Efficient Estimation of Smooth Functionals in Gaussian Shift Models

Vladimir Koltchinskii* **and Mayya Zhilova**†

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

July 15, 2020

Abstract: We study a problem of estimation of smooth functionals of parameter θ of Gaussian shift model

$$X = \theta + \xi, \ \theta \in E,$$

where E is a separable Banach space and X is an observation of unknown vector θ in Gaussian noise ξ with zero mean and known covariance operator Σ. In particular, we develop estimators $T(X)$ of $f(\theta)$ for functionals $f : E \mapsto \mathbb{R}$ of Hölder smoothness $s > 0$ such that

$$\sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \lesssim \Big(\|\Sigma\| \vee (\mathbb{E}\|\xi\|^2)^s\Big) \wedge 1,$$

where $\|\Sigma\|$ is the operator norm of Σ, and show that this mean squared error rate is minimax optimal at least in the case of standard finite-dimensional Gaussian shift model ($E = \mathbb{R}^d$ equipped with the canonical Euclidean norm, $\xi = \sigma Z$, $Z \sim \mathcal{N}(0; I_d)$). Moreover, we determine a sharp threshold on the smoothness s of functional f such that, for all s above the threshold, $f(\theta)$ can be estimated efficiently with a mean squared error rate of the order $\|\Sigma\|$ in a "small noise" setting (that is, when $\mathbb{E}\|\xi\|^2$ is small). The construction of efficient estimators is crucially based on a "bootstrap chain" method of bias reduction. The results could be applied to a variety of special high-dimensional and infinite-dimensional Gaussian models (for vector, matrix and functional data).

Résumé: Dans cet article, nous étudions le problème d'estimation de fonctionnelles lisses d'un paramètre θ dans un modèle gaussien suivant:

$$X = \theta + \xi, \ \theta \in E,$$

où E est un espace de Banach séparable, X est une observation du vecteur θ inconnu et le bruit ξ est gaussien de moyenne nulle et d'opérateur

*Supported in part by NSF Grants DMS-1810958, DMS-1509739 and CCF-1523768
†Supported by NSF Grant DMS-1712990

de covariance Σ connu. En particulier, nous développons des estimateurs $T(X)$ de $f(\theta)$ pour les fonctionnelles $f : E \mapsto \mathbb{R}$ du paramètre de régularité Höldérienne $s > 0$ tels que

$$\sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \lesssim \Big(\|\Sigma\| \vee (\mathbb{E}\|\xi\|^2)^s\Big) \wedge 1,$$

où $\|\Sigma\|$ est la norme d'opérateur de Σ, et nous montrons que cette estimation de l'erreur quadratique moyenne est minimax optimale au moins dans le cas du modèle gaussien de dimension finie avec une matrice de covariance identité ($E = \mathbb{R}^d$ est muni de la norme euclidienne canonique $\xi = \sigma Z$, $Z \sim \mathcal{N}(0; I_d)$). De plus, nous déterminons le seuil exact sur la régularité s de la fonctionnelle f tel que, pour tout s au-dessus de ce seuil, $f(\theta)$ peut-être estimé efficacement avec un erreur quadratique moyen de l'ordre $\|\Sigma\|$ dans le régime de "bruit petit" (i.e. $\mathbb{E}\|\xi\|^2$ est petit). La construction des estimateurs efficaces est basée essentiellement sur une méthode de "chaîne bootstrap" pour la réduction du biais. Les résultats peuvent être appliqués à un grand choix des modèles gaussiens de grande dimension voire même infini-dimensionnels (pour les données vectorielles, matricielles et fonctionnelles).

AMS 2000 subject classifications: Primary 62H12; secondary 62G20, 62H25, 60B20.

Keywords and phrases: Efficiency, Smooth functionals, Gaussian shift model, Bootstrap, Effective rank, Concentration inequalities, Normal approximation.

1. Introduction

The problem of estimation of functionals of "high complexity" parameters of statistical models often occurs both in high-dimensional and in nonparametric statistics, where it is of importance to identify some features of a complex parameter that could be estimated efficiently with a fast (sometimes, parametric) convergence rates. Such problems are very important in the case of vector, matrix or functional parameters in a variety of applications including functional data analysis and kernel machine learning ([34], [5]). In this paper, we study a very basic version of this problem in the case of rather general Gaussian models with unknown mean. Consider the following *Gaussian shift model*

$$X = \theta + \xi,\ \theta \in E, \tag{1.1}$$

where E is a separable Banach space, θ is an unknown parameter and ξ is a mean zero Gaussian random variable in E (the noise) with *known* covariance operator Σ. In other words, an observation $X \sim \mathcal{N}(\theta; \Sigma)$ in Gaussian shift model (1.1) is a Gaussian vector in E with unknown mean θ and known covariance Σ. Recall that Σ is an operator from the dual space E^* into E such that $\Sigma u := \mathbb{E}\langle \xi, u\rangle \xi, u \in E^*$. Here and in what follows, $\langle x, u\rangle$ denotes the value of a linear functional $u \in E^*$ on a vector $x \in E$ (although, in some parts of the paper, with a little abuse of notation, $\langle \cdot, \cdot\rangle$ will also denote the inner product of Euclidean spaces). It is well known that the covariance operator Σ of a Gaussian vector in E is bounded and, moreover, it is nuclear.

Our goal is to study the problem of estimation of $f(\theta)$ for smooth functionals $f : E \mapsto \mathbb{R}$. The problem of estimation of smooth functionals of parameters of infinite-dimensional (nonparametric) models has been studied for several decades. It is considerably harder than in the classical finite-dimensional parametric i.i.d. models, where under standard regularity assumptions, $f(\hat\theta)$ ($\hat\theta$ being the maximum likelihood estimator) is an asymptotically efficient (in the sense of Hàjek-LeCam) estimator of $f(\theta)$ with $\sqrt{n}$-rate for continuously differentiable functions f. In the nonparametric case, classical convergence rates do not necessarily hold in functional estimation problems and minimax optimal convergence rates have to be determined. Moreover, even when the classical convergence rates do hold, the construction of efficient estimator is often a challenging problem. Such problems have been often studied for special models (Gaussian white noise model, nonparametric density estimation model, etc) and for special functionals (with a number of nontrivial results even in the case of linear and quadratic functionals). Early results in this direction are due to Levit [28, 29] and Ibragimov and Khasminskii [15]. Further important references include Ibragimov, Nemirovski and Khasminskii [16], Donoho and Liu [9, 10], Bickel and Ritov [2], Donoho and Nussbaum [11], Nemirovski [31, 32], Birgé and Massart [4], Laurent [26], Lepski, Nemirovski and Spokoiny [30], Cai and Low [6, 7], Klemelä [19] as well as a vast literature on semiparametric efficiency (see, e.g., [3] and references therein). Early results on consistent and asymptotically normal estimation of smooth functionals of high-dimensional parameters are due to Girko [13, 14]. More recently, there has been a lot of interest in efficient and minimax optimal estimation of functionals of parameters of high-dimensional models including a variety of problems related to semiparametric efficiency of regularized estimators (see [36], [17], [37]), on minimax optimal rates of estimation of special functionals (see [8]), on efficient estimation of smooth functionals of covariance in Gaussian models [23, 20].

Throughout the paper, given nonnegative A, B, $A \lesssim B$ means that $A \leq CB$ for a numerical constant C, $A \gtrsim B$ is equivalent to $B \lesssim A$ and $A \asymp B$ is equivalent to $A \lesssim B \lesssim A$. Sometimes signs of relationships $\lesssim, \gtrsim$ and $\asymp$ will be provided with subscripts (say, $A \lesssim_\gamma B$ or $A \asymp_\gamma B$), indicating possible dependence of the constants on the corresponding parameters.

In what follows, exponential bounds on random variables (say, on ζ) are often stated in the following form: there exists a constant $C > 0$ such that, for all $t \geq 1$ with probability at least $1 - e^{-t}$, $\zeta \leq Ct$. The proof could often result in a slightly different bound, for instance, $\zeta \leq Ct$ with probability $1 - 5e^{-t}$. However, replacing constant C with $C' = 2\log(5)C$, it is easy to obtain the probability bound in the initial form $1 - e^{-t}$. In such cases, we say that ,"adjusting the constants" allows us to write the probability as $1 - e^{-t}$ (without providing further details).

We will now briefly discuss the results of Ibragimov, Nemirovski and Khasminskii [16] and follow up results of Nemirovski [31, 32] that are especially close to our approach to the problem. In [16], the following model was studied

$$dX^{(n)}(t) = \theta(t)dt + \frac{1}{\sqrt{n}}dw(t), t \in [0, 1], \tag{1.2}$$

in which a "signal" $\theta \in \Theta \subset L_2([0, 1])$ is observed in a Gaussian white noise (w being a standard Brownian motion on $[0, 1]$). The complexity of the parameter space Θ was characterized by Kolmogorov widths:

$$d_m(\Theta) := \inf_{L \subset L_2([0,1]), \dim(L) \leq m} \sup_{\theta \in \Theta} \|\theta - P_L\theta\|^2,$$

where P_L denotes the orthogonal projection onto subspace L. Assuming that $\Theta \subset U := \{\theta \in L_2([0,1]) : \|\theta\| \leq 1\}$ and, for some $\beta > 0$,

$$d_m(\Theta) \lesssim m^{-\beta}, m \geq 1,$$

the goal of the authors was to determine a "smoothness threshold" $s(\beta) > 0$ such that, for all $s > s(\beta)$ and for all functionals f on $L_2([0,1])$ of smoothness s, $f(\theta)$ could be estimated efficiently with rate $n^{-1/2}$ based on observation $X^{(n)}$ (whereas for $s < s(\beta)$ there exist functionals f of smoothness s such that $f(\theta)$ could not be estimated with parametric rate $n^{-1/2}$). It turned out that the main difficulties in this problem are related to a proper definition of the smoothness of the functional f. In particular, even such simple functional as $f(\theta) = \|\theta\|^2$ could not be estimated efficiently on some sets Θ with $\beta \leq 1/4$. The smoothness of functionals on Hilbert space $L_2([0,1])$ is usually defined in terms of their Hölder type norms that, in turn, depend on a way in which the norm of Fréchet derivatives $f^{(k)}(\theta)$ is defined. The k-th order Fréchet derivative is a symmetric k-linear form on $L_2([0,1])$. The most common definition of the norm of such a form $M(h_1, \dots, h_k), h_1, \dots, h_k \in L_2([0,1])$ is the operator norm: $\|M\| := \sup_{h_1,\dots,h_k \in U} |M(h_1, \dots, h_k)|$. Other possibilities include Hilbert–Schmidt norm $\|M\|_{HS}$ and "hybrid" norms $\|M\|_{(j)} := \sup_{h_1,\dots,h_j \in U} \|M(h_1, \dots, h_j, \cdot, \dots, \cdot)\|_{HS}, 0 \leq j \leq k$. The Hölder classes in [16] were defined in terms of the following norms: for $s = k + \gamma$, $k \geq 0$, $\gamma \in (0,1]$,

$$\|f\|_{\tilde{C}^s} := \max_{0 \leq j \leq k-1} \sup_{x \in 2U} \|f^{(j)}(x)\|_{HS} \bigvee \sup_{x \in 2U} \|f^{(k)}(x)\|_{(1)} \bigvee \sup_{x,x' \in 2U, \theta \neq \theta'} \frac{\|f^{(k)}(x) - f^{(k)}(x')\|}{\|x - x'\|^\gamma}.$$

With this somewhat complicated definition, it was proved that, if $\|f\|_{\tilde{C}^s} < \infty$ and either $\frac{1}{2\beta} + 1 < s \leq 3$, or $s > 3 \vee \frac{1}{2\beta}$, then there exists an asymptotically efficient estimator of $f(\theta)$ with convergence rate $n^{-1/2}$. The construction of such estimators was based on the development of a method of unbiased estimation of Hilbert–Schmidt polynomials on $L_2([0,1])$ and on Taylor expansion of $f(\theta)$ in a neighborhood of an estimator $\hat{\theta}$ of θ with an optimal nonparametric error rate. It was later shown in [31, 32] that the smoothness thresholds described above are optimal.

We will study similar problems for Gaussian shift model (1.1) trying to determine smoothness thresholds for efficient estimation in terms of proper complexity characteristics for this model.

Among the simplest smooth functionals on E are bounded linear functionals $E \ni \theta \mapsto \langle \theta, u\rangle, u \in E^*$. For a straightforward estimator $\langle X, u\rangle$ of such a functional,

$$\mathbb{E}_\theta(\langle X, u\rangle - \langle \theta, u\rangle)^2 = \mathbb{E}\langle \xi, u\rangle^2 = \langle \Sigma u, u\rangle, u \in E^*,$$

and, for functionals u from the unit ball of E^* the largest possible mean squared error is equal to the operator norm of Σ :

$$\|\Sigma\| = \sup_{u,v\in E^*, \|u\|,\|v\|\le 1} \mathbb{E}\langle \xi, u\rangle\langle \xi, v\rangle = \sup_{u\in E^*,\|u\|\le 1} \mathbb{E}\langle \xi, u\rangle^2.$$

It is also not hard to prove the following proposition.

Proposition 1.1. *Let*

$$\hat{T}(X) := \begin{cases} \langle X, u\rangle & \text{for } \|\Sigma\| \le 1 \\ 0 & \text{for } \|\Sigma\| > 1. \end{cases}$$

Then

$$\sup_{\|u\|\le 1} \sup_{\|\theta\|\le 1} \mathbb{E}_\theta(\hat{T}(X) - \langle \theta, u\rangle)^2 \le \|\Sigma\| \wedge 1$$

and

$$\sup_{\|u\|\le 1} \inf_T \sup_{\|\theta\|\le 1} \mathbb{E}_\theta(T(X) - \langle \theta, u\rangle)^2 \gtrsim \|\Sigma\| \wedge 1. \tag{1.3}$$

In what follows, the complexity of estimation problem will be characterized by two parameters of the noise ξ. One is the operator norm $\|\Sigma\|$, which is involved in the minimax mean squared error for estimation of linear functionals. It will be convenient to view $\|\Sigma\|$ as the *weak variance* of ξ. Another complexity parameter is the *strong variance* of ξ defined as

$$\mathbb{E}\|\xi\|^2 = \mathbb{E} \sup_{u,v\in E^*,\|u\|,\|v\|\le 1} \langle \xi, u\rangle\langle \xi, v\rangle = \mathbb{E} \sup_{u\in E^*,\|u\|\le 1} \langle \xi, u\rangle^2.$$

Clearly, $\mathbb{E}\|\xi\|^2 \ge \|\Sigma\|$. The ratio of these two parameters,

$$\mathbf{r}(\Sigma) := \frac{\mathbb{E}\|\xi\|^2}{\|\Sigma\|},$$

is called the *effective rank* of Σ and it was used earlier in concentration bounds for sample covariance operators and their spectral projections [22, 21]. The following properties of $\mathbf{r}(\Sigma)$ are obvious:

$$\mathbf{r}(\Sigma) \ge 1 \text{ and } \mathbf{r}(\lambda\Sigma) = \mathbf{r}(\Sigma), \lambda > 0.$$

Thus, the effective rank is invariant with respect to rescaling of Σ (or rescaling of the noise). In this sense, $\|\Sigma\|$ and $\mathbf{r}(\Sigma)$ can be viewed as complementary parameters of the noise. It is easy to check that, if E is a Hilbert space, then $\mathbf{r}(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|}$, which implies that $\mathbf{r}(\Sigma) \leq \mathrm{rank}(\Sigma) \leq \dim(E)$. Clearly, $\mathbf{r}(\Sigma)$ could be viewed as a way to measure the dimensionality of the noise. In particular, for the maximum likelihood estimator X of θ in the Gaussian shift model (1.1), we have $\mathbb{E}_\theta\|X-\theta\|^2 = \mathbb{E}\|\xi\|^2 = \|\Sigma\|\mathbf{r}(\Sigma)$, resembling a standard formula $\sigma^2 d$ for the risk of estimation of a vector in $\mathbb{R}^d$ observed in a "white noise" with variance σ^2.

We discuss below several simple examples of the general Gaussian shift model (1.1).

Example 1.1. Standard finite-dimensional Gaussian shift model. Let $E = \mathbb{R}^d$ be equipped with the canonical Euclidean inner product and the corresponding norm (the ℓ_2-norm), and let $\xi = \sigma Z$, where $\sigma > 0$ is a known constant and $Z \sim \mathcal{N}(0; I_d)$. In this case, $\Sigma = \sigma^2 I_d$, $\|\Sigma\| = \sigma^2$, $\mathbb{E}\|\xi\|^2 = \sigma^2 d$ and $\mathbf{r}(\Sigma) = d$. Note that the size of effective rank $\mathbf{r}(\Sigma)$ crucially depends on the choice of underlying norm of the linear space. For instance, if instead of the canonical Euclidean inner product used in the case of standard Gaussian shift model, the space $E = \mathbb{R}^d$ is equipped with the ℓ_∞-norm, then we still have $\|\Sigma\| = \sigma^2$, but

$$\mathbb{E}\|\xi\|_{\ell_\infty}^2 \asymp \sigma^2 \log d,$$

implying that $\mathbf{r}(\Sigma) \asymp \log d$.

Example 1.2. Matrix Gaussian shift models. Let E be the space of all symmetric $d \times d$ matrices equipped with the operator norm and let $\xi = \sigma Z$ with known parameter $\sigma > 0$ and Z sampled from the Gaussian orthogonal ensemble (that is, $Z = (Z_{ij})_{i,j=1}^d$ is a symmetric random matrix, $Z_{ij}, i \leq j$ are independent r.v., $Z_{ij} \sim \mathcal{N}(0,1), i < j$, $Z_{ii} \sim \mathcal{N}(0;2)$). In this case, $\|\Sigma\| \asymp \sigma^2$ and

$$\mathbb{E}\|\xi\|^2 = \sigma^2\mathbb{E}\|Z\|^2 \asymp \sigma^2 d,$$

implying that $\mathbf{r}(\Sigma) \asymp d$. As before, the effective rank would be different for a different choice of norm on E. For instance, if E is equipped with the Hilbert–Schmidt norm, then $\mathbf{r}(\Sigma) \asymp d^2$ (compare this with Example 1). One can similarly consider other matrix Gaussian shift models (for instance, for rectangular matrices).

Example 1.3. Gaussian functional data model. Let $E = C([0,1]^d), d \geq 1$ be equipped with the sup-norm $\|\cdot\|_\infty$. Suppose that $\xi := \sigma Z$, where $\sigma > 0$ is a known parameter and Z is a mean zero Gaussian process on $[0,1]^d$ with the sample paths continuous a.s. (and with known distribution). Without loss of generality, assume that $\sup_{t\in[0,1]^d} \mathbb{E}Z^2(t) = 1$. Suppose that, for some $\beta > 0$,

$$\tau^2(t,s) := \mathbb{E}|Z(t) - Z(s)|^2 \lesssim |t-s|^{\beta},\ \ t,s \in [0,1]^d.$$

Then, it is easy to see that the following bound holds for the metric entropy $H_\tau([0,1]^d;\varepsilon)$ of $[0,1]^d$ with respect to metric τ :

$$H_\tau([0,1]^d;\varepsilon) \lesssim_\beta d \log \frac{1}{\varepsilon}.$$

It follows from Dudley's entropy bound that

$$\mathbb{E}\|Z\|_\infty^2 \lesssim_\beta \Bigl(\int_0^1 H_\tau^{1/2}([0,1]^d;\varepsilon) d\varepsilon\Bigr)^2 \lesssim d.$$

Therefore, it is easy to conclude that $\|\Sigma\| \asymp \sigma^2$ and $\mathbb{E}\|\xi\|_\infty^2 \lesssim \sigma^2 d$, implying that $\mathbf{r}(\Sigma) \lesssim d$.

In the following sections, we develop estimators $T(X)$ of $f(\theta)$ in Gaussian shift model with mean squared error of the order

$$\sup_{\|\theta\|\le 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \lesssim \Bigl(\|\Sigma\| \vee (\mathbb{E}\|\xi\|^2)^s\Bigr) \wedge 1,$$

where s is the degree of smoothness of functional f. In this bound, $\|\Sigma\|$ is the weak variance of the noise and it will be shown later that the term $(\mathbb{E}\|\xi\|^2)^s$ provides an upper bound on the bias of estimator $T(X)$. If $\mathbb{E}\|\xi\|^2 < 1$, the bias term becomes smaller than the weak variance $\|\Sigma\|$ for a sufficiently large degree of smoothness s. We show that this error rate is minimax optimal (at least in the case of standard finite dimensional Gaussian shift model). Moreover, we determine a sharp threshold on smoothness s such that, for all s above this threshold and all functionals f of smoothness s, the mean squared error rate of estimation of $f(\theta)$ is of the order $\|\Sigma\| \wedge 1$ (as for linear functionals), and, for all s strictly above the threshold, we prove the efficiency of our estimators in the "small noise" case (when the strong variance $\mathbb{E}\|\xi\|^2$ is small). The key ingredient in the development of such estimators is a *bootstrap chain bias reduction* method introduced in [20] in the problem of estimation of smooth functionals of covariance operators. We will outline this approach in Section 2 and develop it in detail in Section 3 for Gaussian shift models.

2. Overview of Main Results

We will study how the optimal error rate of estimation of $f(\theta)$ for parameter θ of Gaussian shift model (1.1) depends on the smoothness of the functional $f : E \mapsto \mathbb{R}$ as well as on the weak and strong variances, $\|\Sigma\|$ and $\mathbb{E}_\Sigma\|\xi\|^2$, of the noise ξ (or, equivalently, on the parameters $\|\Sigma\|$ and $\mathbf{r}(\Sigma)$). To this end, we define below a Banach space $C^{s,\gamma}(E)$ of functionals $f : E \mapsto \mathbb{R}$ of smoothness $s > 0$ such that f and its derivatives grow as $\|\theta\| \to \infty$ not faster than $\|\theta\|^\gamma$ for some $\gamma \geq 0$.

2.1. Differentiability

For Banach spaces E, F, let $\mathcal{M}_k(E; F)$ be the Banach space of symmetric k-linear forms $M : E \times \cdots \times E \mapsto F$ with bounded operator norm

$$\|M\| := \sup_{\|h_1\|\leq 1,\dots,\|h_k\|\leq 1} \|M(h_1, \dots, h_k)\| < \infty.$$

For $k = 0$, $\mathcal{M}_0(E; F)$ is the space of constants (vectors of F). A function $P : E \mapsto F$ defined by $P(x) := M(x, \dots, x), x \in E$, where $M \in \mathcal{M}_k(E; F), k \geq 0$ is called *a bounded homogeneous k-polynomial* on E with values in F. It is known that P uniquely defines $M \in \mathcal{M}_k(E; F)$. *A bounded polynomial* on E with values in F is an arbitrary function $P : E \mapsto F$ represented as a finite sum $P(x) := \sum_{j\in I} P_j(x), x \in E, I \subset \mathbb{Z}_+$, where P_j is a non-zero bounded homogeneous j-polynomial. For $I = \emptyset$, we set $P := 0$. Polynomials $P_j, j \in I$ are uniquely defined by P. The degree of P is defined as $\deg(P) := \max(I)$ (with $\deg(0) = 0$). If $P_j(x) = M_j(x, \dots, x)$ for $M_j \in \mathcal{M}_j(E; F)$, define

$$\|P\|_{\mathrm{op}} := \sum_{j\in I} \|M_j\|.$$

Recall that a function $f : E \mapsto F$ is called Fréchet differentiable at a point $x \in E$ iff there exists a bounded linear operator $f'(x)$ from E to F (Fréchet derivative) such that

$$f(x + h) - f(x) = f'(x)h + o(\|h\|) \text{ as } h \to 0.$$

Higher order Fréchet derivatives could be defined by induction. The k-th order Fréchet derivative $f^{(k)}(x)$ at point x is defined as the Fréchet derivative of the mapping $E \ni x \mapsto f^{(k-1)}(x) \in \mathcal{M}_{k-1}(E; F)$ (assuming its Fréchet differentiability). It is a bounded linear operator from E to $\mathcal{M}_{k-1}(E; F)$ that could be also viewed as a bounded symmetric k-linear form from the space $\mathcal{M}_k(E;F)$. As always, we call f k-times (Fréchet) continuously differentiable if its k-th order derivative exists and it is a continuous function on E. Clearly, polynomials are k times Fréchet differentiable for any k. If P is a polynomial and $\deg(P) = k$, then $P^{(k)}$ is a constant (a k-linear symmetric form that does not depend on x) and $P^{(k+1)} = 0$.

We will be interested in what follows in classes of smooth functionals $f : E \mapsto \mathbb{R}$ with at most polynomial (with respect to $\|x\|$) growth of their derivatives. To this end, we describe below several useful norms.

First, let $g : E \mapsto F$. For $\gamma \geq 0$, let

$$\|g\|_{L_{\infty,\gamma}} := \sup_{x\in E} \frac{\|g(x)\|}{(1 \vee \|x\|)^{\gamma}}$$

and for $\gamma \geq 0, \rho \in (0, 1]$, let

$$\|g\|_{\mathrm{Lip}_{\rho,\gamma}} := \sup_{x'\neq x''} \frac{\|g(x') - g(x'')\|}{(1 \vee \|x'\| \vee \|x''\|)^{\gamma}\|x' - x''\|^{\rho}}.$$

Assuming that spaces E, F are equipped with their Borel σ-algebras, we define $L_{\infty,\gamma}(E;F)$ as the space of measurable functions $g : E \mapsto F$ with $\|g\|_{L_{\infty,\gamma}} < \infty$. We also define

$$\mathrm{Lip}_{\rho,\gamma}(E;F) := \{g : \|g\|_{\mathrm{Lip}_{\rho,\gamma}} < \infty\}.$$

In the case of $F = \mathbb{R}$, we will write simply $L_{\infty,\gamma}(E)$ and $\mathrm{Lip}_{\rho,\gamma}(E)$; for $\gamma = 0$, we write $L_\infty, \mathrm{Lip}_\rho$ instead of $L_{\infty,0}, \mathrm{Lip}_{\rho,0}$.

For $k \geq 0$, we will define the norm

$$\|g\|_{C^{k,\gamma}} := \max_{0\leq j\leq k} \|g^{(j)}\|_{L_{\infty,\gamma}}$$

and the space $C^{k,\gamma}(E;F) := \{g : \|g\|_{C^{k,\gamma}} < \infty\}$ of k times differentiable functions (with the growth rate of derivatives characterized by γ). Finally, for $s = k + \rho$ with $k \geq 0$ and $\rho \in (0, 1)$, define

$$\|g\|_{C^{s,\gamma}} := \max_{0\leq j\leq k} \|g^{(j)}\|_{L_{\infty,\gamma}} \vee \|g^{(k)}\|_{\mathrm{Lip}_{\rho,\gamma}}$$

and the space $C^{s,\gamma}(E;F) := \{g : \|g\|_{C^{s,\gamma}} < \infty\}$. Note that, in the above definitions, the derivatives $g^{(j)}$ are viewed as functions $E \ni x \mapsto g^{(j)}(x) \in \mathcal{M}_j(E;F)$ from the space E into the space $\mathcal{M}_j(E;F)$ of symmetric j-linear forms equipped with the operator norm. This norm is used in the definitions of $\|g^{(j)}\|_{L_{\infty,\gamma}}$, $\|g^{(k)}\|_{\mathrm{Lip}_{\rho,\gamma}}$. As before, we set $C^s := C^{s,0}$. It is easy to see that for any polynomial P such that $\deg(P) = k$ and for all $s > 0$, $P \in C^{s,k}(E)$.

In what follows, we frequently use bounds on the remainder of the first order Taylor expansion

$$S_g(x;h) := g(x+h) - g(x) - g'(x)(h), x, h \in E$$

of Fréchet differentiable function $g : E \mapsto \mathbb{R}$. We will skip the proof of the following simple lemma.

Lemma 2.1. *Assume that $g : E \mapsto \mathbb{R}$ is Fréchet differentiable in E with $g' \in \mathrm{Lip}_{\rho,\gamma}(E; \mathcal{M}_1(E;F))$. Then*

$$|S_g(x;h)| \lesssim \|g'\|_{\mathrm{Lip}_{\rho,\gamma}} (1 \vee \|x\| \vee \|h\|)^{\gamma} \|h\|^{1+\rho}, x, h \in E$$

and

$$|S_g(x;h') - S_g(x;h)| \lesssim \|g'\|_{\mathrm{Lip}_{\rho,\gamma}} (1 \vee \|x\| \vee \|h\| \vee \|h'\|)^{\gamma} (\|h\| \vee \|h'\|)^{\rho} \|h'-h\|,\ x, h, h' \in E.$$

Remark 2.1. Throughout the paper, the generic notation $\|\cdot\|$ is used for the norms of the Banach space E as well as other Banach spaces, such as the dual space E^*, the space of bounded operators from E^* to E (such as covariance operators) and the spaces of symmetric multilinear forms $\mathcal{M}_k(E;F)$. In the last two examples, $\|\cdot\|$ is the operator norm (by default). We are not providing these norms with any subscripts (to avoid overcomplicated notations). In most of the cases, it should be clear to the reader from the context to which object the norm is applied and which specific norm is used (in more ambiguous cases, we are providing additional clarifications). The subscripts will be used only for the norms of function spaces (such as $L_{\infty,\gamma}$, $\mathrm{Lip}_{\rho,\gamma}$, $C^{s,\gamma}$, etc).

2.2. Definition of estimators and risk bounds

The crucial step in construction of estimator T_k is a bias reduction method developed in detail in Section 3 and briefly outlined here. Consider the following linear operator

$$(\mathcal{T}g)(\theta) := \mathbb{E}_{\theta} g(X) = \mathbb{E} g(\theta + \xi), \theta \in E$$

that is well defined on the spaces $L_{\infty,\gamma}(E)$ for $\gamma \geq 0$. Given a smooth functional $f : E \mapsto \mathbb{R}$, we would like to find a functional g on E such that the bias of estimator $g(X)$ of $f(\theta)$ is small enough. In other words, we would like to find an approximate solution of operator equation $(\mathcal{T}g)(\theta) = f(\theta), \theta \in E$. Under the assumption that the strong variance $\mathbb{E}\|\xi\|^2$ of the noise ξ is small, the operator $\mathcal{T}$ is close to the identity operator $\mathcal{I}$. Define $\mathcal{B} := \mathcal{T} - \mathcal{I}$. Then, at least formally, the solution of the equation $\mathcal{T}g = f$ could be written as a Neumann series:

$$g = (\mathcal{I} + \mathcal{B})^{-1} f = (\mathcal{I} - \mathcal{B} + \mathcal{B}^2 - \mathcal{B}^3 + \dots) f.$$

We will define an estimator $f_k(X)$ in terms of a partial sum of this series:

$$f_k(x) := \sum_{j=0}^{k} (-1)^j (\mathcal{B}^j f)(x), x \in E.$$

It will be proved in Section 3, that, for this estimator, the bias $\mathbb{E}_\theta f_k(X) - f(\theta)$ is of the order $\lesssim (\mathbb{E}^{1/2}\|\xi\|^2)^s$, provided that $f \in C^{s,\gamma}(E)$ for $s = k + 1 + \rho$, $k \geq 0$, $\rho \in (0, 1]$ and $\|\theta\|$ is bounded by a constant.

We will prove in Section 4 the following result.

Theorem 2.1. *Suppose that $f \in C^{s,\gamma}(E)$ for some $s > 0$ and $\gamma \geq 0$. For $s \in (0, 1]$, set $k := 0$ and, for $s > 1$, let $s = k + 1 + \rho$ for some $k \geq 0$, $\rho \in (0, 1]$. Define*

$$T_k(X) := \begin{cases} f_k(X) & \text{if } \|\Sigma\|\mathbf{r}(\Sigma) \leq 1/4 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathbb{E}_\theta (T_k(X) - f(\theta))^2 \lesssim_\gamma s^\gamma \|f\|^2_{C^{s,\gamma}} (1 \vee \|\theta\|)^{2\gamma} \Big(\Big(\|\Sigma\| \vee (\mathbb{E}\|\xi\|^2)^s \Big) \wedge 1 \Big). \tag{2.1}$$

It follows from bound (2.1) that

$$\sup_{\|f\|_{C^{s,\gamma}} \leq 1} \sup_{\|\theta\| \leq 1} \mathbb{E}_\theta (T_k(X) - f(\theta))^2 \lesssim_{s,\gamma} \Big(\Big(\|\Sigma\| \vee (\mathbb{E}\|\xi\|^2)^s \Big) \wedge 1 \Big). \tag{2.2}$$

We will show in Section 7 that, *in the case of standard finite-dimensional Gaussian shift model*, the above bound is optimal in a minimax sense. More precisely, in this case, the following result holds.

Theorem 2.2. *Let $E := \mathbb{R}^d$ (equipped with the standard Euclidean norm) and let $X \sim \mathcal{N}(\theta; \sigma^2 I_d), \theta \in \mathbb{R}^d$ for some $\sigma^2 > 0$. Then, for all $s > 0$,*

$$\sup_{\|f\|_{C^s} \leq 1} \inf_T \sup_{\|\theta\| \leq 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \gtrsim \Big(\|\Sigma\| \vee \Big(\mathbb{E}\|\xi\|^2 \Big)^s \Big) \wedge 1, \tag{2.3}$$

where the infimum is taken over all possible estimators $T(X)$.

At this point, we could not extend the lower bound of Theorem 2.2 to general Gaussian shift models in Banach spaces.

Remark 2.2. In a very recent paper [38], Zhou and Li state a similar result (Theorem 7.2) with Besov $B^s_{\infty,1}$-norm instead of the C^s-norm. There is a mistake in the proof of this result (contrary to the claim of the authors, it is impossible to choose function φ used in the proof so that $\|\tilde{\varphi}\|_{B^s_{\infty,1}} \leq 1$ and other required properties hold). However, if the Besov norm is replaced by the C^s-norms used in our paper, their proof seems to be correct. The method of the proof of Theorem 7.2 in [38] differs from ours.

2.3. Efficiency

Bound (2.2) implies that, if the smoothness s of functional f is sufficiently large, namely if

$$(\mathbb{E}\|\xi\|^2)^s \leq \|\Sigma\|, \tag{2.4}$$

then

$$\sup_{\|f\|_{C^{s,\gamma}}\leq 1} \sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (T_k(X) - f(\theta))^2 \lesssim_{s,\gamma} \|\Sigma\| \wedge 1, \tag{2.5}$$

which coincides with the largest minimax optimal mean squared error for linear functionals from the unit ball in E^*. Assuming that $\|\Sigma\|\mathbf{r}(\Sigma) = \mathbb{E}\|\xi\|^2 < 1$, condition (2.4) can be equivalently written as

$$s \geq 1 + \frac{\log \mathbf{r}(\Sigma)}{\log \frac{1}{\|\Sigma\|} - \log \mathbf{r}(\Sigma)}. \tag{2.6}$$

If $\sigma^2 := \|\Sigma\|$ is a small parameter and $\mathbf{r}(\Sigma) \leq \sigma^{-2\alpha}$ for some $\alpha \in (0,1)$, condition (2.6) would follow from the condition $s \geq \frac{1}{1-\alpha}$. On the other hand, it follows from bound (2.3) that, in the case of standard finite-dimensional Gaussian shift model, the smoothness threshold $\frac{1}{1-\alpha}$ is sharp for estimation with mean squared error rate $\asymp \sigma^2$. Indeed, in this case, $\mathbf{r}(\Sigma) = d$ and, if σ is small and $d \asymp \sigma^{-2\alpha}$ for some $\alpha \in (0,1)$, then, for any $s < \frac{1}{1-\alpha}$, there exists a functional f with $\|f\|_{C^{s,\gamma}} \leq 1$ such that

$$\inf_T \sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \gtrsim \sigma^{2s(1-\alpha)},$$

which is significantly larger than σ^2 as $\sigma \to 0$. Moreover, if $d \asymp \sigma^{-2}$, then, for any $s > 0$, there exists a functional f with $\|f\|_{C^{s,\gamma}} \leq 1$ such that

$$\inf_T \sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \gtrsim_s 1,$$

essentially implying that even consistent estimators of $f(\theta)$ do not exist in this case.

In the case when $\mathbf{r}(\Sigma) \lesssim \sigma^{-2\alpha}$ for some $\alpha \in (0,1)$ and $s > \frac{1}{1-\alpha}$ (or, more generally, when $(\mathbb{E}\|\xi\|^2)^s$ is of a smaller order than $\|\Sigma\|$), it is possible to prove that $f_k(X) - f(\theta)$ is close in distribution to normal and establish the efficiency of estimator $f_k(X)$. More precisely, let

$$\sigma^2_{f,\xi}(\theta) := \mathbb{E}(f'(\theta)(\xi))^2 = \langle \Sigma f'(\theta), f'(\theta)\rangle$$

For $s \geq 1, \gamma \geq 0$, denote

$$K(f;\Sigma;\theta) := K_{s,\gamma}(f;\Sigma;\theta) := \frac{\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma \|\Sigma\|^{1/2}}{\sigma_{f,\xi}(\theta)}.$$

It is easy to see that

$$\sigma_{f,\xi}(\theta) \leq \|\Sigma\|^{1/2}\|f'(\theta)\| \leq \|f'\|_{L_{\infty,\gamma}}(1\vee\|\theta\|)^\gamma\|\Sigma\|^{1/2} \leq \|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma\|\Sigma\|^{1/2},$$

implying that $K_{s,\gamma}(f;\Sigma;\theta) \geq 1$. We also have that

$$K_{s,\gamma}(f;\lambda\Sigma;\theta) = K_{s,\gamma}(f;\Sigma;\theta), \lambda > 0,$$

which means that $K_{s,\gamma}(f;\Sigma;\theta)$ does not depend on the noise level $\|\Sigma\|^{1/2}$. In what follows, it will be assumed that the functional $K_{s,\gamma}(f;\Sigma;\theta)$ is bounded from above by a constant, implying that $\sigma_{f,\xi}(\theta)$ is within a constant from its upper bound $\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma\|\Sigma\|^{1/2}$. This is the case, for instance, when θ is in a bounded set and $\sigma_{f,\xi}(\theta) \gtrsim \|\Sigma\|^{1/2}$ (in other words, the standard deviation $\sigma_{f,\xi}(\theta)$ is not too small comparing with the noise level $\|\Sigma\|^{1/2}$). If we write the noise of the model in the form $\xi = \sigma\xi_0$, where $\sigma > 0$ is a small scaling parameter characterizing the noise level and ξ_0 is a mean zero Gaussian random vector in E with covariance operator Σ_0 such that $\|\Sigma_0\| = 1$, then $\|\Sigma\|^{1/2} = \sigma$ and

$$\sigma_{f,\xi}(\theta) = \sigma_{f,\xi_0}(\theta)\sigma = \Big\langle \Sigma_0 f'(\theta), f'(\theta)\Big\rangle^{1/2}\sigma.$$

In this case, the condition $\sigma_{f,\xi}(\theta) \gtrsim \|\Sigma\|^{1/2}$ just means that $\sigma_{f,\xi_0}(\theta) = \Big\langle \Sigma_0 f'(\theta), f'(\theta)\Big\rangle^{1/2}$ is bounded away from 0 (see also Remark 2.4 below).

The following result will be proved in Section 5.

Theorem 2.3. *Suppose $f \in C^{s,\gamma}(E)$ for some $s > 1$ and $\gamma \geq 0$. Let $s = k + 1 + \rho$ for $k \geq 0$, $\rho \in (0, 1]$. Suppose also that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. Then*

$$\sup_{y\in\mathbb{R}}\left|\mathbb{P}_\theta\left\{\frac{f_k(X) - f(\theta)}{\sigma_{f,\xi}(\theta)} \leq y\right\} - \mathbb{P}\{Z \leq y\}\right| \lesssim_\gamma s^{\gamma/2}K_{s,\gamma}(f; \Sigma; \theta)$$
$$\left((\mathbb{E}^{1/2}\|\xi\|^2)^\rho\sqrt{\log\left(\frac{1}{\|\Sigma\|}\right)} \bigvee \|\Sigma\|^{\rho/2}\log^{(1+\rho)/2}\left(\frac{1}{\|\Sigma\|}\right) \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\right), \tag{2.7}$$

where Z is a standard normal r.v. Moreover,

$$\left\|\frac{f_k(X) - f(\theta)}{\sigma_{f,\xi}(\theta)} - Z\right\|_{L_2(\mathbb{P})}$$
$$\lesssim_\gamma s^{\gamma/2}K_{s,\gamma}(f; \Sigma; \theta)\left((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\right). \tag{2.8}$$

It follows from bound (2.8) that

$$\frac{\mathbb{E}_\theta^{1/2}(f_k(X) - f(\theta))^2}{\sigma_{f,\xi}(\theta)}$$
$$\leq 1 + c_\gamma s^{\gamma/2}K_{s,\gamma}(f; \Sigma; \theta)\left((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\right). \tag{2.9}$$

Assume that θ is in a set $\Theta \subset E$ of parameters where $K_{s,\gamma}(f; \Sigma; \theta)$ is upper bounded by a constant. Then, $\frac{\mathbb{E}_\theta^{1/2}(f_k(X)-f(\theta))^2}{\sigma_{f,\xi}(\theta)}$ is close to 1 uniformly in Θ provided that $\mathbb{E}\|\xi\|^2$ is small and $(\mathbb{E}\|\xi\|^2)^s$ is much smaller than $\|\Sigma\|$ (say, if $\mathbf{r}(\Sigma) \lesssim \sigma^{-2\alpha}$ and $s > \frac{1}{1-\alpha}$).

Finally, in Section 6, we will prove the following minimax lower bound.

Theorem 2.4. *Suppose $f \in C^{s,\gamma}(E)$ for some $s \in (1, 2]$ and $\gamma \geq 0$. Let $\theta_0 \in \overline{\mathrm{Im}(\Sigma)}$.*[1] *Then, there exists a constant $D_\gamma > 0$ such that for all $c > 0$ and all covariance operators Σ satisfying the condition $c\|\Sigma\|^{1/2} \leq 1$, the following bound holds*

$$\inf_T \sup_{\|\theta-\theta_0\|\leq c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta(T(X) - f(\theta))^2}{\sigma^2_{f,\xi}(\theta)} \geq 1 - D_\gamma K^2_{s,\gamma}(f; \Sigma; \theta_0)\left(c^{s-1}\|\Sigma\|^{(s-1)/2} + \frac{1}{c^2}\right),$$

where the infimum is taken over all possible estimators $T(X)$.

[1] $\overline{\mathrm{Im}(\Sigma)}$ denotes the closure of the range of operator Σ.

The bound of Theorem 2.4 shows that, when the noise level $\|\Sigma\|^{1/2}$ is small and $K_{s,\gamma}(f;\Sigma;\theta_0)$ is upper bounded by a constant, the following asymptotic minimax result (in spirit of Hàjek and Le Cam) holds

$$\lim_{c\to\infty}\liminf_{\|\Sigma\|^{1/2}\to 0}\inf_T \sup_{\|\theta-\theta_0\|\leq c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta(T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta)} \geq 1$$

locally in a neighborhood of parameter θ_0 of size commensurate with the noise level. This shows the optimality of the variance $\sigma^2_{f,\xi}(\theta)$ of normal approximation and the efficiency of estimator $f_k(X)$.

Remark 2.3. In the case of matrix Gaussian shift model of Example 1.2 (that is, when E is the space of symmetric $d\times d$ matrices equipped with operator norm and $\xi=\sigma Z$, Z being a random matrix from Gaussian orthogonal ensemble), the results of the paper could be applied to bilinear forms of smooth functions of $d\times d$ symmetric matrices: $f(\theta):=\langle h(\theta)u,v\rangle$, where h is a smooth function in real line and $u,v\in\mathbb{R}^d$. The value $h(\theta)$ of a function h on a matrix (operator) θ is defined via the standard functional calculus. It was shown in [20], Corollary 2 (based on the results of [33], [1]) that the C^s-norms of operator functions $\theta\mapsto h(\theta)$ (on the space E of symmetric matrices equipped with the operator norm) could be controlled in terms of Besov $B^s_{\infty,1}$-norm of the underlying function of real variable h : $\|h\|_{C^s(E)}\lesssim_s \|h\|_{B^s_{\infty,1}}$, $s>0$. This allows one to apply all the results stated above to the functional $f(\theta)$ provided that h is in a proper Besov space. Note that spectral projections of θ that correspond to subsets of its spectrum separated by a positive gap from the rest of the spectrum could be represented as $h(\theta)$ for a sufficiently smooth functions h (namely, a smooth function h taking value 1 on the subset of interest and value 0 on the rest of the spectrum). This allows one to apply the results to bilinear forms of spectral projections (see also [24]). In [20], similar results were obtained for smooth functionals of covariance operators.

Remark 2.4. Obviously, the results of the paper can be applied to the model of i.i.d. observations $X_1,\dots,X_n\sim\mathcal{N}(\theta;\Sigma)$, $\theta\in E$. If $\bar{X}:=\frac{X_1+\cdots+X_n}{n}$, then one can define functions f_k for the Gaussian shift model $\bar{X}=\theta+\bar{\xi}$, where $\bar{\xi}:=\frac{\xi_1+\cdots+\xi_n}{n}$, $\xi_1,\dots,\xi_n$ i.i.d. $\mathcal{N}(0;\Sigma)$. It follows from Theorem 2.1 that

$$\mathbb{E}_\theta(T_k(\bar{X})-f(\theta))^2 \lesssim_\gamma s^\gamma\|f\|^2_{C^{s,\gamma}}(1\vee\|\theta\|)^{2\gamma}\Bigl(\Bigl(\frac{\|\Sigma\|}{n}\bigvee\Bigl(\frac{\|\Sigma\|\mathbf{r}(\Sigma)}{n}\Bigr)^s\bigwedge 1\Bigr), \tag{2.10}$$

where $T_k(\bar{X}) = f_k(\bar{X})$ if $\frac{\|\Sigma\|\mathbf{r}(\Sigma)}{n} \leq \frac{1}{4}$ and $T_k(\bar{X}) = 0$ otherwise. Uniformly in the class of covariances with $\|\Sigma\| \lesssim 1$ and $\mathbf{r}(\Sigma) \lesssim n^{\alpha}$ for some $\alpha \in (0,1)$, this yields a bound on the mean squared error of the order $O(\frac{1}{n})$ provided that $s \geq \frac{1}{1-\alpha}$. Moreover, if $s > \frac{1}{1-\alpha}$, estimator $f_k(\bar{X})$ is asymptotically normal and asymptotically efficient with convergence rate $\sqrt{n}$ and limit variance $\sigma^2_{f,\xi}(\theta)$. Asymptotic efficiency holds in bounded subsets of parameters where $K_{s,\gamma}(f;\Sigma;\theta)$ is uniformly bounded (which essentially means that $\sigma_{f,\xi}(\theta)$ is bounded away from zero).

Remark 2.5. The results of this paper could not be *directly* applied to nonparametric model (1.2) studied in [16, 31, 32]. In this model, the signal $\theta \in L_2([0,1])$ is observed in a standard Gaussian white noise, which is not a random element in the space $L_2([0,1])$ (in fact, the infinite-dimensional standard Gaussian white noise could be viewed as a mean zero Gaussian random element with identity covariance that belongs to a space of generalized functions). In some sense, one can view this model as a version of Gaussian shift model with infinite effective rank $\mathbf{r}(\Sigma) = +\infty$ of the covariance operator $\Sigma = I_{L_2([0,1])}$. However, the bias reduction method studied in our paper could be, in principle, used to develop estimators with optimal error rates for such models as (1.2). For instance, suppose that $\Theta \subset L_2([0,1])$ is the parameter space of the model and $f : \Theta \mapsto \mathbb{R}$ is a smooth functional to be estimated based on the observation $X^{(n)}$. One can choose a finite dimensional subspace $L \subset L_2([0,1])$ that provides a good approximation of signals $\theta \in \Theta$ (under proper "complexity" assumptions, for instance, on the rate of decay of Kolmogorov's diameters of set Θ as in [16, 31, 32]). The functional $f(\theta)$ could be now approximated by $f(P_L\theta)$, where P_L is the orthogonal projection onto L. One can then use a projection type estimator $\hat{\theta}_L$ of $P_L\theta$ based on $X^{(n)}$ to develop estimators $f_k(\hat{\theta}_L)$ of $f(P_L\theta)$ with reduced bias. The choice of an approximating space L should be based on a trade-off between the approximation error $f(\theta) - f(P_L\theta)$ and the error of estimator $f_k(\hat{\theta}_L)$ of $f(P_L\theta)$. This approach could potentially lead to extensions of the results of [16, 31, 32] to broader classes of models under various smoothness assumptions on the functional f and complexity assumptions on the parameter space Θ, but its full development poses further challenges and is beyond the scope of this paper.

For some special functionals f, there exist explicit analytic expressions for functions f_k.

Example 2.1. For instance, consider the standard finite-dimensional Gaussian shift model of Example 1.1. In this case, $E = \mathbb{R}^d$ is equipped with the canonical Euclidean norm and $\Sigma = \sigma^2 I_d$ for some $\sigma^2 > 0$. For $d = 1$ and $\sigma^2 = 1$, it is well known that the unique unbiased estimator of $\theta^k, k \geq 0$ is $H_k(X)$, $H_j, j \geq 0$ being the Hermite polynomials. It easily follows that for $d = 1$ and arbitrary $\sigma^2 > 0$, the unique unbiased estimator of θ^k is $\sigma^k H_k\Big(\frac{X}{\sigma}\Big)$. If now $d \geq 1$ and

$$f(\theta) := \sum_{k_1,\dots,k_d} c_{k_1,\dots,k_d} \theta_1^{k_1} \dots \theta_d^{k_d}, \theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$$

is a polynomial of degree $m+1$, then the unique unbiased estimator of $f(\theta)$ is $g(X)$, where

$$g(\theta) = \sum_{k_1,\dots,k_d} c_{k_1,\dots,k_d} \sigma^{k_1+\dots+k_d} H_{k_1}\Big(\frac{\theta_1}{\sigma}\Big) \dots H_{k_d}\Big(\frac{\theta_d}{\sigma}\Big).$$

It will be shown (see Corollary 3.1) that $f_k(X)$ is an unbiased estimator of $f(\theta)$ for all $k \geq \Big[\frac{m+1}{2}\Big]$. Due to the uniqueness of unbiased estimator for this model, $f_k = g$ for all $k \geq \Big[\frac{m+1}{2}\Big]$.

Example 2.2. Let now E be an arbitrary Banach space and the covariance Σ of the Gaussian shift model be arbitrary. It is easy to see that

$$\mathcal{T} \cos(\langle \cdot, u\rangle) = \exp\Big\{-\langle \Sigma u, u\rangle/2\Big\} \cos(\langle \cdot, u\rangle),$$
$$\mathcal{T} \sin(\langle \cdot, u\rangle) = \exp\Big\{-\langle \Sigma u, u\rangle/2\Big\} \sin(\langle \cdot, u\rangle),\ u \in E^*,$$

which follows from the fact that the same property holds for complex exponentials $\exp\{i\langle \cdot, u\rangle\}, u \in E^*$. Therefore,

$$\mathcal{B}^j \cos(\langle \cdot, u\rangle) = \Big(\exp\Big\{-\langle \Sigma u, u\rangle/2\Big\} - 1\Big)^j \cos(\langle \cdot, u\rangle),$$
$$\mathcal{B}^j \sin(\langle \cdot, u\rangle) = \Big(\exp\Big\{-\langle \Sigma u, u\rangle/2\Big\} - 1\Big)^j \sin(\langle \cdot, u\rangle),\ u \in E^*, j \geq 0.$$

It is easy to conclude that if

$$f(x) = \sum_i \Big[c_i \cos(\langle x, u_i\rangle) + d_i \sin(\langle x, u_i\rangle)\Big], x \in E,$$

where $u_i \in E^*, c_i, d_i \in \mathbb{R}$ and the sum is finite, then

$$f_k(x) = \sum_i \lambda_k(\Sigma, u_i)\Big[c_i \cos(\langle x, u_i\rangle) + d_i \sin(\langle x, u_i\rangle)\Big],\ x \in E,$$

where

$$\lambda_k(\Sigma, u) := \exp\Big\{\langle \Sigma u, u\rangle/2\Big\}\Big[1 - \Big(1 - \exp\Big\{-\langle \Sigma u, u\rangle/2\Big\}\Big)^{k+1}\Big], u \in E^*.$$

3. Bias Reduction

A crucial part of our approach to efficient estimation of smooth functionals of θ is a new bias reduction method based on iterative application of parametric bootstrap. Our goal is to construct an estimator of smooth functional $f(\theta)$ of parameter $\theta \in E$ and, to this end, we construct an estimator of the form $g(X)$ for some functional $g : E \mapsto \mathbb{R}$ for which the bias $\mathbb{E}_\theta g(X) - f(\theta)$ is negligible comparing with the noise level $\|\Sigma\|^{1/2}$. Define the following linear operator:

$$(\mathcal{T}g)(\theta) := \mathbb{E}_\theta g(X) = \mathbb{E}g(\theta + \xi), \theta \in E.$$

Proposition 3.1. *For all $\gamma \geq 0$, $\mathcal{T}$ is a bounded linear operator from the space $L_{\infty,\gamma}(E)$ into itself with*

$$\|\mathcal{T}\|_{L_{\infty,\gamma}(E)\mapsto L_{\infty,\gamma}(E)} \leq 2^\gamma(1 + \mathbb{E}\|\xi\|^\gamma). \tag{3.1}$$

PROOF. Indeed, by the definition of $L_{\infty,\gamma}$-norm,

$$|g(\theta + \xi)| \leq 2^\gamma \|g\|_{L_{\infty,\gamma}}(1 \vee \|\theta\| \vee \|\xi\|)^\gamma.$$

Therefore,

$$|(\mathcal{T}g)(\theta)| \leq \mathbb{E}|g(\theta+\xi)| \leq 2^\gamma\|g\|_{L_{\infty,\gamma}}\mathbb{E}(1\vee\|\theta\|\vee\|\xi\|)^\gamma \leq 2^\gamma[(1\vee\|\theta\|)^\gamma+\mathbb{E}\|\xi\|^\gamma]\|g\|_{L_{\infty,\gamma}},$$

which easily implies that

$$\|\mathcal{T}g\|_{L_{\infty,\gamma}} \leq 2^\gamma(1 + \mathbb{E}\|\xi\|^\gamma)\|g\|_{L_{\infty,\gamma}}. \tag{3.2}$$

Therefore $\mathcal{T}$ is a bounded operator from $L_{\infty,\gamma}(E)$ into itself and bound (3.1) holds. □

The following proposition could be easily proved by induction.

Proposition 3.2. *Let $\xi_1, \xi_2, \dots$ be i.i.d. copies of ξ and let $g \in L_{\infty,\gamma}(E)$ for some $\gamma > 0$. Then, for all $k \geq 1$,*

$$(\mathcal{T}^k g)(\theta) = \mathbb{E}g\biggl(\theta + \sum_{j=1}^k \xi_j\biggr), \theta \in E.$$

Note that, by a simple modification of the proof of bound (3.2), we can derive from Proposition 3.2 that

$$\|\mathcal{T}^k g\|_{L_{\infty,\gamma}} \leq 2^\gamma(1 + k^{\gamma/2}\mathbb{E}\|\xi\|^\gamma)\|g\|_{L_{\infty,\gamma}}. \tag{3.3}$$

To find an estimator $g(X)$ of $f(\theta)$ with a small bias it suffices to solve (approximately) the equation $(\mathcal{T}g)(\theta) = f(\theta), \theta \in E$. Denote $\mathcal{B} =: \mathcal{T} - \mathcal{I}$. For a small level of noise ξ, one can expect operator $\mathcal{B}$ to be "small". The solution of equation $\mathcal{T}g = f$ could be then formally written as a Neumann series:

$$g = (\mathcal{I} + \mathcal{B})^{-1} f = (\mathcal{I} - \mathcal{B} + \mathcal{B}^2 - \dots) f.$$

We use a partial sum of this series as an approximate solution

$$f_k(x) := \sum_{j=0}^{k} (-1)^j (\mathcal{B}^j f)(x), x \in E$$

and consider in what follows the estimator $f_k(X)$ of $f(\theta)$.

Our main goal in this section is to prove the following theorem that provides an upper bound on the bias of estimator $f_k(X)$.

Theorem 3.1. *Let $s = k + 1 + \rho$ for some $\rho \in (0, 1]$ and let $\gamma \geq 0$. Suppose that $f \in C^{s,\gamma}(E)$. Denote by $(\mathfrak{B}_k f)(\theta) := \mathbb{E}_\theta f_k(X) - f(\theta), \theta \in E$ the bias of estimator $f_k(X)$. Then*

$$\|\mathfrak{B}_k f\|_{L_{\infty,\gamma}} \lesssim 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} (1 + k^{\gamma/2} \mathbb{E}^{1/2} \|\xi\|^{2\gamma})(1 + \mathbb{E}^{1/2} \|\xi\|^{2\gamma})(\mathbb{E}^{1/2} \|\xi\|^2)^s.$$

By a straightforward simple computation, the bias of estimator $f_k(X)$ is equal to

$$(\mathfrak{B}_k f)(\theta) = \mathbb{E} f_k(X) - f(\theta) = (-1)^k (\mathcal{B}^{k+1} f)(\theta). \tag{3.4}$$

This leaves us with the problem of bounding $(\mathcal{B}^{k+1} f)(\theta)$ for a sufficiently smooth function f. By Newton's Binomial Formula, for all $k \geq 1$,

$$(\mathcal{B}^k f)(\theta) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (T^j f)(\theta), \theta \in E. \tag{3.5}$$

It follows from representation (3.5) and bound (3.3) that

$$\|\mathcal{B}^k g\|_{L_{\infty,\gamma}} \leq 2^\gamma \sum_{j=0}^{k} \binom{k}{j} (1 + j^{\gamma/2} \mathbb{E}\|\xi\|^\gamma) \|g\|_{L_{\infty,\gamma}} \leq 2^{k+\gamma} \|g\|_{L_{\infty,\gamma}} (1 + k^{\gamma/2} \mathbb{E}\|\xi\|^\gamma). \tag{3.6}$$

Remark 3.1. Define $\hat{\theta}^{(k)} := \theta + \sum_{j=1}^{k} \xi_j, k \geq 1$ and $\hat{\theta}^{(0)} := \theta$. Then $\hat{\theta}^{(1)} = \hat{\theta} = X$ is the maximum likelihood estimator of parameter θ, $\hat{\theta}^{(2)}$ is a parametric bootstrap estimator corresponding to $\hat{\theta}$, and $\hat{\theta}^{(k)}, k \geq 2$ could be viewed as successive iterations of parametric bootstrap for Gaussian shift model $X \sim \mathcal{N}(\theta, \Sigma), \theta \in E$. Similar sequence of bootstrap estimators (that form a Markov chain) was studied in [20] in the case of covariance estimation and it was called *a bootstrap chain*. It immediately follows from (3.5) and Proposition 3.2 that

$$(\mathcal{B}^k f)(\theta) = \mathbb{E}_\theta \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\theta}^{(j)}), \theta \in E, \tag{3.7}$$

which means that $(\mathcal{B}^k f)(\theta)$ is equal to the expectation of the k-th order difference of sequence $f(\hat{\theta}^j), j \geq 0$. This interpretation of the function $(\mathcal{B}^k f)(\theta)$ as an expectation could be used to compute the value of this function for a given θ via Monte Carlo method (namely, by averaging the k-th order differences $\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\theta}^{(j)})$ for a finite number N of i.i.d. copies of bootstrap chain $\{\hat{\theta}^{(j)}, j = 0, \dots, k\}$). As a consequence, Monte Carlo approximation could be also used to compute the estimator $f_k(X)$. The bias reduction method studied in this section is a special case of a more general bootstrap chain bias reduction developed in the case of estimation of functionals of covariance in [20]. Operators similar to $\mathcal{B}^k$ were used also in [18] in the problem of bias reduction in estimation of $f(\theta)$, where θ is the parameter of binomial model. In this case, $\mathcal{T}$ maps function f to the corresponding Bernstein polynomial and bounds on $(\mathcal{B}^k f)(\theta)$ could be obtained using some results in approximation theory.

For sufficiently smooth functions f, we will derive a more convenient integral representation of functions $\mathcal{B}^k f$ that would yield sharper bounds on their $L_{\infty,\gamma}$ norms.

Theorem 3.2. *Suppose $f \in C^{k,\gamma}(E)$ for some $\gamma \geq 0$. Then*

$$(\mathcal{B}^k f)(\theta) = \mathbb{E} f^{(k)}\biggl(\theta + \sum_{j=1}^{k} \tau_j \xi_j\biggr)(\xi_1, \dots, \xi_k), \theta \in E,$$

where $\tau_1, \dots, \tau_k \sim U[0,1]$ are i.i.d. random variables independent of $\xi_1, \dots, \xi_k$.

PROOF. Define

$$\varphi(t_1, \dots, t_k) := f\biggl(\theta + \sum_{i=1}^{k} t_i \xi_i\biggr), (t_1, \dots, t_k) \in [0,1]^k.$$

It immediately follows from Proposition 3.2 that

$$(\mathcal{T}^j f)(\theta) = \mathbb{E}\varphi(t_1, \dots, t_k)$$

for all $j \le k$ and for all $(t_1, \dots, t_k) \in \{0,1\}^k$ with $\sum_{i=1}^k t_i = j$. This allows us to rewrite representation (3.5) as follows:

$$(\mathcal{B}^k f)(\theta) = \mathbb{E} \sum_{(t_1,\dots,t_k)\in\{0,1\}^k} (-1)^{k-\sum_{i=1}^k t_i} \varphi(t_1, \dots, t_k).$$

For functions $\phi : [0,1]^k \mapsto \mathbb{R}$, define the first order difference operators $\Delta^{(i)}, i = 1, \dots, k$:

$$\Delta^{(i)}\phi(t_1, \dots, t_k) \coloneqq \phi(t_1, \dots, t_k)\big|_{t_i=1} - \phi(t_1, \dots, t_k)\big|_{t_i=0}.$$

It is easy to show by induction that

$$\Delta^{(1)} \dots \Delta^{(k)}\phi = \sum_{(t_1,\dots,t_k)\in\{0,1\}^k} (-1)^{k-\sum_{i=1}^k t_i} \phi(t_1, \dots, t_k),$$

implying that

$$(\mathcal{B}^k f)(\theta) = \mathbb{E}\Delta^{(1)} \dots \Delta^{(k)}\varphi.$$

For $f \in C^{k,\gamma}(E)$, the function φ is k times continuously differentiable on $[0,1]^k$ with

$$\frac{\partial^k \varphi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} = f^{(k)}\left(\theta + \sum_{j=1}^k t_j\xi_j\right)(\xi_1, \dots, \xi_k).$$

By generalized Newton-Leibnitz formula,

$$\Delta^{(1)} \dots \Delta^{(k)}\varphi = \int_0^1 \dots \int_0^1 \frac{\partial^k \varphi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k.$$

Therefore,

$$(\mathcal{B}^k f)(\theta) = \mathbb{E} \int_0^1 \dots \int_0^1 f^{(k)}\left(\theta + \sum_{j=1}^k t_j\xi_j\right)(\xi_1, \dots, \xi_k) dt_1 \dots dt_k,$$

which implies the result.

□

Remark 3.2. The representation formula of Theorem 3.2 for functions $(\mathcal{B}^k f)(\theta)$ suggests another approach to approximate computation of the value of these functions at a given point θ. Namely, one can simulate sufficiently large number N of i.i.d. copies of r.v. $\{\tau_j, \xi_j : 1 \leq j \leq k\}$ (recall that the distribution of the noise is known) and then compute the average of r.v. $f^{(k)}\Bigl(\theta + \sum_{j=1}^k \tau_j \xi_j\Bigr)(\xi_1, \dots, \xi_k)$ for these i.i.d. copies. This approach yields another Monte Carlo approximation of estimator $f_k(X)$.

Corollary 3.1. *Let $f : E \mapsto \mathbb{R}$ be a polynomial of degree $m+1 \geq 1$. Then $\mathcal{B}^{m+1} f = 0$ and, as a consequence, $f_m(X)$ is an unbiased estimator of $f(\theta)$. Moreover, $\mathcal{B}^j f = 0$ for all $j > (m+1)/2$ implying that, for all $k \geq \left[\frac{m+1}{2}\right]$, $f_k(X) = f_{[\frac{m+1}{2}]}(X)$ is an unbiased estimator of $f(\theta)$.*

PROOF. Note that $f \in C^{s,m+1}(E)$ for all $s > 0$. Since $f^{(m+1)}(\theta) = M, \theta \in E$ for some $M \in \mathcal{M}_{m+1}(E;\mathbb{R})$, we can use independence of $\xi_1, \dots, \xi_{m+1}$ to get

$$(\mathcal{B}^{m+1} f)(\theta) = \mathbb{E} M(\xi_1, \dots, \xi_{m+1}) = M(\mathbb{E}\xi_1, \dots, \mathbb{E}\xi_{m+1}) = 0$$

and

$$(\mathfrak{B}_m f)(\theta) = (-1)^k (\mathcal{B}^{m+1} f)(\theta) = 0.$$

To prove the second claim, note that, for $j > (m+1)/2$, $f^{(j)}$ is a polynomial of degree $m+1-j < j$. By the Taylor expansion, for a fixed θ, $x \mapsto f^{(j)}(\theta + x)$ is also a polynomial of degree $m+1-j$. Therefore, we can now represent

$$f^{(j)}\Bigl(\theta + \sum_{i=1}^j \tau_i \xi_i\Bigr)(\xi_1, \dots, \xi_j)$$

as a sum of multilinear forms

$$M(\tau_{i_1}\xi_{i_1}, \dots, \tau_{i_l}\xi_{i_l})(\xi_1, \dots, \xi_j),\ l \leq m+1-j < j, 1 \leq i_1, \dots, i_l \leq j.$$

Since $\{i_1, \dots, i_l\}$ is a proper subset of $\{1, \dots, j\}$, we easily get by conditioning that

$$\mathbb{E} M(\tau_{i_1}\xi_{i_1}, \dots, \tau_{i_l}\xi_{i_l})(\xi_1, \dots, \xi_j) = 0,$$

which implies that

$$(\mathcal{B}^j f)(\theta) = \mathbb{E} f^{(j)}\Bigl(\theta + \sum_{i=1}^j \tau_i \xi_i\Bigr)(\xi_1, \dots, \xi_j) = 0, \theta \in E.$$

□

Remark 3.3. Other representations of unbiased estimators of polynomials of parameter θ of Gaussian shift model (especially, in the case of standard model of Example 1.1) could be found in the literature (in particular, see [16]).

Representation of Theorem 3.2 could be now used to provide an upper bound on $L_{\infty,\gamma}$-norm of function $\mathcal{B}^k f$.

Proposition 3.3. *For all $\gamma \geq 0$ and all $f \in C^{k,\gamma}(E)$, the following bound holds:*

$$\|\mathcal{B}^k f\|_{L_{\infty,\gamma}} \leq 2^\gamma \|f^{(k)}\|_{L_{\infty,\gamma}} \Big(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}\Big)(\mathbb{E}^{1/2}\|\xi\|^2)^k.$$

PROOF. Observe that

$$\Big|f^{(k)}\Big(\theta + \sum_{j=1}^k \tau_j\xi_j\Big)(\xi_1,\dots,\xi_k)\Big| \leq \|f^{(k)}\|_{L_{\infty,\gamma}} \bigg(1 \vee \bigg\|\theta + \sum_{j=1}^k \tau_j\xi_j\bigg\|\bigg)^\gamma \|\xi_1\|\dots\|\xi_k\|$$

$$\leq 2^\gamma \|f^{(k)}\|_{L_{\infty,\gamma}} \bigg(1 \vee \|\theta\| \vee \bigg\|\sum_{j=1}^k \tau_j\xi_j\bigg\|\bigg)^\gamma \|\xi_1\|\dots\|\xi_k\|,$$

implying that

$$|(\mathcal{B}^k f)(\theta)| \leq 2^\gamma \|f^{(k)}\|_{L_{\infty,\gamma}} \mathbb{E}^{1/2}\bigg(1 \vee \|\theta\| \vee \bigg\|\sum_{j=1}^k \tau_j\xi_j\bigg\|\bigg)^{2\gamma} \mathbb{E}^{1/2}(\|\xi_1\|^2\dots\|\xi_k\|^2)$$

$$\leq 2^\gamma \|f^{(k)}\|_{L_{\infty,\gamma}} \bigg((1 \vee \|\theta\|)^{2\gamma} + \mathbb{E}\bigg\|\sum_{j=1}^k \tau_j\xi_j\bigg\|^{2\gamma}\bigg)^{1/2} (\mathbb{E}^{1/2}\|\xi\|^2)^k.$$

Next note that, conditionally on $\tau_1,\dots,\tau_k$, the distribution of $\sum_{j=1}^k \tau_j\xi_j$ is the same as the distribution of r.v. $\Big(\sum_{j=1}^k \tau_j^2\Big)^{1/2}\xi$. Therefore,

$$\mathbb{E}\bigg\|\sum_{j=1}^k \tau_j\xi_j\bigg\|^{2\gamma} = \mathbb{E}\Big(\sum_{j=1}^k \tau_j^2\Big)^\gamma \mathbb{E}\|\xi\|^{2\gamma} \leq k^\gamma \mathbb{E}\|\xi\|^{2\gamma},$$

and we get

$$|(\mathcal{B}^k f)(\theta)| \leq 2^\gamma \|f^{(k)}\|_{L_{\infty,\gamma}} \Big((1 \vee \|\theta\|)^\gamma + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}\Big)(\mathbb{E}^{1/2}\|\xi\|^2)^k.$$

This yields the bound of the proposition.

□

The next corollary is immediate.

Corollary 3.2. *Suppose* $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. *For all* $\gamma \geq 0$ *and all* $f \in C^{k,\gamma}(E)$, *the following bound holds:*

$$\|f_k\|_{L_{\infty,\gamma}} \leq 2^{\gamma+1}\|f\|_{C^{k,\gamma}}\Big(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}\Big) \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}.$$

Theorem 3.3. *Suppose* $f \in C^{k+1,\gamma}(E)$ *for some* $\gamma \geq 0$. *Then* $\theta \mapsto (\mathcal{B}^k f)(\theta)$ *is Fréchet differentiable with continuous derivative*

$$(\mathcal{B}^k f)'(\theta)(h) = \mathbb{E}f^{(k+1)}\Big(\theta + \sum_{j=1}^k \tau_j\xi_j\Big)(\xi_1, \dots, \xi_k, h), \theta, h \in E, \tag{3.8}$$

where $\tau_1, \dots, \tau_k$ *are i.i.d. random variables uniformly distributed in* $[0, 1]$ *and independent of* $\xi_1, \dots, \xi_k$.

PROOF. First note that the expression in the right hand side of (3.8) is well defined. This easily follows from the bound

$$\begin{aligned}&\Big\|f^{(k+1)}\Big(\theta + \sum_{j=1}^k \tau_j\xi_j\Big)(\xi_1, \dots, \xi_k)\Big\| \\ &\leq \|f^{(k+1)}\|_{L_{\infty,\gamma}}\Big(1 \vee \|\theta\| \vee \Big\|\sum_{j=1}^k \tau_j\xi_j\Big\|\Big)^\gamma \|\xi_1\| \dots \|\xi_k\|\end{aligned}$$

whose right hand side has finite expectation. By Lebesgue dominated convergence theorem, this also implies the continuity of the function $\theta \mapsto (\mathcal{B}^k f)'(\theta)(h)$ defined by expression (3.8). It remains to show that this expression indeed provides the derivative of $\mathcal{B}^k f$. To this end, observe that

$$\begin{aligned}&f^{(k)}\Big(\theta + h + \sum_{j=1}^k \tau_j\xi_j\Big)(\xi_1, \dots, \xi_k) - f^{(k)}\Big(\theta + \sum_{j=1}^k \tau_j\xi_j\Big)(\xi_1, \dots, \xi_k) \\ &= \int_0^1 f^{(k+1)}\Big(\theta + th + \sum_{j=1}^k \tau_j\xi_j\Big)(\xi_1, \dots, \xi_k, h)dt,\end{aligned}$$

which implies

$$
\begin{aligned}
& f^{(k)}\biggl(\theta + h + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k) - f^{(k)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k) \\
& - f^{(k+1)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k,h) \\
& = \int_0^1 \biggl[f^{(k+1)}\biggl(\theta + th + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k,h) - f^{(k+1)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k,h)\biggr] dt
\end{aligned}
$$

and

$$
\begin{aligned}
& (\mathcal{B}^k f)(\theta + h) - (\mathcal{B}^k f)(\theta) - (\mathcal{B}^k f)'(\theta)(h) \\
& = \mathbb{E} \int_0^1 \biggl[f^{(k+1)}\biggl(\theta + th + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k,h) - f^{(k+1)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)(\xi_1,\dots,\xi_k,h)\biggr] dt.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \Bigl|(\mathcal{B}^k f)(\theta + h) - (\mathcal{B}^k f)(\theta) - (\mathcal{B}^k f)'(\theta)(h)\Bigr| \\
& \le \mathbb{E} \int_0^1 \biggl\|f^{(k+1)}\biggl(\theta + th + \sum_{j=1}^k \tau_j \xi_j\biggr) - f^{(k+1)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)\biggr\| dt \|\xi_1\| \dots \|\xi_k\| \|h\|.
\end{aligned}
$$

It remains to observe that by continuity of $f^{(k+1)}$

$$
\biggl\|f^{(k+1)}\biggl(\theta + th + \sum_{j=1}^k \tau_j \xi_j\biggr) - f^{(k+1)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)\biggr\| \to 0 \text{ as } h \to 0, t \in [0,1],
$$

and to use Lebesgue dominated convergence to conclude that

$$
\mathbb{E} \int_0^1 \biggl\|f^{(k+1)}\biggl(\theta + th + \sum_{j=1}^k \tau_j \xi_j\biggr) - f^{(k+1)}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j\biggr)\biggr\| dt \|\xi_1\| \dots \|\xi_k\| = o(1) \text{ as } h \to 0.
$$

This proves Fréchet differentiability of the function $\theta \mapsto (\mathcal{B}^k f)(\theta)$ along with formula (3.8) for its derivative.

□

The following corollary is immediate.

Corollary 3.3. *Suppose $f \in C^{k+1,\gamma}(E)$ for some $\gamma \geq 0$. Then*

$$S_{\mathcal{B}^k f}(\theta; h) = \mathbb{E} S_{f^{(k)}}\biggl(\theta + \sum_{j=1}^k \tau_j \xi_j; h\biggr)(\xi_1, \dots, \xi_k), \theta, h \in E. \tag{3.9}$$

Proposition 3.4. *Let $s = k + 1 + \rho$ for some $\rho \in (0, 1]$ and let $\gamma \geq 0$. Suppose that $f \in C^{s,\gamma}(E)$. Then, for all $j = 1, \dots, k$,*

$$\|(\mathcal{B}^j f)'\|_{L_\infty,\gamma} \leq 2^\gamma \|f^{(j+1)}\|_{L_\infty,\gamma}(1 + j^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^j. \tag{3.10}$$

Moreover, for all $j = 1, \dots, k-1$

$$\|(\mathcal{B}^j f)'\|_{\mathrm{Lip}_{1,\gamma}} \leq 2^\gamma \|f^{(j+1)}\|_{\mathrm{Lip}_{1,\gamma}}(1 + j^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^j \tag{3.11}$$

and

$$\|(\mathcal{B}^k f)'\|_{\mathrm{Lip}_{\rho,\gamma}} \leq 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^k. \tag{3.12}$$

PROOF. We will prove only the last bound of the proposition. The proof of other bounds is similar. Using representation (3.8), we get

$$\begin{aligned}
&\|(\mathcal{B}^k f)'(\theta_1) - (\mathcal{B}^k f)'(\theta_2)\| \\
&\leq \mathbb{E}\biggl\|f^{(k+1)}\biggl(\theta_1 + \sum_{j=1}^k \tau_j\xi_j\biggr) - f^{(k+1)}\biggl(\theta_2 + \sum_{j=1}^k \tau_j\xi_j\biggr)\biggr\|\|\xi_1\|\dots\|\xi_k\| \\
&\leq 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}}\mathbb{E}\biggl(1 \vee \|\theta_1\| \vee \|\theta_2\| \vee \biggl\|\sum_{j=1}^k \tau_j\xi_j\biggr\|\biggr)^\gamma \|\xi_1\|\dots\|\xi_k\|\|\theta_1 - \theta_2\|^\rho \\
&\leq 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}}\mathbb{E}^{1/2}\biggl(1 \vee \|\theta_1\| \vee \|\theta_2\| \vee \biggl\|\sum_{j=1}^k \tau_j\xi_j\biggr\|\biggr)^{2\gamma} (\mathbb{E}^{1/2}\|\xi\|^2)^k\|\theta_1 - \theta_2\|^\rho
\end{aligned}$$

Next recall that, conditionally on $\tau_1, \dots, \tau_k$, $\sum_{j=1}^k \tau_j\xi_j$ has the same distribution as $\Bigl(\sum_{j=1}^k \tau_j^2\Bigr)^{1/2}\xi$. Therefore,

$$\begin{aligned}
&\mathbb{E}\biggl(1 \vee \|\theta_1\| \vee \|\theta_2\| \vee \biggl\|\sum_{j=1}^k \tau_j\xi_j\biggr\|\biggr)^{2\gamma} = \mathbb{E}\biggl(1 \vee \|\theta_1\| \vee \|\theta_2\| \vee \biggl(\sum_{j=1}^k \tau_j^2\biggr)^{1/2}\|\xi\|\biggr)^{2\gamma} \\
&\leq \mathbb{E}\biggl(1 \vee \|\theta_1\| \vee \|\theta_2\| \vee k^{1/2}\|\xi\|\biggr)^{2\gamma} \leq (1 \vee \|\theta_1\| \vee \|\theta_2\|)^{2\gamma} + k^\gamma\mathbb{E}\|\xi\|^{2\gamma}.
\end{aligned}$$

Hence, we easily get

$$\|(\mathcal{B}^k f)'(\theta_1) - (\mathcal{B}^k f)'(\theta_2)\|$$
$$\le 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} \Big[(1 \vee \|\theta_1\| \vee \|\theta_2\|)^\gamma + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}\Big](\mathbb{E}^{1/2}\|\xi\|^2)^k \|\theta_1 - \theta_2\|^\rho,$$

implying the result. □

Proposition 3.5. *Let $s = k + 1 + \rho$ for some $\rho \in (0, 1]$ and let $\gamma \ge 0$. Suppose that $f \in C^{s,\gamma}(E)$. Then*

$$\|\mathcal{B}^{k+1} f\|_{L_{\infty,\gamma}} \lesssim 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} (1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(1 + \mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^s.$$

PROOF. Note that

$$(\mathcal{B}^{k+1} f)(\theta) = (\mathcal{B}\mathcal{B}^k f)(\theta) = \mathbb{E}(\mathcal{B}^k f)(\theta + \xi) - (\mathcal{B}^k f)(\theta)$$
$$= \mathbb{E}(\mathcal{B}^k f)'(\theta)(\xi) + \mathbb{E}S_{\mathcal{B}^k} f(\theta; \xi) = \mathbb{E}S_{\mathcal{B}^k} f(\theta; \xi).$$

Using the first bound of Lemma 2.1 along with bound (3.12), we get

$$|S_{\mathcal{B}^k f}(\theta; \xi)| \lesssim \|(\mathcal{B}^k f)'\|_{\mathrm{Lip}_{\rho,\gamma}} (1 \vee \|\theta\| \vee \|\xi\|)^\gamma \|\xi\|^{1+\rho}$$
$$\lesssim 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} (1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^k (1 \vee \|\theta\| \vee \|\xi\|)^\gamma \|\xi\|^{1+\rho}.$$

Therefore,

$$|(\mathcal{B}^{k+1} f)(\theta)| \le \mathbb{E}|S_{\mathcal{B}^k f}(\theta; \xi)|$$
$$\lesssim 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} (1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^k \mathbb{E}(1 \vee \|\theta\| \vee \|\xi\|)^\gamma \|\xi\|^{1+\rho}$$
$$\lesssim 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} (1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^k \mathbb{E}^{1/2}(1 \vee \|\theta\| \vee \|\xi\|)^{2\gamma} \mathbb{E}^{1/2}\|\xi\|^{2(1+\rho)}.$$

Since for a centered Gaussian random variable ξ and for $\rho \in (0, 1]$,

$$\mathbb{E}^{1/2}\|\xi\|^{2(1+\rho)} \lesssim (\mathbb{E}^{1/2}\|\xi\|^2)^{1+\rho},$$

we get

$$|(\mathcal{B}^{k+1} f)(\theta)|$$
$$\lesssim 2^\gamma \|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}} (1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}) \Big[(1 \vee \|\theta\|)^\gamma + \mathbb{E}^{1/2}\|\xi\|^{2\gamma}\Big](\mathbb{E}^{1/2}\|\xi\|^2)^{k+1+\rho},$$

implying the claim. □

Theorem 3.1 immediately follows from the bound of Proposition 3.5 and formula (3.4).

4. Concentration

In this section, we prove a concentration inequality for random variable $g(\xi)$, where ξ is a Gaussian random vector in E with mean zero and covariance operator Σ and g is a functional on E satisfying the assumption described below. This inequality will be then used to prove concentration bounds for estimator $f_k(X)$.

Assumption 1. *Suppose $g : E \mapsto \mathbb{R}$ satisfies the following Lipschitz condition:*

$$|g(x) - g(x')| \leq L(\|x\| \vee \|x'\|)\|x - x'\|, x, x' \in E,$$

where $\delta \geq 0 \mapsto L(\delta) \in \mathbb{R}_+$ is a non-decreasing continuous function such that

$$L(a\delta) \lesssim_L L(a)e^{\delta^2/2}, \delta \geq 0, a \geq 0. \tag{4.1}$$

It is easy to see that assumption (4.1) on function L implies that for any constant $c_1 > 0$ there exists a constant $c_2 > 0$ (depending only on L) such that $L(c_1\delta) \leq c_2 L(\delta), \delta \geq 0$. Clearly, (4.1) holds for $L(\delta) := C\delta^\alpha, \delta \geq 0$ for arbitrary $C > 0, \alpha \geq 0$. Also, if functions $L_1, \dots, L_m$ satisfy assumption (4.1), then so do the functions $L_1 + \dots + L_m$, $L_1 \vee \dots \vee L_m$. In particular, this implies that any function of the form

$$L(\delta) := C_1\delta^{\alpha_1} \bigvee \cdots \bigvee C_m\delta^{\alpha_m}, \delta \geq 0,$$

where $m \geq 1$, $C_1 > 0, \dots, C_m > 0$ and $\alpha_1 \geq 0, \dots, \alpha_m \geq 0$ are given constants, satisfy assumption (4.1).

Note that, if $g(0) = 0$, then Assumption 1 implies that

$$|g(x)| \leq L(\|x\|)\|x\|, x \in E.$$

Theorem 4.1. *Suppose Assumption 1 holds. For all $t \geq 1$ with probability at least $1 - e^{-t}$,*

$$|g(\xi) - \mathbb{E}g(\xi)| \lesssim L(\mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})\|\Sigma\|^{1/2}\sqrt{t}. \tag{4.2}$$

PROOF. Without loss of generality, assume that $g(0) = 0$. For $\delta > 0$, define

$$h(x) := g(x)\varphi\left(\frac{\|x\|}{\delta}\right), x \in E,$$

where

$$\varphi(u) = \begin{cases} 2 - x, & u \in (1,2), \\ 1, & u \leq 1, \\ 0, & u \geq 2. \end{cases}$$

Clearly, φ is a Lipschitz function with constant 1. We will now prove a Lipschitz condition for the function $h : E \mapsto \mathbb{R}$.

Lemma 4.1. *Under Assumption 1, for all $x, x' \in E$*

$$|h(x) - h(x')| \leq 3L(3\delta)\|x - x'\|.$$

PROOF. Note that

$$\begin{aligned} |h(x) - h(x')| = {} & |h(x) - h(x')| I\{\|x\| \leq 2\delta, \|x'\| \leq 2\delta\} \\ & + |h(x) - h(x')| I\{\|x\| \leq 2\delta, \|x'\| > 2\delta\} \\ & + |h(x) - h(x')| I\{\|x'\| \leq 2\delta, \|x\| > 2\delta\}. \end{aligned} \tag{4.3}$$

For the first summand in the right hand side of (4.3), we have the following bound for $\|x\|, \|x'\| \leq 2\delta$

$$\begin{aligned} & |h(x) - h(x')| I\{\|x\|, \|x'\| \leq 2\delta\} \\ & \leq |g(x) - g(x')| \varphi\left(\frac{\|x\|}{\delta}\right) + |g(x')| \left| \varphi\left(\frac{\|x\|}{\delta}\right) - \varphi\left(\frac{\|x'\|}{\delta}\right) \right| \\ & \leq L(2\delta)\|x - x'\| + L(2\delta)\|x'\|\|x - x'\|/\delta \\ & \leq 3L(2\delta)\|\|x - x'\|. \end{aligned} \tag{4.4}$$

To bound the second summand in (4.3), observe that for $\|x\| \leq 2\delta, \|x'\| > 2\delta$

$$\begin{aligned} & |h(x) - h(x')| I\{\|x\| \leq 2\delta, \|x'\| > 2\delta\} \\ & = |h(x) - h(x')| I\{\|x\| \leq 2\delta, \|x'\| > 2\delta, \|x - x'\| \geq \delta\} \\ & \quad + |h(x) - h(x')| I\{\|x\| \leq 2\delta, \|x'\| > 2\delta, \|x - x'\| < \delta\}, \end{aligned} \tag{4.5}$$

and bound the first term in the right hand side of (4.5) as follows:

$$\begin{aligned} & |h(x) - h(x')| I\{\|x\| \leq 2\delta, \|x'\| > 2\delta, \|x - x'\| \geq \delta\} \\ & = |g(x)| \varphi\left(\frac{\|x\|}{\delta}\right) I\{\|x\| \leq 2\delta, \|x'\| > 2\delta, \|x - x'\| \geq \delta\} \\ & \leq L(2\delta)\|x\| I\{\|x\| \leq 2\delta, \|x'\| > 2\delta, \|x - x'\| \geq \delta\}. \\ & \leq 2L(2\delta)\|x - x'\|. \end{aligned}$$

For the second term in (4.5), we have

$$\begin{aligned}
&|h(x)-h(x')|I\{\|x\|\le 2\delta, \|x'\|>2\delta, \|x-x'\|<\delta\}\\
&=|h(x)-h(x')|I\{\|x\|\le 2\delta, 2\delta<\|x'\|\le 3\delta, \|x-x'\|<\delta\}\\
&\le |h(x)-h(x')|I\{\|x\|\le 3\delta, \|x'\|\le 3\delta\}\\
&\le 3L(3\delta)\|\|x-x'\|,
\end{aligned}$$

where the last inequality is proved similarly to bound (4.4) (with an obvious change of 2δ to 3δ). Substituting the above bounds in (4.3), leads to the resulting inequality. □

In what follows, we set

$$\delta=\delta(t):=\mathbb{E}\|\xi\|+C\|\Sigma\|^{1/2}\sqrt{t}$$

for $t\ge 1$ with a constant $C>0$ such that

$$\mathbb{P}\{\|\xi\|\ge\delta(t)\}\le e^{-t}, t\ge 1$$

(which holds by the Gaussian concentration inequality, see, e.g., [27]).

Let $M:=\mathrm{Med}(g(\xi))$. Assuming that $t\ge\log(4)$, we get

$$\begin{aligned}
&\mathbb{P}\{h(\xi)\ge M\}\ge\mathbb{P}\{h(\xi)\ge M, \|\xi\|\le\delta(t)\}\ge\mathbb{P}\{g(\xi)\ge M, \|\xi\|\le\delta(t)\}\\
&\ge\mathbb{P}\{g(\xi)\ge M\}-\mathbb{P}\{\|\xi\|\ge\delta(t)\}\ge\frac{1}{2}-e^{-t}\ge\frac{1}{4},
\end{aligned}$$

where we used the fact that, on the event $\{\|\xi\|\le\delta\}$, $h(\xi)=g(\xi)$. Similarly, we have $\mathbb{P}\{h(\xi)\le M\}\ge\frac{1}{4}$. We can now use again Gaussian concentration inequality (in a little bit non-standard fashion, see [22], Section 3 for a similar argument) to prove that with probability at least $1-e^{-t}$

$$|h(\xi)-M|\lesssim L(3\delta(t))\|\Sigma\|^{1/2}\sqrt{t}$$

and, since $h(\xi)$ and $g(\xi)$ coincide on the event of probability at least $1-e^{-t}$, we also have that

$$|g(\xi)-M|\lesssim L(3\delta(t))\|\Sigma\|^{1/2}\sqrt{t}$$

with probability at least $1-2e^{-t}$. Moreover, by adjusting the value of the constant in the above inequality, the probability bound can be written in its standard form $1-e^{-t}$ and the inequality holds for all $t\ge 1$. Using the properties of function L (namely, its monotonicity and condition (4.1)) and the definition of $\delta(t)$, we can also rewrite the above bound as

$$|g(\xi)-M|\le C_L(L(\mathbb{E}\|\xi\|)\|\Sigma\|^{1/2}\sqrt{t}\bigvee L(\|\Sigma\|^{1/2}\sqrt{t})\|\Sigma\|^{1/2}\sqrt{t})=:s(t)$$

for some constant $C_L > 0$. Note that this bound actually holds for all $t \geq 0$ with probability at least $1 - e^{1-t}$. Note also that the function $t \mapsto s(t)$ is strictly increasing on $[0, +\infty)$ with $s(0) = 0$ and $s(+\infty) = +\infty$. Moreover, it easily follows from condition (4.1) that $s(t) = o(e^t)$ as $t \to \infty$. It remains to integrate out the tails of the probability bound:

$$
\begin{aligned}
&|\mathbb{E}g(\xi) - M| \leq \mathbb{E}|g(\xi) - M| \\
&= \int_0^\infty \mathbb{P}\{|g(\xi) - M| \geq s\}ds = \int_0^\infty \mathbb{P}\{|g(\xi) - M| \geq s(t)\}ds(t) \\
&\leq e\int_0^\infty e^{-t}ds(t) = e\int_0^\infty s(t)e^{-t}dt.
\end{aligned}
$$

By condition (4.1),

$$
s(t) \leq C_L L(\mathbb{E}\|\xi\|)\|\Sigma\|^{1/2}\sqrt{t} + C_L' L(\|\Sigma\|^{1/2})\|\Sigma\|^{1/2}\sqrt{t}e^{t/2}, t \geq 0.
$$

Therefore,

$$
\begin{aligned}
&|\mathbb{E}g(\xi) - M| \leq \int_0^\infty s(t)e^{-t}dt \\
&\lesssim_L L(\mathbb{E}\|\xi\|)\|\Sigma\|^{1/2}\int_0^\infty \sqrt{t}e^{-t}dt + L(\|\Sigma\|^{1/2})\|\Sigma\|^{1/2}\int_0^\infty \sqrt{t}e^{-t/2}dt \\
&\lesssim_L L(\mathbb{E}\|\xi\|)\|\Sigma\|^{1/2} \bigvee L(\|\Sigma\|^{1/2})\|\Sigma\|^{1/2},
\end{aligned}
$$

which now allows us to replace the median M by the mean $\mathbb{E}g(\xi)$ in the concentration bound, completing the proof. □

The following corollary is immediate (for the proof, check that Assumption 1 holds with $L(\delta) = C\|g\|_{\mathrm{Lip}_{1,\gamma}}(1 \vee \|\theta\| \vee \delta)^\gamma$ for some $C > 0$).

Corollary 4.1. *Suppose $g \in \mathrm{Lip}_{1,\gamma}(E)$ for some $\gamma \geq 0$. Then, for all $\theta \in E$ and for all $t \geq 1$ with probability at least $1 - e^{-t}$*

$$
|g(\theta + \xi) - \mathbb{E}g(\theta + \xi)| \lesssim \|g\|_{\mathrm{Lip}_{1,\gamma}}(1 \vee \|\theta\| \vee \mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})^\gamma\|\Sigma\|^{1/2}\sqrt{t}.
$$

Another immediate corollary of this theorem is the following concentration bound for the remainder $S_g(\theta;\xi)$ of the first order Taylor expansion of $g(\theta + \xi)$. For the proof, it is enough to observe that, by Lemma 2.1, the function $x \mapsto S_g(\theta; x)$ satisfies Assumption 1 with $L(\delta) = C\|g'\|_{\mathrm{Lip}_{\rho,\gamma}}(1 \vee \|\theta\| \vee \delta)^\gamma\delta^\rho$ for some constant $C > 0$.

Corollary 4.2. *Suppose, for some $\rho \in (0, 1]$ and $\gamma \geq 0$, $\|g'\|_{\mathrm{Lip}_{\rho,\gamma}} < \infty$. Then, for all $\theta \in E$ and for all $t \geq 1$ with probability at least $1 - e^{-t}$,*

$$
|S_g(\theta;\xi) - \mathbb{E}S_g(\theta;\xi)| \lesssim \|g'\|_{\mathrm{Lip}_{\rho,\gamma}}(1 \vee \|\theta\| \vee \mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})^\gamma(\mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})^\rho\|\Sigma\|^{1/2}\sqrt{t}.
$$

We now apply corollaries 4.1 and 4.2 to obtain concentration bounds for estimator $f_k(X)$ and the remainder of its first order Taylor expansion.

Proposition 4.1. *Let $\gamma \geq 0$ and suppose that $f \in C^{k+1,\gamma}(E)$ and that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. Then, for all $t \geq 1$, with probability at least $1 - e^{-t}$*

$$|f_k(\theta+\xi) - \mathbb{E}f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1 \vee \|\theta\| \vee \|\Sigma\|^{1/2}\sqrt{t})^\gamma \|\Sigma\|^{1/2}\sqrt{t}. \tag{4.6}$$

PROOF. Using bound (3.10), we get

$$\begin{aligned}
\|f_k'\|_{L_\infty,\gamma} &\leq \sum_{j=0}^k \|(\mathcal{B}^j f)'\|_{L_\infty,\gamma} \\
&\leq 2^\gamma \sum_{j=0}^k \|f^{(j+1)}\|_{L_\infty,\gamma}(1 + j^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^j \\
&\leq 2^\gamma \|f\|_{C^{k+1,\gamma}}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}) \sum_{j=0}^k (\mathbb{E}^{1/2}\|\xi\|^2)^j \\
&\leq 2^{\gamma+1}\|f\|_{C^{k+1,\gamma}}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}) \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}.
\end{aligned}$$

The result now follows from Corollary 4.1. □

With a little additional work, we get the following modification of concentration bound (4.6).

Corollary 4.3. *Let $\gamma \geq 0$ and suppose that $f \in C^{k+1,\gamma}(E)$ and that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. If $\gamma \leq 1$, then, for all $t \geq 1$, with probability at least $1 - e^{-t}$*

$$|f_k(\theta+\xi) - \mathbb{E}f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1 \vee \|\theta\|)^\gamma \|\Sigma\|^{1/2}\sqrt{t}. \tag{4.7}$$

If $\gamma > 1$, then, for all $t \geq 1$, with the same probability

$$|f_k(\theta+\xi) - \mathbb{E}f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1 \vee \|\theta\|)^\gamma \Big(\|\Sigma\|^{1/2}\sqrt{t} \vee (\|\Sigma\|^{1/2}\sqrt{t})^\gamma\Big). \tag{4.8}$$

PROOF. It follows from Corollary 3.2 that $\|f_k\|_{L_\infty,\gamma} \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}$. This implies that

$$|f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(1 \vee \|\theta\| \vee \|\xi\|)^\gamma,$$

which easily yields the following bounds

$$|\mathbb{E}f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(1\vee\|\theta\|)^\gamma$$

and

$$|f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(1\vee\|\theta\|\vee\|\Sigma\|^{1/2}\sqrt{t})^\gamma$$

(the last bound holds for all $t\geq 1$ with probability at least $1-e^{-t}$). Therefore, for all $t\geq 1$ with probability at least $1-e^{-t}$

$$|f_k(\theta+\xi)-\mathbb{E}f_k(\theta+\xi)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(1\vee\|\theta\|\vee\|\Sigma\|^{1/2}\sqrt{t})^\gamma. \quad (4.9)$$

If $\|\Sigma\|^{1/2}\sqrt{t}\leq 1$, then bound (4.7) follows from bound (4.6) (regardless of what the value of $\gamma\geq 0$ is). If $\|\Sigma\|^{1/2}\sqrt{t}>1$, we use bound (4.9) to get

$$\begin{aligned}
&|f_k(\theta+\xi)-\mathbb{E}f_k(\theta+\xi)|\\
&\lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(1\vee\|\theta\|)^\gamma \bigvee (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(\|\Sigma\|^{1/2}\sqrt{t})^\gamma\\
&\lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k,\gamma}}(1\vee\|\theta\|)^\gamma(\|\Sigma\|^{1/2}\sqrt{t})^\gamma,
\end{aligned}$$

which yields (4.8). □

Given an increasing, convex function $\psi:\mathbb{R}_+\mapsto\mathbb{R}_+$ with $\psi(0)=0$ and $\psi(t)\to+\infty$ as $t\to+\infty$ (in what follows, an Orlicz function), the Orlicz ψ-norm of a r.v. η is defined as

$$\|\eta\|_\psi := \inf\biggl\{C>0:\mathbb{E}\psi\biggl(\frac{|\eta|}{C}\biggr)\leq 1\biggr\}.$$

For $p\geq 1$ and $\psi(t):=t^p, t\geq 0$, this yields the usual L_p-norms. Another popular choice is $\psi_\alpha(t):=e^{t^\alpha}-1, t\geq 0$ for some $\alpha\geq 1$, in particular, ψ_2-norm for subgaussian random variables and ψ_1-norm for subexponential random variables.

We will need the following simple lemma.

Lemma 4.2. *Let Y be a non-negative random variable. Suppose, for some $A_1>0,\dots,A_m>0,\beta_1>0,\dots,\beta_m>0$ and for all $t\geq 1$,*

$$\mathbb{P}\{Y\geq A_1t^{\beta_1}\vee\dots\vee A_mt^{\beta_m}\}\leq e^{-t}.$$

Let $\beta:=\max_{1\leq j\leq m}\beta_j$. Then, for any Orlicz function ψ satisfying the condition $\psi(t)\leq c_1e^{c_2t^{1/\beta}}, t\geq 0$ for some constants $c_1,c_2>0$, we have

$$\|Y\|_\psi \lesssim_\psi A_1\vee\dots\vee A_m.$$

Proposition 4.2. *Let $s = k+1+\rho$ for some $\rho \in (0,1]$ and let $\gamma \geq 0$. Suppose that $f \in C^{s,\gamma}(E)$ and that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. If $\gamma \leq 1$, then, for all $t \geq 1$, with probability at least $1-e^{-t}$*

$$|f_k(\theta+\xi) - f(\theta)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma\Bigl(\|\Sigma\|^{1/2}\sqrt{t}\vee(\mathbb{E}^{1/2}\|\xi\|^2)^s\Bigr). \tag{4.10}$$

If $\gamma > 1$, then, for all $t \geq 1$, with the same probability

$$|f_k(\theta+\xi) - f(\theta)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma\Bigl(\|\Sigma\|^{1/2}\sqrt{t}\vee(\|\Sigma\|^{1/2}\sqrt{t})^\gamma\vee(\mathbb{E}^{1/2}\|\xi\|^2)^s\Bigr). \tag{4.11}$$

Moreover, for all Orlicz functions ψ satisfying the condition $\psi(t) \leq c_1 e^{c_2 t^{2/(\gamma\vee 1)}}$ for all $t \geq 0$ and for some constants $c_1, c_2 > 0$, the following bound holds:

$$\|f_k(\theta+\xi) - f(\theta)\|_\psi \lesssim_{\gamma,\psi} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma\Bigl(\|\Sigma\|^{1/2}\vee(\mathbb{E}^{1/2}\|\xi\|^2)^s\Bigr). \tag{4.12}$$

PROOF. The proof immediately follows from bounds (4.7), (4.8), Lemma 4.2 and bound on the bias of Theorem 3.1.

□

Now it is easy to prove Theorem 2.1 stated in Section 2.

PROOF. If $\|\Sigma\|^{1/2}\mathbf{r}^{1/2}(\Sigma) = \mathbb{E}^{1/2}\|\xi\|^2 > 1/2$, then $T_k(X) = 0$ and the claim of the theorem immediately follows from the bound

$$|f(\theta)| \leq \|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma.$$

Assume now that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. If $s > 1$, the result follows from bound (4.12) (with $\psi(t) = t^2$). If $s \in (0,1]$, then $f_k(X) = f(X)$ and

$$|f(X) - f(\theta)| \lesssim_s \|f\|_{C^{s,\gamma}}(1\vee\|\theta\|^\gamma\vee\|\xi\|^\gamma)\|\xi\|^s,$$

which easily implies

$$\begin{aligned}\mathbb{E}_\theta(f_k(X) - f(\theta))^2 &\lesssim_s \|f\|^2_{C^{s,\gamma}}(1\vee\|\theta\|^{2\gamma}\vee\mathbb{E}^{1/2}\|\xi\|^{4\gamma})\mathbb{E}^{1/2}\|\xi\|^{4s} \\ &\lesssim_{s,\gamma} (1\vee\|\theta\|)^{2\gamma}(\mathbb{E}\|\xi\|^2)^s,\end{aligned}$$

and the claim follows again.

□

Remark 4.1. In the case when the functional $f : E \mapsto \mathbb{R}$ is a bounded polynomial of degree $k+1$, estimator $f_{[\frac{k+1}{2}]}(X)$ is unbiased (see Corollary 3.1) and the following version of bound (4.12) for $\psi(t) = t^2$ holds (the proof follows the same lines as the proof of (4.12) with minor modifications):

$$\mathbb{E}_\theta\Bigl(f_{[\frac{k+1}{2}]}(X) - f(\theta)\Bigr)^2 \lesssim_k \|f\|_{\rm op}^2 (1 \vee \|\theta\|)^{2k} \|\Sigma\| \Bigl(1 \vee (\mathbb{E}\|\xi\|^2)^k\Bigr). \tag{4.13}$$

In the case of standard finite-dimensional Gaussian shift model (see Example 1.1 in Section 1) and $f(\theta) = \|\theta\|^2$ (a polynomial of degree 2), it is easy to check that $f_1(X) = \|X\|^2 - \sigma^2 d$. Then, bound (4.13) yields that

$$\sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (f_1(X) - f(\theta))^2 \lesssim \sigma^2 \Bigl(1 \vee \sigma^2 d\Bigr),$$

which could be also proved by an elementary analysis. The last fact is well known and it is crucial in understanding of so called "elbow effect" in estimation of the quadratic functional $f(\theta) = \|\theta\|^2$ in infinite-dimensional (nonparametric) models (the phase transition in the convergence rate at the smoothness degree 1/4).

The following proposition provides a concentration bound on the remainder $S_{f_k}(\theta;\xi)$ of Taylor expansion of function $f_k(\theta+\xi)$ (at point θ). It will be used in the proof of the efficiency of estimators $f_k(X)$.

Proposition 4.3. *Let $s = k + 1 + \rho$ for some $\rho \in (0, 1]$ and let $\gamma \geq 0$. Suppose that $f \in C^{s,\gamma}(E)$ and that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. Then, for all $t \geq 1$, with probability at least $1 - e^{-t}$*

$$\begin{aligned} &|S_{f_k}(\theta;\xi) - \mathbb{E}S_{f_k}(\theta;\xi)| \\ &\lesssim_\gamma (k+1)^{\gamma/2} \|f\|_{C^{s,\gamma}} (1 \vee \|\theta\| \vee \|\Sigma\|^{1/2}\sqrt{t})^\gamma ((\mathbb{E}\|\xi\|)^\rho \vee (\|\Sigma\|^{1/2}\sqrt{t})^\rho \vee \|\Sigma\|^{1/2}\sqrt{t}) \|\Sigma\|^{1/2}\sqrt{t}. \end{aligned} \tag{4.14}$$

PROOF. It follows from bounds (3.11) of Proposition 3.4 that

$$
\begin{aligned}
&\|f'_{k-1}\|_{\mathrm{Lip}_{1,\gamma}} \le \|f'\|_{\mathrm{Lip}_{1,\gamma}} + \sum_{j=1}^{k-1} \|(\mathcal{B}^j f)'\|_{\mathrm{Lip}_{1,\gamma}} \\
&\le \|f'\|_{\mathrm{Lip}_{1,\gamma}} + 2^{\gamma} \sum_{j=1}^{k-1} \|f^{(j+1)}\|_{\mathrm{Lip}_{1,\gamma}} (1 + j^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^j \\
&\le \|f''\|_{L_{\infty,\gamma}} + 2^{\gamma} \sum_{j=1}^{k-1} \|f^{(j+2)}\|_{L_{\infty,\gamma}} (1 + j^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^j \\
&\le \|f\|_{C^{k+1,\gamma}} \biggl(1 + 2^{\gamma}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma}) \sum_{j=1}^{k-1} (\mathbb{E}^{1/2}\|\xi\|^2)^j\biggr) \\
&\le 2^{\gamma+2}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})\|f\|_{C^{k+1,\gamma}}.
\end{aligned}
$$

Using the bound of Corollary 4.2, we get that for all $t \ge 1$ with probability at least $1 - e^{-t}$,

$$
\begin{aligned}
&|S_{f_{k-1}}(\theta;\xi) - \mathbb{E}S_{f_{k-1}}(\theta;\xi)| \\
&\lesssim 2^{\gamma+2}\|f\|_{C^{k+1,\gamma}}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(1 \vee \|\theta\| \vee \mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})^{\gamma}(\mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})\|\Sigma\|^{1/2}\sqrt{t} \\
&\lesssim_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1 \vee \|\theta\| \vee \|\Sigma\|^{1/2}\sqrt{t})^{\gamma}(\mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})\|\Sigma\|^{1/2}\sqrt{t}.
\end{aligned}
$$

Similarly, using the bound of Corollary 4.2 along with bound (3.12) of Proposition 3.4, we get that with probability at least $1 - e^{-t}$

$$
\begin{aligned}
&|S_{\mathcal{B}^k f}(\theta;\xi) - \mathbb{E}S_{\mathcal{B}^k f}(\theta;\xi)| \\
&\lesssim_{\gamma} (k+1)^{\gamma/2}\|f^{(k+1)}\|_{\mathrm{Lip}_{\rho,\gamma}}(\mathbb{E}^{1/2}\|\xi\|^2)^k(1 \vee \|\theta\| \vee \|\Sigma\|^{1/2}\sqrt{t})^{\gamma}(\mathbb{E}\|\xi\| \vee \|\Sigma\|^{1/2}\sqrt{t})^{\rho}\|\Sigma\|^{1/2}\sqrt{t}.
\end{aligned}
$$

Combining these bounds and adjusting the constants yield bound (4.14). □

5. Normal Approximation Bounds

In this section, we develop normal approximation bounds for $f_k(X) - f(\theta)$ needed to complete the proof of Theorem 2.3. More precisely, it will be shown that $f_k(X) - f(\theta)$ could be approximated by a mean zero normal random variable with variance $\sigma^2_{f,\xi}(\theta) := \mathbb{E}(f'(\theta)(\xi))^2 = \langle \Sigma f'(\theta), f'(\theta)\rangle$. Recall that

$$
K(f;\Sigma;\theta) := K_{s,\gamma}(f;\Sigma;\theta) := \frac{\|f\|_{C^{s,\gamma}}(1 \vee \|\theta\|)^{\gamma}\|\Sigma\|^{1/2}}{\sigma_{f,\xi}(\theta)}.
$$

Theorem 5.1. *Suppose, for some $s = k + 1 + \rho$, $\rho \in (0,1]$ and some $\gamma \geq 0$, $f \in C^{s,\gamma}(E)$. Suppose also that $\mathbb{E}^{1/2}\|\xi\|^2 \leq 1/2$. Then, the following representation holds*

$$f_k(X) - f(\theta) = \sigma_{f,\xi}(\theta)Z + R, \tag{5.1}$$

where Z is a standard normal random variable and R is the remainder satisfying, for all $t \geq 1$ with probability at least $1 - e^{-t}$, the bound

$$\begin{aligned}&|R| \lesssim_\gamma s^{\gamma/2}\|f\|_{C^{s,\gamma}}(1 \vee \|\theta\| \vee \|\Sigma\|^{1/2}\sqrt{t})^\gamma \\ &\Big((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \|\Sigma\|^{1/2}\sqrt{t} \vee (\|\Sigma\|^{1/2}\sqrt{t})^{1+\rho} \vee \|\Sigma\|t \vee (\mathbb{E}^{1/2}\|\xi\|^2)^s\Big).\end{aligned} \tag{5.2}$$

Moreover, for any Orlicz function ψ such that $\psi(t) \lesssim c_1 e^{c_2 t^{2/(2+\gamma)}}, t \geq 0$ for some constants $c_1, c_2 > 0$,

$$\begin{aligned}&\left\|\frac{f_k(X) - f(\theta)}{\sigma_{f,\xi}(\theta)} - Z\right\|_\psi \\ &\lesssim_{\gamma,\psi} s^{\gamma/2} K_{s,\gamma}(f;\Sigma;\theta)\Big((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\Big).\end{aligned} \tag{5.3}$$

Remark 5.1. Note that $\mathbb{E}\|\xi\|^2 = \|\Sigma\|\mathbf{r}(\Sigma)$, where $\mathbf{r}(\Sigma)$ is the effective rank of Σ. Assume that $\|\Sigma\|$ is "small" (that is, the noise level is small) and, for some $\alpha \in (0,1)$, $\mathbf{r}(\Sigma) \lesssim \|\Sigma\|^{-\alpha}$. Then $\mathbb{E}\|\xi\|^2 \lesssim \|\Sigma\|^{1-\alpha}$, which is "small", too. Moreover, under the assumption that $s > \frac{1}{1-\alpha}$,

$$\frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}} \lesssim \sqrt{\frac{\|\Sigma\|^{s(1-\alpha)}}{\|\Sigma\|}} = \sqrt{\|\Sigma\|^{s(1-\alpha)-1}}$$

is also "small", implying that the right hand side of bound (5.3) is "small". The same conclusion holds for the right hand side of bound (2.7) provided that $K_{s,\gamma}(f;\Sigma;\theta) \lesssim 1$.

Remark 5.2. We will also state (without providing a proof) the following bound on the risk of estimator $f_k(X)$ with respect to convex loss functions (under some constraints on their growth rate). Let $\ell : \mathbb{R} \mapsto \mathbb{R}_+$ be a loss function such that $\ell(-t) = \ell(t), t \in \mathbb{R}$, ℓ is an Orlicz function on $\mathbb{R}_+$ and, for some $\delta \in (0,1)$, $\nu < \frac{2}{\gamma \vee 1}$

$$\ell(t) \lesssim e^{(1-\delta)t^\nu}, t \geq 0. \tag{5.4}$$

Suppose also that

$$\Big(\mathbb{E}^{1/2}\|\xi\|^2\Big)^s \le \|\Sigma\|^{1/2}. \tag{5.5}$$

Then

$$\begin{aligned}&\left|\mathbb{E}\ell\bigg(\frac{f_k(X)-f(\theta)}{\sigma_{f,\xi}(\theta)}\bigg)-\mathbb{E}\ell(Z)\right|\\ &\lesssim_{\gamma,\ell,\delta} s^{\gamma/2}K_{s,\gamma,\ell,k}(f;\Sigma;\theta)\bigg((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\bigg),\end{aligned} \tag{5.6}$$

where

$$K_{s,\gamma,\ell,k}(f;\Sigma;\theta) := (k+1)^{\gamma/2}K_{s,\gamma}(f;\Sigma;\theta)\bigg(\ell\bigg(c_{\gamma,\nu}(k+1)^{\frac{\gamma}{2-(\gamma\vee 1)\nu}}K_{s,\gamma}^{\frac{1}{1-(\gamma\vee 1)\nu/2}}(f;\Sigma;\theta)\bigg)+1\bigg).$$

Bound (2.8) of Theorem 2.3 follows from bound (5.3) of Theorem 5.1 (for $\psi(t)=t^2$). We now turn to the proof of Theorem 5.1 and bound (2.7) of Theorem 2.3.

PROOF. Clearly,

$$\begin{aligned}&f_k(X)-f(\theta)\\ &= f_k(X)-\mathbb{E}_\theta f_k(X)+\mathbb{E}_\theta f_k(X)-f(\theta)\\ &= f_k'(\theta)(\xi)+S_{f_k}(\theta;\xi)-\mathbb{E}S_{f_k}(\theta;\xi)+\mathbb{E}_\theta f_k(X)-f(\theta)\\ &= \sigma_{f_k,\xi}(\theta)Z+S_{f_k}(\theta;\xi)-\mathbb{E}S_{f_k}(\theta;\xi)+\mathbb{E}_\theta f_k(X)-f(\theta)\\ &= \sigma_{f,\xi}(\theta)Z+R,\end{aligned}$$

where Z is a standard normal random variable and

$$R := (\sigma_{f_k,\xi}(\theta)-\sigma_{f,\xi}(\theta))Z+S_{f_k}(\theta;\xi)-\mathbb{E}S_{f_k}(\theta;\xi)+\mathbb{E}_\theta f_k(X)-f(\theta) \tag{5.7}$$

is the remainder.

The following lemma will be used to control $\sigma_{f_k,\xi}(\theta)-\sigma_{f,\xi}(\theta)$.

Lemma 5.1. *Suppose that, for some $\gamma\ge 0$, $f\in C^{k+1,\gamma}(E)$ and $\mathbb{E}^{1/2}\|\xi\|^2\le 1/2$. Then*

$$|\sigma_{f_k,\xi}(\theta)-\sigma_{f,\xi}(\theta)| \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1\vee\|\theta\|)^\gamma\|\Sigma\|^{1/2}\mathbb{E}^{1/2}\|\xi\|^2. \tag{5.8}$$

PROOF. Note that

$$|\sigma_{f_k,\xi}(\theta) - \sigma_{f,\xi}(\theta)| \le |\sigma_{f_k - f,\xi}(\theta)| \le \sum_{j=1}^{k} \mathbb{E}^{1/2}\Big|(\mathcal{B}^j f)'(\theta)(\xi)\Big|^2$$
$$= \sum_{j=1}^{k} \Big\langle \Sigma(\mathcal{B}^j f)'(\theta), (\mathcal{B}^j f)'(\theta)\Big\rangle^{1/2} \le \|\Sigma\|^{1/2} \sum_{j=1}^{k} \|(\mathcal{B}^j f)'(\theta)\|.$$

Using bound (3.10), we get

$$\begin{aligned}
&|\sigma_{f_k,\xi}(\theta) - \sigma_{f,\xi}(\theta)| \\
&\le 2^{\gamma}\|\Sigma\|^{1/2}(1\vee\|\theta\|)^{\gamma}\sum_{j=1}^{k}\|f^{(j+1)}\|_{L_{\infty,\gamma}}(1 + j^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})(\mathbb{E}^{1/2}\|\xi\|^2)^j \\
&\le 2^{\gamma}\|f\|_{C^{k+1,\gamma}}\|\Sigma\|^{1/2}(1\vee\|\theta\|)^{\gamma}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})\sum_{j=1}^{k}(\mathbb{E}^{1/2}\|\xi\|^2)^j \\
&\le 2^{\gamma+1}\|f\|_{C^{k+1,\gamma}}\|\Sigma\|^{1/2}(1\vee\|\theta\|)^{\gamma}(1 + k^{\gamma/2}\mathbb{E}^{1/2}\|\xi\|^{2\gamma})\mathbb{E}^{1/2}\|\xi\|^2 \\
&\lesssim_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1\vee\|\theta\|)^{\gamma}\|\Sigma\|^{1/2}\mathbb{E}^{1/2}\|\xi\|^2.
\end{aligned}$$

□

Bound (5.2) follows from representation (5.7), Theorem 3.1, bound (4.14) and bound (5.8).

We now prove bound (5.3). We can easily deduce from (4.14) that:

$$|S_{f_k}(\theta;\xi) - \mathbb{E}S_{f_k}(\theta;\xi)| \lesssim_{\gamma} A_1 t^{1/2} \vee A_2 t^{(1+\gamma)/2} \vee A_3 t^{(1+\rho)/2} \vee A_4 t^{(1+\rho+\gamma)/2} \vee A_5 t \vee A_6 t^{(2+\gamma)/2},$$

where

$$\begin{aligned}
A_1 &\asymp_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^{\gamma}(\mathbb{E}\|\xi\|)^{\rho}\|\Sigma\|^{1/2}, \\
A_2 &\asymp_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(\mathbb{E}\|\xi\|)^{\rho}\|\Sigma\|^{(1+\gamma)/2}, \\
A_3 &\asymp_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^{\gamma}\|\Sigma\|^{(1+\rho)/2}, \\
A_4 &\asymp_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}\|\Sigma\|^{(1+\rho+\gamma)/2}, \\
A_5 &\asymp_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^{\gamma}\|\Sigma\|, \\
A_6 &\asymp_{\gamma} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}\|\Sigma\|^{(2+\gamma)/2}.
\end{aligned}$$

Using Lemma 4.2, we conclude that, for any ψ satisfying the condition $\psi(t) \le c_1 e^{c_2 t^{2/(2+\gamma)}}, t \ge 0$, we have

$$\Big\|S_{f_k}(\theta;\xi) - \mathbb{E}S_{f_k}(\theta;\xi)\Big\|_{\psi} \lesssim_{\gamma,\psi} A_1 \vee \cdots \vee A_m.$$

Using the fact that $\|\Sigma\| \leq \mathbb{E}\|\xi\|^2 \leq 1$, it is easy to check that

$$A_1 \vee \cdots \vee A_m \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma(\mathbb{E}^{1/2}\|\xi\|^2)^\rho\|\Sigma\|^{1/2}.$$

Thus,

$$\Big\|S_{f_k}(\theta;\xi) - \mathbb{E}S_{f_k}(\theta;\xi)\Big\|_\psi \lesssim_{\gamma,\psi} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma(\mathbb{E}^{1/2}\|\xi\|^2)^\rho\|\Sigma\|^{1/2}. \tag{5.9}$$

Using bound (5.8), we get

$$\Big\|(\sigma_{f_k,\xi}(\theta) - \sigma_{f,\xi}(\theta))Z\Big\|_\psi \lesssim_\gamma (k+1)^{\gamma/2}\|f\|_{C^{k+1,\gamma}}(1\vee\|\theta\|)^\gamma\mathbb{E}^{1/2}\|\xi\|^2\|\Sigma\|^{1/2}\|Z\|_\psi,$$

which is dominated by the right hand side of (5.9). Thus, we can conclude that

$$\|R\|_\psi \lesssim_{\gamma,\psi} (k+1)^{\gamma/2}\|f\|_{C^{s,\gamma}}(1\vee\|\theta\|)^\gamma\Big((\mathbb{E}^{1/2}\|\xi\|^2)^\rho\|\Sigma\|^{1/2}\bigvee(\mathbb{E}\|\xi\|)^s\Big),$$

implying bound (5.3).

To prove normal approximation bound (2.7), we need the following elementary lemma.

Lemma 5.2. *For random variables η_1, η_2, denote*

$$\Delta(\eta_1,\eta_2) := \sup_{x\in\mathbb{R}}|\mathbb{P}\{\eta_1\leq x\} - \mathbb{P}\{\eta_2 \leq x\}|$$

and

$$\delta(\eta_1,\eta_2) := \inf_{\delta>0}\Big[\mathbb{P}\{|\eta_1-\eta_2|\geq\delta\} + \delta\Big].$$

Then, for an arbitrary random variable η and a standard normal random variable Z,

$$\Delta(\eta, Z) \leq \delta(\eta; Z).$$

We apply this lemma to random variable $\eta := \frac{f_k(X)-f(\theta)}{\sigma_{f,\xi}(\theta)}$. Using representation (5.1) and bound (5.2), we get that, for all $t\geq 1$ with probability at least $1-e^{-t}$

$$\left|\frac{f_k(X)-f(\theta)}{\sigma_{f,\xi}(\theta)} - Z\right| \lesssim_\gamma (k+1)^{\gamma/2}K_{s,\gamma}(f;\Sigma;\theta)(1\vee\|\Sigma\|^{1/2}\sqrt{t})^\gamma$$
$$\bigg((\mathbb{E}^{1/2}\|\xi\|^2)^\rho\sqrt{t}\bigvee\|\Sigma\|^{\rho/2}t^{(1+\rho)/2}\bigvee\|\Sigma\|^{1/2}t\bigvee\frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\bigg).$$

Let $t := \log\Bigl(\frac{1}{\|\Sigma\|}\Bigr)$. With this choice of t, it is easy to see that

$$\|\Sigma\|^{1/2}\sqrt{t} \lesssim 1 \text{ and } \|\Sigma\|^{1/2}t \lesssim \|\Sigma\|^{\rho/2}t^{(1+\rho)/2}.$$

Thus, with probability at least $1 - \|\Sigma\|$,

$$\Bigl|\frac{f_k(X) - f(\theta)}{\sigma_{f,\xi}(\theta)} - Z\Bigr| \lesssim_\gamma (k+1)^{\gamma/2} K_{s,\gamma}(f;\Sigma;\theta)$$
$$\biggl((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \sqrt{\log\Bigl(\frac{1}{\|\Sigma\|}\Bigr)} \bigvee \|\Sigma\|^{\rho/2}\log^{(1+\rho)/2}\Bigl(\frac{1}{\|\Sigma\|}\Bigr) \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\biggr).$$

It follows from Lemma 5.2 that

$$\Delta(\eta; Z) \le \delta(\eta; Z) \lesssim_\gamma (k+1)^{\gamma/2} K_{s,\gamma}(f;\Sigma;\theta)$$
$$\biggl((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \sqrt{\log\Bigl(\frac{1}{\|\Sigma\|}\Bigr)} \bigvee \|\Sigma\|^{\rho/2}\log^{(1+\rho)/2}\Bigl(\frac{1}{\|\Sigma\|}\Bigr) \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\biggr) + \|\Sigma\|.$$

Since also

$$\|\Sigma\| \le \|\Sigma\|^{\rho/2}\log^{(1+\rho)/2}\Bigl(\frac{1}{\|\Sigma\|}\Bigr),$$

we can conclude that

$$\Delta(\eta; Z) \lesssim_\gamma (k+1)^{\gamma/2} K_{s,\gamma}(f;\Sigma;\theta)$$
$$\biggl((\mathbb{E}^{1/2}\|\xi\|^2)^\rho \sqrt{\log\Bigl(\frac{1}{\|\Sigma\|}\Bigr)} \bigvee \|\Sigma\|^{\rho/2}\log^{(1+\rho)/2}\Bigl(\frac{1}{\|\Sigma\|}\Bigr) \bigvee \frac{(\mathbb{E}^{1/2}\|\xi\|^2)^s}{\|\Sigma\|^{1/2}}\biggr).$$

□

6. The proof of efficiency: a lower bound

Our goal in this section is to prove Theorem 2.4. It will be convenient for our purposes to represent the noise as a sum of a series with i.i.d. standard normal coefficients. To this end, we use the following well known result.

Theorem 6.1 ([25]). *Let $\xi \in E$, $\xi \sim \mathcal{N}(0;\Sigma)$. There exists a sequence $\{g_k\}_{k\in\mathbb{N}}$ of i.i.d. standard normal random variables and a sequence $\{x_k\}_{k\in\mathbb{N}}$ in E such that, for all $k \in \mathbb{N}$, $x_k \notin \overline{\mathrm{span}\{x_j : j \neq k\}}$, $\xi = \sum_{k=1}^\infty x_k g_k$ with the series in the right hand side converging in E a.s., and $\sum_{k=1}^\infty \|x_k\|^2 < \infty$.*

Clearly, $\overline{\mathrm{Im}(\Sigma)} = \overline{\mathrm{span}\{x_j : j \in \mathbb{N}\}}$. In the rest of this section, we provide the proof of Theorem 2.4.

PROOF. First, we will replace $\sigma^2_{f,\xi}(\theta)$ in the lower bound with $\sigma^2_{f,\xi}(\theta_0)$. To this end, we use the following simple lemma.

Lemma 6.1. *For all $\theta \in E$ such that*

$$\|\theta - \theta_0\| \le c\|\Sigma\|^{1/2} < 1,$$

the following bound holds:

$$\left|\frac{\sigma^2_{f,\xi}(\theta)}{\sigma^2_{f,\xi}(\theta_0)} - 1\right| \le 2^{s+2\gamma} K^2_{s,\gamma}(f;\Sigma;\theta_0) c^{s-1} \|\Sigma\|^{(s-1)/2}.$$

PROOF. Here and in what follows, denote $\rho := s - 1$. We have

$$\left|\frac{\sigma^2_{f,\xi}(\theta)}{\sigma^2_{f,\xi}(\theta_0)} - 1\right| = \frac{\Big|\langle \Sigma f'(\theta), f'(\theta)\rangle - \langle \Sigma f'(\theta_0), f'(\theta_0)\rangle\Big|}{\sigma^2_{f,\xi}(\theta_0)}$$
$$\le \frac{\|\Sigma\|\|f'(\theta) - f'(\theta_0)\|(\|f'(\theta)\| + \|f'(\theta_0)\|)}{\sigma^2_{f,\xi}(\theta_0)}$$
$$\le \frac{\|\Sigma\|\|f'\|_{\mathrm{Lip}_{\rho,\gamma}}(1 \vee \|\theta\| \vee \|\theta_0\|)^\gamma \|\theta - \theta_0\|^\rho \|f'\|_{L_{\infty,\gamma}}\Big((1 \vee \|\theta\|)^\gamma + (1 \vee \|\theta_0\|)^\gamma\Big)}{\sigma^2_{f,\xi}(\theta_0)}$$

We then use the condition $\|\theta - \theta_0\| \le 1$ to get

$$\|\theta\| \le \|\theta_0\| + \|\theta - \theta_0\| \le 2(1 \vee \|\theta_0\|).$$

Therefore,

$$\left|\frac{\sigma^2_{f,\xi}(\theta)}{\sigma^2_{f,\xi}(\theta_0)} - 1\right| \le \frac{2^{2\gamma+1}\|\Sigma\|\|f\|^2_{C^{s,\gamma}}(1 \vee \|\theta_0\|)^{2\gamma}\|\theta - \theta_0\|^\rho}{\sigma^2_{f,\xi}(\theta_0)}$$
$$\le 2^{2\gamma+1} K^2(f;\Sigma;\theta_0)\|\theta - \theta_0\|^\rho \le 2^{2\gamma+1+\rho} K^2(f;\Sigma;\theta_0) c^\rho \|\Sigma\|^{\rho/2},$$

concluding the proof. ∎

The bound of Lemma 6.1 implies that

$$\sup_{\|\theta-\theta_0\|\le c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta(T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta)} = \sup_{\|\theta-\theta_0\|\le c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta(T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta_0)} \frac{\sigma^2_{f,\xi}(\theta_0)}{\sigma^2_{f,\xi}(\theta)}$$
$$\ge \sup_{\|\theta-\theta_0\|\le c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta(T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta_0)} \frac{1}{1+\sup_{\|\theta-\theta_0\|\le c\|\Sigma\|^{1/2}}\left|\frac{\sigma^2_{f,\xi}(\theta)}{\sigma^2_{f,\xi}(\theta_0)}-1\right|}$$
$$\ge \sup_{\|\theta-\theta_0\|\le c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta(T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta_0)} \frac{1}{1+2^{s+2\gamma}K^2_{s,\gamma}(f;\Sigma;\theta_0)c^{s-1}\|\Sigma\|^{(s-1)/2}}. \tag{6.1}$$

The rest of the proof is based on a finite-dimensional approximation and an application of van Trees inequality. For a fixed $N\in\mathbb{N}$, let

$$L_N := \mathrm{span}\{x_1,\dots,x_N\}\subset E,$$

and

$$\xi_N := \sum_{k=1}^N x_k g_k \in L_N,\ \xi_N^\perp := \xi-\xi_N = \sum_{k>N} x_k g_k. \tag{6.2}$$

Clearly, random variables ξ_N and $\xi_N^\perp$ are independent.

We define a linear mapping $A_N:\mathbb{R}^N\mapsto L_N$ such that, for all $(\alpha_1,\dots,\alpha_N)\in\mathbb{R}^N$, $A_N(\alpha_1,\dots,\alpha_N):=\sum_{k=1}^N\alpha_k x_k$. Since $x_1,\dots,x_N$ are linearly independent vectors and L_N is an N-dimensional subspace of E, A_N is a bijection between the spaces $\mathbb{R}^N$ and L_N with inverse $A_N^{-1}: L_N\mapsto\mathbb{R}^N$. In what follows, $\mathbb{R}^N$ is viewed as a Euclidean space with canonical inner product. Denote by $L_N^*\supset E^*$ the dual space of L_N and let $A_N^*: L_N^*\mapsto\mathbb{R}^N$ be the adjoint operator of A_N. For $\alpha=(\alpha_1,\dots,\alpha_N)\in\mathbb{R}^N$ and $u\in L_N^*$, we have

$$\langle\alpha, A_N^*u\rangle = \langle A_N\alpha, u\rangle = \sum_{j=1}^N \alpha_j\langle x_j,u\rangle,$$

implying that $A_N^*u=\Big(\langle x_j,u\rangle : j=1,\dots,N\Big)$. With some abuse of notation, we denote by $\langle\cdot,\cdot\rangle$ both the inner product of $\mathbb{R}^N$ (and other inner product spaces) and the action of a linear functional on a vector in a Banach space.

Let $Z_N:=(g_1,\dots,g_N)\sim\mathcal{N}(0,I_N)$. Then $\xi_N=A_NZ_N$. Denote by Σ_N the covariance operator of $\xi_N:\Sigma_N u:=\mathbb{E}\langle\xi_N,u\rangle\xi_N, u\in L_N^*$. Then

$$\Sigma_N u = \sum_{j=1}^N\langle x_j,u\rangle x_j = A_NA_N^*u, u\in L_N^*,$$

implying that

$$\Sigma_N = A_N A_N^*. \tag{6.3}$$

It is easy to check that $\|\Sigma_N - \Sigma\| \to 0$ as $N \to \infty$, which follows from the bound

$$\|(\Sigma - \Sigma_N)u\| = \Big\|\sum_{j>N+1} \langle x_j, u\rangle x_j\Big\| \leq \sum_{j>N+1} \|x_j\|^2\|u\|, u \in E^*$$

and the condition $\sum_{j\in\mathbb{N}} \|x_j\|^2 < \infty$. It is also easy to see that, for all $u \in E^*$, $\langle \Sigma_N u, u\rangle$ monotonically converges to $\langle \Sigma u, u\rangle$ and that $\|\Sigma_N\| \leq \|\Sigma\|, N \geq 1$.

Since $\theta_0 \in \overline{\mathrm{span}\{x_j : j \in \mathbb{N}\}}$, there exists a sequence $\theta_{0,N} \in L_N$ such that $\theta_{0,N} \to \theta_0$ as $N \to \infty$. Therefore,

$$\sigma^2_{f,\xi_N}(\theta_{0,N}) = \langle \Sigma_N f'(\theta_{0,N}), f'(\theta_{0,N})\rangle \to \langle \Sigma f'(\theta_0), f'(\theta_0)\rangle = \sigma^2_{f,\xi}(\theta_0) \text{ as } N \to \infty.$$

By a simple continuity argument, it also follows that

$$K(f; \Sigma_N; \theta_{0,N}) \to K(f; \Sigma; \theta_0) \text{ as } N \to \infty.$$

Thus, for all large enough N,

$$\begin{aligned} &U(\theta_0; c; \Sigma) := \Big\{\theta \in E : \|\theta - \theta_0\| \leq c\|\Sigma\|^{1/2}\Big\} \\ &\supset \Big\{\theta \in L_N : \|\theta - \theta_{0,N}\| \leq \frac{c}{2}\|\Sigma_N\|^{1/2}\Big\} =: U_N(\theta_{0,N}; c/2; \Sigma_N). \end{aligned}$$

Using a simple conditioning argument and Jensen's inequality, this implies

$$\begin{aligned} &\sup_{\theta\in U(\theta_0;c;\Sigma)} \mathbb{E}_\theta(T(X) - f(\theta))^2 \geq \sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \mathbb{E}_\theta(T(X) - f(\theta))^2 \\ &\geq \sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \mathbb{E}\mathbb{E}\{(T(\theta + \xi_N + \xi_N^\perp) - f(\theta))^2|\xi_N\} \\ &\geq \sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \mathbb{E}(\mathbb{E}\{T(\theta + \xi_N + \xi_N^\perp)|\xi_N\} - f(\theta))^2 \\ &= \sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \mathbb{E}_\theta(\tilde{T}(X_N) - f(\theta))^2, \end{aligned}$$

where

$$X_N := \theta + \xi_N \in L_N \text{ and } \tilde{T}(x) := \mathbb{E}T(x + \xi_N^\perp), x \in E.$$

Next, we get

$$\sup_{\theta\in U(\theta_0;c;\Sigma)} \frac{\mathbb{E}_\theta (T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta_0)}$$
$$\geq \sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \frac{\mathbb{E}_\theta (\tilde T(X_N)-f(\theta))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})}\frac{\sigma^2_{f,\xi_N}(\theta_{0,N})}{\sigma^2_{f,\xi}(\theta_0)}. \tag{6.4}$$

To bound

$$\sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \frac{\mathbb{E}_\theta (\tilde T(X_N)-f(\theta))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})}$$

from below, we will use the following lemma whose proof is based on an application of van Trees inequality (see [12]).

Lemma 6.2. *Under the assumptions of Theorem 2.4, for some constant $D'_\gamma > 0$ and for all large enough N, the following bound holds for an arbitrary estimator $T(X_N)$:*

$$\sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)} \frac{\mathbb{E}_\theta (T(X_N)-f(\theta))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})}$$
$$\geq 1 - D'_\gamma K^2_{s,\gamma}(f;\Sigma_N;\theta_{0,N})\Big(c^{s-1}\|\Sigma_N\|^{(s-1)/2}+\frac{1}{c^2}\Big).$$

To complete the proof of Theorem 2.4, use bounds (6.1), (6.4) and the bound of Lemma 6.2 to get

$$\sup_{\|\theta-\theta_0\|\leq c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta (T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta)}$$
$$\geq \frac{1 - D'_\gamma K^2_{s,\gamma}(f;\Sigma_N;\theta_{0,N})\Big(c^{s-1}\|\Sigma_N\|^{(s-1)/2}+\frac{1}{c^2}\Big)}{1+2^{s+2\gamma}K^2_{s,\gamma}(f;\Sigma;\theta_0)c^{s-1}\|\Sigma\|^{(s-1)/2}}\frac{\sigma^2_{f,\xi_N}(\theta_{0,N})}{\sigma^2_{f,\xi}(\theta_0)}.$$

Passing to the limit as $N\to\infty$, we get

$$\sup_{\|\theta-\theta_0\|\leq c\|\Sigma\|^{1/2}} \frac{\mathbb{E}_\theta (T(X)-f(\theta))^2}{\sigma^2_{f,\xi}(\theta)} \geq \frac{1 - D'_\gamma K^2_{s,\gamma}(f;\Sigma;\theta_0)\Big(c^{s-1}\|\Sigma\|^{(s-1)/2}+\frac{1}{c^2}\Big)}{1+2^{s+2\gamma}K^2_{s,\gamma}(f;\Sigma;\theta_0)c^{s-1}\|\Sigma\|^{(s-1)/2}}$$
$$\geq 1-(D'_\gamma+2^{s+2\gamma})K^2_{s,\gamma}(f;\Sigma;\theta_0)\Big(c^{s-1}\|\Sigma\|^{(s-1)/2}+\frac{1}{c^2}\Big),$$

implying the bound of Theorem 2.4.

□

Finally, we prove Lemma 6.2.

PROOF. Let $c' := \frac{c}{K_{s,\gamma}(f;\Sigma_N;\theta_{0,N})}$. For $t \in [-c'/2, c'/2]$, $\theta_{0,N} \in L_N$ and $h \in L_N$, define

$$\theta_t := \theta_{0,N} + th, \quad X_N := \theta_t + \xi_N.$$

Consider a problem of estimation of a function

$$\varphi(t) := f(\theta_t), t \in [-c'/2, c'/2]$$

based on an observation $X_N \sim \mathcal{N}(\theta_t, \Sigma_N), t \in [-c'/2, c'/2]$. Since $A_N : \mathbb{R}^N \mapsto L_N$ is a bijection, an equivalent problem is to estimate $\varphi(t)$ based on an observation

$$A_N^{-1}X := A_N^{-1}\theta_t + Z_N \sim \mathcal{N}(A_N^{-1}\theta_t; I_N).$$

The Fisher information for the model $A_N^{-1}X \sim \mathcal{N}(A_N^{-1}\theta_t; I_N)$ with $t \in [-c'/2, c'/2]$ is equal to

$$I(t) = I = \langle A_N^{-1}h, A_N^{-1}h\rangle.$$

We will choose $h := \frac{\Sigma_N f'(\theta_{0,N})}{\sigma_{f,\xi_N}(\theta_{0,N})}$. For this choice of h,

$$\begin{aligned}
\frac{c'}{2}\|h\| &\le \frac{(c'/2)\|\Sigma_N\|\|f'(\theta_{0,N})\|}{\sigma_{f,\xi_N}(\theta_{0,N})} \\
&\le \frac{c'}{2}\frac{\|f\|_{C^{s,\gamma}}(1 \vee \|\theta_{0,N}\|)^\gamma \|\Sigma_N\|^{1/2}}{\sigma_{f,\xi_N}(\theta_{0,N})}\|\Sigma_N\|^{1/2} \\
&= \frac{c'}{2}K_{s,\gamma}(f;\Sigma_N;\theta_{0,N})\|\Sigma_N\|^{1/2} = \frac{c}{2}\|\Sigma_N\|^{1/2} < 1,
\end{aligned}$$

implying that, for all large enough N, $\theta_t \in U_N(\theta_{0,N}; c/2; \Sigma_N)$, $|t| \le c'/2$ and, as a consequence,

$$\begin{aligned}
&\sup_{\theta \in U_N(\theta_{0,N};c/2;\Sigma_N)} \frac{\mathbb{E}_\theta(T(X_N) - f(\theta))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})} \ge \sup_{t\in[-c'/2,c'/2]} \frac{\mathbb{E}_t(T(X_N) - \varphi(t))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})} \\
&= \sup_{t\in[-c'/2,c'/2]} \frac{\mathbb{E}_t(\hat{T}(A_N^{-1}X_N) - \varphi(t))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})}, \qquad (6.5)
\end{aligned}$$

where $\hat{T}(x) := T(A_N x), x \in \mathbb{R}^N$. We also have

$$\begin{aligned}
I &= \frac{\langle A_N^{-1}\Sigma_N f'(\theta_{0,N}), A_N^{-1}\Sigma_N f'(\theta_{0,N})\rangle}{\sigma^2_{f,\xi_N}(\theta_{0,N})} = \frac{\langle A_N^{-1}A_N A_N^* f'(\theta_{0,N}), A_N^{-1}A_N A_N^* f'(\theta_{0,N})\rangle}{\sigma^2_{f,\xi_N}(\theta_{0,N})} \\
&= \frac{\langle A_N^* f'(\theta_{0,N}), A_N^* f'(\theta_{0,N})\rangle}{\sigma^2_{f,\xi_N}(\theta_{0,N})} = \frac{\langle A_N A_N^* f'(\theta_{0,N}), f'(\theta_{0,N})\rangle}{\sigma^2_{f,\xi_N}(\theta_{0,N})} = \frac{\langle \Sigma_N f'(\theta_{0,N}), f'(\theta_{0,N})\rangle}{\sigma^2_{f,\xi_N}(\theta_{0,N})} = 1.
\end{aligned}$$

Let π be a prior density on $[-1,1]$ with $\pi(-1)=\pi(1)=0$ and such that

$$J_\pi := \int_{-1}^{1} \frac{(\pi'(s))^2}{\pi(s)} ds < \infty.$$

Denote $\pi_{c'}(t) := \frac{2}{c'}\pi\Big(\frac{2t}{c'}\Big), t \in [-c'/2, c'/2]$. Then $J_{\pi_{c'}} = \frac{4J_\pi}{(c')^2}$.

By van Trees inequality, for any estimator $\hat{T}(A_N^{-1}X_N)$ of $\varphi(t)$, it holds that

$$\sup_{t\in[-c'/2,c'/2]} \mathbb{E}_t(\hat{T}(A_N^{-1}X_N) - \varphi(t))^2 \geq \int_{-c'/2}^{c'/2} \mathbb{E}_t(\hat{T}(A_N^{-1}X_N) - \varphi(t))^2 \pi_{c'}(t)dt \geq$$
$$\geq \frac{\big(\int_{-c'/2}^{c'/2} \varphi'(t)\pi_{c'}(t)dt\big)^2}{\int_{-c'/2}^{c'/2} I(t)dt + 4J_\pi/(c')^2} \geq \frac{\big(\int_{-c'/2}^{c'/2} \varphi'(t)\pi_{c'}(t)dt\big)^2}{1 + 4J_\pi/(c')^2}. \tag{6.6}$$

It remains to bound from below $\big(\int_{-c'/2}^{c'/2} \varphi'(t)\pi_{c'}(t)dt\big)^2$. Note that $\varphi'(t) = \langle h, f'(\theta_t)\rangle$ and let

$$I_0 = \int_{-c'/2}^{c'/2} \langle h, f'(\theta_{0,N})\rangle \pi_{c'}(t)dt = \langle h, f'(\theta_{0,N})\rangle,$$
$$I_1 = \int_{-c'/2}^{c'/2} [\varphi'(t) - \varphi'(0)]\pi_{c'}(t)dt.$$

We have

$$\left(\int_{-c'/2}^{c'/2} \varphi'(t)\pi_{c'}(t)dt\right)^2$$
$$= (I_0+I_1)^2 \geq I_0^2 - 2|I_0||I_1| \geq \langle h, f'(\theta_{0,N})\rangle^2 - 2|\langle h, f'(\theta_{0,N})\rangle||I_1|.$$

With $h = \frac{\Sigma_N f'(\theta_{0,N})}{\sigma_{f,\xi_N}(\theta_{0,N})}$, we get

$$\langle h, f'(\theta_{0,N})\rangle^2 = \frac{\langle \Sigma_N f'(\theta_{0,N}), f'(\theta_{0,N})\rangle^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})} = \sigma^2_{f,\xi_N}(\theta_{0,N})$$

and

$$\left(\int_{-c'/2}^{c'/2} \varphi'(t)\pi_{c'}(t)dt\right)^2 \geq \sigma^2_{f,\xi_N}(\theta_{0,N}) - 2\sigma_{f,\xi_N}(\theta_{0,N})|I_1|. \tag{6.7}$$

Finally, we bound $|I_1|$ as follows. Note that

$$\begin{aligned}|\varphi'(t)-\varphi'(0)| &= |\langle h, f'(\theta_t)-f'(\theta_{0,N})\rangle| \\ &\le \|h\|\|f'\|_{\mathrm{Lip}_{\rho,\gamma}}(1\vee\|\theta_{0,N}\|\vee\|\theta_t\|)^\gamma (c'/2)^\rho\|h\|^\rho \\ &\le \|f\|_{C^{s,\gamma}}(1\vee\|\theta_{0,N}\|\vee(\|\theta_{0,N}\|+(c'/2)\|h\|))^\gamma (c'/2)^\rho\|h\|^{1+\rho} \\ &\le 2^{\gamma-\rho}\|f\|_{C^{s,\gamma}}(1\vee\|\theta_{0,N}\|)^\gamma (c')^\rho\|h\|^{1+\rho},\end{aligned}$$

where we used the fact that $(c'/2)\|h\|\le 1$. It follows that

$$\begin{aligned}|I_1| &\le 2^{\gamma-\rho}\|f\|_{C^{s,\gamma}}(1\vee\|\theta_{0,N}\|)^\gamma (c')^\rho\|h\|^{1+\rho} \\ &\le 2^{\gamma-\rho}\frac{\|f\|_{C^{s,\gamma}}(1\vee\|\theta_{0,N}\|)^\gamma}{\sigma^{1+\rho}_{f,\xi_N}(\theta_{0,N})}(c')^\rho\|\Sigma_N\|^{1+\rho}\|f'(\theta_{0,N})\|^{1+\rho} \\ &\le 2^{\gamma-\rho}\sigma_{f,\xi_N}(\theta_{0,N})\frac{\|f\|_{C^{s,\gamma}}(1\vee\|\theta_{0,N}\|)^\gamma}{\sigma^{2+\rho}_{f,\xi_N}(\theta_{0,N})}(c')^\rho\|\Sigma_N\|^{1+\rho}\|f'\|^{1+\rho}_{L_{\infty,\gamma}}(1\vee\|\theta_{0,N}\|)^{\gamma(1+\rho)} \\ &\le 2^{\gamma-\rho}\sigma_{f,\xi_N}(\theta_{0,N})\frac{\|f\|^{2+\rho}_{C^{s,\gamma}}(1\vee\|\theta_{0,N}\|)^{\gamma(2+\rho)}\|\Sigma_N\|^{(2+\rho)/2}}{\sigma^{2+\rho}_{f,\xi_N}(\theta_{0,N})}(c')^\rho\|\Sigma_N\|^{\rho/2} \\ &= 2^{\gamma-\rho}\sigma_{f,\xi_N}(\theta_{0,N})K^{2+\rho}_{s,\gamma}(f;\Sigma_N;\theta_{0,N})(c')^\rho\|\Sigma_N\|^{\rho/2}.\end{aligned}$$

We substitute this bound in (6.7) to get

$$\begin{aligned}&\left(\int_{-c'/2}^{c'/2}\varphi'(t)\pi_{c'}(t)dt\right)^2 \\ &\ge \sigma^2_{f,\xi_N}(\theta_{0,N})\Big(1-2^{\gamma+1-\rho}K^{2+\rho}_{s,\gamma}(f;\Sigma_N;\theta_{0,N})(c')^\rho\|\Sigma_N\|^{\rho/2}\Big). \qquad (6.8)\end{aligned}$$

Using bounds (6.5), (6.6) and (6.8), we conclude that

$$\begin{aligned}&\sup_{\theta\in U_N(\theta_{0,N};c/2;\Sigma_N)}\frac{\mathbb{E}_\theta(T(X_N)-f(\theta))^2}{\sigma^2_{f,\xi_N}(\theta_{0,N})}\ge\frac{1-2^{\gamma+1-\rho}K^{2+\rho}_{s,\gamma}(f;\Sigma_N;\theta_{0,N})(c')^\rho\|\Sigma_N\|^{\rho/2}}{1+4J_\pi/(c')^2} \\ &\ge 1-2^{\gamma+1-\rho}K^2_{s,\gamma}(f;\Sigma_N;\theta_{0,N})c^\rho\|\Sigma_N\|^{\rho/2}-\frac{4J_\pi K^2_{s,\gamma}(f;\Sigma_N;\theta_{0,N})}{c^2},\end{aligned}$$

implying the claim of the lemma. □

7. The proof of minimax lower bound

In this section, we use a modification of the approach developed by Nemirovski [31, 32] to prove minimax lower bounds implying the optimality of smoothness thresholds for efficient estimation. This will be done only in the case of classical Gaussian shift model (see Example 1.1)

$$X = \theta + \sigma Z,\ \theta \in \mathbb{R}^d,\ Z \sim \mathcal{N}(0, I_d)$$

with unknown mean θ and known noise level σ^2. The noise in this model is $\xi := \sigma Z$ with covariance $\Sigma = \sigma^2 I_d$, and the parameter space is the Euclidean space $\mathbb{R}^d$ with canonical inner product. Our main goal is to prove Theorem 2.2 stated in Section 2. Our approach is based on a construction of a set Θ of $2^{d/8}$ 2ε-separated points of the unit ball in $\mathbb{R}^d$ and a set of smooth functionals $f_l(\theta), l = 1, \dots, d$. Assuming the existence of estimators $T_l(X), l = 1, \dots, l$ of these functionals with mean squared error rate δ^2, we show that it is possible to estimate parameter $\theta \in \Theta$ with mean squared error $\lesssim \frac{\delta^2}{\varepsilon^{2(s-1)}}$. We compare this with well known minimax rates of estimation of $\theta \in \Theta$ to prove a lower bound on δ^2.

PROOF. Let h be the Hamming distance on the binary cube $\{-1, 1\}^d$:

$$h(\omega, \omega') := \sum_{j=1}^{d} I(\omega_j \neq \omega_j'), \omega, \omega' \in \{-1, 1\}^d.$$

It follows from Varshamov-Gilbert bound (see [35], Lemma 2.9) that there exists a subset $\Omega \subset \{-1, 1\}^d$ such that $\mathrm{card}(\Omega) \geq 2^{d/8}$ and $h(\omega, \omega') \geq d/8, \omega \neq \omega', \omega, \omega' \in \Omega$. For some $\varepsilon \in (0, 1/8)$, let

$$\theta_\omega := \frac{8\varepsilon}{\sqrt{d}}(\omega_1, \dots, \omega_d), \omega \in \Omega$$

and let $\Theta := \{\theta_\omega : \omega \in \Omega\}$. Note that $\|\theta_\omega\| = 8\varepsilon$ and

$$\|\theta_\omega - \theta_{\omega'}\| = 16\varepsilon \sqrt{\frac{h(\omega, \omega')}{d}}, \omega, \omega' \in \Omega, \tag{7.1}$$

which implies that, for all $\omega \neq \omega'$,

$$\|\theta_\omega - \theta_{\omega'}\| \geq \frac{8}{\sqrt{2}}\varepsilon \geq 2\varepsilon.$$

Let $\varphi : \mathbb{R} \mapsto [0, 1]$ be a C^∞ function with support in $[-1, 1]$, with $\|\tilde{\varphi}\|_{C^s} \leq 1$ for $\tilde{\varphi}(t) := \varphi(\|t\|^2), t \in \mathbb{R}^d$ and $\varphi(0) > 0$ being a constant. Define

$$f_l(\theta) := \sum_{\omega \in \Omega} \omega_l \varepsilon^s \tilde{\varphi}\left(\frac{\theta - \theta_\omega}{\varepsilon}\right), \theta \in \mathbb{R}^d, l = 1, \dots, d.$$

Note that the functions $\varepsilon^s\tilde{\varphi}\Big(\frac{\theta-\theta_\omega}{\varepsilon}\Big), \omega\in\Omega$ have disjoint supports (since the function $\varphi\Big(\frac{\theta-\theta_\omega}{\varepsilon}\Big)$ is supported in a ball of radius ε around θ_ω and points $\theta_\omega, \omega\in\Omega$ are 2ε-separated). It follows that $f_l(\theta_\omega)=\omega_l\varphi(0)\varepsilon^s, \omega\in\Omega, l=1,\dots,d$ and also that $\|f_l\|_{C^s}\leq 1$ (recall that $\|\tilde{\varphi}\|_{C^s}\leq 1$ and $\varepsilon\leq 1/8$).

Define

$$\tau(\theta,\theta') := \left(\frac{1}{d}\sum_{l=1}^d (f_l(\theta)-f_l(\theta'))^2\right)^{1/2}, \theta,\theta'\in\Theta.$$

We will need the following simple lemma:

Lemma 7.1.

$$\tau(\theta,\theta') = \frac{\varphi(0)\varepsilon^{s-1}}{8}\|\theta-\theta'\|, \theta,\theta'\in\Theta. \tag{7.2}$$

PROOF. Indeed, for all $\omega,\omega'\in\Omega$, we have by a straightforward computation that

$$\tau(\theta_\omega,\theta_{\omega'}) = 2\varphi(0)\varepsilon^s\sqrt{\frac{h(\omega,\omega')}{d}}$$

(this is based on the fact that $f_l(\theta_\omega)=\varphi(0)\varepsilon^s\omega_l$). Combining this with (7.1) yields

$$\tau(\theta_\omega,\theta_{\omega'}) = \frac{\varphi(0)\varepsilon^{s-1}}{8}\|\theta_\omega-\theta_{\omega'}\|, \omega,\omega'\in\Omega,$$

which implies the claim.

□

In addition, we will use the following well known fact:

Lemma 7.2. *If $\varepsilon^2\leq c'\sigma^2 d$ for a small enough numerical constant $c'>0$, then*

$$\inf_{\hat{\theta}}\max_{\theta\in\Theta}\mathbb{E}_\theta\|\hat{\theta}(X)-\theta\|^2\geq c''\sigma^2 d, \tag{7.3}$$

where the infimum is taken over all estimators $\hat{\theta}$ and c'' is a numerical constant.

The proof of this fact is quite standard (it could be based, for instance, on Theorem 2.5 in [35]). Note that the lower bound could be also written as $c''\varepsilon^2$ for some numerical constant c''.

Suppose now that, for some $\delta > 0$,

$$\sup_{\|f\|_{C^s}\leq 1} \inf_T \sup_{\|\theta\|\leq 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 < \delta^2. \tag{7.4}$$

This implies that

$$\max_{1\leq l\leq d} \inf_T \max_{\theta\in\Theta} \mathbb{E}_\theta (T(X) - f_l(\theta))^2 < \delta^2$$

and, moreover, for all $l = 1, \dots, d$ there exist estimators $T_l(X)$ such that

$$\max_{\theta\in\Theta} \mathbb{E}_\theta (T_l(X) - f_l(\theta))^2 < \delta^2.$$

It will be convenient to replace estimators $T_l(X)$ by estimators $\tilde{T}_l(X)$ defined as follows: $\tilde{T}_l(X) := \varepsilon^s\varphi(0)$ if $T_l(X) \geq 0$ and $\tilde{T}_l(X) := -\varepsilon^s\varphi(0)$ otherwise. For these modified estimators, it is easy to check that

$$\max_{\theta\in\Theta} \mathbb{E}_\theta (\tilde{T}_l(X) - f_l(\theta))^2 < 4\delta^2. \tag{7.5}$$

Define finally $\tilde{\omega} := (\tilde{\omega}_1, \dots, \tilde{\omega}_d)$, where $\tilde{\omega}_l = \tilde{\omega}_l(X) := \mathrm{sign}(\tilde{T}_l(X))$ and set $\tilde{\theta} = \tilde{\theta}(X) := \theta_{\tilde{\omega}}$. The following identity immediately follows from the definitions and from (7.2):

$$\|\tilde{\theta} - \theta\| = \frac{8}{\varphi(0)\varepsilon^{s-1}} \tau(\tilde{\theta}, \theta) = \frac{8}{\varphi(0)\varepsilon^{s-1}} \left(\frac{1}{d}\sum_{l=1}^d (\tilde{T}_l(X) - f_l(\theta))^2\right)^{1/2}, \theta\in\Theta.$$

Therefore, we can deduce from (7.5)

$$\mathbb{E}_\theta\|\tilde{\theta} - \theta\|^2 = \frac{8^2}{\varphi^2(0)\varepsilon^{2(s-1)}} \frac{1}{d}\sum_{l=1}^d \mathbb{E}_\theta (\tilde{T}_l(X) - f_l(\theta))^2 \leq \frac{4\ 8^2\delta^2}{\varphi^2(0)\varepsilon^{2(s-1)}}, \theta\in\Theta. \tag{7.6}$$

It remains to set $\varepsilon^2 := c'(\sigma^2 d \wedge 1)$ and to use minimax lower bound (7.3) to get

$$\max_{\theta\in\Theta} \mathbb{E}_\theta\|\tilde{\theta} - \theta\|^2 \geq c''\varepsilon^2.$$

Combining this with bound (7.6), we get

$$\frac{4\ 8^2\delta^2}{\varphi^2(0)\varepsilon^{2(s-1)}} \geq c''\varepsilon^2,$$

which implies that $\delta^2 \gtrsim \varepsilon^{2s}$. Therefore,

$$\sup_{\|f\|_{C^s}\le 1} \inf_T \sup_{\|\theta\|\le 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \gtrsim (\sigma^2 d)^s \wedge 1. \tag{7.7}$$

To complete the proof, it remains to show that, for some $c_2 > 0$, the following bound holds:

$$\sup_{\|f\|_{C^s}\le 1} \inf_T \sup_{\|\theta\|\le 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \ge c_2(\sigma^2 \wedge 1). \tag{7.8}$$

This easily follows from the bound of Theorem 2.4. To this end, take $f(\theta) := \langle \theta, u\rangle \varphi(\|\theta\|^2), \theta \in \mathbb{R}^d$, where $\|u\| = \kappa$ for a small enough constant $\kappa > 0$ and $\varphi : \mathbb{R} \mapsto [0,1]$ is a C^∞ function with $\varphi(t) = 1, t \in [0,1]$ and $\varphi(t) = 0, |t| > 2$. It is easy to see that u and φ could be chosen in such a way that $\|f\|_{C^s} \le 1$. For such a function f and for $\|\theta\| \le 1$, $\sigma^2_{f,\xi}(\theta) = \kappa^2\sigma^2$ and $K(f;\Sigma;\theta) \le \frac{1}{\kappa}$. Take also $\theta_0 = 0$. The bound of Theorem 2.4 now easily implies that, for small enough constants $c_3, c_4 > 0$ and for all $\sigma \le c_3$,

$$\inf_T \sup_{\|\theta\|\le 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \ge c_4 \sigma^2. \tag{7.9}$$

If $\sigma > c_3$, then $\sigma^2 d \gtrsim c_3^2$, and bound (7.7) implies that, for some $c_4' > 0$,

$$\inf_T \sup_{\|\theta\|\le 1} \mathbb{E}_\theta (T(X) - f(\theta))^2 \ge c_4'.$$

Together with (7.9), this implies (7.8).

□

Acknowledgement. The authors are very thankful to Martin Wahl for careful reading of the paper and pointing out a number of typos and to anonymous referees for a number of useful suggestions.

References

[1] A. Aleksandrov and V. Peller. Operator Lipschitz Functions. arXiv: 1611.01593.

[2] P. Bickel and Y. Ritov. Estimating integrated square density derivatives: sharp best order of convergence estimates. *Sankhya*, 1988, 50, 381–393.

[3] P.J. Bickel, C.A.J. Klaassen, Y. Ritov and J.A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, 1993.

[4] L. Birgé and P. Massart. Estimation of integral functionals of a density. *Annals of Statistics*, 1995, 23, 11-29.

[5] G. Blanchard, O. Bousquet, and L. Zwald. Statistical properties of kernel principal component analysis. *Machine Learning,* 2007, 66(2-3), 259-294.

[6] T.T. Cai and M. Low. On adaptive estimation of linear functionals. *Annals of Statistics*, 2005, 33, 2311–2343.

[7] T.T. Cai and M. Low. Non-quadratic estimators of a quadratic functional. *Annals of Statistics*, 2005, 33, 2930–2956.

[8] O. Collier, L. Comminges and A. Tsybakov. Minimax estimation of linear and quadratic functionals on sparsity classes. *Annals of Statistics*, 2017, 45, 3, 923–958.

[9] D. Donoho and R. Liu. On minimax estimation of linear functionals. Technical Report N 105. Department of Statistics, UC Berkeley, August 1987.

[10] D. Donoho and R. Liu. Geometrizing rates of convergence, II. *Annals of Statistics*, 1991, 19, 2, 633-667.

[11] D. Donoho and M. Nussbaum. Minimax quadratic estimation of a quadratic functional. *J. Complexity*, 1990, 6, 290–323.

[12] R.D. Gill and B.Y. Levit. Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli,* 1995, 1(1-2), 59–79.

[13] V.L. Girko. Introduction to general statistical analysis. *Theory Probab. Appl.*, 1987, 32, 2: 229–242.

[14] V.L. Girko. *Statistical analysis of observations of increasing dimension.* Springer, 1995.

[15] I. A. Ibragimov and R.Z. Khasminskii. *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York, 1981.

[16] I.A. Ibragimov, A.S. Nemirovski and R.Z. Khasminskii. Some problems of nonparametric estimation in Gaussian white noise. *Theory of Probab. and Appl.*, 1987, 31, 391–406.

[17] A. Javanmard and A. Montanari. Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory*, 2014, 60, 10, 6522–6554.

[18] J. Jiao, Y. Han and T. Weissman (2017) Bias correction with Jackknife, Bootstrap and Taylor Series. *arXiv:1709.06183.*

[19] J. Klemelä. Sharp adaptive estimation of quadratic functionals. *Probability Theory and Related Fields*, 2006, 134, 539–564.

[20] V. Koltchinskii. Asymptotically Efficient Estimation of Smooth Functionals of Covariance Operators. arxiv:1710.09072

[21] V. Koltchinskii and K. Lounici. Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincaré Probab. Statist.*, 2016, 52, 4, 1976–2013.

[22] V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 2017, 23, 1, 110–133.

[23] V. Koltchinskii, M. Löffler and R. Nickl. Efficient Estimation of Linear Functionals of Principal Components. *arXiv:1708.07642*.

[24] V. Koltchinskii and D. Xia. Perturbation of linear forms of singular vectors under Gaussian noise. In: *High Dimensional Probability VII: The Cargèse Volume*, Progress in Probability, vol 71, Birkhäuser, pp. 397–423.

[25] S. Kwapien and B. Szymanski. Some remarks on Gaussian measures on Banach spaces. *Probab. Math. Statist.*, 1980, 1, 1, 59–65.

[26] B. Laurent. Efficient estimation of integral functionals of a density. *Annals of Statistics*, 1996, 24, 659–681.

[27] M. Ledoux. *The Concentration of Measure Phenomenon.* American Mathematical Society. 2001.

[28] B. Levit. On the efficiency of a class of non-parametric estimates. *Theory of Prob. and applications*, 1975, 20(4), 723–740.

[29] B. Levit. Asymptotically efficient estimation of nonlinear functionals. *Probl. Peredachi Inf. (Problems of Information Transmission)*, 1978, 14(3), 65–72.

[30] O. Lepski, A. Nemirovski and V. Spokoiny. On estimation of the L_r norm of a regression function. *Probab. Theory Relat. Fields*, 1999, 113, 221–253.

[31] A. Nemirovski. On necessary conditions for the efficient estimation of functionals of a nonparametric signal which is observed in white noise. *Theory of Probab. and Appl.*, 1990, 35, 94–103.

[32] A. Nemirovski. *Topics in Non-parametric Statistics.* Ecole d'Ete de Probabilités de Saint-Flour. Lecture Notes in Mathematics, v. 1738, Springer, New York, 2000.

[33] V.V. Peller. Hankel operators in the perturbation theory of unitary and self-adjoint operators. *Funk. anal. i ego pril.*, 1985, 19(2), 37–51 (In Russian), English transl.: *Func. Anal. Appl.*, 1985, 19(2), 111–123.

[34] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis.* Springer Series in Statistics. Springer, 2005.

[35] A.B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, 2009.

[36] S. van de Geer, P. Bühlmann, Y. Ritov and R. Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models.

Annals of Statistics, 2014, 42(3), 1166–1202.

[37] C.-H. Zhang and S.S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2014, 76 217–242.

[38] F. Zhou and P. Li. A Fourier Analytical Approach to Estimation of Smooth Functions in Gaussian Shift Model. 2019, *arXiv:1911.02010*.