Exponential inequalities for dependent V-statistics via random Fourier features

Yandi Shen[*]   Fang Han[*]   Daniela Witten[*]

Abstract

We establish exponential inequalities for a class of V-statistics under strong mixing conditions. Our theory is developed via a novel kernel expansion based on random Fourier features and the use of a probabilistic method. This type of expansion is new and useful for handling many notorious classes of kernels.

1 Introduction

Consider the following V-statistic of order m generated by the symmetric kernel f,

$$V_n(f) := \sum_{i_1,\ldots,i_m=1}^{n} f(X_{i_1},\ldots,X_{i_m}), \tag{1.1}$$

where $\{X_i\}_{i=1}^n$ is a stationary sequence with marginal measure P on the d-dimensional real space $\mathbb{R}^d$. The purpose of this paper is to establish exponential-type tail bounds for (1.1) when $\{X_i\}_{i=1}^n$ are weakly dependent.

In (1.1), if the summation is taken over m-tuples $(i_1,\ldots,i_m)$ of distinct indices, the resulting is a U-statistic. In many applications, the techniques of analyzing U- and V-statistics are the same. Non-asymptotic tail bounds and limiting theorems of V- and U-statistics in the i.i.d. case have also been extensively studied [17, 2, 15, 1].

The analysis of V- and U-statistics when the observed data are no longer independent has attracted increasing attention in statistics and probability, with most of the efforts put on deriving limit theorems and bootstrap consistency. See, for instance, [25], [10], [11], [8], [12], [18], [7], [9], [4], [19], [20], [26], [3], among many others. However, there are few results on non-asymptotic concentration bounds for V- and U-statistics.

[*]University of Washington, United States of America. E-mail: ydshen@uw.edu, fanghan@uw.edu, dwitten@uw.edu

Exceptions include [5] and [16], who proved Hoeffding-type inequalities for U- and V-statistics under ϕ-mixing conditions. There, the results either rely on assumptions difficult to verify, or are limited to nondegenerate ones.

In this paper, we show that for a strongly mixing stationary sequence, exponential inequalities hold for a large class of V- and U-statistics. The main theorem is presented in Section 2. We then illustrate the usefulness of our theory with examples and some further extensions in Section 3. Detailed proof of the main theorem is given in Section 4, with the rest of proofs given in Section 5.

Notation used in the rest of the paper is as follows. $L_1(\mathbb{R}^d)$ denotes the class of integrable functions in $\mathbb{R}^d$, and for each $p \geq 1$, $\|f\|_{L_p} := \left\{ \int_{\mathbb{R}^d} |f(x)|^p dx \right\}^{1/p}$. For a real vector $u \in \mathbb{R}^d$, $\|u\|$ denotes its Euclidean norm. For two real numbers a, b, $a \vee b := \max\{a, b\}$.

2 Main results

For two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, the strong mixing coefficient is defined as

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

A stationary sequence $\{X_i\}_{i \in \mathbb{Z}}$ is called *strong mixing* (hereafter also called α-*mixing*) if

$$\alpha(i) := \alpha(\mathcal{M}_0, \mathcal{G}_i) \to 0 \text{ as } i \to \infty,$$

where $\mathcal{M}_0 := \sigma(X_j, j \leq 0)$ and $\mathcal{G}_i := \sigma(X_j, j \geq i)$ for $i \geq 1$ are the σ-algebras generated by $\{X_j, j \leq 0\}$ and $\{X_j, j \geq i\}$ respectively.

We now introduce concepts in V-statistics. Let $\{\widetilde{X}_i\}_{i=1}^n$ be an i.i.d. sequence with $\widetilde{X}_1$ identically distributed as X_1. The mean value of a symmetric kernel f is defined as

$$\theta := \theta(f) := \mathbb{E}f\left(\widetilde{X}_1, \ldots, \widetilde{X}_m\right).$$

The kernel f is called *centered* if $\theta(f) = 0$, and *degenerate of level* $r - 1$ ($2 \leq r \leq m$) if

$$\mathbb{E}f\left(x_1, \ldots, x_{r-1}, \widetilde{X}_r, \ldots, \widetilde{X}_m\right) = \theta$$

for any $(x_1^\top, \ldots, x_{r-1}^\top)^\top \in \mathrm{supp}(P^{r-1})$, the support of the product measure P^{r-1}. The kernel f is called *fully degenerate* if it is degenerate of level $m - 1$.

When f is degenerate of level $r - 1$, its Hoeffding decomposition takes the form

$$f(x_1, \ldots, x_m) - \theta = \sum_{1 \leq i_1 < \ldots < i_r \leq m} f_r(x_{i_1}, \ldots, x_{i_r}) + \ldots + f_m(x_1, \ldots, x_m),$$

where $\{f_p\}_{p=r}^m$ are recursively defined as

$$f_1(x) := g_1(x),$$

$$f_p(x_1, \ldots, x_p) := g_p(x_1, \ldots, x_p) - \sum_{k=1}^p f_1(x_k) - \ldots - \sum_{1 \leq k_1 < \ldots < k_{p-1} \leq p} f_{p-1}\left(x_{k_1}, \ldots, x_{k_{p-1}}\right),$$

(2.1)

for $p = 2, \cdots, m$, with $\{g_p\}_{p=1}^m$ defined as $g_m := f - \theta$, and

$$g_p(x_1, \ldots, x_p) := \mathbb{E}f\left(x_1, \ldots, x_p, \widetilde{X}_{p+1}, \ldots, \widetilde{X}_m\right) - \theta$$

for $1 \leq p \leq m-1$. For each $1 \leq p \leq m$ and $1 \leq k \leq n$, we denote the V-statistic generated by f_p and data $\{X_i\}_{i=1}^k$ by

$$V_k(f_p) := \sum_{i_1,\ldots,i_p=1}^{k} f_p(X_{i_1},\ldots,X_{i_p}).$$

For a real function $g \in L_1(\mathbb{R}^d)$, its Fourier transform is defined as

$$\widehat{g}(u) := \int_{\mathbb{R}^d} g(x) e^{-2\pi i u^\top x} dx,$$

where $dx := dx_1 \ldots dx_d$.

Theorem 2.1. *Suppose $\{X_i\}_{i=1}^n$ in (1.1) is part of a stationary sequence $\{X_i\}_{i\in\mathbb{Z}}$ that is geometrically α-mixing with coefficient*

$$\alpha(i) \leq \gamma_1 \exp(-\gamma_2 i) \quad \text{for all } i \geq 1, \tag{2.2}$$

where γ_1, γ_2 are two positive absolute constants. Suppose $f \in L_1(\mathbb{R}^{md})$ is continuous, and its Fourier transform $\widehat{f}$ satisfies

$$\int_{\mathbb{R}^{md}} \left|\widehat{f}(u)\right| \|u\|^q du < \infty \tag{2.3}$$

for some $q \geq 1$. Then, there exists a positive constant $C = C(m,\gamma_1,\gamma_2)$ such that for each $1 \leq p \leq m$, and any $x > 0$,

$$\mathbb{P}\left\{ n^{-p} \max_{1\leq k\leq n} |V_k(f_p)| \geq x \right\} \leq 6 \exp\left(-\frac{Cnx^{2/p}}{A_{p,n}^{1/p} + x^{1/p} M_{p,n}^{1/p}} \right) \tag{2.4}$$

with

$$A_{p,n} = 2^{2m} \left\|\widehat{f}\right\|_{L_1}^2 \left\{ \frac{64\gamma_1^{1/3}}{1-\exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p \text{ and } M_{p,n} = 2^m \left\|\widehat{f}\right\|_{L_1} (\log n)^{2p}. \tag{2.5}$$

We remark that a maximal-type tail estimate for $V_n(f)$ in (1.1) can be obtained in a straightforward manner by assembling the tail estimate in (2.4) for each $r \leq p \leq m$, where r is the degenerate level of f. Indeed, for any $1 \leq k \leq n$ and $1 \leq p \leq m$, we have by the symmetry of f

$$\sum_{1\leq i_1,\ldots,i_m\leq k} \sum_{1\leq j_1<\ldots<j_p\leq m} f_p(X_{i_{j_1}},\ldots,X_{i_{j_p}}) = \binom{m}{p} \sum_{1\leq i_1,\ldots,i_m\leq k} f_p(X_{i_1},\ldots,X_{i_p})$$

$$= \binom{m}{p} k^{m-p} \sum_{1\leq i_1,\ldots,i_p\leq k} f_p(X_{i_1},\ldots,X_{i_p}).$$

This entails that

$$n^{-m} \max_{1\leq k\leq n} \left| \sum_{1\leq i_1,\ldots,i_m\leq k} \left(f(X_{i_1},\ldots,X_{i_m}) - \theta \right) \right|$$

$$= n^{-m} \max_{1\leq k\leq n} \left| \sum_{1\leq i_1,\ldots,i_m\leq k} \sum_{p=r}^{m} \sum_{1\leq j_1<\ldots<j_p\leq m} f_p(X_{i_{j_1}},\ldots,X_{i_{j_p}}) \right|$$

$$\leq n^{-m} \max_{1\leq k\leq n} \sum_{r\leq p\leq m} \binom{m}{p} k^{m-p} \left| \sum_{1\leq i_1,\ldots,i_p\leq k} f_p(X_{i_1},\ldots,X_{i_p}) \right|$$

$$\leq \sum_{r\leq p\leq m} \binom{m}{p} n^{-p} \max_{1\leq k\leq n} |V_k(f_p)|.$$

Therefore, by adjusting the constant C in (2.4), we obtain that

$$\mathbb{P}\left(n^{-m}\max_{1\le k\le n}\left|\sum_{1\le i_1,\ldots,i_m\le k}\left(f(X_{i_1},\ldots,X_{i_m})-\theta\right)\right|\ge x\right)\le 6\sum_{p=r}^{m}\exp\left(-\frac{Cnx^{2/p}}{A_{p,n}^{1/p}+x^{1/p}M_{p,n}^{1/p}}\right).$$

We now provide a proof sketch of Theorem 2.1 with a focus on the technical novelties. One key step in our proof is to find a uniform approximation $\widetilde{f}=\widetilde{f}(;t,M)$ of the original kernel f under any prescribed accuracy t such that (i) $\left|\widetilde{f}-f\right|\le t$ uniformly over a large enough compact set $[-M,M]^{md}$, and (ii) $\widetilde{f}$ admits the following tensor expansion

$$\widetilde{f}(x_1,\ldots,x_m)=\sum_{j_1,\ldots,j_m=1}^{K}f_{j_1,\ldots,j_m}e_{j_1}(x_1)\ldots e_{j_m}(x_m). \tag{2.6}$$

Here, K is a positive integer that depends on both the approximation error t and the range of approximation M, $\{f_{j_1,\ldots,j_m}\}_{j_1,\ldots,j_m=1}^{K}$ is a real sequence, and $\{e_j(\cdot)\}_{j=1}^{K}$ is a set of uniformly bounded real bases. Once such an $\widetilde{f}$ is found, a truncation argument will yield the proximity between $\{f_p\}_{p=1}^{m}$ and $\{\widetilde{f}_p(;t,M)\}_{p=1}^{m}$, the latter being the degenerate components of $\widetilde{f}(;t,M)$ in its Hoeffding decomposition. Then, using each $\widetilde{f}_p(;t,M)$ as a proxy, standard moment estimates with the aid of exponential inequalities for partial sum processes (cf. Corollary 24 in [21]) will render a tail bound for each $\max_{1\le k\le n}|V_k(f_p)|$.

The problem then boils down to finding such an $\widetilde{f}$ with the tensor structure (2.6). One main difficulty in this step is to construct expansion bases $\{e_j(\cdot)\}_{j=1}^{K}$ that are uniformly bounded. Many classical approaches in multivariate function approximation are unable to provide a satisfactory answer to this problem. For example, uniform polynomial approximation by the Stone-Weierstrass theorem will have very poor performance, since high orders of the polynomials lead to a large upper bound of the bases. The use of Lipschitz-continuous scale and wavelet functions, as exploited in [19], is also inappropriate for the same reason.

Our solution is based on a probabilistic method, and especially, by realizing that the tensor decomposition (2.6) is intrinsically connected to the idea of randomized feature mapping [22] in the kernel learning literature. More specifically, when $f\in L_1(\mathbb{R}^{md})$ is continuous and $\widehat{f}\in L_1(\mathbb{R}^{md})$, the Fourier inversion formula implies that

$$f(x_1,\ldots,x_m)=\int_{\mathbb{R}^{md}}\widehat{f}(u_1,\ldots,u_m)e^{2\pi i(u_1^\top x_1+\ldots+u_m^\top x_m)}du_1\ldots du_m,$$

where the right-hand side can be seen as the expectation of a Fourier basis with random frequency, which follows the sign measure of $\widehat{f}$. Due to the boundedness of the Fourier bases, Hoeffding's inequality guarantees an exponentially fast rate for a sample mean statistic of Fourier bases

$$s_K(x_1,\ldots,x_m):=\frac{1}{K}\sum_{j=1}^{K}\exp\left\{2\pi i(u_{j,1}^\top x_1+\ldots+u_{j,m}^\top x_m)\right\}$$

to approximate f at each fixed point $x\in\mathbb{R}^{md}$. The elements $\exp\{2\pi i(u_{j,1}^\top x_1+\ldots+u_{j,m}^\top x_m)\}$ in $s_K(x_1,\ldots,x_m)$ naturally decompose to bounded basis functions of inputs x_j. An entropy-type argument is then used so that we could prove the existence of a satisfactory set of bases such that the approximation holds uniformly over any compact set $[-M,M]^{md}$. The detailed proof will be given in Section 4.

3 Examples and extensions

Motivated by their wide applications in statistics and machine learning, we will put special focus on shift-invariant symmetric kernels in the case $m = 2$ with $f(x, y) = f_0(x - y)$ for some $f_0 : \mathbb{R}^d \to \mathbb{R}$. We start with a corollary of Theorem 2.1 for such kernels.

Corollary 3.1. *Let $\{X_i\}_{i=1}^n$ be as in Theorem 2.1. Let $m = 2$ and the kernel f be shift-invariant with $f(x, y) = f_0(x - y)$ for some $f_0 : \mathbb{R}^d \to \mathbb{R}$. Suppose that $f_0 \in L_1(\mathbb{R}^d)$ is continuous, and its Fourier transform $\widehat{f_0}$ satisfies*

$$\int_{\mathbb{R}^d} \left| \widehat{f_0}(u) \right| \|u\|^q du < \infty \tag{3.1}$$

for some $q \geq 1$. Then, for $p = 1, 2$, the same tail bound in (2.4) holds with

$$A_{p,n} = 16 \left\| \widehat{f_0} \right\|_{L_1}^2 \left\{ \frac{64 \gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p \ \text{ and } \ M_{p,n} = 4 \left\| \widehat{f_0} \right\|_{L_1} (\log n)^{2p}.$$

In view of Bochner's theorem (cf. Section 1.4.3, [23]), Corollary 3.1 can be further simplified when the kernel is positive definite. Recall that a real function $g_0 : \mathbb{R}^d \to \mathbb{R}$ is said to be *positive definite (PD)* if for any positive integer n and real vectors $\{x_i\}_{i=1}^n \in \mathbb{R}^d$, the matrix $A = (a_{i,j})_{i,j=1}^n$ with $a_{i,j} = g_0(x_i - x_j)$ is positive semi-definite (PSD).

Corollary 3.2. *Let $\{X_i\}_{i=1}^n$ be as in Theorem 2.1. Let $m = 2$ and the kernel f be shift-invariant with $f(x, y) = f_0(x - y)$ for some $f_0 : \mathbb{R}^d \to \mathbb{R}$. Suppose f_0 satisfies the conditions in Corollary 3.1 and is also PD. Then, for $p = 1, 2$, the same tail bound in (2.4) holds with*

$$A_{p,n} = 4 f_0(0)^2 \left\{ \frac{64 \gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p \ \text{ and } \ M_{p,n} = 2 f_0(0)(\log n)^{2p}.$$

Moreover, the same bound holds with the above A_p and M_p if f_0 only satisfies (3.1) for some $0 < q < 1$, but is both PD and Lipschitz continuous.

We now list several commonly-used kernels covered by Theorem 2.1 and the previous two corollaries.

1. The d-dimensional Gaussian kernel $f(x, y) = f_0(x - y) = \exp\left(-\|x - y\|^2/2\right)$ is shift-invariant with f_0 being both Schwartz and PD, and f_0 satisfies (3.1) for arbitrary $q \geq 1$. Thus, f satisfies the conditions of Corollary 3.2.

2. For the d-dimensional Cauchy kernel $f(x, y) = f_0(x - y)$ with $f_0(x) = \prod_{\ell=1}^d 2/\left(1 + x_\ell^2\right)$, f_0 is PD and its Fourier transform $\widehat{f_0}(u) = \exp\left(-\|u\|_1\right)$ satisfies (3.1) for arbitrary $q \geq 1$. Therefore, it satisfies the conditions of Corollary 3.2.

3. The d-dimensional Laplacian kernel $f(x, y) = f_0(x - y) = \exp(-\|x - y\|_1)$ is shift-invariant and PD. The Fourier transform of f_0 is $\widehat{f_0}(u) = \prod_{\ell=1}^d \left\{2/(1 + u_\ell^2)\right\}$, which has fractional moments and thus satisfies (3.1) for any $0 < q < 1$. Since f_0 is both PD and Lipschitz, it satisfies the conditions in Corollary 3.2.

4. The 1-dimensional "hat" kernel: $f(x, y) = f_0(x - y)$ with $f_0(x)$ equal to $x + 1$ for $-1 \leq x \leq 0$, $1 - x$ for $0 \leq x \leq 1$ and 0 otherwise. f_0 is PD and 1-Lipschitz. Its Fourier transform is $\widehat{f_0}(u) = \{1 - \cos(2\pi u)\}/(2\pi^2 u^2)$ and thus has fractional moment. Therefore, f_0 satisfies (3.1) for any $0 < q < 1$, and hence is also covered by Corollary 3.2.

We then discuss extensions to Theorem 2.1. The smoothness assumption (2.3) in Theorem 2.1 could be further relaxed by employing the standard smoothing technique

through mollifiers. More precisely, we resort to an intermediate kernel f_h between f and $\widetilde{f}$. It is constructed by convolving f with the Gaussian mollifier with scale parameter h. The parameter h controls the trade-off between approximation error and smoothness: small h leads to finer approximation of f by f_h, but makes f_h less smooth and thus renders a larger constant in the tail bound. Theorem 2.1 is then applied on this intermediate kernel f_h to obtain the tail bound.

As a particular example, the following corollary deals with Lipschitz kernels considered in [19]. Introduce the following constant from integration with polar coordinates (with convention $(-1)!! = 0!! = 1$):

$$\Gamma(n) := \begin{cases} ((n-2)!!)^{-1}(2\pi)^{\frac{n}{2}} & n \text{ is even} \\ ((n-2)!!)^{-1}2(2\pi)^{\frac{n-1}{2}} & n \text{ is odd.} \end{cases} \tag{3.2}$$

Corollary 3.3. Let $\{X_i\}_{i=1}^n$ be as in Theorem 2.1. Suppose the kernel $f \in L_1(\mathbb{R}^{md})$ is bounded, uniformly continuous, and its Fourier transform satisfies

$$\left|\widehat{f}(u)\right| \leq \frac{L}{1 + \|u\|^{md+\varepsilon}} \tag{3.3}$$

for some $\varepsilon > 0$ and positive constant L. Then, for $1 \leq p \leq m$, the bound in (2.4) holds with

$$A_{p,n} = (1 + \varepsilon^{-1})^2 2^{2m} c_1^2 L^2 \left\{ \frac{64\gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p,$$
$$M_{p,n} = (1 + \varepsilon^{-1}) 2^m c_1 L (\log n)^{2p},$$

where $c_1 = \Gamma(md)$.

The tail condition in (3.3) is in general milder than (2.3) in Theorem 2.1, and naturally arises in Fourier analysis (cf. Chapter 8.4 in [14]). The following is the version of Corollary 3.3 for shift-invariant kernels.

Corollary 3.4. Suppose $\{X_i\}_{i=1}^n$ are as in Theorem 2.1. Let $m = 2$ and the kernel f be shift-invariant with $f(x,y) = f_0(x - y)$ for some $f_0 : \mathbb{R}^d \to \mathbb{R}$. Suppose that f_0 satisfies condition (3.3) (with $m = 1$) for some $\varepsilon > 0$ and positive constant L. Then, for $p = 1, 2$, the bound in (2.4) holds with

$$A_{p,n} = 16(1 + \varepsilon^{-1})^2 c_1^2 L^2 \left\{ \frac{64\gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p \text{ and } M_{p,n} = 4(1 + \varepsilon^{-1}) c_1 L (\log n)^{2p},$$

where $c_1 = \Gamma(d)$.

Corollaries 3.3 and 3.4 cover the cosine kernel, defined as $f(x,y) := f_0(x-y) := \prod_{\ell=1}^d \cos(x_\ell - y_\ell) \mathbb{1}(|x_\ell - y_\ell| \leq \pi/2)$. Consider the simple 1-dimensional case. Here, even though the trigonometric identity $\cos(x - y) = \cos(x)\cos(y) + \sin(x)\sin(y)$ gives a direct expansion of $\cos(x-y)$, there is no trivial expansion of the indicator $\mathbb{1}(|x - y| \leq \pi/2)$. However, letting $f_0(x) = \cos(x)\mathbb{1}(|x| \leq \pi/2)$, it is immediate that f_0 is 1-Lipschitz and thus uniformly continuous. Moreover, its Fourier transform is $\left|\widehat{f_0}(u)\right| = \left|2\cos(\pi^2 u)/(1 - 4\pi^2 u^2)\right|$, and hence f_0 satisfies (3.3) with $\varepsilon = 1$ and $L = 2$.

4 Proof of Theorem 2.1

We will use the following extra notation. For any real-valued function f on $\mathbb{R}^d$, $\nabla_x f$ is the gradient of f. For a set A, $|A|$ indicates its cardinality. For a subset $\mathcal{M}$ in $\mathbb{R}^d$, we will use $\operatorname{diam}(\mathcal{M})$ to denote its diameter, i.e., $\operatorname{diam}(\mathcal{M}) := \sup_{x,y \in \mathcal{M}} \|x - y\|$. For a function

f, we write $f(;\theta)$ to emphasize its dependence on some parameter θ. For a measurable set A, we will use $\mathbb{1}\{A\}$ to denote the indicator variable on the set A. For any positive integer N, we will use $[N]$ to denote the set $\{1,\ldots,N\}$.

As described in the proof sketch after Theorem 2.1, we split the main part of the proof into the following lemmas. The first lemma finds a symmetric kernel $\widetilde{f}$ with tensor decomposition (2.6) that approximates f uniformly over some prescribed set $[-M,M]^{md}$ and accuracy t.

Lemma 4.1. *Suppose the kernel $f \in L_1(\mathbb{R}^{md})$ is continuous and satisfies condition (2.3) for some $q \geq 1$. Then, for any $M > 0$ and $t > 0$, there exists a symmetric function $\widetilde{f} = \widetilde{f}(;t,M)$ such that $\left| f(x_1,\ldots,x_d) - \widetilde{f}(x_1,\ldots,x_d) \right| \leq t$ uniformly over all $(x_1^\top,\ldots,x_d^\top)^\top \in [-M,M]^{md}$, and $\widetilde{f}$ satisfies (2.6) for some positive integer $K = K(t,M)$, $\{f_{j_1,\ldots,j_m}\}_{j_1,\ldots,j_m=1}^K$, and $\{e_j(\cdot)\}_{j=1}^K$ such that*

$$\sum_{j_1,\ldots,j_m=1}^K |f_{j_1,\ldots,j_m}| \leq F \quad and \quad \sup_{1\leq j\leq K} \sup_{x\in\mathbb{R}^d} |e_j(x)| \leq B \tag{4.1}$$

for some constants F,B that do not depend on M and t. In particular, one can take $F = 2^m \left\|\widehat{f}\right\|_{L_1}$ and $B = 1$.

Proof. This proof adapts from that of Claim 1 in [22]. Throughout the proof, $x_1,\ldots,x_m$ and $u_1,\ldots,u_m$ are real vectors in $\mathbb{R}^d$, $dx = dx_1\ldots dx_d$, and x,u will be real vectors in $\mathbb{R}^{md}$. Let $\widehat{f}: \mathbb{R}^{md} \to \mathbb{C}$ be the Fourier transform of f, that is,

$$\widehat{f}(u_1,\ldots,u_m) = \int_{\mathbb{R}^{md}} f(x_1,\ldots,x_m) e^{-2\pi i(u_1^\top x_1 + \ldots + u_m^\top x_m)} dx_1\ldots dx_m.$$

Clearly, Condition (2.3) with some $q \geq 1$ implies that $\widehat{f} \in L_1(\mathbb{R}^{md})$. Since f is continuous, by the Fourier inversion formula (see, for example, Chapter 6 of [24]), we have

$$f(x_1,\ldots,x_m) = \int_{\mathbb{R}^d} \widehat{f}(u_1,\ldots,u_m) e^{2\pi i(u_1^\top x_1 + \ldots + u_m^\top x_m)} du_1\ldots du_m.$$

Note that without the continuity of f, the above equation only holds almost surely with respect to the Lebesgue measure. Let $\widehat{f} = \widehat{g} + i\widehat{h}$ for real-valued functions $\widehat{g}, \widehat{h}$, then since f is real-valued, we have $f = I - II$, where

$$I := \int_{\mathbb{R}^{md}} \widehat{g}(u_1,\ldots,u_m) \cos\left\{ 2\pi\left(u_1^\top x_1 + \ldots + u_m^\top x_m \right) \right\} du_1\ldots du_m,$$

$$II := \int_{\mathbb{R}^{md}} \widehat{h}(u_1,\ldots,u_m) \sin\left\{ 2\pi\left(u_1^\top x_1 + \ldots + u_m^\top x_m \right) \right\} du_1\ldots du_m.$$

We now approximate I and II separately. I can be further written as $I = I_+ - I_-$, where

$$I_+ := \int_{[\widehat{g}>0]} \widehat{g}(u_1,\ldots,u_m) \cos\left\{ 2\pi\left(u_1^\top x_1 + \ldots + u_m^\top x_m \right) \right\} du_1\ldots du_m,$$

$$I_- := \int_{[\widehat{g}<0]} -\widehat{g}(u_1,\ldots,u_m) \cos\left\{ 2\pi\left(u_1^\top x_1 + \ldots + u_m^\top x_m \right) \right\} du_1\ldots du_m.$$

Let $A_g^+ := \int_{[\widehat{g}>0]} \widehat{g}(u)du$ and $A_g^- := \int_{[\widehat{g}<0]} (-\widehat{g}(u))du$, and note that A_g^+ and A_g^- are both nonnegative and satisfy $A_g^+ + A_g^- = \|\widehat{g}\|_{L_1} < \infty$ and $A_g^+ - A_g^- = f(0)$, where we use the fact that $\widehat{g} \in L_1(\mathbb{R}^{md})$ since $\widehat{f} \in L_1(\mathbb{R}^{md})$. Then, we have

$$I = A_g^+ \cdot \mathbb{E}_u\left[\cos\left\{ 2\pi\left(u_1^\top x_1 + \ldots u_m^\top x_m \right) \right\}\right] - A_g^- \cdot \mathbb{E}_v\left[\cos\left\{ 2\pi\left(v_1^\top x_1 + \ldots v_m^\top x_m \right) \right\}\right]$$

$$=: A_g^+ \cdot k_g^+(x_1,\ldots,x_m) - A_g^- \cdot k_g^-(x_1,\ldots,x_m),$$

where $(u_1^\top, \ldots, u_m^\top)^\top$ follows the distribution $\widehat{g}\mathbb{1}\{\widehat{g} > 0\}/A_g^+$, and $(v_1^\top, \ldots, v_m^\top)^\top$ follows the distribution $-\widehat{g}\mathbb{1}\{\widehat{g} < 0\}/A_g^-$. Assume without loss of generality that $A_g^+ > 0$ and $A_g^- > 0$. We now focus on I_+. For any compact subset $\mathcal{M} \subset \mathbb{R}^{md}$, there exist T Euclidean balls with radius r that cover $\mathcal{M}$, where $T \leq \{c\,\mathrm{diam}(\mathcal{M})/r\}^{md}$ with $c = 3\sqrt{md/\pi}$. Denote $\{d_1, \ldots, d_T\}$ as the centers of these balls in $\mathbb{R}^{md}$. Now choose an i.i.d. sample $\{(u_{i1}^\top, \ldots, u_{im}^\top)^\top\}_{i=1}^{D_1}$ from the distribution $\widehat{g}\mathbb{1}\{\widehat{g} > 0\}/A_g^+$ with the sample size D_1 to be specified later. Then, for each center $d = (d_1^\top, \ldots, d_m^\top)^\top$ and any $t > 0$, it holds by Hoeffding's inequality that

$$\mathbb{P}\left\{\left|\frac{1}{D_1}\sum_{i=1}^{D_1}\cos\left\{2\pi\left(u_{i1}^\top d_1 + \ldots + u_{im}^\top d_m\right)\right\} - k_g^+(d_1, \ldots, d_m)\right| \geq \frac{t}{8}\right\} \leq \exp\left(-\frac{D_1 t^2}{128}\right).$$

Let $s_{D_1}(x_1, \ldots, x_m) := \sum_{i=1}^{D_1}\cos\left\{2\pi\left(u_{i1}^\top x_1 + \ldots + u_{im}^\top x_m\right)\right\}/D_1$ so that $k_g^+(x_1, \ldots, x_m) = \mathbb{E}_u\{s_{D_1}(x_1, \ldots, x_m)\}$. Then, for any $q \geq 1$, it holds that

$$\begin{aligned}
\mathbb{E}\left[\sup_x \|\nabla_x\{s_{D_1}(x) - k_g^+(x)\}\|^q\right] &= \mathbb{E}\left[\sup_x \|\nabla_x s_{D_1}(x) - \mathbb{E}\nabla_x s_{D_1}(x)\|^q\right] \\
&\leq \mathbb{E}\left[\sup_x \{\|\nabla_x s_{D_1}(x)\| + \mathbb{E}(\|\nabla_x s_{D_1}(x)\|)\}^q\right] \\
&\leq 2^{q-1}\mathbb{E}\left[\sup_x \|\nabla_x s_{D_1}(x)\|^q + \sup_x \{\mathbb{E}(\|\nabla_x s_{D_1}(x)\|)\}^q\right] \\
&\leq 2^q\mathbb{E}\left(\sup_x \|\nabla_x s_{D_1}(x)\|^q\right),
\end{aligned}$$
(4.2)

where in the first line we use the finiteness of $\int_{\mathbb{R}^{md}}\left|\widehat{f}(u)\right|\|u\|du$ (guaranteed by Condition (2.3)) and dominated convergence theorem to exchange the derivative with expectation. Moreover,

$$\begin{aligned}
\mathbb{E}\left(\sup_x \|\nabla_x s_{D_1}(x)\|^q\right) &= \mathbb{E}\left[\sup_x \left\|\frac{1}{D_1}\sum_{i=1}^{D_1}2\pi u_i\cos\left\{2\pi\left(u_i^\top x\right)\right\}\right\|^q\right] \\
&\leq (2\pi)^q\mathbb{E}\left\{\left(\frac{1}{D_1}\sum_{i=1}^{D_1}\|u_i\|\right)^q\right\} \leq (2\pi)^q\mathbb{E}\left\{\frac{1}{D_1}\sum_{i=1}^{D_1}\|u_i\|^q\right\} = (2\pi)^q\mathbb{E}(\|u_1\|^q),
\end{aligned}$$

where we have used the finiteness of $\mathbb{E}(\|u_1\|^q)$ since $\int_{\mathbb{R}^{md}}\left|\widehat{f}(u)\right|\|u\|^q du < \infty$ and the convexity of the function x^q when $q \geq 1$. Therefore, it holds that

$$\mathbb{E}\left\{\sup_x \|\nabla_x\left(s_{D_1}(x) - k_g^+(x)\right)\|^q\right\} \leq (4\pi)^q\mathbb{E}(\|u_1\|^q),$$

and thus by Markov's inequality,

$$\mathbb{P}\left(\sup_x \|\nabla_x\{s_{D_1}(x) - k_g^+(x)\}\| \geq \frac{t}{8r}\right) \leq \left(\frac{32\pi r}{t}\right)^q\mathbb{E}(\|u_1\|^q).$$

By the triangle inequality, the event $\{\sup_{x \in \mathcal{M}}\left|s_{D_1}(x) - k_g^+(x)\right| \leq t/4\}$ has greater probability than the following event

$$\left\{\left|s_{D_1}(d) - k_g^+(d)\right| \leq t/8, \forall d \in \{d_1, \ldots, d_T\}\right\}\bigcap\left\{\sup_x \|\nabla_x\{s_{D_1}(x) - k_g^+(x)\}\| \leq t/(8r)\right\}.$$

Therefore, we have

$$\mathbb{P}\left\{\sup_{x \in \mathcal{M}}\left|s_{D_1}(x) - k_g^+(x)\right| \geq \frac{t}{4}\right\} \leq \left(\frac{c\,\mathrm{diam}(\mathcal{M})}{r}\right)^{md}\exp\left(-\frac{D_1 t^2}{128}\right) + \left(\frac{32\pi r}{t}\right)^q\mathbb{E}(\|u_1\|^q).$$

Letting the right-hand side of the above inequality be of the form $\kappa_1 r^{-md} + \kappa_2 r^q$, and $r = (\kappa_1/\kappa_2)^{1/(q+md)}$, we have

$$\mathbb{P}\Big(\sup_{x \in \mathcal{M}} \big|s_{D_1}(x) - k_g^+(x)\big| \geq \frac{t}{4}\Big) \leq 2\Big(\frac{32\pi(\mathbb{E}\|u_1\|^q)^{1/q} c \operatorname{diam}(\mathcal{M})}{t}\Big)^{\frac{qmd}{q+md}} \exp\Big(-\frac{D_1 t^2}{128}\frac{q}{q+md}\Big).$$

Now, using the fact

$$\mathbb{E}\|u_1\|^q = \int_{\mathbb{R}^{md}} \|u\|^q \frac{\widehat{g}(u)\mathbb{1}\{\widehat{g}(u) > 0\}}{A_g^+}\, du \leq \frac{1}{A_g^+}\int_{\mathbb{R}^{md}} \|u\|^q |\widehat{g}(u)|\, du \leq \frac{1}{A_g^+}\int_{\mathbb{R}^{md}} \|u\|^q \big|\widehat{f}(u)\big|\, du,$$

we conclude that there exists $\{u_i\}_{i=1}^{D_1} \in \mathbb{R}^{md}$ such that uniformly over $\mathcal{M}$, it holds that

$$A_g^+ \cdot \big|s_{D_1}(x) - k_g^+(x)\big| = \left|\frac{A_g^+}{D_1}\sum_{i=1}^{D_1} \cos\big\{2\pi\big(u_i^\top x\big)\big\} - A_g^+ \cdot k_g^+(x)\right| \leq A_g^+ \frac{t}{4}$$

when D_1 is chosen such that

$$D_1 \geq C_1 \frac{md}{t^2}\log\left\{\frac{\pi c \operatorname{diam}(\mathcal{M})\mu_q\big(\widehat{f}\big)}{(A_g^+)^{1/q}t}\right\}$$

for some sufficiently large constant $C_1 = C_1(t, f, \mathcal{M})$. Equivalently, it holds that

$$\big|A_g^+ \cdot s_{D_1}(x) - A_g^+ \cdot k_g^+(x)\big| \leq t/4$$

when D_1 is chosen such that

$$D_1 \geq C_1 \frac{md(A_g^+)^2}{t^2}\log\left\{\frac{\pi c \operatorname{diam}(\mathcal{M})(A_g^+)^{1-1/q}\mu_q\big(\widehat{f}\big)}{t}\right\}.$$

Similarly, it can be shown that there exists $\{v_i\}_{i=1}^{D_2} \in \mathbb{R}^{md}$ such that

$$\big|A_g^- \cdot s_{D_2}(x) - A_g^- \cdot k_g^-(x)\big| \leq t/4$$

uniformly over $x \in \mathcal{M}$, where

$$s_{D_2}(x) = \frac{1}{D_2}\sum_{i=1}^{D_2} \cos\big\{2\pi\big(v_i^\top x\big)\big\},$$

and D_2 is chosen such that

$$D_2 \geq C_2 \frac{md(A_g^-)^2}{t^2}\log\left\{\frac{8\pi c \operatorname{diam}(\mathcal{M})(A_g^-)^{1-1/q}\mu_q\big(\widehat{f}\big)}{t}\right\}$$

for some sufficiently large constant $C_2 = C_2(t, f, \mathcal{M})$. Repeating this procedure for the approximation of II, then with $A_h^+, A_h^-, k_h^+, k_h^-$ similarly defined as $A_g^+, A_g^-, k_g^+, k_g^-$, we can find s_{D_3} and s_{D_4} which are sample means of sine functions such that

$$\big|A_h^+ \cdot s_{D_3}(x) - A_h^+ \cdot k_h^+(x)\big| \leq t/4 \quad \text{and} \quad \big|A_h^- \cdot s_{D_4}(x) - A_h^- \cdot k_h^-(x)\big| \leq t/4$$

uniformly over all $x \in \mathcal{M}$, when the sample sizes D_3 and D_4 are respectively chosen such that

$$D_3 \geq C_3 \frac{md(A_h^+)^2}{t^2}\log\left\{\frac{\pi c \operatorname{diam}(\mathcal{M})(A_h^+)^{1-1/q}\mu_q\big(\widehat{f}\big)}{t}\right\},$$

$$D_4 \geq C_4 \frac{md(A_h^-)^2}{t^2}\log\left\{\frac{\pi c \operatorname{diam}(\mathcal{M})(A_h^-)^{1-1/q}\mu_q\big(\widehat{f}\big)}{t}\right\}$$

for some sufficiently large constants C_3, C_4 that depend on $t, f, \mathcal{M}$. Putting together the pieces, we obtain that

$$|s_D(x) - f(x)| := \left| \left\{ A_g^+ \cdot s_{D_1}(x) - A_g^- \cdot s_{D_2}(x) - A_h^+ \cdot s_{D_3}(x) + A_h^- \cdot s_{D_4}(x) \right\} - f(x) \right|$$

is smaller than t when D_1-D_4 are chosen as above. Since

$$A_g^+ + A_g^- + A_h^+ + A_h^- = \int_{\mathbb{R}^{md}} \left| \widehat{g} \right| + \left| \widehat{h} \right| \leq \sqrt{2} \int_{\mathbb{R}^{md}} \sqrt{\left| \widehat{g} \right|^2 + \left| \widehat{h} \right|^2} = \sqrt{2} \left\| \widehat{f} \right\|_{L_1}$$

and for each u, $\cos\left\{ 2\pi \left(u^\top x \right) \right\}$ can be written as at most 2^{m-1} linear combinations of the term $z_{u_1}\left(2\pi u_1^\top x_1 \right) \ldots z_{u_m}\left(2\pi u_m^\top x_m \right)$, where $\{z_{u_i}(\cdot)\}_{i=1}^m$ is either the cosine or sine function.

Therefore, taking $\mathcal{M} = [-M, M]^{md}$, it holds that $|s_D - f| \leq t$ uniformly over $[-M, M]^{md}$, and s_D satisfies (4.1) with constants $F = 2^m \left\| \widehat{f} \right\|_{L_1}$ and $B = 1$. Lastly, define the symmetrized version of s_D to be

$$\widetilde{s}_D(x_1, \ldots, x_m) := \frac{1}{m!} \sum_\pi s_D(\pi(x_1), \ldots, \pi(x_m)),$$

where the summation is taken over all $m!$ permutations of $(x_1, \ldots, x_m)$. Then, due to the symmetry of f, $|\widetilde{s}_D - f| \leq t$ uniformly over $[-M, M]^{md}$ and $\widetilde{s}_D$ satisfies (4.1) with the same F, B as s_D. $\qquad \square$

The second lemma builds upon the previous one and guarantees the existence of an approximating kernel $\widetilde{f}$ such that f_p and $\widetilde{f}_p$, the pth term in the Hoeffding decomposition of f and $\widetilde{f}$, are sufficiently close for each $1 \leq p \leq m$.

Lemma 4.2. *Suppose the kernel $f \in L_1(\mathbb{R}^{md})$ is continuous and satisfies condition (2.3) for some $q \geq 1$. Then, for any $M > 0$ and $t > 0$, there exists a symmetric function $\widetilde{f} = \widetilde{f}(; t, M)$ such that $\widetilde{f}$ satisfies all the properties in Lemma 4.1, and moreover, for each $1 \leq p \leq m$,*

$$\left| f_p(x_1, \ldots, x_p) - \widetilde{f}_p(x_1, \ldots, x_p) \right| \leq Ct$$

uniformly over all $(x_1^\top, \ldots, x_p^\top) \in [-M, M]^{pd}$ for some positive constant $C = C(m)$.

Proof. To highlight dependence, for any $t_0 > 0$ and $M_0 > 0$, we will denote the approximating kernel in Lemma 4.1 by $\widetilde{f} = \widetilde{f}(; t_0, M_0)$, so that $\left| f - \widetilde{f} \right| \leq t_0$ uniformly over $[-M_0, M_0]^{md}$ and $\widetilde{f}$ satisfies (4.1) with $F = 2^m \left\| \widehat{f} \right\|_{L_1}$ and $B = 1$. This implies that

$$\sup_{(x_1^\top, \ldots, x_m^\top)^\top \in \mathbb{R}^d} \left| \widetilde{f}(x_1, \ldots, x_d; t_0, M_0) \right| \leq 2^m \left\| \widehat{f} \right\|_{L_1}, \tag{4.3}$$

and, in particular, $\widetilde{f} \in L_1(\mathbb{R}^{md})$ under the product measure P^m. We will prove Lemma 4.2 by choosing a $\widetilde{f}(; t_0, M_0)$ with some t_0 and M_0 to be specified later that only depend on the prescribed t and M. Again, to show clearly the dependence on t_0 and M_0, we write $\widetilde{\theta}(t_0, M_0)$, $\widetilde{f}_p(; t_0, M_0)$, and $\widetilde{g}_p(; t_0, M_0)$ in the Hoeffding decomposition of $\widetilde{f}(; t_0, M_0)$. By definition, in order to show $\left| f_p - \widetilde{f}_p(; t_0, M_0) \right| \leq Ct$ over $[-M, M]^{pd}$ for some $C = C(m)$, it suffices to show that $\left| \theta - \widetilde{\theta}(t_0, M_0) \right| \leq 2t_0$ and $|g_p - \widetilde{g}_p(; t_0, M_0)| \leq 2t_0$ over $[-M_0, M_0]^{pd}$

as long as we choose $t_0 \le t$ and $M_0 \ge M$. For $\left|\theta - \widetilde{\theta}(t_0, M_0)\right|$, recalling that $\{\widetilde{X}_i\}_{i=1}^m$ are i.i.d. with law P, one has

$$
\begin{aligned}
\left|\theta - \widetilde{\theta}(t_0, M_0)\right| &= \left|\mathbb{E}\Big\{f(\widetilde{X}_1, \ldots, \widetilde{X}_m) - \widetilde{f}(\widetilde{X}_1, \ldots, \widetilde{X}_m; t_0, M_0)\Big\}\right| \\
&\le \mathbb{E}\Big\{\left|f - \widetilde{f}(; t_0, M_0)\right| \mathbb{1}\Big\{(\widetilde{X}_1^\top, \ldots, \widetilde{X}_m^\top)^\top \in [-M_0, M_0]^{md}\Big\}\Big\} + \\
&\quad \mathbb{E}\Big\{\left|f - \widetilde{f}(; t_0, M_0)\right| \mathbb{1}\Big\{(\widetilde{X}_1^\top, \ldots, \widetilde{X}_m^\top)^\top \notin [-M_0, M_0]^{md}\Big\}\Big\} \\
&\le t_0 + \mathbb{E}\Big\{\left|f - \widetilde{f}(; t_0, M_0)\right| \mathbb{1}\Big\{(\widetilde{X}_1^\top, \ldots, \widetilde{X}_m^\top)^\top \notin [-M_0, M_0]^{md}\Big\}\Big\}.
\end{aligned}
$$

Now, by (4.3), the variable $\left|f - \widetilde{f}\right| \mathbb{1}\Big\{(\widetilde{X}_1^\top, \ldots, \widetilde{X}_m^\top)^\top \notin [-M_0, M_0]^{md}\Big\}$ has an integrable majorant $|f| + 2^m \left\|\widehat{f}\right\|_{L_1}$ under the product measure P^m, and clearly converges to zero in probability as $M_0 \to \infty$. Thus, by choosing $t_0 = t$ and the dominated convergence theorem, there exists some $M_1 = M_1(t)$ such that for each $M_0 \ge M_1(t)$, $\left|\theta - \widetilde{\theta}(t, M_0)\right| \le 2t$. With a similar argument, there exists some $M_2 = M_2(t)$ such that for any $M_0 \ge M_2(t)$ and $1 \le p \le m$, it holds that

$$
|g_p(x_1, \ldots, x_p) - \widetilde{g}_p(x_1, \ldots, x_p; t, M_0)| \le 2t, \quad \text{for all } (x_1^\top, \ldots, x_p^\top)^\top \in [-M_0, M_0]^{pd}.
$$

Therefore, by choosing $M_0 := M \vee M_1(t) \vee M_2(t)$, one has $\widetilde{f}(; t, M_0)$ satisfies all the desired properties. This completes the proof. $\qquad\square$

The third lemma derives a maximal-type tail bound for each $\max_{1 \le k \le n} |V_k(f_p)|$ when f admits the tensor decomposition (2.6).

Lemma 4.3. *Suppose $\{X_i\}_{i=1}^n$ are as in Theorem 2.1. Suppose the symmetric kernel $f : \mathbb{R}^{md} \to \mathbb{R}$ can be written as*

$$
f(x_1, \ldots, x_m) = \sum_{j_1, \ldots, j_m = 1}^K f_{j_1, \ldots, j_m} e_{j_1}(x_1) \ldots e_{j_m}(x_m),
$$

where K is some positive integer, $\{f_{j_1, \ldots, j_m}\}_{j_1, \ldots, j_m = 1}^K$ is a real sequence, and $\{e_j(\cdot)\}_{j=1}^K$ is a set of real basis functions satisfying

$$
\sum_{j_1, \ldots, j_m = 1}^K |f_{j_1, \ldots, j_m}| \le F \quad \text{and} \quad \sup_{1 \le j \le K} \sup_{x \in \mathbb{R}^d} |e_j(x)| \le B
$$

for some positive constants F and B. Let $\mu_a := \sup_{1 \le j \le K} \left(\mathbb{E}|e_j(X_1)|^a\right)^{1/a}$ for each $a \ge 1$. Then, there exists a positive constant $C = C(m, \gamma_1, \gamma_2)$ such that for any $1 \le p \le m$, and any $x \ge 0$,

$$
\mathbb{P}\Big(n^{-p} \max_{1 \le k \le n} |V_k(f_p)| \ge x\Big) \le 6 \exp\left(-\frac{Cnx^{2/p}}{A_{p,n}^{1/p} + x^{1/p} M_{p,n}^{1/p}}\right),
$$

where

$$
A_{p,n} = \mu_1^{2(m-p)} F^2 \left(\sigma^2 + B^2 (\log n)^4 / n\right)^p, \quad M_{p,n} = \mu_1^{m-p} F B^p (\log n)^{2p},
$$

and $\sigma^2 = 64 \gamma_1^{1/3} \mu_3^2 / (1 - \exp(-\gamma_2/3))$.

Proof. Throughout the proof, let C_i's be positive constants that only depend on m, γ_1, γ_2, and we will use the shorthand $f_{j_{a:b}}$ for $f_{j_a,\ldots,j_b}$ for positive integers $a < b$. We drop the dependence of $A_{p,n}$ and $M_{p,n}$ on n for notational simplicity.

Fix a $1 \leq p \leq m$ and we now derive the tail bound for $\max_{1 \leq k \leq n} |V_k(f_p)|$. For the set of bases $\{e_j(\cdot)\}_{j=1}^{K}$ in the expansion of f, define $\widetilde{e}_j := e_j - \mathbb{E}\{e_j(X_1)\}$ for $j \in [K]$. Since f is symmetric, for any $(x_1^\top, \ldots, x_m^\top)^\top \in \mathbb{R}^{md}$, $f(x_1, \ldots, x_m) = f(\pi(x_1), \ldots, \pi(x_m))$ for any permutation π of $\{x_1, \ldots, x_m\}$. By the definition of $\{f_p\}_{p=1}^{m}$ in (2.1), one can readily check that

$$f_p(x_1, \ldots, x_p) = \sum_{j_1, \ldots, j_m = 1}^{K} f_{j_1, \ldots, j_m} \mathbb{E}(e_{j_1}) \ldots \mathbb{E}(e_{j_{m-p}}) \widetilde{e}_{j_{m-p+1}}(x_1) \ldots \widetilde{e}_{j_m}(x_p),$$

for $r \leq p \leq m$. Thus, we have

$$V_k(f_p) = \sum_{j_1, \ldots, j_m = 1}^{K} f_{j_{1:m}} \mathbb{E}(e_{j_1}) \ldots \mathbb{E}(e_{j_{m-p}}) \left\{ \sum_{i=1}^{k} \widetilde{e}_{j_{m-p+1}}(X_i) \right\} \ldots \left\{ \sum_{i=1}^{k} \widetilde{e}_{j_m}(X_i) \right\}.$$

Define, for each $j \in [K]$ and $k \in [n]$,

$$S_{k,j} := \sum_{i=1}^{k} \widetilde{e}_j(X_i) \quad \text{and} \quad Z_j := \max_{1 \leq k \leq n} |S_{k,j}|.$$

Note that for each $j \in [K]$, $\{\widetilde{e}_j(X_i)\}_{i=1}^{n}$ is also geometrically α-mixing. We now control each even order moment of $\max_{1 \leq k \leq n} |V_k(f_p)|$. Let

$$T_p := n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)|.$$

Define

$$\nu_j = C_3(n\sigma_j^2 + B^2), \quad \nu = C_3(n\sigma^2 + B^2), \quad c = C_4 B(\log n)^2$$

and $\sigma^2 = \sup_{j \in [K]} \sigma_j^2$, with

$$\sigma_j^2 := \text{Var}\{\widetilde{e}_j(X_1)\} + 2 \sum_{i > 1} |\text{Cov}\{\widetilde{e}_j(X_1), \widetilde{e}_j(X_i)\}|.$$

Integrating the tail estimate in Corollary 24 of [21] and using Theorem 2.3 in [6] yield that, for any positive integer N, by choosing C_4 in c to be sufficiently large,

$$\mathbb{E}(Z_j^{2pN}) = (4pN) \int_0^\infty x^{2pN-1} \mathbb{P}(Z_j \geq x) dx$$

$$= 4pN \cdot \left(\int_0^{CB \log n} x^{2pN-1} \mathbb{P}(Z_j \geq x) dx + \int_{CB \log n}^\infty x^{2pN-1} \mathbb{P}(Z_j \geq x) dx \right)$$

$$\leq C(B \log n)^{2pN} + (pN)!(8\nu)^{pN} + (2pN)!(4c)^{2pN}$$

$$\leq (pN)!(8\nu)^{pN} + (2pN)!(5c)^{2pN}.$$

Then, employing a similar argument as in [5] (cf. Equation (12) therein), it holds that

$$\mathbb{E}T_p^{2N} = \mathbb{E} \max_{1 \le k_1,\dots,k_{2N} \le n} \sum_{j_1,\dots,j_{2mN}=1}^{K} f_{j_{1:m}} \dots f_{j_{(2N-1)m+1:2mN}} \cdot \mathbb{E}(e_{j_1}) \dots \mathbb{E}(e_{j_{m-p}}) \dots$$

$$\mathbb{E}(e_{j_{(2N-1)m+1}}) \dots \mathbb{E}(e_{j_{2Nm-p}}) \mathbb{E}(S_{k_1,j_{m-p+1}} \dots S_{k_1,j_m} \dots S_{k_{2N},j_{2Nm-p+1}} \dots S_{k_{2N},j_{2Nm}})$$

$$\le n^{-2Np} \mu_1^{2N(m-p)} \sum_{j_1,\dots,j_{2mN}=1}^{K} |f_{j_{1:m}}| \dots |f_{j_{(2N-1)m+1:2mN}}| \mathbb{E}(|Z_{j_{m+1-p}}| \dots |Z_{j_{2mN}}|)$$

$$\le n^{-2Np} \mu_1^{2N(m-p)} \sum_{j_1,\dots,j_{2mN}=1}^{K} |f_{j_{1:m}}| \dots |f_{j_{(2N-1)m+1:2mN}}| (\mathbb{E}Z_{j_{m+1-p}}^{2pN})^{\frac{1}{2pN}} \dots (\mathbb{E}Z_{j_{2mN}}^{2pN})^{\frac{1}{2pN}}$$

$$\le n^{-2Np} \mu_1^{2N(m-p)} F^{2N} \{(pN)!(8\nu)^{pN} + (2pN)!(4c)^{2pN}\},$$

where in the second inequality we use the generalized Hölder inequality. By Stirling's approximation formula $\sqrt{2\pi} n^{n+1/2} e^{-n} \le n! \le e n^{n+1/2} e^{-n}$, it holds that

$$\{(pN)!\}^{1/p} \le e^{1/p}(pN)^{N+1/2p} e^{-N} \le C_5^N N^{N+1/2} e^{-N} \le C_6^N N!.$$

Similarly, we have $\{(2pN)!\}^{1/p} \le C_6^{2N}(2N)!$. Thus we have

$$\mathbb{E}(T_p)^{\frac{2N}{p}} \le (\mathbb{E}T_p^{2N})^{\frac{1}{p}}$$

$$\le n^{-2N} \mu_1^{\frac{2N(m-p)}{p}} F^{\frac{2N}{p}} \{C_7^N N! \nu^N + C_8^{2N}(2N)! c^{2N}\}.$$

Now we control the Laplace transform of $T_p^{1/p}$,

$$\mathbb{E}(e^{\lambda T_p^{1/p}}) = \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} \mathbb{E}T_p^{N/p} \le 3 \sum_{N=0}^{\infty} \frac{\lambda^{2N}}{(2N)!} \mathbb{E}T_p^{2N/p}$$

$$\le 3 \left\{ \sum_{N=0}^{\infty} \frac{\lambda^{2N}}{(2N)!} C_7^N n^{-2N} N! \mu_1^{\frac{2N(m-p)}{p}} F^{\frac{2N}{p}} \nu^N + \sum_{N=0}^{\infty} \lambda^{2N} n^{-2N} C_8^{2N} \mu_1^{\frac{2N(m-p)}{p}} F^{\frac{2N}{p}} c^{2N} \right\},$$

(4.4)

where in the first inequality we use only the even moments with an absolute constant 3. For the first summand in (4.4), we have

$$\sum_{N=0}^{\infty} \lambda^{2N} \frac{N!}{(2N)!} C_7^N n^{-2N} \mu_1^{2N(m-p)/p} F^{2N/p} \nu^N \le \sum_{N=0}^{\infty} \frac{(\lambda/n)^{2N}}{N!} 2^{-N} C_9^N \mu_1^{2N(m-p)/p} F^{2N/p} \nu^N$$

$$= \exp\left\{ C_{10}(\lambda/n)^2 \mu_1^{2(m-p)/p} F^{2/p} \nu \right\},$$

where in the first line we use the relation $N!/(2N)! \le 2^{-N}/N!$. For the second summand, we have

$$\sum_{N=0}^{\infty} \lambda^{2N} n^{-2N} C_8^{2N} \mu_1^{2N(m-p)/p} F^{2N/p} c^{2N} = 1 + \frac{(\lambda/n)^2 C_8^2 \mu_1^{2(m-p)/p} F^{2/p} c^2}{1 - (\lambda/n)^2 C_8^2 \mu_1^{2(m-p)/p} F^{2/p} c^2}$$

$$\le 1 + \frac{(\lambda/n)^2 C_8^2 \mu_1^{2(m-p)/p} F^{2/p} c^2}{1 - (\lambda/n) C_8 \mu_1^{(m-p)/p} F^{1/p} c}$$

for $\lambda \le n / \{C_8 \mu_1^{(m-p)/p} F^{1/p} c\}$. Now using the relation $e^x + 1 + y \le 2e^{x+y}$ which holds for all positive x, y, we have

$$\mathbb{E}(e^{\lambda T_p^{1/p}}) \le 6 \exp\left[\frac{(\lambda/n)^2 C_{11} \mu_1^{\frac{2(m-p)}{p}} F^{\frac{2}{p}}(\nu + c^2)}{2\{1 - (\lambda/n) C_8 c \mu_1^{\frac{(m-p)}{p}} F^{\frac{1}{p}}\}} \right] = 6 \exp\left\{ \frac{(\lambda/n)^2 C_{11} n A_p^{1/p}}{2(1 - (\lambda/n) C_8 M_p^{1/p})} \right\}.$$

Now, taking $\lambda = nx^{1/p}/(C_{11}nA_p^{1/p} + C_8 M_p^{1/p}x^{1/p})$ in the exponential Markov inequality, we have

$$\mathbb{P}(T_p \geq x) \leq \exp(-\lambda x^{1/p})\mathbb{E}\left(e^{\lambda T_p^{1/p}}\right) \leq 6\exp\left(-\frac{C_{12}nx^{2/p}}{A_p^{1/p} + x^{1/p}M_p^{1/p}}\right).$$

Moreover, taking $\delta = 1$ in Theorem 3 of [13], we obtain

$$\mathrm{Var}\{\widetilde{e}_j(X_1)\} + 2\sum_{i>1}|\mathrm{Cov}\{\widetilde{e}_j(X_1), \widetilde{e}_j(X_i)\}| \leq 2\sum_{i\geq1}|\mathrm{Cov}\{\widetilde{e}_j(X_1), \widetilde{e}_j(X_i)\}|$$

$$\leq 2\left\{\sum_{n=0}^{\infty}8\alpha^{1/3}(n)\right\}\|\widetilde{e}_j(X_1)\|_3\|\widetilde{e}_j(X_1)\|_3 \leq 64\left\{\sum_{n=0}^{\infty}\alpha^{1/3}(n)\right\}\|e_j(X_1)\|_3\|e_j(X_1)\|_3$$

$$\leq 64\gamma_1^{1/3}\mu_3^2\left\{\sum_{n=0}^{\infty}\exp\left(-\gamma_2 n/3\right)\right\} = \frac{64\gamma_1^{1/3}}{1 - \exp\{-\gamma_2/3\}}\mu_3^2.$$

Putting together the pieces completes the proof. □

We now use Lemmas 4.1–4.3 to complete the proof of Theorem 2.1.

Proof of Theorem 2.1. Fix a $1 \leq p \leq m$. Fix some $t > 0$ and $M > 0$, and define the event

$$\mathcal{E} := \{X_i \in [-M, M]^d \text{ for all } 1 \leq i \leq n\}.$$

Then, for the prescribed t and M, Lemma 4.2 implies that there exists a symmetric kernel $\widetilde{f} = \widetilde{f}(;t,M)$ such that $\left|f - \widetilde{f}(;t,M)\right| \leq t$ uniformly over all $(x_1^\top, \ldots, x_m^\top)^\top \in [-M, M]^{md}$, and for each $1 \leq p \leq m$, $\left|f_p - \widetilde{f}_p(;t,M)\right| \leq Ct$ uniformly over $[-M, M]^{pd}$ for some $C = C(m)$. By definition, this implies that $n^{-p}\left|\max_{1\leq k\leq n}|V_k(f_p)| - \max_{1\leq k\leq n}\left|V_k(\widetilde{f}_p)\right|\right| \leq Ct$ on the event $\mathcal{E}$, and thus for any $x > 0$,

$$\mathbb{P}(n^{-p}\max_{1\leq k\leq n}|V_k(f_p)| \geq x + Ct)$$

$$= \mathbb{P}\left(\left\{n^{-p}\max_{1\leq k\leq n}|V_k(f_p)| \geq x + Ct\right\}\bigcap\mathcal{E}\right) + \mathbb{P}\left(\left\{n^{-p}\max_{1\leq k\leq n}|V_k(f_p)| \geq x + Ct\right\}\bigcap\mathcal{E}^c\right)$$

$$\leq \mathbb{P}\left(n^{-p}\max_{1\leq k\leq n}\left|V_k(\widetilde{f}_p(;t,M))\right| \geq x\right) + n\mathbb{P}(X_1 \notin [-M, M]^d).$$

Again by Lemma 4.2, $\widetilde{f}(;t,M)$ satisfies the conditions of Lemma 4.3 with constants $F = 2^m\left\|\widehat{f}\right\|_{L_1}$ and $B = 1$. Therefore, applying the trivial bound that $\mu_3 \leq B = 1$ in Lemma 4.3, we obtain that

$$\mathbb{P}\left(n^{-p}\max_{1\leq k\leq n}|V_k(f_p)| \geq x + Ct\right) \leq 6\exp\left(-\frac{Cnx^{2/p}}{A_p^{1/p} + x^{1/p}M_p^{1/p}}\right) + n\mathbb{P}(X_1 \notin [-M, M]^d),$$

where A_p and M_p are defined in (2.5). Now, note that the first summand on the right hand side does not depend on M or t. Accordingly, by first choosing a large enough M that depends only on x, n, F, since the measure P considered in this paper is always tight, we obtain that the second term is smaller than an arbitrary small proportion of the first term. Lastly, choosing $t = x$ and adjusting the constant finishes the proof. □

5 Proofs of other results

We will only prove Corollaries 3.1–3.3. The proof of Corollary 3.4 is similar to that of Corollary 3.1.

Proof of Corollary 3.1. By inspection of the proof of Theorem 2.1, it suffices to prove that when f_0 satisfies (3.1) for some $q \geq 1$, the conclusion of Lemma 4.1 still holds with $F = 4\left\|\widehat{f_0}\right\|_{L_1}$ and $B = 1$. Now, following the proof of Lemma 4.1 with prescribed range $2M$ and approximation error t, there exists an $\widetilde{f_0}$ with expansion in the cosine bases $\left\{\cos\left(2\pi u^\top x\right)\right\}$ such that $\left|f_0 - \widetilde{f_0}\right| \leq t$ uniformly over $[-2M, 2M]^d$ and $\widetilde{f_0}$ satisfies the (4.1) with constants $F = 2\left\|\widehat{f_0}\right\|_{L_1}$ and $B = 1$. Let $\widetilde{f}(x, y) := \widetilde{f_0}(x - y)$. Then, $\left|f - \widetilde{f}\right| \leq t$ uniformly over $[-M, M]^{2d}$, and by the trigonometric identity

$$\cos\left\{2\pi u^\top (x - y)\right\} = \cos\left(2\pi u^\top x\right)\cos\left(2\pi u^\top y\right) + \sin\left(2\pi u^\top x\right)\sin\left(2\pi u^\top y\right),$$

$\widetilde{f}$ satisfies (4.1) with constants $F = 4\left\|\widehat{f_0}\right\|_{L_1}$ and $B = 1$. This completes the proof. □

Proof of Corollary 3.2. Again, we only need to reprove Lemma 4.1. When f_0 is PD, we have by definition that $f_0(0) \geq 0$ and for each $x, y \in \mathbb{R}^d$, $f_0(x - y) = f_0(y - x)$ implies that $f_0(x) = f_0(-x)$ for any $x \in \mathbb{R}^d$. This implies that the Fourier transform $\widehat{f_0}$ of f_0 is real-valued, and $\widehat{h} = 0$ in the proof of Lemma 4.1. Moreover, since $f_0 \in L_1(\mathbb{R}^d)$ as it satisfies (3.1) for some $q \geq 1$, f equals the inverse Fourier transform of $\widehat{f}$. Thus, by Bochner's theorem (cf. Section 1.4.3, [23]), $\widehat{f_0}$ is nonnegative and we have $f_0 = I = I_+$ with $m = 1$ in the proof of Lemma 4.1. By definition, we have

$$f_0(x) = \int_{\mathbb{R}^d} \widehat{f_0}(u) e^{2\pi i x^\top u} du = \int_{\mathbb{R}^d} \left|\widehat{f_0}(u)\right| e^{2\pi i x^\top u} du.$$

Letting $x = 0$ in the above equation, we obtain $\left\|\widehat{f_0}\right\|_{L_1} = f_0(0)$.

Now consider the case where $\widehat{f_0}$ only has fractional moment. Let $\widetilde{f_0} := f_0/f_0(0)$ and denote the Lipschitz constants of $\widetilde{f_0}$ and f_0 as $L_{\widetilde{f_0}}$ and L_{f_0}, respectively. Then, $L_{\widetilde{f_0}} = L_{f_0}/f_0(0)$. Now we proceed with the proof of Lemma 4.1 until (4.2), and replace it with

$$\mathbb{E}\left\{\sup_x \left\|\nabla_x s_D(x) - \widetilde{f_0}(x)\right\|^q\right\} \leq \mathbb{E}\left\{\sup_x \left(\|\nabla_x s_D(x)\|^q + \left\|\nabla_x \widetilde{f_0}(x)\right\|^q\right)\right\}$$
$$\leq \mathbb{E}\left\{\sup_x \|\nabla_x s_D(x)\|^q\right\} + \sup_x \left\|\nabla_x \widetilde{f_0}(x)\right\|^q$$
$$\leq \mathbb{E}\left\{\sup_x \|\nabla_x s_D(x)\|^q\right\} + L_{\widetilde{f_0}}^q,$$

where $s_D(x) = \sum_{i=1}^D \cos\left(2\pi u_i^\top x\right)/D$ (here we use the notation s_D instead of s_{D_1} since in the PD case we only need to approximate the term I_+ as argued in the first part of the corollary). Note that the original (4.2) in the proof of Lemma 4.1 no longer holds as mere fractional moment does not guarantee the exchange of derivative and expectation in its first step. For the first term in the above inequality, we have

$$\mathbb{E}\left\{\sup_x \|\nabla_x s_D(x)\|^q\right\} = \mathbb{E}\left[\sup_x \left\|\frac{1}{D}\sum_{i=1}^D 2\pi u_i \cos\left\{2\pi\left(u_i^\top x\right)\right\}\right\|^q\right]$$
$$\leq (2\pi)^q \mathbb{E}\left(\left\|\frac{1}{D}\sum_{i=1}^D u_i\right\|^q\right) \leq (2\pi)^q \mathbb{E}\left\{\left(\frac{1}{D}\sum_{i=1}^D \|u_i\|\right)^q\right\}$$
$$\leq (2\pi)^q \mathbb{E}\left\{D^{-q}\sum_{i=1}^D \|u_i\|^q\right\} = (2\pi)^q D^{1-q}\mathbb{E}(\|u_1\|^q).$$

Therefore, it holds that

$$\mathbb{E}\left\{ \sup_x \left\| \nabla_x s_D(x) - \widetilde{f}_0(x) \right\|^q \right\} \le (2\pi)^q D^{1-q} \mathbb{E}(\|u\|^q) + L_{\widetilde{f}_0}^q.$$

Markov inequality now gives

$$\mathbb{P}\left\{ \sup_x \left\| \nabla_x \left(s_D(x) - \widetilde{f}_0(x) \right) \right\| \ge \frac{t}{2r} \right\} \le \left(\frac{2r}{t} \right)^q \left\{ (2\pi)^q D^{1-q} \mathbb{E}(\|u\|^q) + L_{\widetilde{f}_0}^q \right\}.$$

Proceeding with the proof of Lemma 4.1, we obtain

$$\mathbb{P}\left(\sup_{x \in \mathcal{M}} \left| s_D(x) - \widetilde{f}_0 \right| \ge t \right) \le \left(\frac{2r}{t} \right)^q \left\{ (2\pi)^q D^{1-q} \mathbb{E}\|u\|^q + L_{\widetilde{f}_0}^q \right\} + \left(\frac{c \operatorname{diam}(\mathcal{M})}{r} \right)^{md} \exp\left(-\frac{Dt^2}{8} \right).$$

Writing the right-hand side of the above inequality in the form $\kappa_1 r^{-md} + \kappa_2 r^q$ and letting $r = (\kappa_1/\kappa_2)^{1/(q+md)}$, we obtain

$$\mathbb{P}\left(\sup_{x \in \mathcal{M}} \left| s_D(x) - \widetilde{f}_0 \right| \ge t \right)$$
$$\le 2 \left(\frac{2c \operatorname{diam}(\mathcal{M})}{\varepsilon} \right)^{\frac{qmd}{q+md}} \left\{ (2\pi)^q D^{1-q} \mathbb{E}(\|u_1\|^q) + L_{\widetilde{f}_0}^q \right\}^{\frac{md}{q+md}} \exp\left(-\frac{D\varepsilon^2}{8} \frac{q}{q+md} \right).$$

For any $t > 0$, we can choose large enough $D = D(t)$ such that the right-hand side of the above inequality is arbitrarily small. The proof is complete. $\qquad\square$

Proof of Corollary 3.3. Let $K(\cdot) : \mathbb{R}^{md} \to \mathbb{R}$ be the standard md-variate Gaussian density defined as $K(x) := \exp(-\|x\|^2/2)(2\pi)^{-md/2}$, and $K_h(x) = K(x/h)h^{-md}$ for some positive constant h. Define $f_h(x) := (f * K_h)(x)$. Then, it holds that

$$|f_h(x) - f(x)| = \left| \int_{\mathbb{R}^{md}} (2\pi)^{-md/2} \exp\left(-\frac{\|y\|^2}{2} \right) \{ f(x - yh) - f(x) \} dy \right|$$
$$\le \int_{\mathbb{R}^{md}} (2\pi)^{-md/2} \exp\left(-\frac{\|y\|^2}{2} \right) |f(x - yh) - f(x)| dy.$$

Denote the upper bound of f as M_f. Then, for any $t > 0$, there exists some positive constant $A = A(M_f, m, d, t)$ such that

$$\int_{([-A,A]^{md})^c} (2\pi)^{-md/2} \exp\left(-\frac{\|y\|^2}{2} \right) |f(x - yh) - f(x)| dy$$
$$\le 2M_f \int_{([-A,A]^{md})^c} (2\pi)^{-md/2} \exp\left(-\frac{\|y\|^2}{2} \right) dy \le t/4.$$

Inside $[-A, A]^{md}$, using the uniform continuity of f, there exists some $h = h(M_f, m, d, t)$, such that

$$\int_{[-A,A]^{md}} (2\pi)^{-md/2} \exp\left(-\frac{\|y\|^2}{2} \right) |f(x - yh) - f(x)| dy \le t/4.$$

Putting together the pieces, it holds that for any $t > 0$, there exists some $h = h(M_f, m, d, t)$ such that $\|f_h - f\|_\infty \le t/2$. Since both f and K_h belong to $L_1(\mathbb{R}^{md})$, their Fourier transforms exist. It can be readily checked that $\widehat{K}_h(u) = \exp\left(-2\pi^2 h^2 \|u\|^2 \right)$, and thus

$$\widehat{f}_h(u) = \widehat{f}(u) \cdot \widehat{K}_h(u) = \widehat{f}(u) \exp\left(-2\pi^2 h^2 \|u\|^2 \right).$$

Using the relation $\|f * g\|_{L_q} \leq \|f\|_{L_q}\|g\|_{L_1}$ for any $q \geq 1$ and $f \in L_q(\mathbb{R}^{md}), g \in L_1(\mathbb{R}^{md})$ and the fact that $K_h \in L_1(\mathbb{R}^{md})$, it holds that $f_h \in L_1(\mathbb{R}^{md})$. Moreover, it can readily checked that

$$\mu_q^q\left(\widehat{f_h}\right) = \int_{\mathbb{R}^{md}} \left|\widehat{f_h}(u)\right| \|u\|^q du = \int_{\mathbb{R}^{md}} \left|\widehat{f}(u)\right| \|u\|^q \exp\left(-2\pi^2 h^2 \|u\|^2\right) du < \infty$$

for any $q \geq 1$. Therefore, by Lemma 4.1, for any given $M > 0$ and $t > 0$, we can find an approximating kernel $\widetilde{f_h} = \widetilde{f_h}(t)$ such that $\left|\widetilde{f_h} - f_h\right| \leq t/2$ uniformly over $[-M, M]^{md}$, and $\widetilde{f_h}$ further satisfies (4.1) with constants $F = 2^m \left\|\widehat{f_h}\right\|_{L_1}$ and $B = 1$. Choosing $h = h(M_f, m, d, t/2)$, by the triangle inequality, we have

$$\left|f - \widetilde{f_h}\right| \leq t/2 + t/2 = t$$

uniformly over $[-M, M]^{md}$. Lastly, we upper bound the term $\left\|\widehat{f_h}\right\|_{L_1}$. To this end, we have

$$\left\|\widehat{f_h}\right\|_{L_1} = \int_{\mathbb{R}^{md}} \left|\widehat{f}(u)\right| \exp(-2\pi^2 h^2 \|u\|^2) du \leq L \int_{\mathbb{R}^{md}} \frac{1}{1 + \|u\|^{md+\varepsilon}} \exp(-2\pi^2 h^2 \|u\|^2) du.$$

Using polar coordinates, it holds that

$$\begin{aligned}
\left\|\widehat{f_h}\right\|_{L_1} &\leq \Gamma(md) L \int_0^\infty \frac{r^{md-1}}{1 + r^{md+\varepsilon}} \exp(-2\pi^2 h^2 r^2) dr \\
&\leq \Gamma(md) L \left(1 + \int_1^\infty \frac{1}{r^{1+\varepsilon}} dr\right) \\
&= (1 + \varepsilon^{-1}) \Gamma(md) L.
\end{aligned}$$

This completes the proof. □

References

[1] Radosł Adamczak, *Moment inequalities for U-statistics*, The Annals of Probability **34** (2006), no. 6, 2288–2314. MR-2294982

[2] Miguel A Arcones and Evarist Giné, *Limit theorems for U-processes*, The Annals of Probability **21** (1993), no. 3, 1494–1542. MR-1235426

[3] Yves D Atchadé and Matias D Cattaneo, *A martingale decomposition for quadratic forms of markov chains (with applications)*, Stochastic Processes and their Applications **124** (2014), no. 1, 646–677. MR-3131309

[4] Eric Beutner and Henryk Zähle, *Deriving the asymptotic distribution of U-and V-statistics of dependent data using weighted empirical processes*, Bernoulli **18** (2012), no. 3, 803–822. MR-2948902

[5] Igor'Semenovich Borisov and Nadezhda Volodko, *A note on exponential inequalities for the distribution tails of canonical von Mises' statistics of dependent observations*, Statistics and Probability Letters **96** (2015), 287–291. MR-3281778

[6] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, 2013. MR-3185193

[7] Herold Dehling, *Limit theorems for dependent U-statistics*, Dependence in Probability and Statistics (New York), Springer, 2006, pp. 65–86. MR-2283250

[8] Herold Dehling and Murad S Taqqu, *The empirical process of some long-range dependent sequences with an application to U-statistics*, The Annals of Statistics **17** (1989), no. 4, 1767–1783. MR-1026312

[9] Herold Dehling and Martin Wendler, *Central limit theorem and the bootstrap for U-statistics of strongly mixing data*, Journal of Multivariate Analysis **101** (2010), no. 1, 126–137. MR-2557623

[10] Manfred Denker, *Statistical decision procedures and ergodic theory*, Ergodic Theory and Related Topics: Proceedings Vitte/Hiddensee (Horst Michel, ed.), 1981, pp. 35–47. MR-0730767

[11] Manfred Denker and Gerhard Keller, *On U-statistics and v. mise' statistics for weakly dependent processes*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **64** (1983), no. 4, 505–522. MR-0717756

[12] Isha Dewan and BLS Prakasa Rao, *Asymptotic normality for U-statistics of associated random variables*, Journal of Statistical Planning and Inference **97** (2001), no. 2, 201–225. MR-1861150

[13] Paul Doukhan, *Mixing: Properties and examples*, Springer, 1994. MR-1312160

[14] Gerald B Folland, *Real analysis: Modern techniques and their applications*, John Wiley & Sons, 2013. MR-1681462

[15] Evarist Giné, Rafał Latała, and Joel Zinn, *Exponential and moment inequalities for U-statistics*, High Dimensional Probability II, vol. 47, Springer, 2000, pp. 13–38. MR-1857312

[16] Fang Han, *An exponential inequality for U-statistics under mixing conditions*, Journal of Theoretical Probability **31** (2018), no. 1, 556–578. MR-3769824

[17] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), no. 301, 13–30. MR-0144363

[18] Tailen Hsing and Wei Biao Wu, *On weighted U-statistics for stationary processes*, The Annals of Probability **32** (2004), no. 2, 1600–1631. MR-2060311

[19] Anne Leucht, *Degenerate U-and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency*, Bernoulli **18** (2012), no. 2, 552–585. MR-2922461

[20] Anne Leucht and Michael H Neumann, *Degenerate U-and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics*, Annals of the Institute of Statistical Mathematics **65** (2013), no. 2, 349–386. MR-3011626

[21] Florence Merlevède and Magda Peligrad, *Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples*, The Annals of Probability **41** (2013), no. 2, 914–960. MR-3077530

[22] Ali Rahimi and Benjamin Recht, *Random features for large-scale kernel machines*, Advances in Neural Information Processing Systems, 2007, pp. 1177–1184.

[23] Walter Rudin, *Fourier analysis on groups*, Courier Dover Publications, 1962. MR-0152834

[24] Elias M Stein and Rami Shakarchi, *Fourier analysis: An introduction*, Princeton University Press, 2011. MR-1970295

[25] Ken-Ichi Yoshihara, *Limiting behavior of U-statistics for stationary, absolutely regular processes*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **35** (1976), no. 3, 237–252. MR-0418179

[26] Zhou Zhou, *Inference of weighted V-statistics for nonstationary time series and its applications*, The Annals of Statistics **42** (2014), no. 1, 87–114. MR-3161462