigma+H)-\gamma(\Sigma)),\\
A_4 &:= (\gamma(\Sigma+H)-\gamma(\Sigma))(\partial_{T_1}R)(\partial_{T_2\cup\{s_2\}}h(L\bar\Sigma_{1,0}R))(\partial_{T_3}L)\gamma(\bar\Sigma_{1,0}),\\
A_5 &:= \gamma(\Sigma)(\partial_{T_1}R)(\partial_{T_2\cup\{s_2\}}h(L(\Sigma+H)R)-\partial_{T_2\cup\{s_2\}}h(L\Sigma R))(\partial_{T_3}L)\gamma(\bar\Sigma_{1,0}),\\
A_6 &:= \gamma(\Sigma)(\partial_{T_1}R)(\partial_{T_2\cup\{s_2\}}h(L\Sigma R))(\partial_{T_3}L)(\gamma(\Sigma+H)-\gamma(\Sigma)),\\
A_7 &:= (\gamma(\Sigma+H)-\gamma(\Sigma))(\partial_{T_1}R)(\partial_{T_2}h(L\bar\Sigma_{1,0}R))(\partial_{T_3}L)D\gamma(\bar\Sigma_{1,0};H'-H),\\
A_8 &:= \gamma(\Sigma)(\partial_{T_1}R)(\partial_{T_2}h(L(\Sigma+H)R)-\partial_{T_2}h(L\Sigma R))(\partial_{T_3}L)D\gamma(\bar\Sigma_{1,0};H'-H),\\
A_9 &:= \gamma(\Sigma)(\partial_{T_1}R)(\partial_{T_2}h(L\Sigma R))(\partial_{T_3}L)\big(D\gamma(\Sigma+H;H'-H)-D\gamma(\Sigma;H'-H)\big).
\end{aligned}$$

To bound the norms of $A_1, \dots, A_9$, we need several lemmas. We introduce some notation to be used in their proofs. Recall that for a partition $\Delta = (\Delta_1, \dots, \Delta_j)$ of the set $\{t_1, \dots, t_k\}$,

$$\partial_\Delta(L\Sigma R) = (\partial_{\Delta_1}(L\Sigma R), \dots, \partial_{\Delta_j}(L\Sigma R)).$$

We will need some transformations of $\partial_\Delta(L\Sigma R)$. In particular, for $i = 1, \dots, j$ and $H \in \mathcal{B}_{sa}(\mathbb{H})$, denote

$$\begin{aligned}&\partial_\Delta(L\Sigma R)[i:\Sigma\to H]\\ &\quad:= (\partial_{\Delta_1}(L\Sigma R),\dots,\partial_{\Delta_{i-1}}(L\Sigma R),\partial_{\Delta_i}(LHR),\partial_{\Delta_{i+1}}(L\Sigma R),\dots,\partial_{\Delta_j}(L\Sigma R)).\end{aligned}$$

We will also write

$$\begin{aligned}&\partial_\Delta(L\Sigma R)[i:\Sigma\to H; i+1,\dots,j:\Sigma\to\Sigma+H]\\ &:= (\partial_{\Delta_1}(L\Sigma R),\dots,\partial_{\Delta_{i-1}}(L\Sigma R),\partial_{\Delta_i}(LHR),\partial_{\Delta_{i+1}}(L(\Sigma+H)R),\dots,\\ &\qquad\qquad \partial_{\Delta_j}(L(\Sigma+H)R)).\end{aligned}$$

In addition, the following notation will be used:

$$\partial_\Delta(L\Sigma R) \sqcup B := (\partial_{\Delta_1}(L\Sigma R), \dots, \partial_{\Delta_j}(L\Sigma R), B).$$

The meaning of other similar notation should be clear from the context. Finally, recall that $\delta_i = \|W_i - I\|, i \geq 1$.

Lemma 22. *Suppose that, for some $0 \leq m \leq k$, $h \in L_\infty(\mathcal{C}_+(\mathbb{H}); \mathcal{B}_{sa}(\mathbb{H}))$ is $m+1$ times differentiable with uniformly bounded continuous derivatives $D^j h$, $j = 1, \dots, m+1$. For all $T \subset \{t_1, \dots, t_k\}$ with $|T| = m$,*

$$\|\partial_T h(L(\Sigma + H)R) - \partial_T h(L\Sigma R)\|$$
$$\leq 2^{m(k+m+2)+1} \max_{1\leq j\leq m+1} \|D^j h\|_{L_\infty}((\|\Sigma\|+\|H\|)^m \vee 1) \prod_{t_i\in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^{2m+2} \|H\|. \tag{7.10}$$

Proof. By Lemma 15,

$$\partial_T h(L(\Sigma + H)R) - \partial_T h(L\Sigma R)$$
$$= \sum_{j=1}^m \sum_{\Delta\in\mathcal{D}_{j,T}} \left[D^j h(L(\Sigma + H)R)(\partial_\Delta(L(\Sigma + H)R)) - D^j h(L\Sigma R)(\partial_\Delta(L\Sigma R))\right]. \tag{7.11}$$

Obviously,

$$D^j h(L(\Sigma + H)R)(\partial_\Delta(L(\Sigma + H)R)) - D^j h(L\Sigma R)(\partial_\Delta(L\Sigma R))$$
$$= \sum_{i=1}^j D^j h(L(\Sigma + H)R)(\partial_\Delta(L\Sigma R)[i : \Sigma \to H; i+1, \dots, j : \Sigma \to \Sigma + H])$$
$$+ (D^j h(L(\Sigma + H)R) - D^j h(L\Sigma R))(\partial_\Delta(L\Sigma R)).$$

The following bounds hold for all $1 \leq i \leq j$:

$$\|D^j h(L(\Sigma + H)R)(\partial_\Delta(L\Sigma R)[i : \Sigma \to H; i+1, \dots, j : \Sigma \to \Sigma + H])\|$$
$$\leq \|D^j h(L(\Sigma + H)R)\| \prod_{1\leq l<i} \|\partial_{\Delta_l}(L\Sigma R)\| \prod_{i<l\leq j} \|\partial_{\Delta_l}(L(\Sigma + H)R)\| \, \|\partial_{\Delta_i}(LHR)\|.$$

As in the proof of Lemma 17, we get

$$\|D^j h(L(\Sigma + H)R)(\partial_\Delta(L\Sigma R)[i : \Sigma \to H; i+1, \dots, j : \Sigma \to \Sigma + H])\|$$
$$\leq \|D^j h\|_{L_\infty} 2^{kj} \|\Sigma\|^{i-1} \|\Sigma + H\|^{j-i} \prod_{t_i\in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^{2j} \|H\|$$
$$\leq 2^{kj} \|D^j h\|_{L_\infty} (\|\Sigma\| + \|H\|)^{j-1} \prod_{t_i\in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^k (1+\delta_i)^{2j} \|H\|.$$

In addition,

$$
\begin{aligned}
&\|(D^j h(L(\Sigma+H)R) - D^j h(L\Sigma R))(\partial_\Delta(L\Sigma R))\| \\
&\qquad \le \|D^{j+1}h\|_{L_\infty}\|L\|\,\|R\|\,\|H\|\prod_{i=1}^{j}\|\partial_{\Delta_i}(L\Sigma R)\| \\
&\qquad \le 2^{kj}\|D^{j+1}h\|_{L_\infty}\|\Sigma\|^j\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2j+2}\|H\|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\|D^j h(L(\Sigma+H)R)(\partial_\Delta(L(\Sigma+H)R)) - D^j h(L\Sigma R)(\partial_\Delta(L\Sigma R))\| \\
&\qquad \le 2^{kj}\big(j\|D^j h\|_{L_\infty}(\|\Sigma\|+\|H\|)^{j-1} + \|D^{j+1}h\|_{L_\infty}\|\Sigma\|^j\big) \\
&\qquad\quad \times\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2j+2}\|H\|.
\end{aligned}
$$

Substituting the last bound into (7.11) and recalling that $\mathrm{card}(\mathcal{D}_{j,T}) \le j^m$, it is easy to conclude the proof of (7.10). □

Lemma 23. *Suppose that, for some* $0 \le m \le k$, $h \in L_\infty(\mathcal{C}_+(\mathbb{H}); \mathcal{B}_{sa}(\mathbb{H}))$ *is* $m+1$ *times differentiable with uniformly bounded continuous derivatives* $D^j h$, $j = 1,\dots,m+1$. *Then for some constant* $C > 0$ *and for all* $T \subset \{t_1,\dots,t_k\}$ *with* $|T| = m$ *and all* $s_1 \in [0,1]$,

$$
\begin{aligned}
&\|\partial_{T\cup\{s_2\}}h(L\bar{\Sigma}_{s_1,0}R)\| \\
&\quad \le 2^{m(k+m+2)+1}\max_{1\le j\le m+1}\|D^j h\|_{L_\infty}((\|\Sigma\|+\|H\|)^m \vee 1)\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2m+2} \\
&\qquad \times \|H'-H\|.
\end{aligned}
$$

Proof. By Lemma 15,

$$
\partial_{T\cup\{s_2\}}h(L\bar{\Sigma}R) = \sum_{j=1}^{m}\sum_{\Delta\in\mathcal{D}_{j,T}}\partial_{\{s_2\}}D^j h(L\bar{\Sigma}R)(\partial_\Delta(L\bar{\Sigma}R)). \tag{7.13}
$$

Next, we have

$$
\begin{aligned}
\partial_{\{s_2\}}D^j h(L\bar{\Sigma}R)(\partial_\Delta(L\bar{\Sigma}R)) &= D^{j+1}h(L\bar{\Sigma}R)(\partial_\Delta(L\bar{\Sigma}R)\sqcup\partial_{\{s_2\}}(L\bar{\Sigma}R)) \\
&\quad + \sum_{i=1}^{j}D^j h(L\bar{\Sigma}R)(\partial_\Delta((L\bar{\Sigma}R)[i:\partial_{\Delta_i}(L\bar{\Sigma}R)\to\partial_{\Delta_i\cup\{s_2\}}(L\bar{\Sigma}R)]).
\end{aligned}
$$

Note that $\partial_{\{s_2\}}(L\bar{\Sigma}R) = L(H'-H)R$ and $\partial_{\Delta_i\cup\{s_2\}}(L\bar{\Sigma}R) = \partial_{\Delta_i}(L(H'-H)R)$, implying

$$
\begin{aligned}
\partial_{\{s_2\}}D^j h(L\bar{\Sigma}R)(\partial_\Delta(L\bar{\Sigma}R)) &= D^{j+1}h(L\bar{\Sigma}R)(\partial_\Delta(L\bar{\Sigma}R)\sqcup L(H'-H)R) \\
&\quad + \sum_{i=1}^{j}D^j h(L\bar{\Sigma}R)(\partial_\Delta(L\bar{\Sigma}R)[i:\bar{\Sigma}\to H'-H]).
\end{aligned} \tag{7.14}
$$

The following bounds hold:

$$\|D^{j+1}h(L\bar{\Sigma}R)(\partial_{\Delta}(L\bar{\Sigma}R)\sqcup L(H'-H)R)\|$$
$$\le \|D^{j+1}h\|_{L_\infty}\prod_{i=1}^{j}\|\partial_{\Delta_i}(L\bar{\Sigma}R)\|\,\|L\|\,\|R\|\,\|H'-H\|$$

and

$$\|D^{j}h(L\bar{\Sigma}R)(\partial_{\Delta}(L\bar{\Sigma}R)[i:\bar{\Sigma}\to H'-H])\|$$
$$\le \|D^{j}h\|_{L_\infty}\prod_{l\neq i}\|\partial_{\Delta_l}(L\bar{\Sigma}R)\|\,\|\partial_{\Delta_i}(L(H'-H)R)\|.$$

The rest of the proof is based on the bounds almost identical to the ones in the proof of Lemma 22. □

Lemma 24. *Suppose that, for some $0\le m\le k$, $h\in L_\infty(\mathcal{C}_+(\mathbb{H});\mathcal{B}_{\rm sa}(\mathbb{H}))$ is $m+2$ times differentiable with uniformly bounded continuous derivatives $D^j h$, $j=1,\dots,m+2$. For some constant $C>0$ and for all $T\subset\{t_1,\dots,t_k\}$ with $|T|=m$,*

$$\|\partial_{T\cup\{s_2\}}h(L(\Sigma+H)R)-\partial_{T\cup\{s_2\}}h(L\Sigma R)\|$$
$$\le C^{k(m+1)}\max_{1\le j\le m+2}\|D^j h\|_{L_\infty}((\|\Sigma\|+\|H\|)^m\vee 1)\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2m+4}$$
$$\times\|H\|\,\|H'-H\|. \tag{7.15}$$

Moreover, if for some $0\le m\le k$, $h\in L_\infty(\mathcal{C}_+(\mathbb{H});\mathcal{B}_{\rm sa}(\mathbb{H}))$ is $m+1$ times continuously differentiable and, for some $\beta\in(0,1]$, $\|h\|_{C^{m+1+\beta}}<\infty$, then

$$\|\partial_{T\cup\{s_2\}}h(L(\Sigma+H)R)-\partial_{T\cup\{s_2\}}h(L\Sigma R)\|$$
$$\le C^{k(m+1)}\|h\|_{C^{m+1+\beta}}((\|\Sigma\|+\|H\|)^m\vee 1)\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2m+4}$$
$$\times(\|H\|\vee\|H\|^\beta)\|H'-H\|. \tag{7.16}$$

Proof. By (7.13),

$$\partial_{T\cup\{s_2\}}h(L(\Sigma+H)R)-\partial_{T\cup\{s_2\}}h(L\Sigma R)$$
$$=\sum_{j=1}^{m}\sum_{\Delta\in\mathcal{D}_{j,T}}\Big[\partial_{\{s_2\}}D^j h(L\bar{\Sigma}_{1,0}R)(\partial_{\Delta}(L\bar{\Sigma}_{1,0}R))-\partial_{\{s_2\}}D^j h(L\bar{\Sigma}_{0,0}R)(\partial_{\Delta}(L\bar{\Sigma}_{0,0}R))\Big], \tag{7.17}$$

and by (7.14),

$$\partial_{\{s_2\}} D^j h(L\bar{\Sigma}_{1,0}R)(\partial_\Delta(L\bar{\Sigma}_{1,0}R)) - \partial_{\{s_2\}} D^j h(L\bar{\Sigma}_{0,0}R)(\partial_\Delta(L\bar{\Sigma}_{0,0}R))$$
$$= \sum_{i=1}^{j} D^{j+1} h(L(\Sigma+H)R)(\partial_\Delta(L\Sigma R)[i:\Sigma\to H; i+1,\dots,j:\Sigma\to\Sigma+H]\sqcup L(H'-H)R)$$
$$+[D^{j+1}h(L(\Sigma+H)R)-D^{j+1}h(L\Sigma R)](\partial_\Delta(L\Sigma R)\sqcup L(H'-H)R)$$
$$+\sum_{i=1}^{j}\sum_{i'\neq i} D^j h(L(\Sigma+H)R)(\partial_\Delta(L\Sigma R)[i:\Sigma\to H'-H; i':\Sigma\to H; l>i', l\neq i:\Sigma\to\Sigma+H])$$
$$+\sum_{i=1}^{j}[D^j h(L(\Sigma+H)R)-D^j h(L\Sigma R)](\partial_\Delta(L\Sigma R)[i:\Sigma\to H'-H]). \tag{7.18}$$

Similarly to the bounds in the proof of Lemma 22, we get

$$\|D^{j+1}h(L(\Sigma+H)R)(\partial_\Delta(L\Sigma R)[i:\Sigma\to H; i+1,\dots,j:\Sigma\to\Sigma+H]\sqcup L(H'-H)R)\|$$
$$\le 2^{kj}\|D^{j+1}h\|_{L_\infty}(\|\Sigma\|+\|H\|)^{j-1}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2j+2}\|H\|\,\|H'-H\|,$$
$$\|[D^{j+1}h(L(\Sigma+H)R)-D^{j+1}h(L\Sigma R)](\partial_\Delta(L\Sigma R)\sqcup L(H'-H)R)\|$$
$$\le 2^{kj}\|D^{j+2}h\|_{L_\infty}\|\Sigma\|^{j}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2j+4}\|H\|\,\|H'-H\|,$$
$$\|D^j h(L(\Sigma+H)R)(\partial_\Delta(L\Sigma R)[i:\Sigma\to H'-H; i':\Sigma\to H; l>i', l\neq i:\Sigma\to\Sigma+H])\|$$
$$\le 2^{kj}\|D^j h\|_{L_\infty}(\|\Sigma\|+\|H\|)^{j-2}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2j}\|H\|\,\|H'-H\|$$

and

$$\|[D^j h(L(\Sigma+H)R)-D^j h(L\Sigma R)](\partial_\Delta(L\Sigma R)[i:\Sigma\to H'-H])\|$$
$$\le 2^{kj}\|D^{j+1}h\|_{L_\infty}\|\Sigma\|^{j-1}\prod_{t_i\in T}\frac{\delta_i}{1+\delta_i}\prod_{i=1}^{k}(1+\delta_i)^{2j}\|H\|\,\|H'-H\|.$$

These bounds along with formulas (7.17), (7.18) imply that (7.15) holds. The proof of (7.16) is similar. □

We now get back to bounding the operators $A_1,\dots,A_9$ on the right hand side of (7.9). It easily follows from Lemmas 16, 17, 22, 23 and 24 as well as from the bounds

$$\|\gamma(\Sigma)\| \le \|\Sigma\|^{1/2},$$
$$\|\gamma(\Sigma+H)-\gamma(\Sigma)\| \le 2\|\gamma\|_{B^1_{\infty,1}}\|H\| \lesssim \frac{\log(2/\delta)}{\sqrt{\delta}}\|H\|,$$
$$\|D\gamma(\Sigma;H)\| \le 2\|\gamma\|_{B^1_{\infty,1}}\|H\| \lesssim \frac{\log(2/\delta)}{\sqrt{\delta}}\|H\|$$

and

$$\|D\gamma(\Sigma+H;H'-H)-D\gamma(\Sigma;H'-H)\| \lesssim \|\gamma\|_{B^2_{\infty,1}}\|H\|\,\|H'-H\| \lesssim \frac{\log^2(2/\delta)}{\delta}\|H\|\,\|H'-H\|$$

that for some constant $C_1 > 0$ and for all $l = 1, \ldots, 9$,

$$|A_l| \le C_1^{k^2}\frac{\log^2(2/\delta)}{\delta}\|Dg\|_{C^{k+1+\beta}}((\|\Sigma\|+\|H\|)^{k+1/2}\vee 1)\prod_{i=1}^k \delta_i(1+\delta_i)^{2k+5} \times (\|H\|\vee\|H\|^{\beta})\|H'-H\|.$$

It then follows from (7.9) that

$$\left\|\frac{\partial^{k+1}\phi(t_1,\ldots,t_k,1,0)}{\partial t_1\ldots\partial t_k\partial s_2}-\frac{\partial^{k+1}\phi(t_1,\ldots,t_k,0,0)}{\partial t_1\ldots\partial t_k\partial s_2}\right\| \le C^{k^2}\frac{\log^2(2/\delta)}{\delta}\|Dg\|_{C^{k+1+\beta}}((\|\Sigma\|+\|H\|)^{k+1/2}\vee 1)\prod_{i=1}^k \delta_i(1+\delta_i)^{2k+5} \times (\|H\|\vee\|H\|^{\beta})\|H'-H\| \tag{7.19}$$

with some constant $C > 0$. Similarly to (6.18), we have that for $k \le d$ with some constant $C_2 \ge 1$,

$$\mathbb{E}\prod_{i=1}^k \delta_i(1+\delta_i)^{2k+5} \le C_2^{k^2}(d/n)^{k/2}.$$

Using this together with (7.19) to bound the expectation in (7.3) yields (7.5). □

Theorem 9 immediately follows from Lemmas 19 and 20. □

We will now derive a bound on the bias of the estimator $\mathcal{D}g_k(\hat{\Sigma})$ that improves the bounds of Section 6 under stronger smoothness assumptions on g.

Corollary 7. *Suppose $g \in L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ is $k+2$ times continuously differentiable for some $k \le d \le n$ and, for some $\beta \in (0,1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. In addition, suppose that for some $\delta > 0$, $\sigma(\Sigma) \subset [\delta, 1/\delta]$. Then, for some constant $C > 0$,*

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma})-\mathcal{D}g(\Sigma)\| \le C^{k^2}\frac{\log^2(2/\delta)}{\delta}\|Dg\|_{C^{k+1+\beta}}(\|\Sigma\|\vee 1)^{k+3/2}\|\Sigma\|\left(\frac{d}{n}\right)^{(k+1+\beta)/2}. \tag{7.20}$$

Proof. First note that

$$\mathcal{B}^{k+1}\mathcal{D}g(\Sigma) = \mathbb{E}_\Sigma B_k(\hat{\Sigma}) - B_k(\Sigma)$$
$$= \mathbb{E}_\Sigma DB_k(\Sigma; \hat{\Sigma} - \Sigma) + \mathbb{E}_\Sigma S_{B_k}(\Sigma; \hat{\Sigma} - \Sigma) = \mathbb{E}_\Sigma S_{B_k}(\Sigma; \hat{\Sigma} - \Sigma).$$

It follows from (7.1) (with $H' = \hat{\Sigma} - \Sigma$ and $H = 0$) that

$$\|S_{B_k}(\Sigma; \hat{\Sigma} - \Sigma)\| \leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} \left(\frac{d}{n}\right)^{k/2} \gamma_{\beta,k}(\Sigma; \|\hat{\Sigma} - \Sigma\|)\|\hat{\Sigma} - \Sigma\|. \quad (7.21)$$

Since

$$\gamma_{\beta,k}(\Sigma; \|\hat{\Sigma} - \Sigma\|)\|\hat{\Sigma} - \Sigma\| \leq (\|\Sigma\| \vee 1)^{k+1/2}(\|\hat{\Sigma} - \Sigma\|^{1+\beta} + \|\hat{\Sigma} - \Sigma\|^2) + \|\hat{\Sigma} - \Sigma\|^{k+\beta+3/2} + \|\hat{\Sigma} - \Sigma\|^{k+5/2},$$

we can use the bound $\mathbb{E}^{1/p}\|\hat{\Sigma} - \Sigma\|^p \lesssim \|\Sigma\|(\sqrt{d/n} \vee \sqrt{p/n})$ to deduce that for some constant $C_1 > 0$ and for $k \leq d \leq n$,

$$\mathbb{E}\gamma_{\beta,k}(\Sigma; \|\hat{\Sigma} - \Sigma\|)\|\hat{\Sigma} - \Sigma\| \leq C_1^k(\|\Sigma\| \vee 1)^{k+3/2}\|\Sigma\|\left(\frac{d}{n}\right)^{(1+\beta)/2}.$$

Therefore, for some constant $C > 0$,

$$\|\mathcal{B}^{k+1}\mathcal{D}g(\Sigma)\| \leq \mathbb{E}\|S_{B_k}(\Sigma; \hat{\Sigma} - \Sigma)\| \leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}}(\|\Sigma\| \vee 1)^{k+3/2}\|\Sigma\|\left(\frac{d}{n}\right)^{(k+1+\beta)/2}.$$

Since $\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma) = (-1)^k\mathcal{B}^{k+1}\mathcal{D}g(\Sigma)$, the result follows. □

8. Normal approximation bounds for estimators with reduced bias

In this section, our goal is to prove bounds showing that, for sufficiently smooth orthogonally invariant functions g, for large enough k and for an operator B with nuclear norm bounded by a constant, the distribution of the random variables

$$\frac{\sqrt{n}\,(\langle \mathcal{D}g_k(\hat{\Sigma}), B\rangle - \langle \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma; B)} \quad (8.1)$$

is close to the standard normal distribution as $n \to \infty$ and $d = o(n)$. It will be shown that this holds true with

$$\sigma_g^2(\Sigma; B) = 2\|\Sigma^{1/2}(D\mathcal{D}g(\Sigma))^* B\Sigma^{1/2}\|_2^2, \quad (8.2)$$

where $(D\mathcal{D}g(\Sigma))^*$ is the adjoint operator of $D\mathcal{D}g(\Sigma)$:

$$\langle D\mathcal{D}g(\Sigma)H_1, H_2\rangle = \langle H_1, (D\mathcal{D}g(\Sigma))^* H_2\rangle.$$

We will prove the following result.

Theorem 10. *Suppose that, for some* $s > 0$, $g \in C^{s+1}(\mathcal{B}_{sa}(\mathbb{H})) \cap L_\infty^O(\mathcal{C}_+(\mathbb{H}))$ *is an orthogonally invariant function. Suppose that* $d \geq 3\log n$ *and, for some* $\alpha \in (0, 1)$, $d \leq n^\alpha$. *Suppose also that* Σ *is nonsingular and, for a small enough constant* $c > 0$,

$$d \leq \frac{cn}{(\|\Sigma\| \vee \|\Sigma^{-1}\|)^4}. \tag{8.3}$$

Finally, suppose that $s > \frac{1}{1-\alpha}$ *and let* k *be an integer such that* $\frac{1}{1-\alpha} < k + 1 + \beta \leq s$ *for some* $\beta \in (0, 1]$. *Then there exists a constant* C *such that*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sqrt{n}\,(\langle \mathcal{D}g_k(\hat{\Sigma}), B\rangle - \langle \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma; B)} \leq x \right\} - \Phi(x) \right|$$
$$\leq C^{k^2} L_g(B; \Sigma)\big[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}\big] + C/\sqrt{n}, \tag{8.4}$$

where

$$L_g(B; \Sigma) := \frac{\|B\|_1 \|Dg\|_{C^s}}{\sigma_g(\Sigma; B)} (\|\Sigma\| \vee \|\Sigma^{-1}\|) \log^2(2(\|\Sigma\| \vee \|\Sigma^{-1}\|)) \|\Sigma\| (\|\Sigma\| \vee 1)^{k+3/2}.$$

We will also need the following exponential upper bound on the r.v. (8.1).

Proposition 7. *Under the assumptions of Theorem 10, there exists a constant* C *such that for all* $t \geq 1$, *with probability at least* $1 - e^{-t}$,

$$\left| \frac{\sqrt{n}\,(\langle \mathcal{D}g_k(\hat{\Sigma}), B\rangle - \langle \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma; B)} \right| \leq C^{k^2}(L_g(B; \Sigma) \vee 1)\sqrt{t}. \tag{8.5}$$

Our main application is to the problem of estimation of the functional $\langle f(\Sigma), B\rangle$ for a given smooth function f and a given operator B. We will use $\langle f_k(\hat{\Sigma}), B\rangle$ as its estimator, where $f_k(\Sigma) := \sum_{j=0}^k (-1)^j \mathcal{B}^j f(\Sigma)$. Denote

$$\sigma_f^2(\Sigma; B) = 2\|\Sigma^{1/2} Df(\Sigma; B)\Sigma^{1/2}\|_2^2.$$

Theorem 11. *Suppose that* $f \in B^s_{\infty,1}(\mathbb{R})$ *for some* $s > 0$. *Suppose that* $d \geq 3\log n$ *and* $d \leq n^\alpha$ *for some* $\alpha \in (0, 1)$. *Suppose also that* Σ *is nonsingular and, for a small enough constant* $c = c_s > 0$,

$$d \leq \frac{cn}{(\|\Sigma\| \vee \|\Sigma^{-1}\|)^4}. \tag{8.6}$$

Finally, suppose that $s > \frac{1}{1-\alpha}$ *and let* k *be an integer such that* $\frac{1}{1-\alpha} < k + 1 + \beta \leq s$ *for some* $\beta \in (0, 1]$. *Then there exists a constant* C *such that*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sqrt{n}\,(\langle f_k(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma; B)} \leq x \right\} - \Phi(x) \right|$$
$$\leq C^{k^2} M_f(B; \Sigma)\big[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}\big] + C/\sqrt{n}, \tag{8.7}$$

where

$$M_f(B;\Sigma) := \frac{\|B\|_1 \|f\|_{B^s_{\infty,1}}}{\sigma_f(\Sigma;B)} (\|\Sigma\| \vee \|\Sigma^{-1}\|)^{2+s} \log^2(2(\|\Sigma\| \vee \|\Sigma^{-1}\|)) \|\Sigma\| (\|\Sigma\| \vee 1)^{k+3/2}.$$

Proposition 8. *Under the assumptions of Theorem 11, there exists a constant C such that for all $t \geq 1$, with probability at least $1 - e^{-t}$,*

$$\left| \frac{\sqrt{n}\,(\langle f_k(\hat\Sigma), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma;B)} \right| \leq C^{k^2} (M_f(B;\Sigma) \vee 1)\sqrt{t}. \tag{8.8}$$

Proof of Theorem 10 and Proposition 7. Recall the notation $D_k(\Sigma) := \mathcal{D}g_k(\Sigma)$. For a given operator B, define $\mathfrak{d}_k(\Sigma) := \langle D_k(\Sigma), B\rangle$, recall that

$$\mathfrak{d}_k(\hat\Sigma) - \mathbb{E}\mathfrak{d}_k(\hat\Sigma) = \langle D\mathfrak{d}_k(\Sigma), \hat\Sigma - \Sigma\rangle + S_{\mathfrak{d}_k}(\Sigma;\hat\Sigma - \Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma;\hat\Sigma-\Sigma),$$

and consider the following representation:

$$\frac{\sqrt{n}\,(\langle \mathcal{D}g_k(\hat\Sigma), B\rangle - \langle \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma;B)} = \frac{\sqrt{n}\,\langle D\mathfrak{d}_k(\Sigma), \hat\Sigma - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} + \zeta, \tag{8.9}$$

with the remainder $\zeta := \zeta_1 + \zeta_2 + \zeta_3$, where

$$\zeta_1 := \frac{\sqrt{n}\,(\langle \mathbb{E}\mathcal{D}g_k(\hat\Sigma) - \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma;B)},$$

$$\zeta_2 := \frac{\sqrt{n}\,(S_{\mathfrak{d}_k}(\Sigma;\hat\Sigma-\Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma;\hat\Sigma-\Sigma))}{\sigma_g(\Sigma;B)},$$

$$\zeta_3 := \frac{\sqrt{n}\,\langle D\mathfrak{d}_k(\Sigma), \hat\Sigma - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} \frac{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2 - \sigma_g(\Sigma;B)}{\sigma_g(\Sigma;B)}.$$

Step 1. By Lemma 9,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sqrt{n}\,\langle D\mathfrak{d}_k(\Sigma), \hat\Sigma - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} \leq x \right\} - \Phi(x) \right| \lesssim \left(\frac{\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_3}{\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} \right)^3 \frac{1}{\sqrt{n}} \lesssim \frac{\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|}{\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} \frac{1}{\sqrt{n}} \lesssim \frac{1}{\sqrt{n}}. \tag{8.10}$$

Note also that

$$\frac{\sqrt{n}\,\langle D\mathfrak{d}_k(\Sigma), \hat\Sigma - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} \overset{d}{=} \frac{\sum_{j=1}^n \sum_{i\geq 1} \lambda_i (Z_{i,j}^2 - 1)}{\sqrt{2n}\,(\sum_{i\geq 1} \lambda_i^2)^{1/2}},$$

where $\{Z_{i,j}\}$ are i.i.d. standard normal random variables and $\{\lambda_i\}$ are the eigenvalues of $\mathcal{D}\mathfrak{d}_k(\Sigma)$ (see the proof of Lemma 9). To provide an upper bound on the right hand side, we use Lemma 11 to infer that with probability at least $1 - e^{-t}$,

$$\left| \frac{\sqrt{n}\,\langle D\mathfrak{d}_k(\Sigma), \hat{\Sigma} - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2} \right| \lesssim \sqrt{t} \vee \frac{t}{\sqrt{n}}. \tag{8.11}$$

We will now control each of the random variables $\zeta_1, \zeta_2, \zeta_3$ separately.

Step 2. To bound ζ_1, we observe that, for $\delta = \frac{1}{\|\Sigma\| \vee \|\Sigma^{-1}\|}$, we have $\sigma(\Sigma) \subset [\delta, 1/\delta]$ and use inequality (7.20) that yields

$$\begin{aligned} |\zeta_1| &\le \frac{\sqrt{n}\,\|\mathbb{E}\mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\|\,\|B\|_1}{\sigma_g(\Sigma; B)} \\ &\le C^{k^2}\Lambda_{k,\beta}(g; \Sigma; B)(\|\Sigma\| \vee 1)^{k+3/2}\|\Sigma\|\sqrt{n}\,(d/n)^{(k+1+\beta)/2}, \end{aligned} \tag{8.12}$$

where

$$\Lambda_{k,\beta}(g; \Sigma; B) := \frac{\|B\|_1\|Dg\|_{C^{k+1+\beta}}}{\sigma_g(\Sigma; B)}(\|\Sigma\| \vee \|\Sigma^{-1}\|)\log^2(2(\|\Sigma\| \vee \|\Sigma^{-1}\|)).$$

Since $d \le n^{\alpha}$ for some $\alpha \in (0, 1)$, the last bound implies that

$$|\zeta_1| \le C^{k^2}\Lambda_{k,\beta}(g; \Sigma; B)(\|\Sigma\| \vee 1)^{k+3/2}\|\Sigma\|n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}}, \tag{8.13}$$

which tends to 0 for $k + 1 + \beta > \frac{1}{1-\alpha}$.

Step 3. To bound ζ_2, recall Theorem 6 and Corollary 6. It follows from these statements that, under the assumptions $\|g\|_{C^{k+2+\beta}} < \infty$ and $d \le n/2$, for all $t \ge 1$ with probability at least $1 - e^{-t}$,

$$\begin{aligned} |\zeta_2| &\le \frac{\sqrt{n}\,|S_{\mathfrak{d}_k}(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma; \hat{\Sigma} - \Sigma)|}{\sigma_g(\Sigma; B)} \\ &\le C^{k^2}\Lambda_{k,\beta}(g; \Sigma; B)\gamma_{\beta,k}(\Sigma; \delta_n(\Sigma; t))(\sqrt{\|\Sigma\|} + \sqrt{\delta_n(\Sigma; t)})\sqrt{\|\Sigma\|}\sqrt{t}, \end{aligned} \tag{8.14}$$

where

$$\begin{aligned} \delta_n(\Sigma; t) &:= \|\Sigma\|\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right) \\ &\le \|\Sigma\|\left(\sqrt{\frac{d}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right) =: \bar{\delta}_n(\Sigma; t). \end{aligned}$$

Recall that $\gamma_{\beta,k}(\Sigma; u) = (\|\Sigma\| \vee u \vee 1)^{k+1/2}(u \vee u^{\beta})$ for $u > 0$. For $d \le n$ and $t \le n$, we have $\delta_n(\Sigma; t) \le \|\Sigma\|$ and $\gamma_{\beta,k}(\Sigma; \delta_n(\Sigma; t)) \le (\|\Sigma\| \vee 1)^{k+3/2}$, which implies that, for some $C > 1$ and for all $t \in [1, n]$, with probability at least $1 - e^{-t}$,

$$|\zeta_2| \le C^{k^2}\Lambda_{k,\beta}(g; \Sigma; B)\|\Sigma\|(\|\Sigma\| \vee 1)^{k+3/2}\sqrt{t}. \tag{8.15}$$

Let now $t = 3\log n$. For $d \geq 3\log n$, $d \leq n$, we have $\bar{\delta}_n(\Sigma; t) \leq \|\Sigma\|\sqrt{d/n} \leq \|\Sigma\|$ and

$$\gamma_{\beta,k}(\Sigma; \delta_n(\Sigma; t)) \leq \gamma_{\beta,k}(\Sigma; \bar{\delta}_n(\Sigma; t)) \leq (\|\Sigma\| \vee 1)^{k+3/2}(d/n)^{\beta/2}.$$

In addition,

$$(\sqrt{\|\Sigma\|} + \sqrt{\delta_n(\Sigma; t)})\sqrt{\|\Sigma\|}\sqrt{t} \lesssim \|\Sigma\|\sqrt{\log n}.$$

Thus, for $d \geq 3\log n$, $d \leq n^\alpha$, it follows from (8.14) that with some constant $C \geq 1$ and with probability at least $1 - n^{-3}$,

$$\begin{aligned}|\zeta_2| &\leq C^{k^2}\Lambda_{k,\beta}(g; \Sigma; B)\|\Sigma\|(\|\Sigma\| \vee 1)^{k+3/2}(d/n)^{\beta/2}\sqrt{\log n}\\ &\leq C^{k^2}\Lambda_{k,\beta}(g; \Sigma; B)\|\Sigma\|(\|\Sigma\| \vee 1)^{k+3/2}n^{-(1-\alpha)\beta/2}\sqrt{\log n}.\end{aligned} \tag{8.16}$$

Step 4. Finally, we need to bound ζ_3. To this end, denote $\mathfrak{b}_k(\Sigma) := \langle B_k(\Sigma), B\rangle$. Then $\mathfrak{b}_0(\Sigma) = \langle \mathcal{D}g(\Sigma), B\rangle$ and $\mathfrak{d}_k(\Sigma) = \sum_{j=0}^k (-1)^j \mathfrak{b}_j(\Sigma)$. Observe that

$$\langle D\mathfrak{b}_j(\Sigma), H\rangle = D\mathfrak{b}_j(\Sigma; H) = \langle DB_j(\Sigma)H, B\rangle = \langle H, (DB_j(\Sigma))^* B\rangle,$$

implying $D\mathfrak{b}_j(\Sigma) = (DB_j(\Sigma))^* B$. Therefore, we have

$$\begin{aligned}\|D\mathfrak{b}_j(\Sigma)\|_2 &= \sup_{\|H\|_2 \leq 1} |\langle DB_j(\Sigma)H, B\rangle| \leq \sup_{\|H\| \leq 1} |\langle DB_j(\Sigma)H, B\rangle|\\ &\leq \|B\|_1 \sup_{\|H\| \leq 1} \|DB_j(\Sigma)H\|.\end{aligned}$$

To bound the right hand side we use Lemma 21 that yields

$$\sup_{\|H\| \leq 1} \|DB_j(\Sigma)H\| \leq C^{j^2} \max_{1 \leq j \leq j+2} \|D^j g\|_{L_\infty}(\|\Sigma\|^{j+1/2} \vee 1)(d/n)^{j/2}.$$

Therefore, for all $j = 1, \ldots, k$,

$$\|D\mathfrak{b}_j(\Sigma)\|_2 \leq C^{k^2}\|B\|_1 \max_{1 \leq j \leq k+2} \|D^j g\|_{L_\infty}(\|\Sigma\|^{k+1/2} \vee 1)(d/n)^{j/2}$$

and

$$\begin{aligned}\|\mathcal{D}\mathfrak{b}_j(\Sigma)\|_2 &= \|\Sigma^{1/2} D\mathfrak{b}_j(\Sigma)\Sigma^{1/2}\|_2 \leq \|\Sigma\|\, \|D\mathfrak{b}_j(\Sigma)\|_2\\ &\leq C^{k^2}\|B\|_1 \max_{1 \leq j \leq k+2} \|D^j g\|_{L_\infty}(\|\Sigma\|^{k+3/2} \vee \|\Sigma\|)(d/n)^{j/2}.\end{aligned}$$

Since also

$$\sqrt{2}\,\|\mathcal{D}\mathfrak{b}_0(\Sigma)\|_2 = \sqrt{2}\,\|\Sigma^{1/2}(D\mathcal{D}g(\Sigma))^* B\Sigma^{1/2}\|_2 = \sigma_g(\Sigma; B),$$

we get

$$\begin{aligned}|\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2 - \sigma_g(\Sigma; B)| &\leq \sqrt{2}\sum_{j=1}^k \|\mathcal{D}\mathfrak{b}_j(\Sigma)\|_2\\ &\leq \sqrt{2}\,C^{k^2}\|B\|_1 \max_{1 \leq j \leq k+2} \|D^j g\|_{L_\infty}(\|\Sigma\|^{k+3/2} \vee \|\Sigma\|)\sum_{j=1}^k (d/n)^{j/2},\end{aligned}$$

implying that, under the assumption $d \le n/4$,

$$\left|\frac{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2 - \sigma_g(\Sigma; B)}{\sigma_g(\Sigma; B)}\right| \le \frac{2\sqrt{2}\,C^{k^2}\|B\|_1}{\sigma_g(\Sigma; B)} \max_{1\le j\le k+2} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+3/2} \vee \|\Sigma\|)\sqrt{d/n}. \quad (8.17)$$

It follows from (8.17) and (8.11) that with some $C > 1$ and with probability at least $1 - e^{-t}$,

$$|\zeta_3| \le \frac{C^{k^2}\|B\|_1}{\sigma_g(\Sigma; B)} \max_{1\le j\le k+2} \|D^j g\|_{L_\infty} \|\Sigma\|(\|\Sigma\| \vee 1)^{k+1/2}\sqrt{\frac{d}{n}}\Big(\sqrt{t} \vee \frac{t}{\sqrt{n}}\Big). \quad (8.18)$$

For $d \ge 3\log n$, $d \le n^\alpha$ and $t = 3\log n$, this yields

$$|\zeta_3| \le \frac{C^{k^2}}{\sigma_g(\Sigma; B)} \max_{1\le j\le k+2} \|D^j g\|_{L_\infty} \|\Sigma\|(\|\Sigma\| \vee 1)^{k+1/2} n^{-(1-\alpha)/2}\sqrt{\log n} \quad (8.19)$$

for some $C \ge 1$ with probability at least $1 - n^{-3}$.

Step 5. It follows from (8.13), (8.16) and (8.19) that for some $C \ge 1$, with probability at least $1 - 2n^{-3}$,

$$|\zeta| \le C^{k^2} \Lambda_{k,\beta}(g; \Sigma; B)\|\Sigma\|(\|\Sigma\| \vee 1)^{k+3/2} \times \big[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n} + n^{-(1-\alpha)/2}\sqrt{\log n}\big],$$

which implies that with the same probability and with a possibly different $C \ge 1$,

$$|\zeta| \le C^{k^2} L_g(B; \Sigma)\big[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}\big].$$

It follows from the last bound that

$$\delta(\xi, \eta) \le 2n^{-3} + C^{k^2} L_g(B; \Sigma)\big[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}\big],$$

where

$$\xi := \frac{\sqrt{n}\,(\langle \mathcal{D}g_k(\hat{\Sigma}), B\rangle - \langle \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma; B)}, \qquad \eta := \frac{\sqrt{n}\,\langle D\mathfrak{d}_k(\Sigma), \hat{\Sigma} - \Sigma\rangle}{\sqrt{2}\,\|\mathcal{D}\mathfrak{d}_k(\Sigma)\|_2},$$

$\xi - \eta = \zeta$ and $\delta(\xi, \eta)$ is defined in Lemma 10. It follows from (8.10) and Lemma 10 that, for some $C \ge 1$, the bound (8.4) holds.

Step 6. It remains to prove Proposition 7. When $t \in [1, n]$, the bound (8.5) immediately follows from (8.9), (8.11), (8.13), (8.15) and (8.18). To prove it for $t > n$, first observe that

$$|\langle \mathcal{D}g(\Sigma), B\rangle| \le \|\Sigma^{1/2} Dg(\Sigma)\Sigma^{1/2}\|\,\|B\|_1 \le \|Dg\|_{L_\infty}\|\Sigma\|\,\|B\|_1. \quad (8.20)$$

We will also prove that for some constant $C > 1$,

$$|\langle \mathcal{D}g_k(\Sigma), B\rangle| \le C^k \|Dg\|_{L_\infty} \|\Sigma\| \, \|B\|_1. \tag{8.21}$$

To this end, note that, by (5.13),

$$\begin{aligned}\|\mathcal{D}\mathcal{B}^k g(\Sigma)\| &\le \mathbb{E} \sum_{I\subset\{1,\dots,k\}} \|\Sigma^{1/2} A_I Dg(A_I^*\Sigma A_I) A_I^* \Sigma^{1/2}\| \\ &\le \sum_{I\subset\{1,\dots,k\}} \|\Sigma\| \, \|Dg\|_{L_\infty} \mathbb{E}\|A_I\|^2 \le \|\Sigma\| \, \|Dg\|_{L_\infty} \sum_{I\subset\{1,\dots,k\}} \mathbb{E}\prod_{i\in I}\|W_i\| \\ &\le \|\Sigma\| \, \|Dg\|_{L_\infty} \sum_{j=0}^k \binom{k}{j} (\mathbb{E}\|W\|)^j = \|\Sigma\| \, \|Dg\|_{L_\infty} (1+\mathbb{E}\|W\|)^k. \end{aligned} \tag{8.22}$$

For $d \le n$, we have $\mathbb{E}\|W - I\| \lesssim \sqrt{d/n} \le C'$ for some $C' > 0$. Thus,

$$1 + \mathbb{E}\|W\| \le 2 + \mathbb{E}\|W - I\| \le 2 + C' =: C.$$

Therefore, $\|\mathcal{D}\mathcal{B}^k g(\Sigma)\| \le C^k \|Dg\|_{L_\infty} \|\Sigma\|$. In view of the definition of g_k, this implies that (8.21) holds with some $C > 1$. It follows from (8.20) and (8.21) that, for some $C > 1$,

$$\left| \frac{\sqrt{n}\,(\langle \mathcal{D}g_k(\hat\Sigma), B\rangle - \langle \mathcal{D}g(\Sigma), B\rangle)}{\sigma_g(\Sigma; B)} \right| \le \frac{C^k \|B\|_1 \|Dg\|_{L_\infty} \|\Sigma\| \sqrt{n}}{\sigma_g(\Sigma; B)}. \tag{8.23}$$

For $t > n$, the right hand side of (8.23) is smaller than the right hand side of (8.5). Thus, (8.5) holds for all $t \ge 1$. □

Proof of Theorem 11 and Proposition 8. First suppose that, for some $\delta > 0$, $\sigma(\Sigma) \subset [2\delta, \infty)$. Let $\gamma_\delta(x) = \gamma(x/\delta)$, where $\gamma : \mathbb{R} \to [0, 1]$ is a nondecreasing C^∞ function, $\gamma(x) = 0$ for $x \le 1/2$ and $\gamma(x) = 1$ for $x \ge 1$. Define $f_\delta(x) = f(x)\gamma_\delta(x)$ for $x \in \mathbb{R}$. Then $f(\Sigma) = f_\delta(\Sigma)$, which also implies that, for all Σ with $\sigma(\Sigma) \subset [2\delta, \infty)$, $Df(\Sigma) = Df_\delta(\Sigma)$ and $\sigma_f(\Sigma; B) = \sigma_{f_\delta}(\Sigma; B)$.

Let $\varphi(x) := \int_0^x \frac{f_\delta(t)}{t}\,dt$ for $x \ge 0$ and $\varphi(x) = 0$ for $x < 0$. Clearly, $f_\delta(x) = x\varphi'(x)$ for $x \in \mathbb{R}$. Let $g(C) := \mathrm{tr}(\varphi(C))$ for $C \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$. Then g is clearly an orthogonally invariant function, $Dg(C) = \varphi'(C)$ for $C \in \mathcal{B}_{\mathrm{sa}}(\mathbb{H})$ and

$$\mathcal{D}g(C) = C^{1/2}\varphi'(C)C^{1/2} = f_\delta(C), \qquad C \in \mathcal{C}_+(\mathbb{H}).$$

It is also easy to see that $\mathcal{D}g_k(C) = (f_\delta)_k(C)$ for $C \in \mathcal{C}_+(\mathbb{H})$. Using Corollary 2 of Section 2, standard bounds for pointwise multipliers of functions in Besov spaces [Tr, Section 2.8.3] and the characterization of Besov norms in terms of difference operators [Tr, Section 2.5.12], it is easy to check that

$$\begin{aligned}\|Dg\|_{C^s} &\le 2^{k+1}\|\varphi'\|_{B^s_{\infty,1}} = 2^{k+1}\left\|\frac{f(x)\gamma_\delta(x)}{x}\right\|_{B^s_{\infty,1}} \lesssim 2^{k+1}\left\|\frac{\gamma_\delta(x)}{x}\right\|_{B^s_{\infty,1}} \|f\|_{B^s_{\infty,1}} \\ &\lesssim 2^{k+1}\frac{1}{\delta}\left\|\frac{\gamma(x/\delta)}{x/\delta}\right\|_{B^s_{\infty,1}} \|f\|_{B^s_{\infty,1}} \lesssim 2^{k+1}(\delta^{-1-s} \vee \delta^{-1})\|f\|_{B^s_{\infty,1}}.\end{aligned}$$

Denote

$$\eta := \frac{\sqrt{n}\,(\langle (f_\delta)_k(\hat\Sigma), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma; B)}.$$

It follows from Theorem 10 that

$$\Delta(\eta; Z) \le C^{k^2} M_{f_\delta}(B;\Sigma)\big[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}\big] + C/\sqrt{n} \tag{8.24}$$

with $M_{f_\delta}(B;\Sigma) \lesssim 2^{k+1} M_{f,\delta}(B;\Sigma)$ and

$$M_{f,\delta}(B;\Sigma) := \frac{\|B\|_1(\delta^{-1-s}\vee\delta^{-1})\|f\|_{B^s_{\infty,1}}}{\sigma_f(\Sigma;B)} \times (\|\Sigma\|\vee\|\Sigma^{-1}\|)\log^2(2(\|\Sigma\|\vee\|\Sigma^{-1}\|))\|\Sigma\|(\|\Sigma\|\vee 1)^{k+3/2}.$$

It will be shown that, under the assumption $\sigma(\Sigma)\subset[2\delta,\infty)$, the estimator $(f_\delta)_k(\hat\Sigma) = \mathcal{D}g_k(\hat\Sigma)$ can be replaced by the estimator $f_k(\hat\Sigma)$. To this end, the following lemma will be proved.

Lemma 25. *Suppose that* $\sigma(\Sigma)\subset[2\delta,\infty)$ *for some* $\delta>0$, *and for a sufficiently large constant* $C_1>1$,

$$d \le \frac{\log^2(1+\delta/\|\Sigma\|)}{C_1^2(k+1)^2} n =: \bar d. \tag{8.25}$$

Then, with probability at least $1-e^{-\bar d}$,

$$\|f_k(\hat\Sigma) - (f_\delta)_k(\hat\Sigma)\| \le (k^2 2^{k+1}+2)e^{-\bar d}\|f\|_{L_\infty}. \tag{8.26}$$

Proof. Recall that, by (5.2),

$$\mathcal{B}^k f(\Sigma) = \mathbb{E}_\Sigma \sum_{j=0}^k (-1)^{k-j}\binom{k}{j} f(\hat\Sigma^{(j)}),$$

implying that

$$\mathcal{B}^k f(\hat\Sigma) - \mathcal{B}^k f_\delta(\hat\Sigma) = \mathbb{E}_{\hat\Sigma}\sum_{j=0}^k (-1)^{k-j}\binom{k}{j}[f(\hat\Sigma^{(j+1)}) - f_\delta(\hat\Sigma^{(j+1)})].$$

Note also that $f(\hat\Sigma^{(j+1)}) = f_\delta(\hat\Sigma^{(j+1)})$ provided that $\sigma(\hat\Sigma^{(j+1)})\subset[\delta,\infty)$ (since $f(x) = f_\delta(x)$ for $x\ge\delta$). This easily implies

$$\|\mathcal{B}^k f(\hat\Sigma) - \mathcal{B}^k f_\delta(\hat\Sigma)\| \le 2^{k+1}\|f\|_{L_\infty}\mathbb{P}_{\hat\Sigma}\{\exists j=1,\dots,k+1 : \sigma(\hat\Sigma^{(j)})\not\subset[\delta,\infty)\}. \tag{8.27}$$

To control the probability of the event $G := \{\exists j=1,\dots,k+1 : \sigma(\hat\Sigma^{(j)})\not\subset[\delta,\infty)\}$, consider the following event:

$$E := \{\|\hat\Sigma^{(j+1)} - \hat\Sigma^{(j)}\| < C_1\|\hat\Sigma^{(j)}\|\sqrt{d/n},\ j=1,\dots,k\}.$$

It follows from (1.6) (applied conditionally on $\hat{\Sigma}^{(j)}$) that, for a proper choice of $C_1 > 0$,

$$\mathbb{P}_{\hat{\Sigma}}(E^c) \le \mathbb{E}_{\hat{\Sigma}} \sum_{j=1}^{k} \mathbb{P}_{\hat{\Sigma}^{(j)}}\{\|\hat{\Sigma}^{(j+1)} - \hat{\Sigma}^{(j)}\| \ge C_1 \|\hat{\Sigma}^{(j)}\| \sqrt{d/n}\} \le k e^{-d}. \tag{8.28}$$

Note that, on the event E, $\|\hat{\Sigma}^{(j+1)}\| \le \|\hat{\Sigma}^{(j)}\|(1+C_1\sqrt{d/n})$, which implies by induction that

$$\|\hat{\Sigma}^{(j)}\| \le \|\hat{\Sigma}\|(1 + C_1\sqrt{d/n})^{j-1}, \qquad j = 1, \dots, k+1.$$

This also shows that, on the event E,

$$\begin{aligned}
\|\hat{\Sigma}^{(j)} - \hat{\Sigma}\| &\le \sum_{i=1}^{j-1} \|\hat{\Sigma}^{(i+1)} - \hat{\Sigma}^{(i)}\| \le \sum_{i=1}^{j-1} \|\hat{\Sigma}^{(i)}\| C_1 \sqrt{d/n} \\
&\le \|\hat{\Sigma}\| C_1 \sqrt{d/n} \sum_{i=1}^{j-1} (1 + C_1\sqrt{d/n})^{i-1} \\
&\le \|\hat{\Sigma}\|[(1 + C_1\sqrt{d/n})^{j-1} - 1], \qquad j = 1, \dots, k+1.
\end{aligned}$$

Consider also the event $F := \{\|\hat{\Sigma} - \Sigma\| \le C_1\|\Sigma\|\sqrt{d/n}\}$, which holds with probability at least $1 - e^{-d}$ for a proper choice of C_1. On this event, $\|\hat{\Sigma}\| \le \|\Sigma\|(1 + C_1\sqrt{d/n})$. Therefore, on the event $E \cap F$,

$$\begin{aligned}
\|\hat{\Sigma}^{(j)} - \Sigma\| &\le \|\Sigma\|(1 + C_1\sqrt{d/n})[(1 + C_1\sqrt{d/n})^{j-1} - 1] + \|\Sigma\| C_1 \sqrt{d/n} \\
&= \|\Sigma\|[(1 + C_1\sqrt{d/n})^{j} - 1], \qquad j = 1, \dots, k+1.
\end{aligned}$$

Note that

$$\|\Sigma\|[(1 + C_1\sqrt{d/n})^{k+1} - 1] \le \|\Sigma\|(\exp\{C_1(k+1)\sqrt{d/n}\} - 1) \le \delta$$

provided that condition (8.25) holds. Therefore, on the event $E \cap F$, $\|\hat{\Sigma}^{(j)} - \Sigma\| \le \delta$, $j = 1, \dots, k+1$. Since $\sigma(\Sigma) \subset [2\delta, \infty)$, this implies that $\sigma(\hat{\Sigma}^{(j)}) \subset [\delta, \infty)$, $j = 1, \dots, k+1$. In other words, $E \cap F \subset G^c$. The bound (8.27) implies that

$$\begin{aligned}
\|\mathcal{B}^k f(\hat{\Sigma}) - \mathcal{B}^k f_\delta(\hat{\Sigma})\| I_F &\le 2^{k+1}\|f\|_{L_\infty} I_F \mathbb{E}_{\hat{\Sigma}} I_G \\
&\le 2^{k+1}\|f\|_{L_\infty} I_F \mathbb{E}_{\hat{\Sigma}} I_{F^c \cup E^c} \le 2^{k+1}\|f\|_{L_\infty} I_F (I_{F^c} + \mathbb{E}_{\hat{\Sigma}} I_{E^c}) \\
&= 2^{k+1}\|f\|_{L_\infty} I_F \mathbb{P}_{\hat{\Sigma}}(E^c) \le k 2^{k+1} e^{-d} \|f\|_{L_\infty} I_F.
\end{aligned}$$

This proves that on the event F of probability at least $1 - e^{-d}$,

$$\|\mathcal{B}^k f(\hat{\Sigma}) - \mathcal{B}^k f_\delta(\hat{\Sigma})\| \le k 2^{k+1} e^{-d} \|f\|_{L_\infty}.$$

Moreover, the same bound also holds for $\|\mathcal{B}^j f(\hat{\Sigma}) - \mathcal{B}^j f_\delta(\hat{\Sigma})\|$ for all $j = 1, \dots, k$, and the dimension d in the above argument can be replaced by an arbitrary upper bound d'

satisfying condition (8.25) (in particular, by $d' = \bar{d}$). Thus, under condition (8.25), with probability at least $1 - e^{-\bar{d}}$,

$$\|\mathcal{B}^j f(\hat{\Sigma}) - \mathcal{B}^j f_\delta(\hat{\Sigma})\| \le j 2^{j+1} e^{-\bar{d}} \|f\|_{L_\infty}, \quad j = 1, \dots, k,$$

and also $\|f(\hat{\Sigma}) - f_\delta(\hat{\Sigma})\| \le 2e^{-\bar{d}}\|f\|_{L_\infty}$. This immediately implies that, under the assumption $\sigma(\Sigma) \subset [2\delta, \infty)$ and condition (8.25), with probability at least $1 - e^{-\bar{d}}$, the bound (8.26) holds. □

Define

$$\xi := \frac{\sqrt{n}\,(\langle f_k(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma; B)}.$$

It follows from (8.26) that with probability at least $1 - e^{-\bar{d}}$,

$$|\xi - \eta| \le \frac{(k^2 2^{k+1} + 2)\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma; B)} \sqrt{n}\, e^{-\bar{d}},$$

and we can conclude that, under the conditions $d \ge 3\log n$ and (8.25), for some $C > 1$,

$$\delta(\xi, \eta) \le \frac{(k^2 2^{k+1} + 2)\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma; B)} \sqrt{n}\, e^{-d} + e^{-d} \le C^k \frac{\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma; B)} n^{-2} + n^{-3}.$$

Combining this with (8.24) and using Lemma 10 shows, with some $C > 1$,

$$\Delta(\xi, Z) \le C^{k^2} M_{f,\delta}(B; \Sigma)[n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}] + \frac{C}{\sqrt{n}} + C^k \frac{\|f\|_{L_\infty}\|B\|_1}{\sigma_f(\Sigma; B)} n^{-2}. \tag{8.29}$$

It remains to choose $\delta := \frac{1}{2\|\Sigma^{-1}\|}$ (which implies that $\sigma(\Sigma) \subset [2\delta, \infty)$). Since

$$\frac{\log^2(1 + \delta/\|\Sigma\|)}{C_1^2(k+1)^2} \ge \frac{\log^2(1 + 1/(2(\|\Sigma\| \vee \|\Sigma^{-1}\|)^2))}{C_1^2 s^2} \ge \frac{c_{1,s}}{(\|\Sigma\| \vee \|\Sigma^{-1}\|)^4} \tag{8.30}$$

for a sufficiently small constant $c_{1,s}$, condition (8.25) follows from assumption (8.6) on d. The bound of Theorem 11 immediately follows from (8.29).

It remains to prove Proposition 8. It follows from (8.26) that, for $t \in [1, \bar{d}]$ and $3\log n \le d \le \bar{d}$, with probability at least $1 - e^{-t}$,

$$\|f_k(\hat{\Sigma}) - (f_\delta)_k(\hat{\Sigma})\| \le (k^2 2^{k+1} + 2) n^{-3}\|f\|_{L_\infty} \le (k^2 2^{k+1} + 2)\|f\|_{L_\infty}\sqrt{t/n}. \tag{8.31}$$

Due to the trivial bound $\|f_k(\hat{\Sigma}) - (f_\delta)_k(\hat{\Sigma})\| \le 2^{k+1}\|f\|_{L_\infty}$ and (8.30), we get, for $t \ge \bar{d}$,

$$\|f_k(\hat{\Sigma}) - (f_\delta)_k(\hat{\Sigma})\| \le \frac{2^{k+1}\|f\|_{L_\infty}}{\sqrt{\bar{d}}}\sqrt{t} \le \frac{2^{k+1}\|f\|_{L_\infty}(\|\Sigma\| \vee \|\Sigma\|^{-1})^2}{\sqrt{c_{1,s}}}\sqrt{\frac{t}{n}}. \tag{8.32}$$

It follows from (8.31) and (8.32) that there exists a constant $C > 1$ such that for all $t \geq 1$, with probability at least $1 - e^{-t}$,

$$\|f_k(\hat{\Sigma}) - (f_\delta)_k(\hat{\Sigma})\| \leq C^k \|f\|_{L_\infty}(\|\Sigma\| \vee \|\Sigma\|^{-1})^2 \sqrt{t/n}.$$

This implies that for some $C > 1$, with the same probability,

$$|\xi - \eta| \leq C^k \frac{\|B\|_1 \|f\|_{L_\infty}(\|\Sigma\| \vee \|\Sigma\|^{-1})^2}{\sigma_f(\Sigma; B)} \sqrt{t}. \tag{8.33}$$

Applying the bound of Proposition 7 to $\mathcal{D}g_k = (f_\delta)_k$, we find that for some constant $C > 1$ and all $t \geq 1$, with probability at least $1 - e^{-t}$ we have $|\eta| \leq C^{k^2} M_{f,\delta}(B; \Sigma)\sqrt{t}$. Combining this with (8.33) shows that for some $C > 1$ and all $t \geq 1$, with probability at least $1 - e^{-t}$ we have $|\xi| \leq C^{k^2} M_{f,\delta}(B; \Sigma)\sqrt{t}$, which, taking into account that $\delta = \frac{1}{2\|\Sigma^{-1}\|}$, completes the proof of Proposition 8. □

Proof of Theorem 3. If $d < 3\log n$, the claims of Theorem 3 easily follow from Corollary 1. For $d \geq 3\log n$, (1.10) immediately follows from the bound of Theorem 11 (take the supremum over the class of covariances $S(d_n; a) \cap \{\sigma_f(\Sigma; B) \geq \sigma_0\}$ and over all the operators B with $\|B\|_1 \leq 1$, and pass to the limit as $n \to \infty$).

To prove (1.11), we apply Lemmas 13 and 14 to $\xi := \xi(\Sigma) := \frac{\sqrt{n}\,(\langle f_k(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma; B)}$ and $\eta := Z$. Using (8.8) and (4.16), we get $\mathbb{E}\ell^2(\xi) \leq 2e\sqrt{2\pi}\, c_1^2 e^{2c_2^2\tau^2}$, where $\tau := 2C^{k^2} M_f(B; \Sigma)$. Using (4.17), (4.16), easy bounds on $\mathbb{E}\ell^2(Z)$, $\mathbb{P}\{|Z| \geq A\}$, and the bound of Theorem 11, we get

$$\begin{aligned}
&|\mathbb{E}\ell(\xi) - \mathbb{E}\ell(Z)| \\
&\quad \leq 4c_1^2 e^{2c_2 A^2}\left[C^{k^2} M_f(B; \Sigma)\big(n^{-\frac{k+\beta-\alpha(k+1+\beta)}{2}} + n^{-(1-\alpha)\beta/2}\sqrt{\log n}\big) + \frac{C}{\sqrt{n}}\right] \\
&\qquad + \sqrt{2e}\,(2\pi)^{1/4} c_1 e^{c_2^2\tau^2} e^{-A^2/(2\tau^2)} + c_1 e^{c_2^2} e^{-A^2/4}.
\end{aligned}$$

It remains to take the supremum over the class of covariances $S(d_n; a) \cap \{\sigma_f(\Sigma; B) \geq \sigma_0\}$ and over all the operators B with $\|B\|_1 \leq 1$, and to let first $n \to \infty$ and then $A \to \infty$. □

9. Lower bounds

Our main goal in this section is to prove Theorem 4 stated in Subsection 1.2.

The main part of the proof is based on an application of the van Trees inequality and follows the same lines as the proof of a minimax lower bound for estimation of linear functionals of principal components in [KLN]. We will need the following lemma (possibly of independent interest) showing the Lipschitz property of the function $\Sigma \mapsto \sigma_f^2(\Sigma; B)$. It holds for an arbitrary separable Hilbert space $\mathbb{H}$ (not necessarily finite-dimensional).

Lemma 26. *Suppose $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$. Then*

$$
\begin{aligned}
&|\sigma_f^2(\Sigma+H;B)-\sigma_f^2(\Sigma;B)| \\
&\quad \le \|f'\|_{L_\infty}(2\|\Sigma\|+\|H\|)\|B\|_1^2\big[2\|f'\|_{L_\infty}\|H\|+8\|f\|_{B^s_{\infty,1}}\|\Sigma\|\,\|H\|^{s-1}\big].
\end{aligned} \tag{9.1}
$$

Proof. Note that

$$
\begin{aligned}
\sigma_f^2(\Sigma;B) &= 2\|\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}\|_2^2 \\
&= 2\,\mathrm{tr}\big(\Sigma^{1/2}Df(\Sigma;B)\Sigma Df(\Sigma;B)\Sigma^{1/2}\big) = 2\,\mathrm{tr}\big(\Sigma Df(\Sigma;B)\Sigma Df(\Sigma;B)\big).
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&\sigma_f^2(\Sigma+H;B)-\sigma_f^2(\Sigma;B) \\
&\quad = 2\,\mathrm{tr}\big(HDf(\Sigma+H;B)(\Sigma+H)Df(\Sigma+H;B)\big) \\
&\qquad + 2\,\mathrm{tr}\big(\Sigma(Df(\Sigma+H;B)-Df(\Sigma;B))(\Sigma+H)Df(\Sigma+H;B)\big) \\
&\qquad + 2\,\mathrm{tr}\big(\Sigma Df(\Sigma;B)HDf(\Sigma+H;B)\big) \\
&\qquad + 2\,\mathrm{tr}\big(\Sigma Df(\Sigma;B)\Sigma(Df(\Sigma+H;B)-Df(\Sigma;B)\big).
\end{aligned} \tag{9.2}
$$

Then

$$
\begin{aligned}
&\big|2\,\mathrm{tr}\big(HDf(\Sigma+H;B)(\Sigma+H)Df(\Sigma+H;B)\big)\big| \\
&\qquad \le 2\|Df(\Sigma+H;B)(\Sigma+H)Df(\Sigma+H;B)\|_1\|H\| \\
&\qquad \le 2\|\Sigma+H\|\,\|Df(\Sigma+H;B)Df(\Sigma+H;B)\|_1\|H\| \\
&\qquad \le 2\|\Sigma+H\|\,\|Df(\Sigma+H;B)\|_2^2\|H\| \\
&\qquad \le 2\|f'\|_{L_\infty}^2(\|\Sigma\|+\|H\|)\|H\|\,\|B\|_2^2.
\end{aligned} \tag{9.3}
$$

Similarly, it can be shown that

$$
\big|2\,\mathrm{tr}\big(\Sigma Df(\Sigma;B)HDf(\Sigma+H;B)\big)\big| \le 2\|f'\|_{L_\infty}^2\|\Sigma\|\,\|H\|\,\|B\|_2^2. \tag{9.4}
$$

Also, we have

$$
\begin{aligned}
&2\,\mathrm{tr}\big(\Sigma(Df(\Sigma+H;B)-Df(\Sigma;B))(\Sigma+H)Df(\Sigma+H;B)\big) \\
&\quad = \langle (Df(\Sigma+H)-Df(\Sigma))(B),C\rangle = \langle (Df(\Sigma+H)-Df(\Sigma))(C),B\rangle \\
&\quad = \langle Df(\Sigma+H;C)-Df(\Sigma;C),B\rangle,
\end{aligned}
$$

where $C := (\Sigma+H)Df(\Sigma;B)\Sigma + \Sigma Df(\Sigma;B)(\Sigma+H)$. Using (2.25), this implies

$$
\begin{aligned}
&\big|2\,\mathrm{tr}\big(\Sigma(Df(\Sigma+H;B)-Df(\Sigma;B))(\Sigma+H)Df(\Sigma+H;B)\big)\big| \\
&\quad = |\langle Df(\Sigma+H;C)-Df(\Sigma;C),B\rangle| \le \|Df(\Sigma+H;C)-Df(\Sigma;C)\|\,\|B\|_1 \\
&\quad \le 4\|f\|_{B^s_{\infty,1}}\|C\|\,\|H\|^{s-1}\|B\|_1 \le 8\|f\|_{B^s_{\infty,1}}\|\Sigma\|\,\|\Sigma+H\|\,\|Df(\Sigma;B)\|\,\|H\|^{s-1}\|B\|_1 \\
&\quad \le 8\|f'\|_{L_\infty}\|f\|_{B^s_{\infty,1}}\|\Sigma\|(\|\Sigma\|+\|H\|)\|H\|^{s-1}\|B\|_1^2.
\end{aligned} \tag{9.5}
$$

Similarly,

$$\begin{aligned}&\big|2\,\mathrm{tr}\big(\Sigma Df(\Sigma;B)\Sigma(Df(\Sigma+H;B)-Df(\Sigma;B)\big)\big|\\&\qquad\le 8\|f'\|_{L_\infty}\|f\|_{B^s_{\infty,1}}\|\Sigma\|^2\|H\|^{s-1}\|B\|_1^2.\end{aligned}\tag{9.6}$$

Substituting (9.3)–(9.6) into (9.2), we get (9.1). □

For given $a'\in(1,a)$ and $\sigma_0'>\sigma_0$, assume that $\mathfrak{B}_f(d_n;a';\sigma_0')\neq\emptyset$ (otherwise, the proof becomes trivial) and, for B with $\|B\|_1\le 1$ such that $\mathring{\mathcal{S}}_{f,B}(d_n;a';\sigma_0')\neq\emptyset$, consider $\Sigma_0\in\mathring{\mathcal{S}}_{f,B}(d_n;a';\sigma_0')$. For $H\in\mathcal{B}_{\mathrm{sa}}(\mathbb{H})$ and $c>0$, define

$$\Sigma_t:=\Sigma_0+\frac{tH}{\sqrt{n}}\quad\text{and}\quad\mathcal{S}_{c,n}(\Sigma_0,H):=\{\Sigma_t:t\in[-c,c]\}.$$

In what follows, H will be chosen so that

$$\|H\|\le\|f'\|_{L_\infty}a^2.\tag{9.7}$$

Recall that the set $\mathring{\mathcal{S}}_{f,B}(d_n;a;\sigma_0)$ is open in the operator norm topology, so Σ_0 is its interior point. Moreover, let $\delta>0$ and suppose that $\|\Sigma-\Sigma_0\|<\delta$. If $\delta<a-a'$, then $\|\Sigma\|<a$. If $\delta<\frac{a-a'}{2a^2}$, then it is easy to check that $\|\Sigma^{-1}\|<a$. Also, using the bound of Lemma 26, it is easy to show that, for B with $\|B\|_1\le 1$, the condition

$$\|f'\|_{L_\infty}(2a+\delta)[2\|f'\|_{L_\infty}\delta+8\|f\|_{B^s_{\infty,1}}a\delta^{s-1}]\le(\sigma_0')^2-\sigma_0^2\tag{9.8}$$

implies that $\sigma(\Sigma;B)>\sigma_0$. Thus, for a small enough $\delta=\delta(f,s,a,a',\sigma_0,\sigma_0')\in(0,1)$ satisfying $\delta<\frac{a-a'}{2a^2}$ and (9.8), we have

$$B(\Sigma_0;\delta):=\{\Sigma:\|\Sigma-\Sigma_0\|<\delta\}\subset\mathring{\mathcal{S}}_{f,B}(d_n;a;\sigma_0).$$

For given c and δ, for H satisfying (9.7) and for all large enough n (more specifically, for $n>c^2a^4\|f'\|^2_{L_\infty}/\delta^2$), we have

$$c\|H\|/\sqrt{n}<\delta,\tag{9.9}$$

implying that $\mathcal{S}_{c,n}(\Sigma_0,H)\subset B(\Sigma_0;\delta)\subset\mathring{\mathcal{S}}_{f,B}(d_n;a;\sigma_0)$. Define

$$\varphi(t):=\langle f(\Sigma_t),B\rangle,\qquad t\in[-c,c].$$

Clearly, φ is continuously differentiable with

$$\varphi'(t)=\frac{1}{\sqrt{n}}\langle Df(\Sigma_t;H),B\rangle,\qquad t\in[-c,c].\tag{9.10}$$

Consider the following parametric model:

$$X_1,\dots,X_n\ \text{i.i.d.}\sim N(0;\Sigma_t),\qquad t\in[-c,c].\tag{9.11}$$

It is well known that the Fisher information matrix for model $X\sim N(0;\Sigma)$ with non-singular covariance Σ is $I(\Sigma)=\frac{1}{2}(\Sigma^{-1}\otimes\Sigma^{-1})$ (see, e.g., [Eat]). This implies that the Fisher information for model $X\sim N(0,\Sigma_t)$, $t\in[-c,c]$, is $I(t)=\langle I(\Sigma_t)\Sigma_t',\Sigma_t'\rangle=$

$\frac{1}{n}\langle I(\Sigma_t)H, H\rangle$, and for (9.11) it is

$$I_n(t) = nI(t) = \langle I(\Sigma_t)H, H\rangle = \tfrac{1}{2}\langle(\Sigma_t^{-1}\otimes\Sigma_t^{-1})H, H\rangle$$
$$= \tfrac{1}{2}\langle\Sigma_t^{-1}H\Sigma_t^{-1}, H\rangle = \tfrac{1}{2}\,\mathrm{tr}(\Sigma_t^{-1}H\Sigma_t^{-1}H) = \tfrac{1}{2}\|\Sigma_t^{-1/2}H\Sigma_t^{-1/2}\|_2^2.$$

We will now use the well known van Trees inequality (see, e.g., [GL]) that provides a lower bound on the average risk of an arbitrary estimator $T(X_1,\dots,X_n)$ of a smooth function $\varphi(t)$ of parameter t of model (9.11) with respect to a smooth prior density π_c on $[-c, c]$ such that $J_{\pi_c} := \int_{-c}^{c}\frac{(\pi_c'(t))^2}{\pi_c(t)}\,dt < \infty$ and $\pi_c(c) = \pi_c(-c) = 0$. It follows from that inequality that

$$\sup_{t\in[-c,c]}\mathbb{E}_t\big(T_n(X_1,\dots,X_n) - g(t)\big)^2$$
$$\geq \int_{-c}^{c}\mathbb{E}_t\big(T_n(X_1,\dots,X_n) - g(t)\big)^2\pi_c(t)\,dt \geq \frac{(\int_{-c}^{c}\varphi'(t)\pi_c(t)\,dt)^2}{\int_{-c}^{c}I_n(t)\pi_c(t)\,dt + J_{\pi_c}}. \tag{9.12}$$

A common choice of prior is $\pi_c(t) := c^{-1}\pi(t/c)$ for a smooth density π on $[-1, 1]$ with $\pi(1) = \pi(-1) = 0$ and $J_\pi := \int_{-1}^{1}\frac{(\pi'(t))^2}{\pi(t)}\,dt < \infty$. In this case, $J_{\pi_c} = c^{-2}J_\pi$. Next we provide bounds on the numerator and the denominator of the right hand side of (9.12).

For the numerator, we get

$$\Big(\int_{-c}^{c}\varphi'(t)\pi_c(t)\,dt\Big)^2 = \Big(\int_{-c}^{c}[\varphi'(0) + (\varphi'(t) - \varphi'(0))]\pi(t/c)\,dt/c\Big)^2$$
$$\geq (\varphi'(0))^2 + 2\varphi'(0)\int_{-c}^{c}(\varphi'(t) - \varphi'(0))\pi(t/c)\,dt/c$$
$$\geq (\varphi'(0))^2 - 2|\varphi'(0)|\int_{-c}^{c}|\varphi'(t) - \varphi'(0)|\pi(t/c)\,dt/c.$$

Using (9.10) along with the bound (based on (2.25))

$$|\varphi'(t) - \varphi'(0)| \leq \frac{1}{\sqrt{n}}\|Df(\Sigma_t; H) - Df(\Sigma_0; H)\|\,\|B\|_1$$
$$\leq \frac{4}{\sqrt{n}}\|f\|_{B^s_{\infty,1}}\|\Sigma_t - \Sigma_0\|^{s-1}\|H\|\,\|B\|_1 \leq \frac{4}{n^{s/2}}\|f\|_{B^s_{\infty,1}}\|H\|^s\|B\|_1|t|^{s-1},$$

we get

$$\Big(\int_{-c}^{c}\varphi'(t)\pi_c(t)\,dt\Big)^2 \geq \frac{1}{n}\langle Df(\Sigma_0; H), B\rangle^2$$
$$-\frac{2}{\sqrt{n}}|\langle Df(\Sigma_0; H), B\rangle|\frac{4}{n^{s/2}}\|f\|_{B^s_{\infty,1}}\|H\|^s\|B\|_1\int_{-c}^{c}|t|^{s-1}\pi(t/c)\,dt/c$$
$$= \frac{1}{n}\langle Df(\Sigma_0; H), B\rangle^2 - \frac{8\|f\|_{B^s_{\infty,1}}\|H\|^s\|B\|_1c^{s-1}}{n^{(1+s)/2}}|\langle Df(\Sigma_0; H), B\rangle|\int_{-1}^{1}|t|^{s-1}\pi(t)\,dt$$
$$\geq \frac{1}{n}\langle Df(\Sigma_0; H), B\rangle^2 - \frac{8\|f\|_{B^s_{\infty,1}}\|H\|^s\|B\|_1c^{s-1}}{n^{(1+s)/2}}|\langle Df(\Sigma_0; H), B\rangle|. \tag{9.13}$$

Observing that

$$\langle Df(\Sigma_0; H), B\rangle = \langle Df(\Sigma_0; B), H\rangle = \langle \Sigma_0^{-1/2} D \Sigma_0^{-1/2}, \Sigma_0^{-1/2} H \Sigma_0^{-1/2}\rangle,$$

where $D := \Sigma_0 Df(\Sigma_0; B)\Sigma_0$, we can rewrite (9.13) as

$$\Big(\int_{-c}^{c} \varphi'(t)\pi_c(t)\,dt\Big)^2 \geq \frac{1}{n}\langle \Sigma_0^{-1/2} D \Sigma_0^{-1/2}, \Sigma_0^{-1/2} H \Sigma_0^{-1/2}\rangle^2 - |\langle \Sigma_0^{-1/2} D \Sigma_0^{-1/2}, \Sigma_0^{-1/2} H \Sigma_0^{-1/2}\rangle| \frac{8\|f\|_{B^s_{\infty,1}}\|H\|^s\|B\|_1 c^{s-1}}{n^{(1+s)/2}}. \quad (9.14)$$

To bound the denominator, we need to control $I_n(t) = \frac{1}{2}\operatorname{tr}(\Sigma_t^{-1} H \Sigma_t^{-1} H)$ in terms of $I_n(0) = \frac{1}{2}\operatorname{tr}(\Sigma_0^{-1} H \Sigma_0^{-1} H)$. To this end, note that

$$\Sigma_t^{-1} = \Sigma_0^{-1} + C\Sigma_0^{-1},$$

where $C := \big(I + \frac{t\Sigma_0^{-1}H}{\sqrt{n}}\big)^{-1} - I$. Suppose H satisfies

$$\frac{c\|\Sigma_0^{-1}H\|}{\sqrt{n}} \leq \frac{1}{2}, \quad (9.15)$$

which also implies that $\|C\| \leq 2|t|\,\|\Sigma_0^{-1}H\|/\sqrt{n} \leq 1$. Note also that

$$\operatorname{tr}(\Sigma_t^{-1} H \Sigma_t^{-1} H) = \operatorname{tr}(\Sigma_0^{-1} H \Sigma_0^{-1} H) + 2\operatorname{tr}(C\Sigma_0^{-1} H \Sigma_0^{-1} H) + \operatorname{tr}(C\Sigma_0^{-1} H C \Sigma_0^{-1} H).$$

Therefore,

$$\begin{aligned} I_n(t) &\leq I_n(0) + \|C\|\,\|\Sigma_0^{-1} H \Sigma_0^{-1} H\|_1 + \tfrac{1}{2}\|C\Sigma_0^{-1}H\|_2\|H\Sigma_0^{-1}C\|_2 \\ &\leq I_n(0) + (\|C\| + \|C\|^2/2)\|\Sigma_0^{-1}H\|_2^2 \leq I_n(0) + 3|t|\,\|\Sigma_0^{-1}H\|_2^3/\sqrt{n} \end{aligned}$$

and

$$\begin{aligned} \int_{-c}^{c} I_n(t)\pi(t/c)\,dt/c &\leq I_n(0) + 3\frac{\|\Sigma_0^{-1}H\|_2^3}{\sqrt{n}}\int_{-c}^{c}|t|\pi(t/c)\,dt/c \\ &\leq \tfrac{1}{2}\|\Sigma_0^{-1/2} H \Sigma_0^{-1/2}\|_2^2 + 3c\frac{\|\Sigma_0^{-1}H\|_2^3}{\sqrt{n}}. \end{aligned} \quad (9.16)$$

Substituting (9.14) and (9.16) into (9.12), we get

$$\sup_{t\in[-c,c]} n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2 \geq$$

$$\frac{\langle \Sigma_0^{-1/2} D \Sigma_0^{-1/2}, \Sigma_0^{-1/2} H \Sigma_0^{-1/2}\rangle^2 - |\langle \Sigma_0^{-1/2} D \Sigma_0^{-1/2}, \Sigma_0^{-1/2} H \Sigma_0^{-1/2}\rangle| \frac{8\|f\|_{B^s_{\infty,1}}\|H\|^s\|B\|_1 c^{s-1}}{n^{(s-1)/2}}}{\frac{1}{2}\|\Sigma_0^{-1/2} H \Sigma_0^{-1/2}\|_2^2 + 3c\frac{\|\Sigma_0^{-1}H\|_2^3}{\sqrt{n}} + \frac{J_\pi}{c^2}}. \quad (9.17)$$

Note that

$$\|\Sigma_0^{-1/2} D\Sigma_0^{-1/2}\|_2^2 = \|\Sigma_0^{1/2} Df(\Sigma_0; B)\Sigma_0^{1/2}\|_2^2 = \tfrac{1}{2}\sigma_f^2(\Sigma_0; B).$$

In what follows, we use $H := D$, which clearly satisfies (9.7) since, for $\Sigma_0 \in \mathring{\mathcal{S}}_{f,B}(d_n; a; \sigma_0)$ and $\|B\|_1 \le 1$,

$$\begin{aligned}\|D\| &= \|\Sigma_0 Df(\Sigma_0; B)\Sigma_0\| \le \|\Sigma_0\|^2 \|Df(\Sigma_0; B)\| \\ &\le \|f'\|_{L_\infty} \|B\|_2 \|\Sigma\|_0^2 \le a^2 \|f'\|_{L_\infty} \|B\|_2 \le a^2 \|f'\|_{L_\infty}.\end{aligned} \tag{9.18}$$

We also have

$$\begin{aligned}&\|\Sigma_0^{-1} D\|_2^2 = \mathrm{tr}(Df(\Sigma_0; B)\Sigma_0^2 Df(\Sigma_0, B)) \\ &\quad\le \|\Sigma_0\|^2 \|Df(\Sigma_0; B)\|_2^2 \le \|\Sigma_0\|^2 \|f'\|_{L_\infty}^2 \|B\|_2^2 \le a^2 \|f'\|_{L_\infty}^2 \|B\|_2^2 \le a^2 \|f'\|_{L_\infty}^2,\end{aligned} \tag{9.19}$$

implying that (9.9) and (9.15) hold for $H = D$ provided that

$$n > 4c^2 a^4 \|f'\|_{L_\infty}^2 / \delta^2. \tag{9.20}$$

With this choice of H, (9.17) implies

$$\sup_{t\in[-c,c]} \frac{n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2}{\sigma_f^2(\Sigma_0; B)} \ge 1 - \frac{3c\frac{\|\Sigma_0^{-1}D\|_2^3}{\sqrt{n}} + \frac{4\|f\|_{B^s_{\infty,1}} \|D\|^s \|B\|_1 c^{s-1}}{n^{(s-1)/2}} + \frac{J_\pi}{c^2}}{\frac{1}{4}\sigma_f^2(\Sigma_0; B) + 3c\frac{\|\Sigma_0^{-1}D\|_2^3}{\sqrt{n}} + \frac{J_\pi}{c^2}}. \tag{9.21}$$

It follows from (9.21), (9.18) and (9.19) that for B satisfying $\|B\|_1 \le 1$,

$$\sup_{t\in[-c,c]} \frac{n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2}{\sigma_f^2(\Sigma_0; B)} \ge 1 - \gamma_{n,c}(f; a; \sigma_0) \tag{9.22}$$

where

$$\gamma_{n,c}(f; a; \sigma_0) := \frac{\frac{3a^3\|f'\|_{L_\infty}^3 c}{\sqrt{n}} + \frac{4a^{2s}\|f\|_{B^s_{\infty,1}} \|f'\|_{L_\infty}^s c^{s-1}}{n^{(s-1)/2}} + \frac{J_\pi}{c^2}}{\frac{1}{4}\sigma_0^2}.$$

Denote $\sigma^2(t) := \sigma_f^2(\Sigma_t; B)$, $t \in [-c, c]$. By Lemma 26,

$$\begin{aligned}|\sigma^2(t) - \sigma^2(0)| &\le \|f'\|_{L_\infty}\left(2\|\Sigma_0\| + \frac{|t|\,\|D\|}{\sqrt{n}}\right)\|B\|_1^2 \\ &\quad\times \left[2\|f\|_{L_\infty}\frac{|t|\,\|D\|}{\sqrt{n}} + 8\|f\|_{B^s_{\infty,1}}\|\Sigma_0\|\frac{|t|^{s-1}\|D\|^{s-1}}{n^{(s-1)/2}}\right].\end{aligned}$$

Note that assumption (9.15) on $H = D$ implies that

$$\frac{c\|D\|}{\sqrt{n}} = \frac{c\|\Sigma_0\Sigma_0^{-1}H\|}{\sqrt{n}} \le \frac{c\|\Sigma_0^{-1}H\|\,\|\Sigma_0\|}{\sqrt{n}} \le \frac{\|\Sigma_0\|}{2}.$$

Using bound (9.18), we get that, for all $t \in [-c, c]$ and B with $\|B\|_1 \le 1$,

$$|\sigma^2(t) - \sigma^2(0)| \le \frac{6ca^3\|f'\|^3_{L_\infty}}{n^{1/2}} + \frac{24c^{s-1}a^{2s}\|f'\|^s_{L_\infty}\|f\|_{B^s_{\infty,1}}}{n^{(s-1)/2}} =: \lambda_{n,c}(f; a).$$

which implies that

$$\sup_{t\in[-c,c]} \frac{\sigma^2(t)}{\sigma^2(0)} \le 1 + \frac{\lambda_{n,c}(f; a)}{\sigma_0^2}. \tag{9.23}$$

It follows from (9.22) and (9.23) that

$$\begin{aligned}
&\sup_{t\in[-c,c]} \frac{n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2}{\sigma^2(t)}\left(1 + \frac{\lambda_{n,c}(f; a)}{\sigma_0^2}\right)\\
&\quad \ge \sup_{t\in[-c,c]} \frac{n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2}{\sigma^2(t)} \sup_{t\in[-c,c]} \frac{\sigma^2(t)}{\sigma^2(0)}\\
&\quad \ge \sup_{t\in[-c,c]} \frac{n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2}{\sigma_f^2(\Sigma_0; B)} \ge 1 - \gamma_{n,c}(f; a; \sigma_0),
\end{aligned}$$

which implies that for all $B \in \mathfrak{B}_f(d_n; a'; \sigma_0')$,

$$\begin{aligned}
&\sup_{\Sigma\in\mathring{\mathcal{S}}_{f,B}(d_n;a;\sigma_0)} \frac{n\mathbb{E}_\Sigma(T_n(X_1,\dots,X_n) - \langle f(\Sigma), B\rangle)^2}{\sigma_f^2(\Sigma; B)}\\
&\quad \ge \sup_{t\in[-c,c]} \frac{n\mathbb{E}_t(T_n(X_1,\dots,X_n) - g(t))^2}{\sigma^2(t)} \ge \frac{1 - \gamma_{n,c}(f; a; \sigma_0)}{1 + \lambda_{n,c}(f; a)/\sigma_0^2}.
\end{aligned} \tag{9.24}$$

It remains to observe that

$$\lim_{c\to\infty} \limsup_{n\to\infty} \gamma_{n,c}(f; a; \sigma_0) = 0 \quad \text{and} \quad \lim_{c\to\infty} \limsup_{n\to\infty} \lambda_{n,c}(f; a) = 0$$

to complete the proof. □

Remark 10. It follows from the proof that the following local version of (1.12) also holds: for all $a' \in (1, a)$ and $\sigma_0' > \sigma_0$,

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \inf_{T_n} \inf_{B\in\mathfrak{B}_f(d_n;a';\sigma_0')} \inf_{\Sigma_0\in\mathring{\mathcal{S}}_{f,B}(d_n;a';\sigma_0')} \sup_{\|\Sigma-\Sigma_0\|<\delta} \frac{n\mathbb{E}_\Sigma(T_n - \langle f(\Sigma), B\rangle)^2}{\sigma_f^2(\Sigma; B)} \ge 1. \tag{9.25}$$

Acknowledgments. The author is grateful to Richard Nickl, Alexandre Tsybakov and Mayya Zhilova for several helpful conversations and to Anna Skripka for careful reading of Section 2 of the paper and making several useful suggestions. A part of this work was done while on leave to the University of Cambridge. The author wishes to thank the Department of Pure Mathematics and Mathematical Statistics of the university for its hospitality. Finally, the author is grateful to the referees for numerous helpful comments. This research was partially supported by NSF Grants DMS-1810958, DMS-1509739 and CCF-1523768.

References

[AP1] Aleksandrov, A., Peller, V.: Functions of perturbed unbounded self-adjoint operators. Operator Bernstein type inequalities. Indiana Univ. Math. J. **59**, 1451–1490 (2010) Zbl 1223.47015 MR 2815039

[AP2] Aleksandrov, A., Peller, V.: Operator Lipschitz functions. arXiv:1611.01593 (2016)

[A] Anderson, T. W.: An Introduction to Multivariate Statistical Analysis. Wiley-Interscience, Hoboken, NJ (2003) Zbl 1039.62044 MR 1990662

[ACDS] Azamov, N. A., Carey, A. L., Dodds, P. G., Sukochev, F. A.: Operator integrals, spectral shift, and spectral flow. Canad. J. Math. **61**, 241–263 (2009) Zbl 1163.47008 MR 2504014

[BaiS] Bai, Z. D., Silverstein, J. W.: CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. **32**, 553–605 (2004) Zbl 1063.60022 MR 2040792

[Bh] Bhatia, R.: Matrix Analysis. Springer (1997) Zbl 0863.15001 MR 1477662

[BKRW] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., Wellner, J. A.: Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins Univ. Press, Baltimore, MD (1993) Zbl 0786.62001 MR 1623559

[BJNP] Birnbaum, A., Johnstone, I. M., Nadler, B., Paul, D.: Minimax bounds for sparse PCA with noisy high-dimensional data. Ann. Statist. **41**, 1055–1084 (2013) Zbl 1292.62071 MR 3113803

[CLZ] Cai, T. T., Liang, T., Zhou, H.: Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. J. Multivariate Anal. **137**, 161–172 (2015) Zbl 1329.62255 MR 3332804

[CL1] Cai, T. T., Low, M.: On adaptive estimation of linear functionals. Ann. Statist. **33**, 2311–2343 (2005) Zbl 1086.62031 MR 2211088

[CL2] Cai, T. T., Low, M.: Nonquadratic estimators of a quadratic functional. Ann. Statist. **33**, 2930–2956 (2005) Zbl 1085.62055 MR 2253108

[CCT] Collier, O., Comminges, L., Tsybakov, A.: Minimax estimation of linear and quadratic functionals on sparsity classes. Ann. Statist. **45**, 923–958 (2017) Zbl 1368.62191 MR 3662444

[DPR] Dauxois, J., Pousse, A., Romain, Y.: Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J. Multivariate Anal. **12**, 136–154 (1982) Zbl 0539.62064 MR 0650934

[Dav] Davies, E. B.: Linear Operators and their Spectra. Cambridge Univ. Press, Cambridge (2007) Zbl 1138.47001 MR 2359869

[Eat] Eaton, M. L.: Multivariate Statistics: A Vector Space Approach. Wiley, New York (1983) Zbl 0587.62097 MR 0716321

[FRW] Fan, J., Rigollet, P., Wang, W.: Estimation of functionals of sparse covariance matrices. Ann. Statist. **43**, 2706–2737 (2015) Zbl 1327.62338 MR 3405609

[FK] Faraut, J., Korányi, A.: Analysis on Symmetric Cones. Clarendon Press, Oxford (1994) Zbl 0841.43002 MR 1446489

[GaoZ] Gao, C., Zhou, H.: Bernstein–von Mises theorems for functionals of the covariance matrix. Electron. J. Statist. **10**, 1751–1806 (2016) Zbl 1346.62059 MR 3522660

[GL] Gill, R. D., Levit, B. Y.: Applications of the van Trees inequality: a Bayesian Cramér–Rao bound. Bernoulli **1**, 59–79 (1995) Zbl 0830.62035 MR 1354456

[GN] Giné, E., Nickl, R.: Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Univ. Press (2016) Zbl 1358.62014 MR 3588285

[Gir] Girko, V. L.: Introduction to general statistical analysis. Theory Probab. Appl. **32**, 229–242 (1987) Zbl 0653.62043 MR 0902754

[Gir1] Girko, V. L.: Statistical Analysis of Observations of Increasing Dimension. Springer (1995) Zbl 0867.62047 MR 1473719

[GonZ] Gonska, H., Zhou, X.-L.: Approximation theorems for the iterated Boolean sums of Bernstein operators. J. Comput. Appl. Math. **53**, 21–31 (1994) Zbl 0816.41020 MR 1305965

[GLM] Graczyk, P., Letac, G., Massam, H.: The complex Wishart distribution and the symmetric group. Ann. Statist. **31**, 287–309 (2003) Zbl 1019.62047 MR 1962508

[GLM1] Graczyk, P., Letac, G., Massam, H.: The hyperoctahedral group, symmetric group representations and the moments of the real Wishart distribution. J. Theoret. Probab. **18**, 1–42 (2005) Zbl 1067.60003 MR 2132270

[IKh] Ibragimov, I. A., Khasminskii, R. Z.: Statistical Estimation: Asymptotic Theory. Springer, New York (1981) Zbl 0467.62026 MR 0620321

[IKhN] Ibragimov, I. A., Nemirovski, A. S., Khasminskii, R. Z.: Some problems of nonparametric estimation in Gaussian white noise. Theory Probab. Appl. **31**, 391–406 (1987) Zbl 0623.62028 MR 0866866

[James] James, A. T.: The distribution of the latent roots of the covariance matrix. Ann. Math. Statist. **31**, 151–158 (1960) Zbl 0201.52401 MR 0126901

[James1] James, A. T.: Zonal polynomials of the real positive definite symmetric matrices. Ann. of Math. **74**, 456–469 (1961) Zbl 0104.02803 MR 0140741

[James2] James, A. T.: Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist. **35**, 475–501 (1964) Zbl 0121.36605 MR 0181057

[JG] Janková, J., van de Geer, S.: Semi-parametric efficiency bounds for high-dimensional models. Ann. Statist. **46**, 2336–2359 (2018) Zbl 1420.62308 MR 3845020

[JaMont] Javanmard, A., Montanari, A.: Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. IEEE Trans. Information Theory **60**, 6522–6554 (2014) Zbl 1360.62074 MR 3265038

[JHW] Jiao, J., Han, Y. and Weissman, T.: Bias correction with jackknife, bootstrap and Taylor series. arXiv:1709.06183 (2017)

[Jo] Johnstone, I. M.: On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist. **29**, 295–327 (2001) Zbl 1016.62078 MR 1863961

[JoLu] Johnstone, I. M., Lu, A. Y.: On consistency and sparsity for principal components analysis in high dimensions. J. Amer. Statist. Assoc. **104**, 682–693 (2009) Zbl 1388.62174 MR 2751448

[Kato] Kato, T: Perturbation Theory for Linear Operators. Springer, New York (1980) Zbl 0435.47001 MR 1335452

[KS] Kissin, E., Shulman, V. S.: Classes of operator-smooth functions. II. Operator-differentiable functions. Integral Equations Operator Theory **49**, 165–210 (2004) Zbl 1061.47013 MR 2060371

[KLN] Koltchinskii, V., Löffler, M., Nickl, R.: Efficient estimation of linear functionals of principal components. Ann. Statist. **48**, 464–490 (2020) Zbl 1440.62232 MR 4065170

[KL1] Koltchinskii, V., Lounici, K.: Asymtotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. Ann. Inst. H. Poincaré Probab. Statist. **52**, 1976–2013 (2016) Zbl 1353.62053 MR 3573302

[KL2] Koltchinskii, V., Lounici, K.: Concentration inequalities and moment bounds for sample covariance operators. Bernoulli **23**, 110–133 (2017) Zbl 1366.60057 MR 3556768

[KL3] Koltchinskii, V., Lounici, K.: Normal approximation and concentration of spectral projectors of sample covariance. Ann. Statist. **45**, 121–157 (2017) Zbl 1367.62175 MR 3611488

[KL4] Koltchinskii, V., Lounici, K.: New asymptotic results in principal component analysis. Sankhya A **79**, 254–297 (2017) Zbl 06822893 MR 3707422

[KZh] Koltchinskii, V., Zhilova, M.: Efficient estimation of smooth functionals in Gaussian shift models. arXiv:1810.02767 (2018)

[KoVa] Kong, W., Valiant, G.: Spectrum estimation from samples. Ann. Statist. **45**, 2218–2247 (2017) Zbl 06821124 MR 3718167

[Led] Ledoux, M.: The Concentration of Measure Phenomenon. Amer. Math. Soc. (2001) Zbl 0995.60002 MR 1849347

[LetMas] Letac, G., Massam, H.: All invariant moments of the Wishart distribution. Scand. J. Statist. **31**, 295–318 (2004) Zbl 1063.62081 MR 2066255

[Lev1] Levit, B.: On the efficiency of a class of non-parametric estimates. Theory Probab. Appl. **20**, 723–740 (1975) Zbl 0367.62041 MR 0403052

[Lev2] Levit, B.: Asymptotically efficient estimation of nonlinear functionals. Problemy Peredachi Informatsii **14**, no. 3, 65–72 (1978) (in Russian) Zbl 0422.62034 MR 0533450

[LP] Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. Ann. Probab. **37**, 1778–1840 (2009) Zbl 1180.15029 MR 2561434

[NSU] Naumov, A., Spokoiny, V., Ulyanov, V.: Bootstrap confidence sets for spectral projectors of sample covariance. Probab. Theory Related Fields **174**, 1091–1132 (2019) Zbl 1420.62073 MR 3980312

[Nem1] Nemirovski, A.: On necessary conditions for the efficient estimation of functionals of a nonparametric signal which is observed in white noise. Theory Probab. Appl. **35**, 94–103 (1990) Zbl 0721.62043 MR 1050056

[Nem2] Nemirovski, A.: Topics in non-parametric statistics. In: Lectures on Probability Theory and Statisyics (Saint-Flour, 1998), Lecture Notes in Math. 1738, Springer, New York, 85–277 (2000) Zbl 0998.62033 MR 1775640

[Nik] Nikol'skii, S. M.: Inequalities for entire functions of finite degree and their applications in the theory of differentiable functions of several variables. Trudy Mat. Inst. Steklova **38**, 244–278 (1951) (in Russian) MR 0048565

[Paul] Paul, D.: Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statist. Sinica **17**, 1617–1642 (2007) Zbl 1134.62029 MR 2399865

[Pel] Peller, V. V.: Hankel operators in the perturbation theory of unitary and self-adjoint operators. Funktsional. Anal. i Prilozhen. **19**, no. 2, 37–51 (1985) (in Russian), English transl.: Funct. Anal. Appl. **19**, 111–123 (1985) Zbl 0587.47016 MR 0800919

[Pel1] Peller, V. V.: Multiple operator integrals and higher operator derivatives. J. Funct. Anal. **233**, 515–544 (2006) Zbl 1102.46024 MR 2214586

[Pel2] Peller, V. V.: Multiple operator integrals in perturbation theory. Bull. Math. Sci. **6**, 15–88 (2016) Zbl 1348.46048 MR 3472849

[Pet] Petrov, V. V.: Sums of Independent Random Variables. Springer (1975) Zbl 0322.60043 MR 0388499

[RW] Reiss, M., Wahl, M.: Non-asymptotic upper bounds for the reconstruction error of PCA. Ann. Statist. **48**, 1098–1123 (2020) Zbl 07241582 MR 4102689

[Skr] Skripka, A.: Taylor approximations of operator functions. In: Operator Theory in Harmonic and Non-Commutative Analysis, Oper. Theory Adv. Appl. 240, Birkhäuser, Basel, 243–256 (2014) Zbl 1343.47018 MR 3134555

[SW] Sosoe, P., Wong, P.: Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices. Adv. Math. **249**, 37–87 (2013) Zbl 1292.60008 MR 3116567

[Tot] Totik, V.: Approximation by Bernstein polynomials. Amer. J. Math. **116**, 995–1018 (1994) Zbl 0812.41018 MR 1287945

[Tr] Triebel, H.: Theory of Function Spaces. Birkhäuser, Basel (1983) Zbl 0546.46028 MR 0781540

[GBRD] van de Geer, S., Bühlmann, P., Ritov, Y., Dezeure, R.: On asymptotically optimal confidence regions and tests for high-dimensional models. Ann. Statist. **42**, 1166–1202 (2014) Zbl 1305.62259 MR 3224285

[Ver] Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. In: Compressed Sensing, Cambridge Univ. Press, Cambridge, 210–268 (2012) MR 2963170

[ZZ] Zhang, C.-H., Zhang, S. S.: Confidence intervals for low dimensional parameters in high dimensional linear models. J. Roy. Statist. Soc. Ser. B Statist. Methodol. **76**, 217–242 (2014) Zbl 1411.62196 MR 3153940