Wasserstein-p bounds in the central limit theorem under local dependence

Tianle Liu[*]   Morgane Austern[†]

Abstract

The central limit theorem (CLT) is one of the most fundamental results in probability; and establishing its rate of convergence has been a key question since the 1940s. For independent random variables, a series of recent works established optimal error bounds under the Wasserstein-p distance (with $p \geq 1$). In this paper, we extend those results to locally dependent random variables, which include m-dependent random fields and U-statistics. Under conditions on the moments and the dependency neighborhoods, we derive optimal rates in the CLT for the Wasserstein-p distance. Our proofs rely on approximating the empirical average of dependent observations by the empirical average of *i.i.d.* random variables. To do so, we expand the Stein equation to arbitrary orders by adapting the Stein's dependency neighborhood method. Finally we illustrate the applicability of our results by obtaining efficient tail bounds.

1 Introduction

The central limit theorem (CLT) is one of the most fundamental theorems in probability theory. Initially formulated for independent and identically distributed random variables, it has since then been generalized to triangular arrays [18], martingales [28], U-statistics [23], locally dependent random variables [24, 21, 34], and mixing random fields [37, 9]. Let (I_n) be an increasing sequence of subsets $I_1 \subseteq I_2 \subseteq \cdots \subseteq I$, whose sizes increase to infinity $|I_n| \to \infty$. Set $(X_i)_{i \in I}$ to be (dependent) centered random variables. Under certain conditions on the moments of (X_i) and on its dependence structure, the CLT states that the scaled sum is asymptotically normal, i.e.,

$$W_n := \sigma_n^{-1} \sum_{i \in I_n} X_i \xrightarrow{d} \mathcal{N}(0, 1),$$

---

[*]Harvard University. E-mail: tianleliu@fas.harvard.edu
[†]Harvard University. E-mail: maustern@fas.harvard.edu

where we write $\sigma_n^2 = \text{Var}\big(\sum_{i \in I_n} X_i\big)$. Starting with the work of Berry and Esseen in 1940s, there is a long history of quantifying how far W_n is from being normally distributed. One of the most important metrics to do so is the Wasserstein-p distance originated in optimal transport theory [44]. For two probability measures ν and μ over the real line $\mathbb{R}$, we denote by $\Gamma(\nu, \mu)$ the set of all couplings of ν and μ, and the Wasserstein-p distance between ν and μ is defined as

$$\mathcal{W}_p(\nu, \mu) := \inf_{\gamma \in \Gamma(\nu, \mu)} \Big(\mathbb{E}_{(X,Y) \sim \gamma} \left[|X - Y|^p \right] \Big)^{1/p}.$$

When the observations (X_i) are independent, [1, 15] established that for $p = 1$ the convergence rate for the CLT is $\mathcal{O}\big(|I_n|^{-\frac{1}{2}}\big)$. Extending such results to $p > 1$ remained for a while an open question. The first bounds for $p > 1$ obtained by [6] dating back to the 1970s were sub-optimal in terms of the sample size $|I_n|$ as they decrease at a slower rate of $\mathcal{O}\big(|I_n|^{-\frac{1}{2}+\frac{1}{p}}\big)$. [36] obtained that, for $1 \le p \le 2$, the Wasserstein distance converges at the optimal rate $\mathcal{O}\big(|I_n|^{-\frac{1}{2}}\big)$ under some additional necessary moment conditions, and they conjectured that such a rate would be extendable to arbitrary $p \ge 1$. This was recently proven to be true by [8, 10] using a series of methods including the Edgeworth expansion and the exchangeable pair method. They showed that if $\max_i \|X_i\|_{p+2} < \infty$ and if $\text{Var}(X_1) = \text{Var}(X_i) = 1$, then there is a constant $K_p < \infty$ such that

$$\mathcal{W}_p\big(\mathcal{L}(W_n), \mathcal{N}(0, 1)\big) \le \frac{K_p \|X_1\|_{p+2}^{1+2/p}}{\sqrt{|I_n|}},$$

where $\mathcal{L}(\cdot)$ designates the distribution of the given random variable. It is however crucial to note that these rates were obtained under the key assumption of independence of the (X_i). In this paper, we aim to generalize this beyond the assumption of independence which is restrictive for many applications.

An important class of dependent observations (X_i) are locally dependent random variables. Intuitively, we say that (X_i) are locally dependent if for every finite group of random variables $(X_i)_{i \in J}$, where $J \subset I$, there exists a subset $N(J) \subset I$ such that $(X_i)_{i \in J}$ is independent from $(X_i)_{i \in I \setminus N(J)}$. The subset $N(J)$ is often called the dependency neighborhood of J. Examples of such random variables include m-dependent random fields, U-statistics, and subgraph count statistics in the Erdős–Rényi random graphs. Under general conditions on the sizes of the dependency neighborhoods the central limit theorem is known to hold and its rate of convergence in Wasserstein-1 distance was established by [3, 5]. This was extended to Wasserstein-2 bounds by [16] by relating it to Zolotarev's metrics and cleverly exploiting Stein's method. Drawing inspiration from [10], sub-optimal rates were also achieved in [17] for arbitrary $p \ge 1$ under more technical conditions. Nevertheless, an optimal rate bound for general Wasserstein-p distances ($p \ge 1$) remains unknown. This is the gap that we fill in this paper. We consider locally dependent (not necessarily identically distributed) random variables (X_i), and consider the empirical average $W_n := \sigma_n^{-1} \sum_{i \in I_n} X_i$ where $\sigma_n^2 := \sum_{i \in I_n} X_i$. For all $p \ge 1$ we obtain bounds for the $\mathcal{W}_p$ distance $\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0, 1))$. We do so under the assumption that the variances (σ_n) are nondegenerate, and under moment conditions and on the sizes of dependency neighborhoods. Notably if the size of the dependency neighborhoods is uniformly bounded we obtain bounds that decrease at the optimal rate (see Theorem 3.4)

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0, 1)) = \mathcal{O}\Big(\frac{1}{\sqrt{|I_n|}} \Big).$$

We further generalize our results to triangular arrays where the random variables $\big(X_i^{(n)}\big)$ are allowed to change with n. Finally, we demonstrate how those bounds can be exploited to obtain non-uniform Berry–Esseen type bounds that have polynomial decay.

The key idea of our proofs is to approximate the empirical average W_n by an empirical average V_n of *i.i.d.* random variables for which Wasserstein's bounds are already known. To do this we establish an Edgeworth-type expansion of the Stein equation in terms of the cumulants of the W_n. Indeed, in Lemma 8.7 we prove that if h is a function smooth enough (made precise later) and $Z \sim \mathcal{N}$ is a standard random variable then

$$
\mathbb{E}[h(W_n)] - \mathbb{E}[h(Z)] = \mathbb{E}[f'_h(W_n) - W_n f_h(W_n)]
$$
$$
= \sum_{(r,s_{1:r}) \in \Gamma(\lceil p \rceil - 1)} (-1)^r \prod_{j=1}^{r} \frac{\kappa_{s_j+2}(W_n)}{(s_j+1)!} \mathcal{N}\left[\prod_{j=1}^{r} (\partial^{s_j+1}\Theta)\, h \right] + \text{ Remainders,} \tag{1.1}
$$

where f_h is the solution of the Stein equation eq. (8.2) and where $(\kappa_j(W_n))$ designates the cumulants of W_n (the other notations will be made explicit in the next few sections). This generalizes a similar well-known result for *i.i.d.* observations established in [4]. To guarantee that our choice of V_n is a good approximation of W_n we utilize this expansion and exploit the Hamburger moment problem to choose V_n to be such that its first $\lceil p \rceil + 1$ cumulants match the ones of W_n.

1.1 Related literature

[1, 15] established that the convergence rate in the central limit theorem is $\mathcal{O}\big(|I_n|^{-\frac{1}{2}}\big)$ in terms of the Wasserstein-1 distance. Since then it has been tightened and generalized to dependent observations. Notably, the Stein's method offers a series of powerful techniques for obtaining Wasserstein-1 bounds in the dependence setting. See [38] for a survey of those methods. [3, 5] obtained Wasserstein-1 bounds under local dependence conditions.

[6] proposed a rate of $\mathcal{O}\big(|I_n|^{-\frac{1}{2}+\frac{1}{p}}\big)$ for the Wasserstein-p distance under the hypothesis that the random variables have finite exponential moments. [39] obtained a similar rate but only required the existence of p-th moments. [35, 36] showed that in order to obtain a convergence rate of $\mathcal{O}\big(|I_n|^{-\frac{1}{2}}\big)$, it is necessary to require finite $(p+2)$-th moments of the random variables. They also obtained the optimal rate for $1 \leq p \leq 2$ and conjectured that a similar rate should be valid for any arbitrary $p > 2$. This conjecture was demonstrated to be true by [8, 10]. Those two papers took different approaches. [8] used an Edgeworth expansion argument. [10], on the other hand, used the Ornstein-Uhlenbeck interpolation combined with a Stein exchangeable pair argument and their methods further applied to multivariate settings. Previous to that, [27] had already obtained the optimal rate for the Wasserstein-p distance using the Ornstein-Uhlenbeck interpolation but needed significantly stronger assumptions on the distribution of the random variables by requiring the existence of a Stein kernel. Moreover, for the special case $p = 2$, the celebrated HWI inequality [32] and Talagrand quadratic transport inequality [42] can help obtain Wasserstein-2 bounds by relating it to the Kullback-Leibler divergence.

Contrary to the independent case, much less is known for the general Wasserstein-p distance for dependent data. [16] adapted the Stein's method to obtain Wasserstein-2 bounds for locally dependent variables. [17] modified the approach of [10] and obtained a sub-optimal rate $\mathcal{O}\big(|I_n|^{-\frac{1}{2}} \log |I_n|\big)$ for the Wasserstein-p distance under local dependence. Our results propose significant extensions to both of those results by generalizing the optimal rate to arbitrary $p \geq 1$.

Our proofs also rely on the Stein's method and a result of [36] that allows to upper the Wasserstein-p distance by an integral probability metric [45]. As those metrics are defined as the supremum of expected differences over a certain class of functions, the Stein's method lends itself nicely to this problem. The Stein's method was first introduced in [41] as a new method to obtain a Berry–Esseen bound and prove the

central limit theorem for weakly dependent data. It has since then become one of the most popular and powerful tools to prove asymptotic normality for dependent data, and different adaptations of it have been proposed, notably the dependency neighborhoods, the exchangeable pairs, the zero-bias coupling, and the size-bias coupling [38]. In addition to being used to prove the central limit theorem, it has also been adapted to obtain limit theorems with the Poisson distribution [12] or the exponential distribution [11, 33]. Moreover, it has been used for comparing different univariate distributions [29]. Our use of the Stein's method is closely related to the dependency neighborhood method described in [38].

1.2 Paper outline

In Section 2 we clarify some notations that we use throughout the paper. Then we present our results under two different local dependence conditions in Section 3. In Section 4 and Section 5 we respectively apply our results to m-dependent random fields and to U-statistics. In Section 6 we apply our results to obtain non-uniform Berry–Esseen bounds with polynomial decay. In Section 7, we make an overview of our proof techniques. In Section 8 we present the main lemmas (notably Lemmas 8.7 and 8.8) and use them to prove the main result Theorem 3.1. Those lemmas and additional results are proved in Sections 9 to 11.

2 General notations

Notations concerning integers and sets   In this paper, we will write $\lceil x \rceil$ to denote the smallest integer that is bigger or equal to x and $\lfloor x \rfloor$ denotes the largest integer smaller or equal to x. We use $\mathbb{N}$ to denote the set of non-negative integers and let $\mathbb{N}_+$ be the set of positive integers. For any $n \in \mathbb{N}_+$, denote $[n] := \{\ell \in \mathbb{N}_+ : 1 \leq \ell \leq n\}$. Moreover, for a finite set B we denote by $|B|$ its cardinality.

Notations for sequences   Given a sequence (x_i) we will shorthand $x_{1:\ell} = (x_1, \ldots, x_\ell)$ and similarly for any subset $B \subseteq \mathbb{N}_+$ we denote $x_B := (x_i)_{i \in B}$.

Notations for functions   For any real valued functions $f(\,\cdot\,), g(\,\cdot\,) : \mathbb{N}_+ \to \mathbb{R}$, we write $f(n) \lesssim g(n)$ or $f(n) = \mathcal{O}(g(n))$ if there exists some constant C (with dependencies that are fixed in the contexts) and an integer $N > 0$ such that the inequality $f(n) \leq Cg(n)$ holds for all $n \geq N$. We further write $f(n) \asymp g(n)$ as shorthand for $f(n) \lesssim g(n)$ and $g(n) \lesssim f(n)$.

Notations for probability distributions   For a random variable X we write by $\mathcal{L}(X)$ the distribution of X.

3 Main theorems

Let $p \geq 1$ be a positive real number, we write $\omega := p + 1 - \lceil p \rceil \in [0, 1]$. We choose I to be an infinite index set and $(I_n)_{n=1}^\infty$ to be an increasing sequence of finite subsets of $I_1 \subseteq I_2 \subseteq \cdots \subsetneq I$ that satisfy $|I_n| \xrightarrow{n \to \infty} \infty$.

Let $\big(X_i^{(n)}\big)_{i \in I_n}$ be a triangular array of random variables, each row indexed by $i \in I_n$ ($n = 1, 2, \ldots$), we define W_n to be the following empirical average

$$W_n := \sigma_n^{-1} \sum_{i \in I_n} X_i^{(n)}, \quad \text{with } \sigma_n^2 := \mathrm{Var}\Big(\sum_{i \in I_n} X_i^{(n)}\Big).$$

Under the hypothesis that the random variables $\left(X_i^{(n)}\right)$ are locally dependent we will, in this section, bound the Wasserstein-p distance between W_n and its normal limit. The bound we obtain depends on the size of the index set I_n, the moments of the random variables and the structure of local dependence in question.

To formally state our conditions on the dependency structure of $(X_i^{(n)})$, we first define the notion of dependency neighborhoods similarly as in [38].

Given random variables $(Y_i)_{i \in I}$ and given $J \subseteq I$, we say that $N(J) \subset I$ is a **dependency neighborhood** of J if $\{Y_j : j \notin N(J)\}$ is independent of $\{Y_j : j \in J\}$. To state our theorem, we impose that such dependency neighborhoods can be defined for $\left(X_i^{(n)}\right)$. More formally, we assume that there is a sequence $(N_n(i_{1:q}))_q$ of subsets of I_n that satisfy the following conditions:

[LD-1]: For each $i_1 \in I_n$, the subset $N_n(i_1) \subseteq I_n$ is such that $\left\{X_j^{(n)} : j \notin N_n(i_1)\right\}$ is independent of $X_{i_1}^{(n)}$.

[LD-q] ($q \geq 2$): For each $i_1 \in I_n, i_2 \in N_n(i_1), \ldots, i_q \in N_n(i_{1:(q-1)})$, the subset $N_n(i_{1:q}) \subset I_n$ is such that $\left\{X_j^{(n)} : j \notin N_n(i_{1:q})\right\}$ is independent of $\left(X_{i_1}^{(n)}, \ldots, X_{i_q}^{(n)}\right)$.

We remark that the sequence of subsets $(N_n(i_{1:q}))_q$ is increasing, i.e., $N_n(i_{1:(q-1)}) \subseteq N_n(i_{1:q})$ in q; and that the neighborhoods $N_n(i_{1:q})$ are allowed to be different for different values of n–which reflects the triangular array structure of our problem. The condition of dependency neighborhoods here generalizes the one in [38] and was also adopted in [16], inspired by [5, 13]. [5] obtained a Wasserstein-1 bound under "decomposable" conditions similar to [LD-1] and [LD-2], and [13] showed a Berry–Esseen type result under slightly stronger assumptions for local dependence, while finally [16] obtained a Wasserstein-2 bound.

In order to define the remainder terms that will appear in our bounds, we introduce the following notions. Given $t \in \mathbb{N}_+$, and $\ell \in \mathbb{N}_+$ such that $t \geq 2$, we say that the tuple $(\eta_1, \eta_2, \ldots, \eta_\ell)$ is an **integer composition** of t if and only if $\eta_{1:\ell}$ are positive integers such that $\eta_1 + \eta_2 + \cdots + \eta_\ell = t$. We denote by $C(t)$ the set of all possible integer compositions

$$C(t) := \left\{\ell, \eta_{1:\ell} \in \mathbb{N}_+ : \textstyle\sum_{j=1}^\ell \eta_j = t\right\}.$$

Moreover, for any random variables $(Y_i)_{i=1}^t$, we define the order-t **compositional expectation** with respect to $\eta_{1:\ell}$ as

$$[\eta_1, \ldots, \eta_\ell] \triangleright (Y_1, \ldots, Y_t) := \mathbb{E}\left[Y_1 \cdots Y_{\eta_1}\right] \mathbb{E}\left[Y_{\eta_1+1} \cdots Y_{\eta_1+\eta_2}\right] \cdots \mathbb{E}\left[Y_{\eta_1+\cdots+\eta_{\ell-1}+1} \cdots Y_t\right].$$
(3.1)

Note that if $\eta_\ell = 1$, the last expectation reduces to $\mathbb{E}[Y_t]$. For any positive integer k and real value $\omega \in (0, 1]$, we define

$$
\begin{aligned}
R_{k,\omega,n} := &\sum_{(\ell,\eta_{1:\ell}) \in C^*(k+2)} \sum_{i_1 \in I_n} \sum_{i_2 \in N_n(i_1)} \cdots \sum_{i_{k+1} \in N_n(i_{1:k})} \\
&[\eta_1, \ldots, \eta_\ell] \triangleright \left(\left|X_{i_1}^{(n)}\right|, \ldots, \left|X_{i_{k+1}}^{(n)}\right|, \left(\sum_{i_{k+2} \in N_n(i_{1:(k+1)})} \left|X_{i_{k+2}}^{(n)}\right|\right)^\omega\right),
\end{aligned}
$$
(3.2)

where $C^*(k+2)$ is given by

$$C^*(t) := \left\{(\ell, \eta_{1:\ell}) \in C(t) : \eta_j \geq 2 \text{ for } 1 \leq j \leq \ell-1,\right\} \subseteq C(t).$$

The terms $(R_{k,\omega,n})$ are remainder terms that appear in our bound of the Wasserstein-p distance between W_n and its normal limit.

Theorem 3.1. *Let $\left(X_i^{(n)}\right)_{i \in I_n}$ be a triangular array of mean zero random variables and suppose that they satisfy [LD-1] to [LD-($\lceil p \rceil$+1)]. Let $\sigma_n^2 := \mathrm{Var}\left(\sum_{i \in I_n} X_i^{(n)}\right)$ and define $W_n := \sigma_n^{-1} \sum_{i \in I_n} X_i^{(n)}$. Further suppose for any $j \in \mathbb{N}_+$ such that $j \leq \lceil p \rceil - 1$, it holds that $R_{j,1,n} \overset{n \to \infty}{\longrightarrow} 0$ as $n \to \infty$. Then there exists an integer $N \in \mathbb{N}_+$ such that for all $n \geq N$, we have the following Wasserstein bounds:*

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq C_p \left(\sum_{j=1}^{\lceil p \rceil - 1} R_{j,1,n}^{1/j} + \sum_{j=1}^{\lceil p \rceil} R_{j,\omega,n}^{1/(j+\omega-1)} \right), \tag{3.3}$$

where $\omega = p + 1 - \lceil p \rceil$ and C_p is a constant that only depends on p.

Remark 3.2. We note that the condition that the remainder terms $R_{j,1,n}$ shrink to 0 for all $j \leq \lceil p \rceil - 1$ impose an implicit constraint on the size of the sets $N_n(i_{1:q})$.

In particular, for $p = 1, 2$ we have

$$\mathcal{W}_1(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq C_1 R_{1,1,n}, \tag{3.4}$$

$$\mathcal{W}_2(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq C_2 \left(R_{1,1,n} + R_{2,1,n}^{1/2} \right). \tag{3.5}$$

where the remainders are given by

$$R_{1,1,n} = \sigma_n^{-3} \sum_{i \in I_n} \sum_{j \in N_n(i)} \sum_{k \in N_n(i,j)} \left(\mathbb{E}\left[|X_i^{(n)} X_j^{(n)} X_k^{(n)}| \right] + \mathbb{E}\left[|X_i^{(n)} X_j^{(n)}| \right] \mathbb{E}\left[|X_k^{(n)}| \right] \right),$$

$$R_{2,1,n} = \sigma_n^{-4} \sum_{i \in I_n} \sum_{j \in N_n(i)} \sum_{k \in N_n(i,j)} \sum_{\ell \in N_n(i,j,k)} \left(\mathbb{E}\left[|X_i^{(n)} X_j^{(n)} X_k^{(n)} X_\ell^{(n)}| \right] \right.$$
$$\left. + \mathbb{E}\left[|X_i^{(n)} X_j^{(n)} X_k^{(n)}| \right] \mathbb{E}\left[|X_\ell^{(n)}| \right] + \mathbb{E}\left[|X_i^{(n)} X_j^{(n)}| \right] \mathbb{E}\left[|X_k^{(n)} X_\ell^{(n)}| \right] \right).$$

Note that (3.4) was proven by [5] and (3.5) is a corollary of Theorem 2.1, [16]. The bound (3.3) with an integer p was also proposed as a conjecture in [16]. As p grows, the right-hand side of (3.3) becomes more and more complicated, which suggests the necessity of new assumptions in order to obtain a simplified result. We further remark that the choice of $N_n(i_{1:q})$ might not be unique (even if we require that it has the smallest cardinality among all possible index sets that fulfill the assumption [LD-q]). Therefore, to be able to obtain more interpretable upper-bounds for the remainder terms $(R_{j,\omega,n})$, we impose a slightly stronger assumption on the dependence structure:

[LD*]: We suppose that there exists a graph $G_n = (V_n, E_n)$, with $V_n := I_n$ being the vertex set and E_n being the edge set, such that for any two disjoint subsets $J_1, J_2 \subseteq I_n$ if there is no edge between J_1 and J_2, then $\left\{ X_j^{(n)} : j \in J_1 \right\}$ is independent of $\left\{ X_j^{(n)} : j \in J_2 \right\}$.

Introduced by [34] the graph G_n defined above is known as the **dependency graph** and was later adopted in [25, 3, 38]. Please refer to [19] for a detailed discussion.

If [LD*] is satisfied, for any subset $J \subseteq I_n$, we define $N_n(J)$ to be the set of vertices in the neighborhood of $J \subseteq I_n$ in the graph G. To be precise, this is

$$N_n(J) := J \cup \{i \in I_n : e(i,j) \in E_n \text{ for some } j \in J\},$$

where $e(i,j)$ denotes an edge between the vertices i and j. To simplify the notations, we further denote $N_n(J)$ by $N_n(i_{1:q})$ if $J = \{i_1, \ldots, i_q\}$ for any $1 \leq q \leq \lceil p \rceil + 1$. Then $(N_n(i_{1:q}))$ not only satisfies [LD-1] to [LD-($\lceil p \rceil$+1)], but has the following properties as well:

(a) $N_n(i_{1:q}) = N_n(i_{\pi(1)}, \ldots, i_{\pi(q)})$ for any permutation π on $\{1, \ldots, q\}$;

(b) $i_q \in N_n(i_{1:(q-1)}) \Leftrightarrow i_1 \in N_n(i_{2:q})$.

We point out that by definition of the dependency graph even if $\{X_j^{(n)} : j \in J_1\}$ is independent of $\{X_j^{(n)} : j \in J_2\}$, there can still be edges between the vertex sets J_1 and J_2. In fact, there might not exist G_n such that there is no edge between J_1 and J_2 as long as $\{X_j^{(n)} : j \in J_1\}$ is independent of $\{X_j^{(n)} : j \in J_2\}$ since pairwise independence does not imply joint dependence.

The condition [LD*] provides us with a tractable bound on $R_{k,\omega,n}$, which is applicable in most of the commonly encountered settings, including m-dependent random fields and U-statistics.

Proposition 3.3. *Given $M \in \mathbb{N}_+$ and real number $\omega \in (0,1]$, suppose that $(X_i^{(n)})_{i \in I_n}$ satisfies [LD*] and that the cardinality of $N_n(i_{1:(k+1)})$ is upper-bounded by $M < \infty$ for any $i_1, \ldots, i_{k+1} \in I_n$. Then there exists a constant $C_{k+\omega}$ only depending on $k + \omega$ such that*

$$R_{k,\omega,n} \leq C_{k+\omega} M^{k+\omega} \sum_{i \in I_n} \sigma_n^{-(k+1+\omega)} \mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big].$$

We remark that the upper bound on $(R_{k,\omega,n})$ depends on the moments of the random variables $(X_i^{(n)})$ and the maximum size of the dependency neighborhoods. The results of Proposition 3.3 can be used to propose a more interpretable upper bound for the Wasserstein-p distance.

Theorem 3.4. *Suppose that $(X_i^{(n)})$ is a triangular array of mean zero random variables satisfying [LD*], and that the cardinality of index set $N_n(i_{1:(\lceil p \rceil+1)})$ is upper-bounded by $M_n < \infty$ for any $i_1, \ldots, i_{\lceil p \rceil+1} \in I_n$. Furthermore, assume that*

$$M_n^{1+\omega} \sigma_n^{-(\omega+2)} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{\omega+2}\big] \to 0, \quad M_n^{p+1} \sigma_n^{-(p+2)} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big] \to 0.$$

Then there is N such that for all $n \geq N$ we have

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1))$$
$$\leq C_p \Big(M_n^{1+\omega} \sigma_n^{-(\omega+2)} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{\omega+2}\big]\Big)^{1/\omega} + C_p \Big(M_n^{p+1} \sigma_n^{-(p+2)} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big]\Big)^{1/p},$$
(3.6)

for some constant C_p that only depends on p.

We notably remark that if the moments are nicely behaved in the sense that

$$B_1 := \sup_{i \in I_n, n \in \mathbb{N}_+} \frac{\sqrt{|I_n|} \cdot \|X_i^{(n)}\|_{p+2}}{\sigma_n} < \infty,$$

and that the size of the dependency neighborhood are universally bounded, i.e.,

$$B_2 := \sup_{i_{1:(\lceil p \rceil+1)} \in I_n, n \in \mathbb{N}_+} \big|N_n(i_{1:(\lceil p \rceil+1)})\big| < \infty,$$

then there is a constant K_p that only depends on B_1, B_2 and $p \geq 1$ such that for n large enough we have

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq \frac{K_p}{\sqrt{|I_n|}}.$$

The rate of convergence matches the known rate for independent random variables (see [8]).

4 Results for m-dependent random fields

Let $d \in \mathbb{N}_+$ be a positive integer, in this section, we study d-dimensional random fields.

Definition 4.1 (m-Dependent Random Field). *A random field $(X_i)_{i \in T}$ on $T \subseteq \mathbb{Z}^d$ is m-dependent if and only if for any subsets $U_1, U_2 \subseteq \mathbb{Z}^d$, the random variables $(X_{i_1})_{i_1 \in U_1 \cap T}$ and $(X_{i_2})_{i_2 \in U_2 \cap T}$ are independent whenever $\|i_1 - i_2\| > m$ for all $i_1 \in U_1$ and $i_2 \in U_2$.*

Here $\|\cdot\|$ denotes the maximum norm on $\mathbb{Z}^d$, that is $\|\boldsymbol{z}\| = \max_{1 \le j \le d} |z_j|$ for $\boldsymbol{z} = (z_1, \ldots, z_d)$.

Now we consider an increasing sequence $T_1 \subseteq T_2 \subseteq \cdots$ of finite subsets of $\mathbb{Z}^d$ that satisfy $|T_n| \xrightarrow{n \to \infty} \infty$. We have the following result as a corollary of Theorem 3.4.

Corollary 4.2. *Let $p \in \mathbb{N}_+$ and $m \in \mathbb{N}_+$ be positives integer. Suppose that $\left(X_i^{(n)}\right)$ is a triangular array where each row is an m-dependent random field indexed by finite subsets $T_n \subseteq \mathbb{Z}^d$ such that $|T_n| \xrightarrow{n \to \infty} \infty$. Let $\sigma_n^2 := \mathrm{Var}\left(\sum_{i \in T_n} X_i^{(n)}\right)$ and define $W_n := \sigma_n^{-1} \sum_{i \in T_n} X_i^{(n)}$. Further suppose that $\mathbb{E}\left[X_i^{(n)}\right] = 0$ for any $i \in T_n$ and that the following conditions hold:*

- *Moment condition: $\sigma_n^{-(p+2)} \sum_{i \in T_n} \mathbb{E}\left[\left|X_i^{(n)}\right|^{p+2}\right] \to 0 \quad$ as $n \to \infty$;*

- *Non-degeneracy condition: $\limsup_n \sigma_n^{-2} \sum_{i \in T_n} \mathbb{E}\left[\left|X_i^{(n)}\right|^2\right] \le M < \infty$ for some $M \ge 1$.*

Then for n large enough, we have

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \le C_{p,d} m^{\frac{(1+\omega)d}{\omega}} M^{\frac{p-\omega}{p\omega}} \sigma_n^{-\frac{p+2}{p}} \left(\sum_{i \in T_n} \mathbb{E}\left[\left|X_i^{(n)}\right|^{p+2}\right]\right)^{1/p}, \qquad (4.1)$$

where $C_{p,d}$ only depends on p and d.

In particular, for a triangular array of m-dependent stationary random fields, suppose that we have $\sup_n \mathbb{E}\left[\left|X_i^{(n)}\right|^{p+2}\right] < \infty$, and that the non-degeneracy condition $\liminf_n \sigma_n^2 / |T_n| > 0$ holds. Then we have

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) = \mathcal{O}(|T_n|^{-1/2}).$$

5 Application to U-statistics

Definition 5.1 (U-Statistic). *Let $(X_i)_{i=1}^n$ be a sequence of i.i.d. random variables. Fix $m \in \mathbb{N}_+$ such that $m \ge 2$. Let $h : \mathbb{R}^m \to \mathbb{R}$ be a fixed Borel-measurable function. The Hoeffding U-statistic is defined as*

$$\sum_{1 \le i_1 \le \cdots \le i_m \le n} h\left(X_{i_1}, \ldots, X_{i_m}\right).$$

Corollary 5.2. *Given $p \ge 1$, suppose that the U-statistic of an i.i.d. sequence $(X_i)_{i=1}^n$ induced by a symmetric function $h : \mathbb{R}^m \to \mathbb{R}$ satisfies the following conditions*

- *Mean zero: $\mathbb{E}\left[h(X_1, \ldots, X_m)\right] = 0$;*

- *Moment condition: $\mathbb{E}\left[\left|h(X_1, \ldots, X_m)\right|^{p+2}\right] < \infty$;*

- *Non-degeneracy condition: $\mathbb{E}[g(X_1)^2] > 0$, where $g(x) := \mathbb{E}\left[h(X_1, \ldots, X_m) \mid X_1 = x\right]$.*

If we let

$$W_n := \frac{1}{\sigma_n} \sum_{1 \leq i_1 \leq \cdots \leq i_m \leq n} h\big(X_{i_1}, \ldots, X_{i_m}\big),$$

where

$$\sigma_n^2 := \mathrm{Var}\left(\sum_{1 \leq i_1 \leq \cdots \leq i_m \leq n} h\big(X_{i_1}, \ldots, X_{i_m}\big)\right),$$

the following Wasserstein bound holds:

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) = \mathcal{O}(n^{-1/2}).$$

6 Application to non-uniform berry–Esseen bounds

In this section, we show a specific application of our results to non-uniform Berry–Esseen bounds with polynomial decay of any order. Mirroring the classical literature, [13] established Berry–Esseen bounds for locally dependent random variables. Notably, their Theorem 2.4 showed that if the random variables $(X_i^{(n)})$ satisfy some boundedness condition on the dependency neighborhoods, then there is a constant $C > 0$ such that

$$\sup_t \big|\mathbb{P}(W_n \geq t) - \Phi^c(t)\big| \leq C \sum_{i \in I_n} \big\|X_i^{(n)}\big\|_3^3 / \sigma_n^3,$$

where $\Phi^c(t) := \mathbb{P}(Z \geq t)$ with $Z \sim \mathcal{N}(0,1)$. This extends the classical Berry–Esseen bound to locally dependent random variables, and can potentially be used to construct Kolmogorov–Smirnov tests under local dependence in nonparametric inference. However, one of the drawbacks of this inequality is that it does not depend on t. One would imagine that for large t we could find tighter bounds for $\big|\mathbb{P}(W_n \geq t) - \Phi^c(t)\big|$. Non-uniform Berry–Esseen bounds establish this. Notably [13] (Theorem 2.5) showed that under the above conditions, there exists some universal constant C' such that

$$\big|\mathbb{P}(W_n \geq t) - \Phi^c(t)\big| \leq \frac{C'}{1 + |t|^3} \sum_{i \in I_n} \big\|X_i^{(n)}\big\|_3^3 / \sigma_n^3, \quad \forall t \in \mathbb{R}.$$

This bound does decrease as $|t|$ increases and does so at a rate of $|t|^{-3}$. However one would expect that this rate could be tightened if additional assumptions were made about the moments of $(X_i^{(n)})$. If the random variables admit some exponential moments then [30] demonstrated that locally dependent random variables admit moderate deviation inequalities. In this section, we show how $\mathcal{W}_p$ bound can help us obtain bounds that decrease polynomially fast in t at a small price in its dependence on $|I_n|$, and do so without assuming infinite moments.

Theorem 6.1. *We assume that the conditions of Theorem 3.4 are satisfied. There is a constant $C > 0$ such that for all $\beta > 0$ and $t > 0$ satisfying*

$$(\sqrt{2\pi}p)^{\frac{1}{p+1}}\Big(1 - \frac{\sqrt{2\beta \log t}}{t}\Big)t^{1 - \frac{\beta}{p+1}} \geq \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)),$$

we have

$$-\frac{C}{t}\varphi\Big(t\Big(1 - \frac{1}{p+1}\Big)\Big) \leq \frac{\mathbb{P}(W_n \geq t) - \Phi^c(t)}{\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1))^{1 - \frac{1}{p+1}}} \leq \frac{C}{t^{1 + \beta\big(1 - \frac{1}{p+1}\big)}},$$

where φ is the density function of $\mathcal{N}(0,1)$.

We can see from this result that the quantity $|\mathbb{P}(W_n \geq t) - \Phi^c(t)|$ decays in both t and n. Notably given any $p, r \geq 1$ assuming that the $(p+2)$-th moments of X_i's and the dependency neighborhoods are bounded in the sense that $\sup_{n \in \mathbb{N}^+, i_{1:(p+1)} \in I_n} |N_n(i_{1:(p+1)})| < \infty$, we have $|\mathbb{P}(W_n \geq t) - \Phi^c(t)| = o\left(t^{-r}|I_n|^{-\frac{p}{2(p+1)}}\right)$ for t and n large enough. In particular, for $p \in \mathbb{N}^+$ Theorems 3.4 and 6.1 imply the uniform Berry–Esseen bound by taking the supremum over t:

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(W_n \geq t) - \Phi^c(t)| \leq C\left(\sum_{i \in I_n} \|X_i^{(n)}\|_3^3 / \sigma_n^3 + \sum_{i \in I_n} \|X_i^{(n)}\|_{p+2}^{1+2/p} / \sigma_n^{1+2/p}\right).$$

Note that it recovers the uniform Berry–Esseen bound in [13] with $p = 1$.

7 Overview of the proofs

The key idea of our proofs is to approximate the sum of weakly dependent random variables $\left(X_i^{(n)}\right)_{i \in I_n}$ by the empirical average of q_n *i.i.d.* random variables $\xi_1^{(n)}, \ldots, \xi_{q_n}^{(n)}$ which we denote $V_n := \frac{1}{\sqrt{q_n}} \sum_{i=1}^{q_n} \xi_i^{(n)}$. More specifically we aim for the Wasserstein-p distance between them $\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{L}(V_n))$ to be as small as possible. To establish the desired result we then exploit the triangle inequality that guarantees that

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{L}(V_n)) + \mathcal{W}_p(\mathcal{L}(V_n), \mathcal{N}(0,1)),$$

and we use previously known bounds for $\mathcal{W}_p(\mathcal{L}(V_n), \mathcal{N}(0,1))$ (Lemma 8.6).

To be able to show that such random variables $\xi_1^{(n)}, \ldots, \xi_{q_n}^{(n)}$ exist, we first show (Lemma 8.8) that as long as the third and higher-order cumulants of W_n decay then there exist integers (q_n) and *i.i.d.* random variables such that the first k ($k \in \mathbb{N}_+$) cumulants of

$$V_n := \frac{1}{\sqrt{q_n}} \sum_{i=1}^{q_n} \xi_i^{(n)}$$

matches those of W_n for n large enough. The decay of the cumulants can be proven to hold by exploiting the local dependence assumptions (see Corollary 9.6).

As a reminder, our goal is to establish that the Wasserstein distance $\mathcal{W}_p(\mathcal{L}(V_n), \mathcal{L}(W_n))$ is small. We relate this to the cumulants thanks to the fact that the Wasserstein-p distance can be upper-bounded by integral probability metrics (Lemma 8.4) and the well-known Stein equation. Indeed for *i.i.d.* random variables $\left(\xi_i^{(n)}\right)_{i=1}^{q_n}$, [4] showed that the following approximation holds (restated in Lemma 8.5)

$$\mathbb{E}[h(V_n)] - \mathcal{N}h = \mathbb{E}[f_h'(V_n) - V_n f_h(V_n)]$$
$$= \sum_{(r, s_{1:r}) \in \Gamma(\lceil p \rceil - 1)} (-1)^r \prod_{j=1}^r \frac{\kappa_{s_j+2}(V_n)}{(s_j+1)!} \mathcal{N}\left[\prod_{j=1}^r (\partial^{s_j+1}\Theta) h\right] + \text{ Remainders}, \tag{7.1}$$

where f_h is the solution of the Stein equation (8.2) and $\kappa_j(\cdot)$ denotes the j-th cumulant of a random variable. (All the other notations in (7.1) will be made clear in Section 8.) We show that we can obtain similar expansions for $\mathbb{E}[f'(W_n) - W_n f(W_n)]$ (see Lemma 8.7):

$$\mathbb{E}[h(W_n)] - \mathcal{N}h = \mathbb{E}[f_h'(W_n) - W_n f_h(W_n)]$$
$$= \sum_{(r, s_{1:r}) \in \Gamma(\lceil p \rceil - 1)} (-1)^r \prod_{j=1}^r \frac{\kappa_{s_j+2}(W_n)}{(s_j+1)!} \mathcal{N}\left[\prod_{j=1}^r (\partial^{s_j+1}\Theta) h\right] + \text{ Remainders}. \tag{7.2}$$

As mentioned in the previous paragraph, q_n and $\xi_i^{(n)}$ can be chosen to be such that $\kappa_j(V_n) = \kappa_j(W_n)$ for $j = 1, \ldots, \lceil p \rceil + 1$. Thus, by taking the difference of (7.1) and (1.1),

we get an upper bound on $\big|\mathbb{E}[h(W_n)] - \mathbb{E}[h(V_n)]\big|$ for a large class of function h. As shown in Lemma 8.4, this allows us to obtain an upper bound of the Wasserstein-p distance between $\mathcal{L}(W_n)$ and $\mathcal{L}(V_n)$ for a general $p \geq 1$. The desired result is therefore implied by the triangle inequality of the Wasserstein-p distance

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{L}(V_n)) + \mathcal{W}_p(\mathcal{L}(V_n), \mathcal{N}(0,1)),$$

and the already known Wasserstein-p bounds for *i.i.d.* random variables (Lemma 8.6).

To be able to show that (1.1) holds, we develop new techniques to obtain such expansions, which will be carefully elaborated and discussed in Section 9.

8 Adapting Stein's method for Wasserstein-p bounds

In this section, we provide the proofs of Theorems 3.1 and 3.4 using Stein's method. We first introduce some background definitions and lemmas before showing the proofs of the main theorems.

8.1 Preliminaries and notations

Definition 8.1 (Hölder Space). *For any $k \in \mathbb{N}$ and real number $\omega \in (0, 1]$, the Hölder space $\mathcal{C}^{k,\omega}(\mathbb{R})$ is defined as the class of k-times continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that the k-times derivative of f is ω-Hölder continuous, i.e.,*

$$|f|_{k,\omega} := \sup_{x \neq y \in \mathbb{R}} \frac{|\partial^k f(x) - \partial^k f(y)|}{|x - y|^\omega} < \infty,$$

where ∂ denotes the differential operator. Here ω is called the Hölder exponent and $|f|_{k,\omega}$ is called the Hölder coefficient.

Using the notions of Hölder spaces, we define the Zolotarev's ideal metrics, which are related to the Wasserstein-p distances via Lemma 8.4.

Definition 8.2 (Zolotarev Distance). *Suppose μ and ν are two probability distributions on $\mathbb{R}$. For any $p > 0$ and $\omega := p + 1 - \lceil p \rceil \in (0, 1]$, the Zolotarev-$p$ distance between μ and ν is defined by*

$$\mathcal{Z}_p(\mu, \nu) := \sup_{f \in \Lambda_p} \left(\int_{\mathbb{R}} f(x)\, \mathrm{d}\mu(x) - \int_{\mathbb{R}} f(x)\, \mathrm{d}\nu(x) \right),$$

where $\Lambda_p := \{f \in \mathcal{C}^{\lceil p \rceil - 1, \omega}(\mathbb{R}) : |f|_{\lceil p \rceil - 1, \omega} \leq 1\}$

We will see in Lemma 8.4 how the Zolotarev distance can be used to obtain $\mathcal{W}_p(\cdot, \cdot)$ rates. To bound $\mathcal{Z}_p(\cdot, \cdot)$ we rely on the Stein's method which was introduced by [41] in order to prove the central limit theorem for dependent data. It has been widely adapted to all kinds of normal approximation problems. See [38] for a detailed exposition.

Stein equation and its solutions

Let $Z \sim \mathcal{N}(0,1)$ be a standard normal random variable. For any measurable function $h : \mathbb{R} \to \mathbb{R}$, if $h(Z) \in \mathcal{L}^1(\mathbb{R})$, we write $\mathcal{N}h := \mathbb{E}[h(Z)]$. Thus, $h(Z) \in \mathcal{L}^1(\mathbb{R})$ if and only if $\mathcal{N}|h| < \infty$. Moreover, we define $f_h(\cdot)$ by

$$f_h(w) := \int_{-\infty}^{w} e^{(w^2 - t^2)/2}(h(t) - \mathcal{N}h)\, \mathrm{d}t = -\int_{w}^{\infty} e^{(w^2 - t^2)/2}(h(t) - \mathcal{N}h)\, \mathrm{d}t. \qquad (8.1)$$

We remark that $f_h(\cdot)$ is a *solution* of the Stein equation meaning that it satisfies

$$f_h'(w) - w f_h(w) = h(w) - \mathcal{N}h, \qquad \forall w \in \mathbb{R}. \qquad (8.2)$$

Bounding $\left|\mathbb{E}(f_h'(W_n) - W_n f_h(W_n))\right|$ therefore allows to control $\left|\mathbb{E}(h(W_n)) - \mathcal{N}h\right|$. If we do this for a large class of functions h we can therefore upper-bound the Zolotarev distance between $\mathcal{L}(W_n)$ and a normal distribution. This is the key idea behind the Stein's method. For notational convenience, we denote by Θ the operator that maps h to f_h for any h such that $\mathcal{N}|h| < \infty$, i.e.,

$$\Theta h = f_h.$$

Note that $\Theta h(\,\cdot\,)$ is a function. If $h \in \Lambda_p$, then we see in Lemma 8.3 that Θh can be bounded.

Lemma 8.3 (Part of Lemma 6 of [4])**.** *For any $p > 0$, let $h \in \Lambda_p$ be as defined in Definition 8.2. Then $\Theta h = f_h$ in (8.1) is a solution to (8.2). Moreover, $\Theta h \in \mathcal{C}^{\lceil p \rceil - 1, \omega}(\mathbb{R}) \cap \mathcal{C}^{\lceil p \rceil, \omega}(\mathbb{R})$ and the Hölder coefficients $|\Theta h|_{\lceil p \rceil - 1, \omega}$ and $|\Theta h|_{\lceil p \rceil, \omega}$ are bounded by some constant only depending on p.*

8.2 Key lemmas

First, we present an important result that states that the Wasserstein-p distance can be controlled in terms of the Zolotarev distance.

Lemma 8.4 (Theorem 3.1 of [36])**.** *For any $p \geq 1$, there exists a positive constant C_p, such that for any pair of distributions μ, ν on $\mathbb{R}$ with finite absolute moments of order p such that*

$$\mathcal{W}_p(\mu, \nu) \leq C_p \big(\mathcal{Z}_p(\mu, \nu)\big)^{1/p}.$$

In particular, $\mathcal{W}_1(\mu, \nu) = \mathcal{Z}_1(\mu, \nu)$ by Kantorovich–Rubinstein duality.

In the next two lemmas, we present already-known results for the normal approximation of sums of independent random variables. Firstly Lemma 8.5 provides an expansion for the difference between $\mathbb{E}[h(S_n)]$, where S_n is an empirical average and $\mathcal{N}h$. This lemma will allow us to relate the Zolotarev distance to the cummulants.

Lemma 8.5 (Theorem 1 of [4])**.** *For any $p > 0$, let $h \in \Lambda_p$ and $S_n := \sum_{i=1}^n X_i$ where $\{X_1, \ldots, X_n\}$ are independent, with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[S_n^2] = 1$. Then it follows that*

$$\mathbb{E}[h(S_n)] - \mathcal{N}h =$$

$$\sum_{(r, s_{1:r}) \in \Gamma(\lceil p \rceil - 1)} (-1)^r \prod_{j=1}^r \frac{\kappa_{s_j + 2}(S_n)}{(s_j + 1)!} \mathcal{N}\left[\prod_{j=1}^r (\partial^{s_j + 1} \Theta)\, h\right] + \mathcal{O}\left(\sum_{i=1}^n \mathbb{E}[|X_i|^{p+2}]\right), \tag{8.3}$$

where the first sum is over $\Gamma(\lceil p \rceil - 1) := \big\{r, s_{1:r} \in \mathbb{N}_+ : \sum_{j=1}^r s_j \leq \lceil p \rceil - 1\big\}$.

Note that there is a slight abuse of notation in (8.3). The last $\prod$ indicates the composition of the operators in the parentheses rather than the product.

Secondly Lemma 8.5 gives an upper bound on the Wasserstein distance between the distribution of this empirical average S_n, and the standard normal distribution. This lemma will guarantee that if an approximation of W_n by a sum of independent random variables V_n can be obtained then V_n is approximately normally distributed.

Lemma 8.6 (Theorem 1.1 of [8])**.** *For any $p \geq 1$, let $S_n := \sum_{i=1}^n X_i$ where $\{X_1, \ldots, X_n\}$ are independent and satisfy that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[S_n^2] = 1$. Then it follows that*

$$\mathcal{W}_p(\mathcal{L}(S_n), \mathcal{N}(0,1)) \leq C_p \left(\sum_{i=1}^n \mathbb{E}[|X_i|^{p+2}]\right)^{1/p}, \tag{8.4}$$

where C_p continuously depends on p.

We now introduce two new lemmas crucial in the proof of Theorem 3.1. They will be proven in Section 9 and Section 10. The first lemma generalizes Lemma 8.6 to the dependent setting.

Lemma 8.7 (Local Expansion). *Suppose that $\left(X_i^{(n)}\right)_{i \in I_n}$ is a triangular array of random variables with dependency neighborhoods satisfying the local dependence conditions [LD-1] to [LD-($\lceil p \rceil$+1)]. Let $W_n := \sum_{i \in I_n} X_i^{(n)}$ with $\mathbb{E}\left[X_i^{(n)}\right] = 0$, $\mathbb{E}[W_n^2] = 1$. Then for any $p > 0$ and $h \in \Lambda_p$, we have*

$$
\begin{aligned}
\mathbb{E}[h(W_n)] - \mathcal{N}h &= \sum_{(r,s_{1:r}) \in \Gamma(\lceil p \rceil - 1)} (-1)^r \prod_{j=1}^{r} \frac{\kappa_{s_j+2}(W_n)}{(s_j+1)!} \mathcal{N}\left[\prod_{j=1}^{r} (\partial^{s_j+1}\Theta)\, h\right] \\
&\quad + \mathcal{O}\left(\sum_{j=1}^{\lceil p \rceil - 1} R_{j,1,n}^{p/j} + \sum_{j=1}^{\lceil p \rceil} R_{j,\omega,n}^{p/(j+\omega-1)}\right),
\end{aligned}
\tag{8.5}
$$

where the first sum is over $\Gamma(\lceil p \rceil - 1) := \left\{r, s_{1:r} \in \mathbb{N}_+ : \sum_{j=1}^{r} s_j \leq \lceil p \rceil - 1\right\}$.

We can see that Lemmas 8.5 and 8.7 look quite similar to one another with the only differences being the dependence structures of $\left(X_i^{(n)}\right)$ and the remainder terms in the expansions. This similarity inspires the proof of Theorem 3.1. To illustrate this, imagine that there would exist some *i.i.d.* random variables $\left(\xi_i^{(n)}\right)_{i=1}^{q_n}$ and a large sample size q_n such that the first $\lceil p \rceil + 1$ cumulants of $V_n := q_n^{-1/2} \sum_{i=1}^{q_n} \xi_i^{(n)}$ match with those of W_n, then the expansion (8.5) and in (8.3) would be almost identical, and the difference between those would be controlled by the remainder terms $(R_{j,1,n})$ and $(R_{j,\omega,n})$. If those remainder terms are small then we could exploit the asymptotic normality of V_n to obtain the asymptotic normality of W_n. We show that such a sequence exists when $|I_n|$ is large.

Lemma 8.8 (Cumulant Matching). *Let $p \geq 1$ and $k := \lceil p \rceil$. If $p > 1$, let $\left(u_j^{(n)}\right)_{j=1}^{k-1}$ be a sequence of real numbers. Suppose that for any $j = 1, \ldots, k-1$, we have $u_j^{(n)} \to 0$ as $n \to \infty$. Then there exist constants C_p, C_p' only depending on p and a positive value $N > 0$ (that might depend on $\left(u_j^{(n)}\right)$) such that for any $n > N$, there exists $q_n \in \mathbb{N}_+$ and a random variable $\xi^{(n)}$ such that*

(a) $\mathbb{E}[\xi^{(n)}] = 0, \quad \mathbb{E}[(\xi^{(n)})^2] = 1$;

(b) $\kappa_{j+2}(\xi^{(n)}) = q_n^{j/2} u_j^{(n)}$ *for* $j = 1, \ldots, k-1$;

(c) *Either* $\max_{1 \leq j \leq k-1}\left|\kappa_{j+2}(\xi^{(n)})\right| = 0$ *or* $\max_{1 \leq j \leq k-1}\left|\kappa_{j+2}(\xi^{(n)})\right| \geq C_p > 0$;

(d) $\mathbb{E}[|\xi^{(n)}|^{p+2}] \leq C_p'$.

Furthermore, q_n can be chosen to be such that $q_n \to \infty$ as $|I| \to \infty$.

We note that the condition that $u_j^{(n)} \to 0$ as $n \to \infty$ is crucial. Lemma 8.8 is an asymptotic statement in the sense that for a given $n \leq N$, q_n and $\xi^{(n)}$ might not exist.

Intuitively, Lemma 8.8a and Lemma 8.8b determines the cumulants of $\xi^{(n)}$ and relates them to the cumulants of W_n. Lemma 8.8c requires that the maximum $\max_{1 \leq j \leq k}\left|\kappa_{j+2}(\xi^{(n)})\right|$ is either 0 or bounded away from 0 as n grows. And Lemma 8.8d indicates that the $(p+2)$-th absolute moment is upper-bounded.

8.3 Proof of Theorem 3.1

The proof of Theorem 3.1 works in three stages:

1. Using Lemma 8.8 we find a sequence of *i.i.d.* random variables $\left(\xi_\ell^{(n)}\right)_\ell$ and a sample size q_n such that the first $k+1$ cumulants of W_n match the first $k+1$ cumulants of $V_n := q_n^{-1/2} \sum_{i=1}^{q_n} \xi_i^{(n)}$;

2. Using Lemma 8.4 we remark that we can bound the Wasserstein distance between the distributions of W_n and an empirical average, V_n, of *i.i.d.* observations in terms

of $\left|\mathbb{E}[h(W_n)] - \mathbb{E}[h(V_n)]\right|$ for a large class of functions h. We do so by exploiting Lemmas 8.5 and 8.7;

3. We remark that Lemma 8.6 provides us with the bound on the Wasserstein distance between the distribution of V_n and the standard normal.

Then Theorem 3.1 follows from the triangle inequality of the Wasserstein metric:

$$\mathcal{W}_p(W_n, \mathcal{N}(0,1)) \leq \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{L}(V_n)) + \mathcal{W}_p(\mathcal{L}(V_n), \mathcal{N}(0,1)).$$

Proof of Theorem 3.1. Without loss of generality, we assume $\sigma_n = 1$ and denote $W_n := \sum_{i \in I_n} X_i^{(n)}$.

Firstly, we remark that according to Corollary 9.6, for all $1 \leq j \leq k-1$ we have $\left|\kappa_{j+2}(W_n)\right| \lesssim R_{j,1,n}$. Moreover, by assumption we have $R_{j,1,n} \to 0$ as $n \to \infty$. Therefore, $\left|\kappa_{j+2}(W_n)\right| \to 0$ as $n \to \infty$ and the assumptions of Lemma 8.8 hold, which implies that there exist constants C_p and C_p' such that for any n large enough there are positive integers (q_n) and random variables $(\xi^{(n)})$ such that

(a) $\mathbb{E}[\xi^{(n)}] = 0, \quad \mathbb{E}[(\xi^{(n)})^2] = 1$;

(b) $\kappa_{j+2}(\xi^{(n)}) = q_n^{j/2} \kappa_{j+2}(W_n)$ for $j = 1, \dots, k-1$;

(c) Either $\max_{1 \leq j \leq k-1} \left|\kappa_{j+2}(\xi^{(n)})\right| = 0$ or $\max_{1 \leq j \leq k-1} \left|\kappa_{j+2}(\xi^{(n)})\right| \geq C_p > 0$;

(d) $\mathbb{E}[|\xi^{(n)}|^{p+2}] \leq C_p'$.

Furthermore, we know that (q_n) satisfies that $q_n \to \infty$ as $n \to \infty$.

As presented in the proof sketch we will use this to bound the distance between the distribution of W_n to the one of an empirical average of at least q_n *i.i.d.* random variables. Note that when $\max_{1 \leq j \leq k-1} |\kappa_{j+2}(\xi^{(n)})| > 0$ then we can obtain (by combining items (b) and (c)) a lower bound on q_n which will be crucial in our arguments as it will allow us to control the distance between this empirical average and its normal limit. When $\kappa_3(W_n) = \cdots = \kappa_{k+1}(W_n) = 0$, such a lower bound on q_n cannot be obtained in a similar way. Thus, we introduce an alternative sequence $(\widetilde{q}_n)$ by setting $\widetilde{q}_n := |I_n|^{2(p+1)/p} \vee q_n$ if $\kappa_3(W_n) = \cdots = \kappa_{k+1}(W_n) = 0$, and $\widetilde{q}_n := q_n$ otherwise. We remark that the sequence $(\widetilde{q}_n)$ still respects $\widetilde{q}_n \to \infty$ as $n \to \infty$.

Let $\xi_1^{(n)}, \dots, \xi_{\widetilde{q}_n}^{(n)}$ be *i.i.d.* copies of $\xi^{(n)}$. Define $V_n := \widetilde{q}_n^{-1/2} \sum_{i=1}^{\widetilde{q}_n} \xi_i^{(n)}$. By construction, for any $j \in \mathbb{N}_+$ such that $j \leq k-1 = \lceil p \rceil - 1$ we have

$$\kappa_{j+2}(V_n) \overset{(*)}{=} \widetilde{q}_n^{-(j+2)/2} \sum_{i=1}^{\widetilde{q}_n} \kappa_{j+2}(\xi_i^{(n)}) = \widetilde{q}_n^{-j/2} \kappa_{j+2}(\xi^{(n)}) = \kappa_{j+2}(W_n).$$

Here in $(*)$ we have used the fact that cumulants are cumulative for independent random variables, which is directly implied by their definition. For more details on this, please refer to [31].

Thus, by Lemma 8.5 and Lemma 8.7, for any $h \in \Lambda_p$ we have

$$\left|\mathbb{E}[h(W_n)] - \mathbb{E}[h(V_n)]\right| \lesssim \sum_{j=1}^{k-1} R_{j,1,n}^{p/j} + \sum_{j=1}^{k} R_{j,\omega,n}^{p/(j+\omega-1)} + \widetilde{q}_n^{-(p+2)/2} \sum_{i=1}^{\widetilde{q}_n} \mathbb{E}[|\xi_i^{(n)}|^{p+2}]. \quad (8.6)$$

To be able to have this upper bound not depend on $\xi_i^{(n)}$ we will upper-bound

$$\widetilde{q}_n^{-(p+2)/2} \sum_{i=1}^{\widetilde{q}_n} \mathbb{E}[|\xi_i^{(n)}|^{p+2}]$$

in terms of the remainders $(R_{j,1,n})$ and $(R_{j,\omega,n})$. To do so we use the lower bounds on $(\widetilde{q}_n)$ implied by the specific form we chose.

If $\max_{1 \leq j \leq k-1} |\kappa_{j+2}(W_n)| > 0$, item (c) implies that

$$C_p \leq \max_{1 \leq j \leq k-1} |\kappa_{j+2}(\xi^{(n)})| \overset{(*)}{=} \max_{1 \leq j \leq k-1} \{\widetilde{q}_n^{j/2} |\kappa_{j+2}(W_n)|\} \overset{(**)}{\lesssim} \max_{1 \leq j \leq k-1} \{\widetilde{q}_n^{j/2} R_{j,1,n}\}.$$

where to get $(*)$ we used item (b) and to get $(**)$ we used Corollary 9.6. Thus, the following holds

$$\widetilde{q}_n^{-p/2} = (\widetilde{q}_n^{-j_0/2})^{p/j_0} \lesssim R_{j_0,1,n}^{p/j_0} \leq \sum_{j=1}^{k-1} R_{j,1,n}^{p/j},$$

where j_0 is the integer satisfying that $|\kappa_{j_0+2}(\xi^{(n)})| = \max_{1 \leq j \leq k-1} |\kappa_{j+2}(\xi^{(n)})|$.

On the other hand, if $\kappa_{j+2}(W_n) = 0$ for all $1 \leq j \leq k-1$, then by definitions we have $\widetilde{q}_n \geq |I_n|^{2(p+1)/p}$, and therefore, $\widetilde{q}_n^{-p/2} \leq |I_n|^{-(p+1)}$. Moreover, by Hölder's inequality we know that the following holds

$$\sum_{i \in I_n} \mathbb{E}\big[|X_i^{(n)}|^2\big] \leq |I_n|^{p/(p+2)} \Big(\sum_{i \in I_n} \mathbb{E}\big[|X_i^{(n)}|^{p+2}\big]\Big)^{2/(p+2)}. \tag{8.7}$$

and

$$\Big(\sum_{i \in I_n} X_i^{(n)}\Big)^2 \leq |I_n| \sum_{i \in I_n} |X_i^{(n)}|^2. \tag{8.8}$$

Since $\mathbb{E}\Big[\Big(\sum_{i \in I_n} X_i^{(n)}\Big)^2\Big] = \sigma_n^2 = 1$, we have

$$\widetilde{q}_n^{-p/2} \leq |I_n|^{-(p+1)} \Big(\mathbb{E}\Big[\Big(\sum_{i \in I} X_i^{(n)}\Big)^2\Big]\Big)^{(p+2)/2}$$

$$\overset{(*)}{\leq} |I_n|^{-p/2} \Big(\sum_{i \in I} \mathbb{E}\big[|X_i^{(n)}|^2\big]\Big)^{(p+2)/2}$$

$$\overset{(**)}{\leq} \sum_{i \in I_n} \mathbb{E}\big[|X_i^{(n)}|^{p+2}\big] \leq R_{k,\omega,n},$$

where to obtain $(*)$ we used (8.8) and to obtain $(**)$ we used (8.7).

Thus, using item (d) and the fact that $\xi_1^{(n)}, \ldots, \xi_{\widetilde{q}_n}^{(n)}$ are *i.i.d.*, we obtain

$$\widetilde{q}_n^{-(p+2)/2} \sum_{i=1}^{\widetilde{q}_n} \mathbb{E}\big[|\xi_i^{(n)}|^{p+2}\big] \leq C_p' \widetilde{q}_n^{-p/2} \lesssim \sum_{j=1}^{k-1} R_{j,1,n}^{p/j} + \sum_{j=1}^{k} R_{j,\omega,n}^{p/(j+\omega-1)}. \tag{8.9}$$

Therefore, by combining this with (8.6) we obtain that there is a constant $K > 0$ that does not depend on h such that

$$\big|\mathbb{E}[h(W_n)] - \mathbb{E}[h(V_n)]\big| \leq K \Big(\sum_{j=1}^{k} R_{j,1,n}^{p/j} + \sum_{j=1}^{k+1} R_{j,\omega,n}^{p/(j+\omega-1)}\Big).$$

By taking supremum over $h \in \Lambda_p$ and by Lemma 8.4 we obtain that

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{L}(V_n)) \lesssim \sup_{h \in \Lambda_p} \big|\mathbb{E}[h(W_n)] - \mathbb{E}[h(V_n)]\big|^{1/p} \lesssim \sum_{j=1}^{k-1} R_{j,1,n}^{1/j} + \sum_{j=1}^{k} R_{j,\omega,n}^{1/(j+\omega-1)}.$$

Moreover, by combining Lemma 8.6 and (8.9) we have

$$\mathcal{W}_p(\mathcal{L}(V_n), \mathcal{N}(0,1)) \lesssim \left(\widetilde{q}_n^{-(p+2)/2} \sum_{i=1}^{\widetilde{q}_n} \mathbb{E}\big[|\xi_i^{(n)}|^{p+2}\big]\right)^{1/p} \lesssim \sum_{j=1}^{k-1} R_{j,1,n}^{1/j} + \sum_{j=1}^{k} R_{j,\omega,n}^{1/(j+\omega-1)}.$$

Therefore, as the Wasserstein distance $\mathcal{W}_p$ satisfies the triangle inequality we conclude that

$$\begin{aligned}
\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq & \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{L}(V_n)) + \mathcal{W}_p(\mathcal{L}(V_n), \mathcal{N}(0,1)) \\
\lesssim & \sum_{j=1}^{k-1} R_{j,1,n}^{1/j} + \sum_{j=1}^{k} R_{j,\omega,n}^{1/(j+\omega-1)}. \qquad \square
\end{aligned}$$

9 Proof of Lemma 8.7

For ease of notation, when there is no ambiguity we will drop the dependence on n in our notation and write W, $N(\cdot)$, σ, X_i, I and $R_{j,\omega}$ for respectively W_n, $N_n(\cdot)$, σ_n, $X_i^{(n)}$, I_n and $R_{j,\omega,n}$.

9.1 Example and roadmap

Given the form of expression in Lemma 8.7, it is natural to consider performing induction on $\lceil p \rceil$. In fact, [4] used a similar induction idea to prove Lemma 8.5, the analogous result to Lemma 8.7 for independent variables. As [16] suggested, the key of each inductive step is the following expansion of $\mathbb{E}[Wf(W)]$.

Proposition 9.1 (Expansion of $\mathbb{E}[Wf(W)]$)**.** *Denote by $\kappa_j(W)$ the j-th cumulant of W. Given $k \in \mathbb{N}_+$ and real number $\omega \in (0,1]$, for any $f \in \mathcal{C}^{k,\omega}(\mathbb{R})$, we have*

$$\mathbb{E}[Wf(W)] = \sum_{j=1}^{k} \frac{\kappa_{j+1}(W)}{j!} \mathbb{E}[\partial^j f(W)] + \mathcal{O}(|f|_{k,\omega} R_{k,\omega}). \tag{9.1}$$

The case $k = \omega = 1$ is a well-known result in the literature of Stein's method (for example see [5, 38]). The case $k = 2, \omega = 1$ was first proven by [16], and they also conjectured that it was true for any positive integer k with $\omega = 1$. Inspired by [16]'s method, we confirm that this conjecture is correct by proving Proposition 9.1.

To help better understand the intuition behind our proof for the general settings, let's first consider the simplest case with $k = \omega = 1$. Given a positive integer m, suppose that $(X_i)_{i=1}^{n}$ is an m-dependent random sequence (the special case of $d = 1$ in Definition 4.1). We let $W := \sum_{i=1}^{n} X_i$ and require that $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[W^2] = 1$. For simplicity, we further assume $f \in \mathcal{C}^2(\mathbb{R}) \cap \mathcal{C}^{1,1}(\mathbb{R})$ meaning that f'' is a continuous and bounded function.

For any indexes $i, j \in [n]$ (by convention $[n] := \{1, 2, \ldots, n\}$), we write

$$N(i) = \{\ell \in [n] : |\ell - i| \leq m\}, \quad N(i,j) := \{\ell \in [n] : |\ell - i| \leq m \text{ or } |\ell - j| \leq m\}.$$

Denote $W_{i,m} := \sum_{j \notin N(i)} X_j$ and $W_{i,j,m} := \sum_{\ell \notin N(i,j)} X_\ell$. The idea is that for each i, we split W into two parts, $W_{i,m}$ and $W - W_{i,m}$. The former is independent of X_i while the latter is the sum of X_j's in the neighborhood of X_i and will converge to 0 when n grows to ∞. Thus, we perform the Taylor expansion for $f(W)$ around $W_{i,m}$.

We have

$$\mathbb{E}\big[Wf(W) - f'(W)\big] \tag{9.2}$$

$$= \sum_{i=1}^{n} \mathbb{E}\big[X_i\big(f(W) - f(W_{i,m}) - f'(W_{i,m})(W - W_{i,m})\big)\big]$$

$$+ \sum_{i=1}^{n} \mathbb{E}[X_i f(W_{i,m})] + \sum_{i=1}^{n} \mathbb{E}\big[X_i(W - W_{i,m})f'(W_{i,m})\big] - \mathbb{E}[f'(W)]$$

$$= \sum_{i=1}^{n} \mathbb{E}\big[X_i\big(f(W) - f(W_{i,m}) - f'(W_{i,m})(W - W_{i,m})\big)\big]$$

$$+ \sum_{i=1}^{n} \mathbb{E}[X_i]\, \mathbb{E}[f(W_{i,m})] + \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j f'(W_{i,m})\big] - \mathbb{E}[f'(W)]$$

$$= \sum_{i=1}^{n} \mathbb{E}\big[X_i\big(f(W) - f(W_{i,m}) - f'(W_{i,m})(W - W_{i,m})\big)\big]$$

$$+ \bigg(\sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j f'(W_{i,m})\big] - \mathbb{E}[f'(W)]\bigg) =: E_1 + E_2.$$

By assumption, $\|f''\|$ is bounded and we have

$$|E_1| = \bigg| \sum_{i=1}^{n} \mathbb{E}\big[X_i\big(f(W) - f(W_{i,m}) - f'(W_{i,m})(W - W_{i,m})\big)\big] \bigg| \tag{9.3}$$

$$\leq \frac{\|f''\|}{2} \sum_{i=1}^{n} \mathbb{E}\big[\big|X_i(W - W_{i,m})^2\big|\big] = \frac{\|f''\|}{2} \sum_{i=1}^{n} \mathbb{E}\Big[|X_i|\Big(\sum_{j \in N(i)} X_j\Big)^2\Big]$$

$$= \frac{\|f''\|}{2} \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{\ell \in N(i)} \mathbb{E}[|X_i X_j X_\ell|] \leq \frac{\|f''\|}{2} \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{\ell \in N(i,j)} \mathbb{E}[|X_i X_j X_\ell|].$$

Now we bound E_2.

$$E_2 = \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j f'(W_{i,m})\big] - \mathbb{E}[f'(W)] \tag{9.4}$$

$$= \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j\big(f'(W_{i,m}) - f'(W_{i,j,m})\big)\big]$$

$$+ \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j f'(W_{i,j,m})\big] - \mathbb{E}[f'(W)]$$

$$\overset{(*)}{=} \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j\big(f'(W_{i,m}) - f'(W_{i,j,m})\big)\big]$$

$$+ \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}[X_i X_j]\, \mathbb{E}[f'(W_{i,j,m})] - \mathbb{E}[f'(W)]$$

$$\overset{(**)}{=} \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j\big(f'(W_{i,m}) - f'(W_{i,j,m})\big)\big]$$

$$+ \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}[X_i X_j]\, \mathbb{E}\big[f'(W_{i,j,m}) - f'(W)\big]$$

$$= (t_1) + (t_2),$$

where to obtain $(*)$ we have used the fact that $W_{i,j,m}$ is independent of (X_i, X_j) in the second equation and to obtain $(**)$ we have assumed hat $\mathbb{E}(W^2) = 1$.

The first term (t_1), can be upper-bounded by the mean value theorem as

$$\left| \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}\big[X_i X_j \big(f'(W_{i,m}) - f'(W_{i,j,m}) \big) \big] \right|$$

$$\leq \sum_{i=1}^{n} \sum_{j \in N(i)} \|f''\| \, \mathbb{E}\big[\big| X_i X_j (W_{i,m} - W_{i,j,m}) \big| \big]$$

$$\leq \|f''\| \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{\ell \in N(i,j)} \mathbb{E}[|X_i X_j X_\ell|].$$

By another application of the mean-value theorem, we remark that the second term (t_2), is controlled by

$$\left| \sum_{i=1}^{n} \sum_{j \in N(i)} \mathbb{E}[X_i X_j] \, \mathbb{E}\big[f'(W_{i,j,m}) - f'(W) \big] \right|$$

$$\leq \sum_{i=1}^{n} \sum_{j \in N(i)} \|f''\| \, \mathbb{E}[|X_i X_j|] \, \mathbb{E}\big[\big| W_{i,j,m} - W \big| \big]$$

$$\leq \|f''\| \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{\ell \in N(i,j)} \mathbb{E}[|X_i X_j|] \, \mathbb{E}[|X_\ell|].$$

Thus,

$$\big| \mathbb{E}\big[W f(W) - f'(W) \big] \big|$$

$$\leq \|f''\| \sum_{i=1}^{n} \sum_{j \in N(i)} \sum_{\ell \in N(i,j)} \left(\frac{3}{2} \mathbb{E}[|X_i X_j X_\ell|] + \mathbb{E}[|X_i X_j|] \, \mathbb{E}[|X_\ell|] \right)$$

$$\leq \frac{3\|f''\|}{2} R_{1,1}.$$

This gives us a bound that matches with (9.1).

For $k \geq 2$, we would like to carry out the expansion in the same spirit. However, it would be too tedious to write out every sum in the process. Thus, in Section 9.2, we introduce the terms called $\mathcal{S}$-sums, $\mathcal{T}$-sums, and $\mathcal{R}$-sums, which serve as useful tools in tracking different quantities when we approximate $\mathbb{E}[f'(W) - Wf(W)]$ with respect to locally dependent random variables. Instead of performing the expansion to get (9.1) for $\mathbb{E}[Wf(W)]$, we first do it for any $\mathcal{T}$-sum and use induction to prove a more general result for the existence of such expansions (see Theorem 9.3). In the general situation of $\mathcal{T}$-sums, the cumulants are replaced by other constants that only depend on the specific $\mathcal{T}$-sum in consideration and the joint distribution of $(X_i)_{i \in I}$. Finally, we prove that in particular, those constants for $\mathbb{E}[Wf(W)]$ are precisely the cumulants of W. This will be done by direct calculation when f is a polynomial and then extended to more general f's by applying Lemma 9.5.

9.2 Notations and definitions

As in Section 3, given an integer $k \geq 1$, suppose $(X_i)_{i \in I}$ is a class of mean zero random variables indexed by I that satisfy the local dependence assumptions [LD-1] to [LD-k]. Without loss of generality, we always assume that $\sigma^2 := \mathrm{Var}\left(\sum_{i \in I} X_i \right) = 1$. We denote $W := \sigma^{-1} \sum_{i \in I} X_i = \sum_{i \in I} X_i$.

$\mathcal{S}$-sums

Fix $k \in \mathbb{N}_+$ and $t_1, \ldots, t_k \in \mathbb{Z}$ be integers such that $|t_j| \leq j - 1$ for any $j \in [k]$. We set $t_1 = 0$. Let $z = \big|\{j : t_j > 0\}\big|$ be the number of indexes j for which t_j is strictly positive. If $z \geq 1$, we write $\{j : t_j > 0\} = \{q_1, \ldots, q_z\}$. Without loss of generality, we suppose that the sequence $2 \leq q_1 < \cdots < q_z \leq k$ is increasing. We further let $q_0 := 1$ and $q_{z+1} := k + 1$. We define an order-k $\mathcal{S}$-**sum** with respect to the sequence $t_{1:k}$ as

$$\mathcal{S}[t_1, \ldots, t_k] := \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} [q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright (X_{i_1}, \ldots, X_{i_k}) \qquad (9.5)$$

$$= \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\big[X_{i_{q_0}} \cdots X_{i_{q_1-1}}\big] \, \mathbb{E}\big[X_{i_{q_1}} \cdots X_{i_{q_2-1}}\big] \cdots \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_{q_{z+1}-1}}\big],$$

where $N_1 := I$, and for $j \in \mathbb{N}_+$ such that $j \geq 2$, we let

$$N_j := \begin{cases} N(i_{1:|t_j|}) = N(i_1, \ldots, i_{|t_j|}) & \text{if } t_j \neq 0 \\ \emptyset & \text{if } t_j = 0 \end{cases}.$$

Note that N_j depends on t_j and the sequence $i_{1:(j-1)}$. For ease of notation, we do not explicitly write out the dependencies on $i_{1:(j-1)}$ when there is no ambiguity. Further note that if any t_j, that is not t_1, is null then $N_j = \emptyset$ therefore, the $\mathcal{S}$-sum $\mathcal{S}[t_1, \ldots, t_k] = 0$.

By definition all $\mathcal{S}$-sums are deterministic quantities, the value of which only depends on $t_{1:k}$, and the joint distribution of $(X_i)_{i \in I}$. We also remark that the signs of t_j's determine how an $\mathcal{S}$-sum factorizes into different expectations. Notably if $z = 0$ (meaning that all the t_j are negative) then the $\mathcal{T}$-sum is

$$\mathcal{T}_{f,s}[t_1, \ldots, t_k] = \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\Big[X_{i_1} \cdots X_{i_k} \partial^{k-1} f\big(W_{i.}[k-s]\big)\Big].$$

Since by assumption, X_i's are centered random variables, the $\mathcal{S}$-sum vanishes if $q_{j+1} = q_j + 1$ for some $0 \leq j \leq z$:

$$\mathcal{S}[t_1, \ldots, t_k] = \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\big[X_{i_{q_0}} \cdots X_{i_{q_1-1}}\big] \cdot$$
$$\mathbb{E}\big[X_{i_{q_1}} \cdots X_{i_{q_2-1}}\big] \cdots \mathbb{E}\big[X_{i_{q_j}}\big] \cdots \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_{q_{z+1}-1}}\big] = 0. \qquad (9.6)$$

Furthermore, the absolute value of t_j's influences the range of running indexes. The bigger $|t_j|$ is the larger the set N_j is. The largest possible index set for i_{j+1} is $N(i_{1:(j-1)})$, which corresponds to the case $|t_j| = j - 1$. On the other hand, if $t_j = 0$, the sum is over an empty set and vanishes. In particular, if we require that the $\mathcal{S}$-sum is not always zero, then t_2 is always taken to be -1 and $i_2 \in N(i_1)$.

$\mathcal{T}$-sums

For any function $f \in \mathcal{C}^{k-1}(\mathbb{R})$ and integer $s \in \mathbb{N}$ such that $s \leq k$, the order-k $\mathcal{T}$-**sum**, with respect to the sequence $t_{1:k}$, is defined as

$$\mathcal{T}_{f,s}[t_1, \ldots, t_k] :=$$
$$\sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} [q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright \big(X_{i_1}, \ldots, X_{i_{k-1}}, X_{i_k} \partial^{k-1} f(W_{i.}[k-s])\big)$$

$$= \begin{cases} \displaystyle\sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\Big[X_{i_1} \cdots X_{i_k} \partial^{k-1} f\big(W_{i.}[k-s]\big)\Big] & \text{if } z = 0 \\[2em] \displaystyle\sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\big[X_{i_{q_0}} \cdots X_{i_{q_1-1}}\big] \cdots \mathbb{E}\big[X_{i_{q_{(z-1)}}} \cdots X_{i_{q_z-1}}\big] \cdot & \\ \qquad \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_k} \partial^{k-1} f\big(W_{i.}[k-s]\big)\big] & \text{if } z \geq 1, \end{cases} \qquad (9.7)$$

where $N_{1:k}, z, q_{0:(z+1)}$ are defined as in the definition of $\mathcal{S}$-sums and $W_{i.}[j]$ is defined as

$$W_{i.}[j] := \begin{cases} W & \text{if } j = 0 \\ \sum_{i \in I \setminus N(i_{1:j})} X_i & \text{if } 1 \le j \le k \end{cases}.$$

Note that the bigger s is, the larger the set $I \setminus N(i_{1:(k-s)})$ is, which means that $W_{i.}[k-s]$ is the sum of more X_i's. Again we remark that the values of $\mathcal{T}$-sums can depend on the values of s and the sequences $t_{1:k}$. In particular, if $s = 0$, then we have $W_{i.}[k-s] = W_{i.}[k] = \sum_{i \in I \setminus N(i_{1:k})} X_i$, which implies that $W_{i.}[k-s]$ is independent of $X_{i_1}, \ldots, X_{i_k}$ by the assumption **[LD-k]**. Thus, we have

$$\mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_k} \partial^{k-1} f\big(W_{i.}[k-s]\big)\big] = \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_k}\big] \, \mathbb{E}\big[\partial^{k-1} f\big(W_{i.}[k-s]\big)\big].$$

By definitions (9.5) and (9.7) we get

$$\mathcal{T}_{f,0}[t_1, \ldots, t_k] = \mathcal{S}[t_1, \ldots, t_k] \, \mathbb{E}[\partial^{k-1} f(W_{i.}[k])]. \tag{9.8}$$

This equation will be useful in our discussion later. In general if $z > 0$ then

$$\mathcal{T}_{f,s}[t_1, \ldots, t_k] = \mathcal{S}[t_1, \ldots, t_{q_z - 1}] \sum_{i_{q_z} \in N_{q_z}} \sum_{i_{q_z+1} \in N_{q_z+1}} \cdots \sum_{i_k \in N_k}$$
$$\mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_k} \partial^{k-1} f\big(W_{i.}[k-s]\big)\big].$$

$\mathcal{R}$-sums

For $k \ge 2$ and given a real number $\omega \in (0, 1]$, we further define an order-k $\mathcal{R}$-**sum** with respect to the sequence $t_{1:k}$ as

$$\mathcal{R}_\omega[t_1, \ldots, t_k] :=$$
$$\sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_{k-1} \in N_{k-1}} [q_1 - q_0, \ldots, q_{z+1} - q_z] \rhd \Big(|X_{i_1}|, \ldots, |X_{i_{k-1}}|, \big(\sum_{i_k \in N_k} |X_{i_k}|\big)^\omega\Big)$$
$$= \begin{cases} \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\Big[X_{i_1} \cdots X_{i_{k-1}} \big(\sum_{i_k \in N_k} |X_{i_k}|\big)^\omega\Big] & \text{if } z = 0 \\[2mm] \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_k \in N_k} \mathbb{E}\big[X_{i_{q_0}} \cdots X_{i_{q_1-1}}\big] \cdots \mathbb{E}\big[X_{i_{q_{(z-1)}}} \cdots X_{i_{q_z-1}}\big] \cdot \\[1mm] \quad \mathbb{E}\Big[X_{i_{q_z}} \cdots X_{i_{k-1}} \big(\sum_{i_k \in N_k} |X_{i_k}|\big)^\omega\Big] & \text{if } z \ge 1. \end{cases} \tag{9.9}$$

We again remark that if $z \ge 1$ then

$$\mathcal{R}_\omega[t_1, \ldots, t_k] = \mathcal{R}_1[t_1, \ldots, t_{q_z-1}] \sum_{i_{q_z} \in N_{q_z}} \sum_{i_{q_z+1} \in N_{q_z+1}} \cdots \sum_{i_k \in N_k}$$
$$\mathbb{E}\Big[X_{i_{q_z}} \cdots X_{i_{k-1}} \big(\sum_{i_k \in N_k} |X_{i_k}|\big)^\omega\Big].$$

We call ω the exponent of the $\mathcal{R}$-sum. If $\omega = 1$, the only difference between an $\mathcal{R}$-sum and an $\mathcal{S}$-sum is that the X_{i_j}'s in (9.5) are replaced by $|X_{i_j}|$'s in (9.9). Thus, an $\mathcal{S}$-sum is always upper-bounded by the corresponding compositional 1-sum, i.e.,

$$\big|\mathcal{S}[t_1, \ldots, t_k]\big| \le \mathcal{R}_1[t_1, \ldots, t_k]. \tag{9.10}$$

Another important observation is that we can compare the values of $\mathcal{R}$-sums with respect to two different sequences $t_1, \ldots, t_k$ and $t_1', \ldots, t_k'$ in certain situations. In specific, if for any $j \in [k]$ we have that if t_j and t_j' are of the same sign and $|t_j| \leq |t_j'|$, then

$$\mathcal{R}_\omega[t_1, \ldots, t_k] \leq \mathcal{R}_\omega[t_1', \ldots, t_k']. \tag{9.11}$$

In fact, the sequences (t_j) and (t_j') having the same sign indicates that $\{j : t_j > 0\} = \{j : t_j' > 0\}$. Thus, we can write

$$\mathcal{R}_\omega[t_1', \ldots, t_k'] = \sum_{i_1 \in N_1'} \sum_{i_2 \in N_2'} \cdots \sum_{i_{k-1} \in N_{k-1}'} [q_1 - q_0, \ldots, q_{z+1} - q_z] \rhd \tag{9.12}$$
$$\left(|X_{i_1}|, \ldots, |X_{i_{k-1}}|, \Big(\sum_{i_k \in N_k'} |X_{i_k}| \Big)^\omega \right),$$

where we note that $N_1' = I = N_1$ and for $j = 2, \ldots, k$ we have

$$N_j' = N(i_1, \ldots, i_{|t_j'|}) \supseteq N(i_1, \ldots, i_{|t_j|}) = N_j.$$

By comparing (9.9) with (9.12), we obtain (9.11).

Re-expression of the remainder terms $R_{k,\omega}$

Using the notion of $\mathcal{R}$-sums, we rewrite the $R_{k,\omega}$ in Section 3 as

$$R_{k,\omega} := \sum_{(\ell, \eta_{1:\ell}) \in C^*(k+2)} \sum_{i_1 \in N_1'} \sum_{i_2 \in N_2'} \cdots \sum_{i_{k+1} \in N_{k+1}'}$$
$$[\eta_1, \ldots, \eta_\ell] \rhd \left(|X_{i_1}|, \ldots, |X_{i_{k+1}}|, \Big(\sum_{i_{k+2} \in N_{k+2}'} |X_{i_{k+2}}| \Big)^\omega \right) \tag{9.13}$$
$$= \sum_{t_{1:(k+2)} \in \mathcal{M}_{1,k+2}} \mathcal{R}_\omega[t_1, t_2, \ldots, t_{k+2}],$$

where $N_1' := I$ and $N_j' := N(i_{1:(j-1)})$ for $j \geq 2$. $C^*(k+2)$ and $\mathcal{M}_{1,k+2}$ are given by

$$C^*(k+2) = \left\{ \ell, \eta_{1:\ell} \in \mathbb{N}_+ : \eta_j \geq 2 \ \forall j \in [\ell-1], \ \sum_{j=1}^\ell \eta_j = k+2 \right\},$$

and

$$\mathcal{M}_{1,k+2} := \left\{ t_{1:(k+2)} : \ t_{j+1} = \pm j \ \& \ t_j \wedge t_{j+1} < 0 \ \forall 1 \leq j \leq k+1 \right\}.$$

Note that $t_j \wedge t_{j+1} < 0$ for any $j \in [k+1]$ means that there is at least one -1 in any two consecutive signs, which corresponds to the requirement that $\eta_j \geq 2$ for $j \in [\ell-1]$ in (9.13).

9.3 Proofs of Proposition 9.1 and Lemma 8.7

In this section, we carry out the local expansion technique and prove Proposition 9.1 and Lemma 8.7.

Firstly, we establish the following lemma, which will be crucial in the inductive step of proving the main theorem.

Lemma 9.2. *Fix $k \in \mathbb{N}_+$. For any $s \in [k] \cup \{0\}$ and $f \in \mathcal{C}^{k,\omega}(\mathbb{R})$, we have*

$$\left| \mathcal{T}_{f,s}[t_1, \ldots, t_{k+1}] - \mathcal{S}[t_1, \ldots, t_{k+1}] \, \mathbb{E}[\partial^k f(W)] \right|$$
$$\leq |f|_{k,\omega} \big(-(\mathbb{I}(t_{k+1} < 0) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1] + \mathbb{I}(s \geq 1) \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)]) \big). \tag{9.14}$$

Given any $\ell \in [k]$ and $s \in [\ell] \cup \{0\}$, we further have

$$
\Big| \mathcal{T}_{f,s}[t_1, \ldots, t_\ell] - \mathcal{S}[t_1, \ldots, t_\ell] \, \mathbb{E}[\partial^{\ell-1} f(W)] \tag{9.15}
$$

$$
- \mathbb{I}(s \geq 1) \cdot \sum_{j=1}^{k-\ell+1} \sum_{h=0}^{j} (-1)^h \frac{1}{h!(j-h)!} \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \underbrace{s-\ell, \ldots, s-\ell}_{h \text{ times}}, \underbrace{-\ell, \ldots, -\ell}_{(j-h) \text{ times}}]
$$

$$
- \Big(\mathbb{I}(t_\ell < 0) \sum_{j=1}^{k-\ell+1} \frac{1}{j!} \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(j-1) \text{ times}}] \Big|
$$

$$
\leq \frac{|f|_{k,\omega}}{(k-\ell+1)!} \Big(-\mathbb{I}(t_\ell < 0) \cdot \mathcal{R}_\omega[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(k-\ell+1) \text{ times}}]
$$

$$
+ \mathbb{I}(s \geq 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_\ell, \underbrace{-\ell, \ldots, -\ell}_{(k-\ell+2) \text{ times}}] \Big).
$$

Proof. Firstly, we remark that the definition of Hölder continuity implies that

$$
\big| \partial^k f(y) - \partial^k f(x) \big| \leq |f|_{k,\omega} |y - x|^\omega, \tag{9.16}
$$

where ω is the Hölder exponent of f and $|f|_{k,\omega}$ is the Hölder constant (see Definition 8.1). Let $z = \big| \{ j \in [k+1] : t_j > 0 \} \big|$ be the number of positive indexes (t_j). If $z \geq 1$, we write $\{ j \in [k+1] : t_j > 0 \} = \{q_1, \ldots, q_z\}$. Without loss of generality, we suppose that the sequence $2 \leq q_1 < \cdots < q_z \leq k+1$ is increasing. We further let $q_0 := 1$ and $q_{z+1} := k+2$. Applying (9.16) we have

$$
\Big| \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_{k+1}} \partial^k f\big(W_{i\cdot}[k+1-s] \big) \big] - \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_{k+1}} \partial^k f\big(W_{i\cdot}[k+1] \big) \big] \Big| \tag{9.17}
$$

$$
\leq |f|_{k,\omega} \mathbb{E}\Big[\big| X_{i_{q_z}} \cdots X_{i_{k+1}} \big| \cdot \big| W_{i\cdot}[k+1-s] - W_{i\cdot}[k+1] \big|^\omega \Big]
$$

$$
\leq |f|_{k,\omega} \mathbb{E}\Big[\big| X_{i_{q_z}} \cdots X_{i_{k+1}} \big| \cdot \Big| \sum_{i \in N(i_{1:(k+1)}) \setminus N(i_{1:(k+1-s)})} X_i \Big|^\omega \Big]
$$

$$
\leq |f|_{k,\omega} \mathbb{E}\Big[\big| X_{i_{q_z}} \cdots X_{i_{k+1}} \big| \cdot \Big| \sum_{i \in N(i_{1:(k+1)})} X_i \Big|^\omega \Big],
$$

where in the last inequality we have used the fact that $N(i_{1:(k+1)}) \setminus N(i_{1:(k+1-s)}) \subseteq N(i_{1:(k+1)})$. If $z = 0$, this directly implies that

$$
\Big| \mathcal{T}_{f,s}[t_1, \ldots, t_{k+1}] - \mathcal{T}_{f,0}[t_1, \ldots, t_{k+1}] \Big| \leq \mathbb{I}(s \geq 1) \cdot |f|_{k,\omega} \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)]. \tag{9.18}
$$

If $z \geq 1$, by definition (9.7) we have for $s \geq 1$

$$
\Big| \mathcal{T}_{f,s}[t_1, \ldots, t_{k+1}] - \mathcal{T}_{f,0}[t_1, \ldots, t_{k+1}] \Big|
$$

$$
= \Big| \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_{k+1} \in N_{k+1}} \mathbb{E}\big[X_{i_{q_0}} \cdots X_{i_{q_1-1}} \big] \cdots \mathbb{E}\big[X_{i_{q_{z-1}}} \cdots X_{i_{q_z-1}} \big] \cdot
$$

$$
\mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_{k+1}} \big(\partial^k f(W_{i\cdot}[k+1-s]) - \partial^k f(W_{i\cdot}[k+1]) \big) \big] \Big|
$$

$$
\leq \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_{k+1} \in N_{k+1}} \mathbb{E}\big[\big| X_{i_{q_0}} \cdots X_{i_{q_1-1}} \big| \big] \cdots \mathbb{E}\big[\big| X_{i_{q_{z-1}}} \cdots X_{i_{q_z-1}} \big| \big] \cdot
$$

$$
\Big| \mathbb{E}\big[X_{i_{q_z}} \cdots X_{i_{k+1}} \partial^k f(W_{i\cdot}[k+1-s]) - \partial^k f(W_{i\cdot}[k+1]) \big] \Big|
$$

$$\overset{(9.17)}{\leq} |f|_{k,\omega} \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_{k+1} \in N_{k+1}} \mathbb{E}\big[|X_{i_{q_0}} \cdots X_{i_{q_1-1}}|\big] \cdots \mathbb{E}\big[|X_{i_{q_{z-1}}} \cdots X_{i_{q_z-1}}|\big] \cdot$$

$$\mathbb{E}\bigg[|X_{i_{q_z}} \cdots X_{i_{k+1}}| \cdot \Big|\sum_{i \in N(i_{1:(k+1)})} X_i\Big|^\omega\bigg]$$

$$= |f|_{k,\omega} \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)].$$

Here the last equality is due to the definition (9.9). Thus, (9.18) is proven for both $z = 0$ and $z \geq 1$. Next we show that

$$\Big|\mathcal{S}[t_1, \ldots, t_{k+1}]\big(\mathbb{E}[\partial^k f(W)] - \mathbb{E}[\partial^k f(W_{i\cdot}[k+1])]\big)\Big|$$
$$\leq -\mathbb{I}(t_{k+1} < 0)|f|_{k,\omega} \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1]. \tag{9.19}$$

In this goal, we first note that if $t_{k+1} \geq 0$, by definition (9.5) we know that $q_z = k+1$ and therefore, according to (9.6) we know that

$$\mathcal{S}[t_1, \ldots, t_{k+1}] = 0,$$

and so (9.19) holds. Otherwise, we note that we have

$$\Big|\mathbb{E}\big[\partial^k f(W)\big] - \mathbb{E}\big[\partial^k f(W_{i\cdot}[k+1])\big]\Big| \leq |f|_{k,\omega} \mathbb{E}\big[|W_{i\cdot}[k+1-s] - W_{i\cdot}[k+1]|^\omega\big]$$
$$\leq |f|_{k,\omega} \mathbb{E}\bigg[\Big|\sum_{i \in N(i_{1:(k+1)}) \setminus N(i_{1:(k+1-s)})} X_i\Big|^\omega\bigg] \leq |f|_{k,\omega} \mathbb{E}\bigg[\Big|\sum_{i \in N(i_{1:(k+1)})} X_i\Big|^\omega\bigg]. \tag{9.20}$$

This implies that

$$\Big|\mathcal{S}[t_1, \ldots, t_{k+1}]\big(\mathbb{E}[\partial^k f(W)] - \mathbb{E}[\partial^k f(W_{i\cdot}[k+1])]\big)\Big|$$
$$\leq \Big|\mathcal{S}[t_1, \ldots, t_{k+1}]\Big| \cdot \Big|\mathbb{E}\big[\partial^k f(W)\big] - \mathbb{E}\big[\partial^k f\big(W_{i\cdot}[k+1]\big)\big]\Big|$$
$$\overset{(*)}{\leq} |f|_{k,\omega} \mathcal{R}_1[t_1, \ldots, t_{k+1}] \mathbb{E}\bigg[\Big|\sum_{i \in N(i_{1:(k+1)})} X_i\Big|^\omega\bigg]$$
$$= |f|_{k,\omega} \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_{k+1} \in N_{k+1}} [q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright$$
$$\big(|X_{i_1}|, \ldots, |X_{i_{k+1}}|\big) \cdot \mathbb{E}\bigg[\Big|\sum_{i \in N(i_{1:(k+1)})} X_i\Big|^\omega\bigg]$$
$$= |f|_{k,\omega} \sum_{i_1 \in N_1} \sum_{i_2 \in N_2} \cdots \sum_{i_{k+1} \in N_{k+1}} [q_1 - q_0, \ldots, q_{z+1} - q_z, 1] \triangleright$$
$$\Big(|X_{i_1}|, \ldots, |X_{i_{k+1}}|, \big(\sum_{i_{k+2} \in N(i_{1:(k+1)})} |X_{i_{k+2}}|\big)^\omega\Big)$$
$$= |f|_{k,\omega} \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1],$$

where $(*)$ is due to (9.10) and (9.20). Taking the difference of (9.18) and (9.19), we obtain (9.14) by applying the equation (9.8).

For $\ell \leq k$, we apply the Taylor expansion with remainders taking the integral form and obtain that

$$\partial^{\ell-1} f(y) - \partial^{\ell-1} f(x) = \sum_{j=1}^{m-\ell} \frac{1}{j!}(y-x)^j \partial^{\ell-1+j} f(x) \tag{9.21}$$

$$+ \frac{1}{(k-\ell+1)!}(y-x)^{k-\ell+1}\int_0^1 (k-\ell+1)v^{k-\ell}\partial^k f(vx+(1-v)y)\,\mathrm{d}v$$

$$\overset{(*)}{=} \sum_{j=1}^{k-\ell+1}\frac{1}{j!}(y-x)^j\partial^{\ell-1+j}f(x)$$

$$+ \frac{1}{(k-\ell+1)!}(y-x)^{k-\ell+1}\int_0^1 (k-\ell+1)v^{k-\ell}\big(\partial^k f(vx+(1-v)y)-\partial^k f(x)\big)\,\mathrm{d}v,$$

where to obtain $(*)$ we added and subtracted $\frac{(y-x)^{k-\ell+1}}{(k-\ell+1)!}\partial^k f(x)$. Moreover, using the fact that $\partial^k f(\cdot)$ is assumed to be Hölder continuous we obtain that

$$\big|\partial^k f(vx+(1-v)y)-\partial^k f(x)\big| \leq |f|_{k,\omega}(1-v)^\omega\,|y-x|^\omega \leq |f|_{k,\omega}|y-x|^\omega. \tag{9.22}$$

Therefore, as $\int_0^1 (k-\ell+1)v^{k-\ell}\,\mathrm{d}v = 1$, by combining (9.22) with (9.21) we get that

$$\left|\partial^{\ell-1}f(y)-\partial^{\ell-1}f(x)-\sum_{j=1}^{k-\ell+1}\frac{1}{j!}(y-x)^j\partial^{\ell-1+j}f(x)\right| \leq \frac{|f|_{k,\omega}}{(k-\ell+1)!}|y-x|^{k-\ell+1+\omega}. \tag{9.23}$$

We prove that the following inequality holds:

$$\left|\mathcal{T}_{f,s}[t_1,\ldots,t_\ell]-\mathcal{T}_{f,0}[t_1,\ldots,t_\ell]\right.$$

$$\left. - \mathbb{I}(s\geq 1)\cdot\sum_{j=1}^{k-\ell+1}\sum_{h=0}^{j}(-1)^h\frac{1}{h!(j-h)!}\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}};\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}}]\right| \tag{9.24}$$

$$\leq \frac{\mathbb{I}(s\geq 1)\cdot|f|_{k,\omega}}{(k-\ell+1)!}\mathcal{R}_\omega[t_1,\ldots,t_\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+2)\text{ times}}].$$

First, let's establish (9.24). Let $z = \big|\{j\in[\ell] : t_j > 0\}\big|$. If $z \geq 1$, we write $\{j\in[\ell] : t_j > 0\} = \{q_1,\ldots,q_z\}$. Without loss of generality, we suppose that the sequence $2 \leq q_1 < \cdots < q_z \leq \ell$ is increasing. We further let $q_0 := 1$ and $q_{z+1} := \ell+1$. Applying (9.23) we have

$$\left|\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}\partial^{\ell-1}f\big(W_{i.}[\ell-s]\big)\big]-\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}\partial^{\ell-1}f\big(W_{i.}[\ell]\big)\big]\right.$$

$$\left. - \sum_{j=1}^{k-\ell+1}\frac{1}{j!}\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}(W_{i.}[\ell-s]-W_{i.}[\ell])^j\partial^{\ell-1+j}f(W_{i.}[\ell])\big]\right| \tag{9.25}$$

$$\leq \frac{|f|_{k,\omega}}{(k-\ell+1)!}\mathbb{E}\big[\big|X_{i_{q_z}}\cdots X_{i_\ell}\big|\cdot\big|W_{i.}[\ell-s]-W_{i.}[\ell]\big|^{k-\ell+1+\omega}\big].$$

For convenience let

$$E_1 := \sum_{i_{q_z}\in N_{q_z}}\cdots\sum_{i_\ell\in N_\ell}\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}\partial^{\ell-1}f\big(W_{i.}[\ell-s]\big)\big]-\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}\partial^{\ell-1}f\big(W_{i.}[\ell]\big)\big],$$

$$E_{2,j} := \sum_{i_{q_z}\in N_{q_z}}\cdots\sum_{i_\ell\in N_\ell}\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}(W_{i.}[\ell-s]-W_{i.}[\ell])^j\partial^{\ell-1+j}f(W_{i.}[\ell])\big],$$

$$E_3 := \sum_{i_{q_z}\in N_{q_z}}\cdots\sum_{i_\ell\in N_\ell}\mathbb{E}\big[\big|X_{i_{q_z}}\cdots X_{i_\ell}\big|\cdot\big|W_{i.}[\ell-s]-W_{i.}[\ell]\big|^{k-\ell+1+\omega}\big].$$

Then (9.25) reduces to

$$\left|E_1-\sum_{j=1}^{k-\ell+1}E_{2,j}/j!\right| \leq |f|_{k,\omega}E_3/(k-\ell+1)!. \tag{9.26}$$

Then we observe that by definition of $W_{i.}[\cdot]$ we have

$$
\mathbb{E}\big[X_{i_{q_z}}\cdots X_{i_\ell}(W_{i.}[\ell-s]-W_{i.}[\ell])^j\partial^{\ell-1+j}f(W_{i.}[\ell])\big]
$$
$$
=\mathbb{E}\Big[X_{i_{q_z}}\cdots X_{i_\ell}\Big(\sum_{i\in N(i_{1:\ell})}X_i-\sum_{i\in N(i_{1:\ell-s})}X_i\Big)^j\partial^{\ell-1+j}f(W_{i.}[\ell])\Big]
\tag{9.27}
$$
$$
=\sum_{h=0}^{j}(-1)^h\binom{j}{h}\mathbb{E}\Big[X_{i_{q_z}}\cdots X_{i_\ell}\Big(\sum_{i\in N(i_{1:\ell-s})}X_i\Big)^h\Big(\sum_{i\in N(i_{1:\ell})}X_i\Big)^{j-h}\partial^{\ell-1+j}f(W_{i.}[\ell])\Big],
$$

and that

$$
\mathbb{E}\big[\big|X_{i_{q_z}}\cdots X_{i_\ell}\big|\cdot\big|W_{i.}[\ell-s]-W_{i.}[\ell]\big|^{k-\ell+1+\omega}\big]
\tag{9.28}
$$
$$
\leq\mathbb{E}\Big[\big|X_{i_{q_z}}\cdots X_{i_{k+1}}\big|\cdot\Big|\sum_{i\in N(i_{1:(k+1)})\setminus N(i_{1:(k+1-s)})}X_i\Big|^{k-\ell+1+\omega}\Big]
$$
$$
\leq\mathbb{E}\Big[\big|X_{i_{q_z}}\cdots X_{i_{k+1}}\big|\cdot\Big|\sum_{i\in N(i_{1:(k+1)})}X_i\Big|^{k-\ell+1+\omega}\Big]
$$
$$
\leq\mathbb{E}\Big[\big|X_{i_{q_z}}\cdots X_{i_{k+1}}\big|\cdot\Big(\sum_{i\in N(i_{1:(k+1)})}|X_i|\Big)^{k-\ell+1}\cdot\Big|\sum_{i\in N(i_{1:(k+1)})}X_i\Big|^{\omega}\Big].
$$

If $z=0$, we take the sum of (9.27) or (9.28) over $i_{q_z}\in N_{q_z},\ldots,i_\ell\in N_\ell$. By definition (9.7) and (9.9) we have

$$
E_1=\mathcal{T}_{f,s}[t_1,\ldots,t_\ell]-\mathcal{T}_{f,0}[t_1,\ldots,t_\ell],
$$
$$
E_{2,j}=\sum_{h=0}^{j}(-1)^h\binom{j}{h}\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}},\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}}],
\tag{9.29}
$$
$$
E_3\leq\mathcal{R}_\omega[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+1)\text{ times}}].
$$

Combining (9.29) and (9.26), we have for $s\geq 1$

$$
\Big|\mathcal{T}_{f,s}[t_1,\ldots,t_\ell]-\mathcal{T}_{f,0}[t_1,\ldots,t_\ell]
$$
$$
-\sum_{j=1}^{k-\ell+1}\sum_{h=0}^{j}(-1)^h\frac{1}{h!(j-h)!}\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}},\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}}]\Big|
$$
$$
\overset{(9.29)}{=}\Big|E_1-\sum_{j=1}^{k-\ell+1}E_{2,j}/j!\Big|\overset{(9.26)}{\leq}|f|_{k,\omega}E_3/(k-\ell+1)!
$$
$$
\overset{(9.29)}{\leq}\frac{|f|_{k,\omega}}{(k-\ell+1)!}\mathcal{R}_\omega[t_1,\ldots,t_\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+2)\text{ times}}].
$$

Thus, (9.24) holds for $z=0$.

If $z\geq 1$, similar to (9.28) we have

$$
\mathcal{S}[t_1,\ldots,t_{q_z-1}]\cdot E_1=\mathcal{T}_{f,s}[t_1,\ldots,t_\ell]-\mathcal{T}_{f,0}[t_1,\ldots,t_\ell],
$$
$$
\mathcal{S}[t_1,\ldots,t_{q_z-1}]\cdot E_{2,j}=\sum_{h=0}^{j}(-1)^h\binom{j}{h}\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}},\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}}],
\tag{9.30}
$$
$$
\mathcal{R}_1[t_1,\ldots,t_{q_z-1}]\cdot E_3\leq\mathcal{R}_\omega[t_1,\ldots,t_\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+2)\text{ times}}].
$$

Combining (9.30) and (9.26) we get for $s \geq 1$

$$\left| \mathcal{T}_{f,s}[t_1, \ldots, t_\ell] - \mathcal{T}_{f,0}[t_1, \ldots, t_\ell] \right.$$
$$\left. - \sum_{j=1}^{k-\ell+1} \sum_{h=0}^{j} (-1)^h \frac{1}{h!(j-h)!} \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \underbrace{s-\ell, \ldots, s-\ell}_{h \text{ times}}, \underbrace{-\ell, \ldots, -\ell}_{(j-h) \text{ times}}] \right|$$

$$\overset{(9.30)}{=} \left| \mathcal{S}[t_1, \ldots, t_{q_z-1}] \right| \cdot \left| E_1 - \sum_{j=1}^{k-\ell+1} E_{2,j}/j! \right| \overset{(9.10)}{\leq} \mathcal{R}_1[t_1, \ldots, t_{q_z-1}] \cdot \left| E_1 - \sum_{j=1}^{k-\ell+1} E_{2,j}/j! \right|$$

$$\overset{(9.26)}{\leq} \mathcal{R}_1[t_1, \ldots, t_{q_z-1}] \cdot |f|_{k,\omega} E_3/(k-\ell+1)! \overset{(9.30)}{\leq} \frac{|f|_{k,\omega}}{(k-\ell+1)!} \mathcal{R}_\omega[t_1, \ldots, t_\ell, \underbrace{-\ell, \ldots, -\ell}_{(k-\ell+2) \text{ times}}].$$

Thus, we have shown (9.24) for both $z = 0$ and $z \geq 1$.

Next we prove that the following inequality holds:

$$\left| \mathcal{S}[t_1, \ldots, t_\ell] \big(\mathbb{E}[\partial^{\ell-1} f(W)] - \mathbb{E}[\partial^{\ell-1} f(W_i.[\ell])] \big) \right.$$
$$\left. - \mathbb{I}(t_\ell < 0) \cdot \sum_{j=1}^{k-\ell+1} \frac{1}{j!} \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(j-1) \text{ times}}] \right| \qquad (9.31)$$

$$\leq \frac{\mathbb{I}(t_\ell < 0) \cdot |f|_{k,\omega}}{(k-\ell+1)!} \mathcal{R}_\omega[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(k-\ell+1) \text{ times}}].$$

For (9.31), we apply (9.23) again and get that

$$\left| \mathbb{E}\big[\partial^k f(W)\big] - \mathbb{E}\big[\partial^k f(W_i.[\ell])\big] \right.$$
$$\left. - \sum_{j=1}^{k-\ell+1} \frac{1}{j!} \mathbb{E}\big[(W_i. - W_i.[\ell])^j \partial^{\ell-1+j} f(W_i.[\ell])\big] \right| \leq \frac{|f|_{k,\omega}}{(k-\ell+1)!} \mathbb{E}\big[|W - W_i.[\ell]|^{k-\ell+1+\omega}\big]. \qquad (9.32)$$

For convenience let

$$E_4 := \mathbb{E}[\partial^{\ell-1} f(W)] - \mathbb{E}[\partial^{\ell-1} f(W_i.[\ell])],$$
$$E_{5,j} := \mathbb{E}\big[(W_i. - W_i.[\ell])^j \partial^{\ell-1+j} f(W_i.[\ell])\big],$$
$$E_6 := \mathbb{E}\big[|W - W_i.[\ell]|^{k-\ell+1+\omega}\big].$$

Then (9.32) reduces to

$$\left| E_4 - \sum_{j=1}^{k-\ell+1} E_{5,j}/j! \right| \leq |f|_{k,\omega} E_6/(k-\ell+1)!. \qquad (9.33)$$

We first note that if $t_\ell \geq 0$ then $\mathcal{S}[t_1, \ldots, t_\ell] = 0$ therefore, (9.31) holds. Moreover, similar to (9.30), we have for $t_\ell < 0$

$$\mathcal{S}[t_1, \ldots, t_\ell] \cdot E_4 = \mathcal{S}[t_1, \ldots, t_\ell] \big(\mathbb{E}[\partial^{\ell-1} f(W)] - \mathbb{E}[\partial^{\ell-1} f(W_i.[\ell])] \big),$$
$$\mathcal{S}[t_1, \ldots, t_\ell] \cdot E_{5,j} = \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(j-1) \text{ times}}], \qquad (9.34)$$
$$\mathcal{R}_1[t_1, \ldots, t_\ell] \cdot E_6 \leq \mathcal{R}_\omega[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(k-\ell+1) \text{ times}}].$$

Combining (9.34) and (9.33), we have

$$\left| \mathcal{S}[t_1,\ldots,t_\ell]\big(\mathbb{E}[\partial^{\ell-1}f(W)] - \mathbb{E}[\partial^{\ell-1}f(W_i.[\ell])]\big) - \sum_{j=1}^{k-\ell+1} \frac{1}{j!}\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(j-1)\text{ times}}] \right|$$

$$\overset{(9.34)}{=} \big|\mathcal{S}[t_1,\ldots,t_\ell]\big| \cdot \big|E_4 - \sum_{j=1}^{k-\ell+1} E_{5,j}/j!\big| \overset{(9.10)}{\leq} \mathcal{R}_1[t_1,\ldots,t_\ell] \cdot \big|E_4 - \sum_{j=1}^{k-\ell+1} E_{5,j}/j!\big|$$

$$\overset{(9.33)}{\leq} \mathcal{R}_1[t_1,\ldots,t_\ell] \cdot |f|_{k,\omega} E_6/(k-\ell+1)! \overset{(9.34)}{\leq} \frac{|f|_{k,\omega}}{(k-\ell+1)!}\mathcal{R}_\omega[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+1)\text{ times}}].$$

Therefore, we have established both (9.24) and (9.31). Taking their difference and applying (9.8), we obtain (9.15). $\qquad\square$

Equipped with the tools in Lemma 9.2, we approximate any $\mathcal{T}$-sum $\mathcal{T}_{f,s}[t_1,\ldots,t_\ell]$ by order-j $\mathcal{S}$-sums ($j=\ell,\ldots,k+1$) with remainder terms being order-$(k+2)$ $\mathcal{R}$-sums.

Theorem 9.3. *Fix $k \in \mathbb{N}_+$. For any $\ell \in [k+1], s \in [\ell] \cup \{0\}$, and $t_1,\ldots,t_\ell \in \mathbb{Z}$ such that $|t_j| \leq j-1$ for any $j \in [\ell]$, there exist $Q_\ell,\ldots,Q_{k+1}$ (which depend on s and $t_{1:\ell}$ and the joint distribution of $(X_i)_{i\in I}$) and a constant $C_{k,\ell}$ ($C_{k,\ell} \leq 4^{k-\ell+1}$) such that for any $f \in \mathcal{C}^{k,\omega}(\mathbb{R})$, we have*

$$\left| \mathcal{T}_{f,s}[t_1,\ldots,t_\ell] - \sum_{j=\ell}^{k+1} Q_j \mathbb{E}[\partial^{j-1}f(W)] \right| \leq C_{k,\ell}|f|_{k,\omega} R_{k,\omega}. \qquad (9.35)$$

Note that by (9.13) $R_{k,\omega}$ is given as

$$R_{k,\omega} = \sum_{t_{1:(k+2)}\in\mathcal{M}_{1,k+2}} \mathcal{R}_\omega[t_1,t_2,\ldots,t_{k+2}],$$

where

$$\mathcal{M}_{1,k+2} := \Big\{ t_{1:(k+2)} : \; t_{j+1} = \pm j \;\; \& \;\; t_j \wedge t_{j+1} < 0 \;\; \forall 1 \leq j \leq k+1 \Big\}.$$

Proof. If there exists an integer $2 \leq j \leq \ell$ such that $t_j = 0$ or there exists $j \in [\ell-1]$ such that t_j and t_{j+1} are both positive, then $\mathcal{T}_{f,s}[t_1,\ldots,t_\ell] = 0$ by definition and the theorem already holds by setting $Q_j = \cdots = Q_{k+1} = 0$.

Otherwise, we claim:

Claim 9.4. *Let $\mathcal{T}_{f,s}[t_1,\ldots,t_\ell]$ be a $\mathcal{T}$-sum. For any $j = \ell+1,\ldots,k+1$, let*

$$\mathcal{E}_{\ell+1,j} := \big\{ t_{(\ell+1):j} : \; |t_{h+1}| \leq h \;\; \& \;\; t_h \wedge t_{h+1} \;\; \forall \ell \leq h \leq j-1 \big\}.$$

For all $j = \ell+1,\ldots,k+1$, $\nu \in [j] \cup \{0\}$, and $(t_{\ell+1},\ldots,t_j) \in \mathcal{E}_{\ell,j}$, there are coefficients $a_{j,\nu,t_{(\ell+1):j}}$ (additionally depending on s) such that if we write

$$Q_j = \sum_{t_{(l+1):j}\in\mathcal{E}_{\ell,j}} \sum_{\nu=0}^{j} a_{j,\nu,t_{(\ell+1):j}} \mathcal{T}_{f,\nu}[t_1,\ldots,t_\ell,t_{\ell+1},\ldots,t_j], \qquad (9.36)$$

then the following holds

$$\left| \mathcal{T}_{f,s}[t_1,\ldots,t_\ell] - \sum_{j=\ell}^{k+1} Q_j \mathbb{E}[\partial^{j-1}f(W)] \right|$$

$$\leq 4^{k-\ell+1}|f|_{k,\omega} \sum_{t_{(\ell+1):(k+2)}\in\mathcal{M}_{\ell,k+1}} \mathcal{R}_\omega[t_1,\ldots,t_\ell,\ldots,t_{k+2}], \qquad (9.37)$$

where

$$\mathcal{M}_{\ell+1,k+2} := \left\{ t_{(\ell+1):(k+2)} : \ t_{j+1} = \pm j \ \& \ t_j \wedge t_{j+1} < 0 \ \ \forall \ell \le j \le k+1 \right\}.$$

We establish this claim by performing induction on ℓ with ℓ taking the value $k + 1, k, \ldots, 1$ in turn.

For $\ell = k + 1$, by (9.14) we have

$$\left| \mathcal{T}_{f,s}[t_1, \ldots, t_{k+1}] - \mathcal{S}[t_1, \ldots, t_{k+1}] \, \mathbb{E}[\partial^k f(W)] \right|$$
$$\le |f|_{k,\omega} \big(\mathbb{I}(t_{k+1} < 0) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1] + \mathbb{I}(s \ge 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)] \big).$$

If there exists $j \in [k]$ such that t_j and t_{j+1} are both positive, then $\mathcal{T}_{f,s}[t_1, \ldots, t_{k+1}] = 0$ and the claim holds with all $a_{j,\nu,t_{\ell:(k+1)}} = 0$. Otherwise, for all $j \le k$ either t_j is negative or t_{j+1} is negative for $j \in [k]$. If $t_{k+1} < 0$, then we have

$$\mathbb{I}(t_{k+1} < 0) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1] + \mathbb{I}(s \ge 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)]$$
$$= \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1] + \mathbb{I}(s \ge 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)]$$
$$\overset{(*)}{\le} \mathcal{R}_\omega[0, \mathrm{sgn}(t_2), 2\,\mathrm{sgn}(t_3), \ldots, k \cdot \mathrm{sgn}(t_{k+1}), k+1]$$
$$\qquad + \mathcal{R}_\omega[0, \mathrm{sgn}(t_2), 2\,\mathrm{sgn}(t_3), \ldots, k \cdot \mathrm{sgn}(t_{k+1}), -(k+1)]$$
$$\le \sum_{\substack{t_{k+2} = \pm(k+1): \\ t_{k+1} \wedge t_{k+2} < 0}} \mathcal{R}_\omega\big[t_1, \ldots, t_{k+1}, t_{k+2}\big],$$

where $(*)$ is a consequence of (9.11) and $\mathrm{sgn}(x) = 0, 1,$ or -1 denotes the sign of a real number x.

Further note that if $t_{k+1} > 0$, then $\mathbb{I}(t_{k+1} < 0) = 0$ and we get

$$\mathbb{I}(t_{k+1} < 0) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1] + \mathbb{I}(s \ge 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)]$$
$$= \mathbb{I}(s \ge 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)]$$
$$\overset{(*)}{\le} \mathcal{R}_\omega[0, \mathrm{sgn}(t_2), 2\,\mathrm{sgn}(t_3), \ldots, k \cdot \mathrm{sgn}(t_{k+1}), -(k+1)]$$
$$\le \sum_{\substack{t_{k+2} = \pm(k+1): \\ t_{k+1} \wedge t_{k+2} < 0}} \mathcal{R}_\omega\big[t_1, \ldots, t_{k+1}, t_{k+2}\big],$$

where $(*)$ is a consequence of (9.11). Thus, we have shown that

$$\left| \mathcal{T}_{f,s}[t_1, \ldots, t_{k+1}] - \mathcal{S}[t_1, \ldots, t_{k+1}] \, \mathbb{E}[\partial^k f(W)] \right|$$
$$\le |f|_{k,\omega} \big(\mathbb{I}(t_{k+1} < 0) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, k+1] + \mathbb{I}(s \ge 1) \cdot \mathcal{R}_\omega[t_1, \ldots, t_{k+1}, -(k+1)] \big)$$
$$\le |f|_{k,\omega} \sum_{\substack{t_{k+2} = \pm(k+1): \\ t_{k+1} \wedge t_{k+2} < 0}} \mathcal{R}_\omega\big[t_1, \ldots, t_{k+1}, t_{k+2}\big].$$

Now suppose the claim holds for $\ell + 1$ and consider the case of ℓ. By (9.15) we have

$$\left| \mathcal{T}_{f,s}[t_1, \ldots, t_\ell] - \mathcal{S}[t_1, \ldots, t_\ell] \, \mathbb{E}[\partial^{\ell-1} f(W)] \right.$$
$$- \mathbb{I}(s \ge 1) \cdot \sum_{j=1}^{k-\ell+1} \sum_{h=0}^{j} (-1)^h \frac{1}{h!(j-h)!} \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \underbrace{s-\ell, \ldots, s-\ell}_{h \text{ times}}, \underbrace{-\ell, \ldots, -\ell}_{(j-h) \text{ times}}]$$
$$\left. + \mathbb{I}(t_\ell < 0) \sum_{j=1}^{k-\ell+1} \frac{1}{j!} \mathcal{T}_{f,j}[t_1, \ldots, t_\ell, \ell, \underbrace{-\ell, \ldots, -\ell}_{(j-1) \text{ times}}] \right|$$

$$\leq \frac{|f|_{k,\omega}}{(k-\ell+1)!}\big(\mathbb{I}(t_\ell < 0)\cdot \mathcal{R}_\omega[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+1)\text{ times}}]$$

$$+\,\mathbb{I}(s\geq 1)\cdot \mathcal{R}_\omega[t_1,\ldots,t_\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+2)\text{ times}}]\big).$$

Note that $\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}},\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}}]$ and $\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(j-1)\text{ times}}]$ are $\mathcal{T}$-sums of order at least $\ell+j$ $(j\geq 1)$. Therefore, we can apply inductive hypothesis on them. In specific, the remainder term ($\mathcal{R}$-sums) in the expansion of

$$\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}},\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}}]$$

is given by this

$$4^{k-\ell-j+1}|f|_{k,\omega}\sum_{t_{(\ell+j+1):(k+2)}\in\mathcal{M}_{\ell+j+1,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,\underbrace{s-\ell,\ldots,s-\ell}_{h\text{ times}},\underbrace{-\ell,\ldots,-\ell}_{(j-h)\text{ times}},t_{\ell+j+1},\ldots,t_{k+2}\big]$$

$$\overset{(9.11)}{\leq} 4^{k-\ell-j+1}|f|_{k,\omega}\sum_{t_{(\ell+j+1):(k+2)}\in\mathcal{M}_{\ell+j+1,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,-\ell,-(\ell+1),\ldots,-(\ell+j-1),t_{\ell+j+1},\ldots,t_{k+2}\big]$$

$$\leq 4^{k-\ell-j+1}|f|_{k,\omega}\sum_{t_{(\ell+2):(k+2)}\in\mathcal{M}_{\ell+2,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,-\ell,t_{\ell+2},\ldots,t_{k+2}\big] =: 4^{k-\ell-j+1}|f|_{k,\omega}\cdot U_1.$$

Similarly, the remainder term in the expansion of $\mathcal{T}_{f,j}[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(j-1)\text{ times}}]$ is given by

$$4^{k-\ell-j+1}|f|_{k,\omega}\sum_{t_{(\ell+j+1):(k+2)}\in\mathcal{M}_{\ell+j+1,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(j-1)\text{ times}},t_{\ell+j+1},\ldots,t_{k+2}\big]$$

$$\leq 4^{k-\ell-j+1}|f|_{k,\omega}\sum_{t_{(\ell+2):(k+2)}\in\mathcal{M}_{\ell+2,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,\ell,t_{\ell+2},\ldots,t_{k+2}\big] =: 4^{k-\ell-j+1}|f|_{k,\omega}\cdot U_2.$$

Note that $U_1 + \mathbb{I}(t_\ell < 0)\cdot U_2$ is controlled by

$$U_1 + \mathbb{I}(t_\ell < 0)\cdot U_2 \qquad\qquad (9.38)$$

$$= \sum_{t_{(\ell+2):(k+2)}\in\mathcal{M}_{\ell+2,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,-\ell,t_{\ell+2},\ldots,t_{k+2}\big]$$

$$+\,\mathbb{I}(t_\ell < 0)\cdot\sum_{t_{(\ell+2):(k+2)}\in\mathcal{M}_{\ell+2,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,\ell,t_{\ell+2},\ldots,t_{k+2}\big]$$

$$\leq \sum_{t_{(\ell+1):(k+2)}\in\mathcal{M}_{\ell+1,k+2}}\mathcal{R}_\omega\big[t_1,\ldots,t_\ell,t_{\ell+1},\ldots,t_{k+2}\big].$$

As we mentioned above, by inductive hypothesis we have that there exist coefficients Q_j satisfying (9.36) such that

$$\left|\mathcal{T}_{f,s}[t_1,\ldots,t_\ell] - \sum_{j=\ell}^{k+1}Q_j\,\mathbb{E}[\partial^{j-1}f(W)]\right|$$

$$\leq \sum_{j=1}^{k-\ell+1}\sum_{h=0}^{j}\frac{1}{h!(j-h)!}4^{k-\ell-j+1}|f|_{k,\omega}\cdot U_1 + \mathbb{I}(t_\ell<0)\sum_{j=1}^{k-\ell+1}\frac{1}{j!}4^{k-\ell-j+1}|f|_{k,\omega}\cdot U_2$$

$$+\,\frac{|f|_{k,\omega}}{(k-\ell+1)!}\big(\mathbb{I}(t_\ell<0)\cdot\mathcal{R}_\omega[t_1,\ldots,t_\ell,\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+1)\text{ times}}] + \mathcal{R}_\omega[t_1,\ldots,t_\ell,\underbrace{-\ell,\ldots,-\ell}_{(k-\ell+2)\text{ times}}]\big).$$

Noting that $\sum_{h=0}^{j} 1/(h!(j-h)!) = 2^j/j!$, we have

$$\left| \mathcal{T}_{f,s}[t_1, \ldots, t_\ell] - \sum_{j=\ell}^{k+1} Q_j \, \mathbb{E}[\partial^{j-1}f(W)] \right|$$

$$\leq \sum_{j=1}^{k-\ell+1} \frac{2^j \cdot 4^{k-\ell-j+1}}{j!} |f|_{k,\omega} \cdot \big(U_1 + \mathbb{I}(t_\ell < 0) \cdot U_2\big)$$

$$+ \frac{|f|_{k,\omega}}{(k-\ell+1)!} \big(\mathbb{I}(t_\ell < 0) \cdot U_2 + U_1\big)$$

$$\leq \big(1 + \sum_{j=1}^{k-\ell+1} 2^{2k-2\ell-j+2}\big)|f|_{k,\omega}\big(U_1 + \mathbb{I}(t_\ell < 0) \cdot U_2\big)$$

$$\leq 4^{k-\ell+1}|f|_{k,\omega}\big(U_1 + \mathbb{I}(t_\ell < 0) \cdot U_2\big)$$

$$\overset{(9.38)}{\leq} 4^{k-\ell+1}|f|_{k,\omega} \sum_{t_{(\ell+1):(k+2)} \in \mathcal{M}_{\ell+1,k+2}} \mathcal{R}_\omega\big[t_1, \ldots, t_\ell, t_{\ell+1}, \ldots, t_{k+2}\big].$$

Thus, we have shown (9.37).

Finally, we note that for all $t_{1:\ell} \in \mathcal{M}_{1,\ell}$ and then by (9.11) we have

$$\sum_{t_{(\ell+1):(k+1)} \in \mathcal{M}_{\ell+1,k+2}} \mathcal{R}_\omega\big[t_1, \ldots, t_\ell, \ldots, t_{k+2}\big]$$

$$\leq \sum_{t_{(\ell+1):(k+2)} \in \mathcal{M}_{\ell+1,k+2}} \mathcal{R}_\omega\big[0, \mathrm{sgn}(t_2), 2\,\mathrm{sgn}(t_3), \ldots, (\ell-1)\,\mathrm{sgn}(t_\ell), t_{\ell+1}, \ldots, t_{k+2}\big]$$

$$\leq \sum_{t_{1:(k+2)} \in \mathcal{M}_{1,k+2}} \mathcal{R}_\omega[t_1, t_2, \ldots, t_{k+2}] = R_{k,\omega}. \qquad \square$$

We remark that if f is a polynomial of degree at most k, then the Hölder constant $|f|_{k,\omega} = 0$ and hence the remainder $C_{k,\ell}|f|_{k,\omega}R_{k,\omega}$ vanishes.

For any $\mathcal{T}$-sum, we have established the existence of expansions in Theorem 9.3. Next we show the uniqueness of such expansions.

Lemma 9.5 (Uniqueness). *Under the same settings as Theorem 9.3, suppose that there exist two sets of coefficients $Q_\ell, \ldots, Q_{k+1}$ and $Q'_\ell, \ldots, Q'_{k+1}$ only depending on s and $t_{1:\ell}$, and the joint distribution of $(X_i)_{i \in I}$ such that for any polynomial f of degree at most ℓ, we have*

$$\mathcal{T}_{f,s}[t_1, \ldots, t_\ell] = Q_\ell \mathbb{E}[\partial^{\ell-1}f(W)] + \cdots + Q_{k+1}\mathbb{E}[\partial^k f(W)]$$

$$= Q'_\ell \mathbb{E}[\partial^{\ell-1}f(W)] + \cdots + Q'_{k+1}\mathbb{E}[\partial^k f(W)].$$

Then $Q_j = Q'_j$ for any $j = \ell, \ldots, k+1$.

Proof. We prove this lemma by contradiction.

Let j be the smallest number such that $Q_j \neq Q'_j$. Since the coefficients $Q_\ell, \ldots, Q_{k+1}$ do not depend on f, we can choose $f(x) = cx^{j-1}$ such that $\partial^{j-1}f(x) = c(j-1)! \neq 0$. But $Q_{j+1}\mathbb{E}[\partial^j f(W)] = \cdots = Q_{k+1}\mathbb{E}[\partial^k f(W)] = 0$, which implies $cQ_j = cQ'_j$. This is a contradiction. Therefore, $Q_j = Q'_j$ for any $j = \ell, \ldots, k+1$. $\qquad \square$

Proof of Proposition 9.1. Applying Theorem 9.3 with $\ell = 1$, and $s = t_1 = t_2 = 0$, we have for any $f \in \mathcal{C}^{k,\omega}(\mathbb{R})$,

$$\mathbb{E}[Wf(W)] = \sum_{i_1 \in I} \mathbb{E}[X_{i_1}f(W)] = \mathcal{T}_{f,0}[0] = \sum_{j=1}^{k+1} Q_j \mathbb{E}[\partial^{j-1}f(W)] + \mathcal{O}(|f|_{k,\omega}R_{k,\omega}),$$

for some $Q_1, \ldots, Q_{k+1}$ that only depend on the distribution of $(X_i)_{i \in I}$ and where $R_{k,\omega}$ is defined in (9.13). Suppose that f is a polynomial of degree at most k, then we observe that $f \in \mathcal{C}^{k,\omega}(\mathbb{R})$ and $|f|_{k,\omega} = 0$. Thus, this implies that

$$\mathcal{T}_{f,0}[0] = \mathbb{E}[Wf(W)] = \sum_{j=1}^{k+1} Q_j \mathbb{E}[\partial^{j-1} f(W)]. \tag{9.39}$$

On the other hand, for any random variable, the moments $(\mu_j)_{j \geq 0}$ and cumulants $(\kappa_j)_{j \geq 0}$, provided that they exist, are connected through the following relations [40]:

$$\mu_n = \sum_{j=1}^{n} \binom{n-1}{j-1} \kappa_j \mu_{n-j}. \tag{9.40}$$

Using this we will obtain a similar expansion to (9.39) by using the cumulants (κ_j). In this goal, we first remark that if $f(x) = x^j$ where $j \leq k$, then by using (9.40) we obtain that

$$\mathbb{E}[Wf(W)] = \mu_{j+1}(W) = \sum_{h=1}^{j+1} \binom{j}{h-1} \kappa_h(W) \mu_{j+1-h}(W)$$

$$= \sum_{h=0}^{j} \binom{j}{h} \kappa_{h+1}(W) \mu_{j-h}(W) = \sum_{h=0}^{k} \frac{\kappa_{h+1}(W)}{h!} \mathbb{E}[\partial^h f(W)].$$

Moreover, we remark that this can be generalized to arbitrary polynomials f of degree k. Indeed, any polynomial f of degree k can be written as $f(x) = \sum_{j=0}^{k} a_j x^j$ for certain coefficients (a_j). By the linearity of expectations, we know that

$$\mathbb{E}[Wf(W)] = \sum_{j=0}^{k} \frac{\kappa_{j+1}(W)}{j!} \mathbb{E}[\partial^j f(W)].$$

Compare this to (9.39) and apply Lemma 9.5. We conclude that $Q_j = \kappa_j(W)/(j-1)!$ for any $j \in [k+1]$. In particular, $Q_1 = 0 = \kappa_1(W)$. $\qquad \square$

Next we upper-bound the cumulants of W using $R_{k,1}$.

Corollary 9.6 (Bounds for Cumulants). *For any $k \in \mathbb{N}_+$, there exists a constant C_k that only depends on k such that $\left|\kappa_{k+2}(W)\right| \leq C_k R_{k,1}$.*

Proof. Let $f(x) = x^{k+1}/(k+1)!$. We remark that $f \in \Lambda_{k+1}$ where $\Lambda_{k+1} := \{f \in \mathcal{C}^{k,1}(\mathbb{R}) : |f|_{k,1} \leq 1\}$. Moreover, by using Proposition 9.1 we have

$$\mathbb{E}[Wf(W)] = \sum_{j=1}^{k} \frac{\kappa_{j+1}(W)}{j!} \mathbb{E}[\partial^j f(W)] + \mathcal{O}(R_{k,1}).$$

Here the constant dropped from the big $\mathcal{O}$ analysis is controlled by 4^k. On the other hand, by (9.40) we have

$$\mathbb{E}[Wf(W)] = \frac{1}{(k+1)!} \mu_{k+2}(W)$$

$$= \sum_{j=1}^{k+1} \binom{k+1}{j} \kappa_{j+1}(W) \mu_{k+1-j}(W)$$

$$= \sum_{j=1}^{k} \frac{\kappa_{j+1}(W)}{j!} \mathbb{E}[\partial^j f(W)] + \frac{\kappa_{k+2}(W)}{(k+1)!}.$$

Thus, there exists C_k such that $\left|\kappa_{k+2}(W)\right| \leq C_k R_{k,1}$. $\qquad \square$

Finally, we are able to prove Lemma 8.7 based on Proposition 9.1 and Corollary 9.6.

Proof of Lemma 8.7. We perform induction on $k := \lceil p \rceil$. We start with $k = 1$. In this goal, we first remark that by Lemma 8.3, we have $f = \Theta h \in \mathcal{C}^{1,\omega}(\mathbb{R})$ and that $|f|_{1,\omega}$ is bounded by a constant. Moreover, as $f = \Theta h$ is the solution to the Stein equation (8.2). By Proposition 9.1 we obtain that

$$\mathbb{E}[h(W)] - \mathcal{N}h = \mathbb{E}[f'(W)] - \mathbb{E}[Wf(W)] = \mathcal{O}(R_{1,\omega}).$$

Therefore, the desired result is established for 1. Suppose that the proposition holds for $1, \ldots, k-1$, we want to prove that it will also hold for k. Let $f = \Theta h$, then by Lemma 8.3 we know that $f \in \mathcal{C}^{k,\omega}(\mathbb{R})$ and that $|f|_{k,\omega}$ is bounded by some constant that only depends on k, ω. Thus, by Proposition 9.1, we have

$$\mathbb{E}[Wf(W)] = \sum_{j=1}^{k} \frac{\kappa_{j+1}(W)}{j!} \mathbb{E}[\partial^j f(W)] + \mathcal{O}(R_{k,\omega}).$$

Hence we have the following expansion of the Stein equation

$$\mathbb{E}[h(W)] - \mathcal{N}h = \mathbb{E}[f'(W)] - \mathbb{E}[Wf(W)] \tag{9.41}$$

$$= -\sum_{j=2}^{k} \frac{\kappa_{j+1}(W)}{j!} \mathbb{E}[\partial^j f(W)] + \mathcal{O}(R_{k,\omega})$$

$$= -\sum_{j=1}^{k-1} \frac{\kappa_{j+2}(W)}{(j+1)!} \mathbb{E}[\partial^{j+1} \Theta h(W)] + \mathcal{O}(R_{k,\omega}).$$

Noting that $\partial^{j+1}\Theta h \in \mathcal{C}^{k-j-1,\omega}(\mathbb{R})$ and $|\partial^{j+1}\Theta h|_{k-j-1,\omega}$ is bounded by a constant only depending on k, ω, then by inductive hypothesis we obtain that

$$\mathbb{E}[\partial^{j+1}\Theta h(W)] - \mathcal{N}[\partial^{j+1}\Theta h] \tag{9.42}$$

$$= \sum_{(r,s_{1:r})\in\Gamma(k-j-1)} (-1)^r \prod_{\ell=1}^{r} \frac{\kappa_{s_\ell+2}(W)}{(s_\ell+1)!} \mathcal{N}\left[\prod_{\ell=1}^{r}(\partial^{s_\ell+1}\Theta)\circ \partial^{j+1}\Theta\, h\right]$$

$$+ \mathcal{O}\left(\sum_{\ell=1}^{k-j-1} R_{\ell,1}^{(k-j-1+\omega)/\ell} + \sum_{\ell=1}^{k-j} R_{\ell,\omega}^{(k-j-1+\omega)/(\ell+\omega-1)}\right),$$

where we denoted $\Gamma(k-j-1) := \left\{r, s_{1:r} \in \mathbb{N}_+ : \sum_{\ell=1}^{r} s_\ell \leq k-j-1\right\}$.

By Corollary 9.6 and Young's inequality, we have

$$\left|\kappa_{j+2}(W)R_{\ell,\omega}^{\frac{k-j+\omega-1}{\ell+\omega-1}}\right| \lesssim R_{j,1}R_{\ell,\omega}^{\frac{k-j+\omega-1}{\ell+\omega-1}} \leq \frac{j}{k+\omega-1}R_{j,1}^{\frac{k+\omega-1}{j}} + \frac{k-j+\omega-1}{k+\omega-1}R_{\ell,\omega}^{\frac{k+\omega-1}{\ell+\omega-1}},$$

$$\left|\kappa_{j+2}(W)R_{\ell,1}^{\frac{k-j+\omega-1}{\ell}}\right| \lesssim R_{j,1}R_{\ell,1}^{\frac{k-j+\omega-1}{\ell}} \leq \frac{j}{k+\omega-1}R_{j,1}^{\frac{k+\omega-1}{j}} + \frac{k-j+\omega-1}{k+\omega-1}R_{\ell,1}^{\frac{k+\omega-1}{\ell}}. \tag{9.43}$$

Thus, we derive that

$$\mathbb{E}[h(W)] - \mathcal{N}h$$

$$\overset{(9.41)}{=} -\sum_{j=1}^{k-1} \frac{\kappa_{j+2}(W)}{(j+1)!} \mathbb{E}[\partial^{j+1}\Theta h(W)] + \mathcal{O}(R_{k,\omega})$$

$$\overset{(9.42)}{=} -\sum_{j=1}^{k-1}\frac{\kappa_{j+2}(W)}{(j+1)!}\mathcal{N}[\partial^{j+1}\Theta h] + \sum_{j=1}^{k-1}\frac{\kappa_{j+2}(W)}{(j+1)!}\cdot$$

$$\sum_{(r,s_{1:r})\in\Gamma(k-j-1)}(-1)^r\prod_{\ell=1}^{r}\frac{\kappa_{s_\ell+2}(W)}{(s_\ell+1)!}\mathcal{N}\left[\prod_{\ell=1}^{r}(\partial^{s_\ell+1}\Theta)\circ\partial^{j+1}\Theta\, h\right]$$

$$+\mathcal{O}\Bigg(R_{k,\omega} + \sum_{j=1}^{k-1}|\kappa_{j+2}(W)|\sum_{\ell=1}^{k-j-1}R_{\ell,1}^{(k+\omega-j-1)/\ell}$$

$$+\sum_{j=1}^{k-1}|\kappa_{j+2}(W)|\sum_{\ell=1}^{k-j}R_{\ell,\omega}^{(k+\omega-j-1)/(\ell+\omega-1)}\Bigg)$$

$$\overset{(9.43)}{=} -\sum_{j=1}^{k-1}\frac{\kappa_{j+2}(W)}{(j+1)!}\mathcal{N}[\partial^{j+1}\Theta h] + \sum_{j=1}^{k-1}\frac{\kappa_{j+2}(W)}{(j+1)!}\cdot$$

$$\sum_{(r,s_{1:r})\in\Gamma(k-j-1)}(-1)^r\prod_{\ell=1}^{r}\frac{\kappa_{s_\ell+2}(W)}{(s_\ell+1)!}\mathcal{N}\left[\prod_{\ell=1}^{r}(\partial^{s_\ell+1}\Theta)\circ\partial^{j+1}\Theta\, h\right]$$

$$+\mathcal{O}\Bigg(R_{k,\omega} + \sum_{j=1}^{k-1}R_{j,1}^{(k+\omega-1)/j} + \sum_{j=1}^{k-1}\sum_{\ell=1}^{k-j-1}R_{\ell,1}^{(k+\omega-1)/\ell}$$

$$+\sum_{j=1}^{k-1}\sum_{\ell=1}^{k-j}R_{\ell,\omega}^{(k+\omega-1)/(\ell+\omega-1)}\Bigg)$$

$$= \sum_{(r,s_{1:r})\in\Gamma(k-1)}(-1)^r\prod_{\ell=1}^{r}\frac{\kappa_{s_\ell+2}(W)}{(s_\ell+1)!}\mathcal{N}\left[\prod_{\ell=1}^{r}(\partial^{s_\ell+1}\Theta)\, h\right]$$

$$+\mathcal{O}\Bigg(\sum_{\ell=1}^{k-1}R_{\ell,1}^{(k+\omega-1)/\ell} + \sum_{\ell=1}^{k}R_{\ell,\omega}^{(k+\omega-1)/(\ell+\omega-1)}\Bigg).$$

Therefore, the desired property was established by induction. □

10 Proof of Lemma 8.8

In Lemma 8.8, we would like to find a random variable with a given sequence of real numbers as its cumulants. Constructing a random variable from its cumulants can be difficult in practice. However, there is a rich literature on establishing the existence of a random variable given the moment sequence. And it is well-known that the moments can be recovered from the cumulants, and vice versa. The explicit expression between moments μ_n and cumulants κ_n is achieved by using the Bell polynomials, i.e.,

$$\mu_n = B_n(\kappa_1,\ldots,\kappa_n) = \sum_{j=1}^{n}B_{n,j}(\kappa_1,\ldots,\kappa_{n-j+1}), \tag{10.1}$$

$$\kappa_n = \sum_{j=1}^{n}(-1)^{j-1}(j-1)!B_{n,j}(\mu_1,\ldots,\mu_{n-j+1}), \tag{10.2}$$

where B_n and $B_{n,j}$ are the exponential Bell polynomial defined by

$$B_n(x_1,\ldots,x_n) := \sum_{j=1}^{n}B_{n,j}(x_1,x_2,\ldots,x_{n-j+1}),$$

$$B_{n,j}(x_1,x_2,\ldots,x_{n-j+1}) := \sum\frac{n!}{i_1!i_2!\cdots i_{n-j+1}!}\left(\frac{x_1}{1!}\right)^{i_1}\left(\frac{x_2}{2!}\right)^{i_2}\cdots\left(\frac{x_{n-j+1}}{(n-j+1)!}\right)^{i_{n-j+1}}.$$
$$\tag{10.3}$$

The sum here is taken over all sequences $i_1, \ldots, i_{n-j+1}$ of non-negative integers such that the following two conditions are satisfied:

$$i_1 + i_2 + \cdots + i_{n-j+1} = j,$$
$$i_1 + 2i_2 + \cdots + (n-j+1)i_{n-j+1} = n.$$

In mathematics, the classical *moment problem* is formulated as follows: Given a sequence $(\mu_i)_{i \geq 0}$, does there exist a random variable defined on a given interval such that $\mu_j = \mathbb{E}[X^j]$ for any non-negative integer j? There are three essentially different types of (closed) intervals. Either two end-points are finite, one end-point is finite, or no end-points are finite, which corresponds to the *Hamburger*, *Hausdorff*, and *Stieltjes* moment problem respectively. See [2, 7] or [43] for a detailed discussion. For our purpose, there is no restriction on the support of random variables. Thus, the following lemma for the Hamburger moment problem is all we need.

Lemma 10.1. *The Hamburger moment problem is solvable, i.e., $(\mu_j)_{j \geq 0}$ is a sequence of moments if and only if $\mu_0 = 1$ and the corresponding Hankel kernel*

$$H = \begin{pmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots \\ \mu_1 & \mu_2 & \mu_3 & \cdots \\ \mu_2 & \mu_3 & \mu_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{10.4}$$

is positive definite, i.e.,

$$\sum_{j,k \geq 0} \mu_{j+k} c_j c_k \geq 0$$

for every real sequence $(c_j)_{j \geq 0}$ with finite support, i.e., $c_j = 0$ except for finitely many j's.

If we define the $(j+1)$-th upper-left determinant of a Hankel matrix by

$$H_j(x_0, x_1, \ldots, x_{2j}) := \begin{vmatrix} x_0 & x_1 & \cdots & x_j \\ x_1 & x_2 & \cdots & x_{j+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_j & x_{j+1} & \cdots & x_{2j} \end{vmatrix}, \tag{10.5}$$

by Sylvester's criterion in linear algebra [20], the positive-definite condition above is equivalent to $H_j(\mu_0, \ldots, \mu_{2j}) > 0$ for any $j \in \mathbb{N}_+$.

In order to prove Lemma 8.8, we construct a Hankel matrix from given values of cumulants and ensure that the upper-left determinants of (10.4) are all positive. Then by Lemma 10.1, there exists a random variable that has matched moments with the ones in (10.4) and hence it also has the required cumulants by (10.2).

For convenience, we write

$$L_j(x_1, \ldots, x_{2j}) := H_j(1, B_1(x_1), B_2(x_1, x_2), \ldots, B_{2j}(x_1, \ldots, x_{2j})).$$

Taking $x_1 = 0$, from the definitions (10.3) and (10.5), there is an expansion

$$L_j(0, x_2, \ldots, x_{2j}) = H_j(1, 0, B_2(0, x_2), \ldots, B_{2j}(0, x_2, \ldots, x_{2j})) = \sum a_{t_2, \ldots, t_{2j}}^{(j)} x_2^{t_2} \cdots x_{2j}^{t_{2j}}, \tag{10.6}$$

where the sum is taken over

$$t_2 + t_3 + \cdots + t_{2j} \geq j,$$
$$2t_2 + 3t_3 + \cdots + (2j)t_{2j} = j(j+1). \tag{10.7}$$

To illustrate, set $B_0 = 1$. By (10.5) each term in $L_j(x_1, \ldots, x_{2j}) = H_j(B_0, B_1, \ldots, B_{2j})$ is a product of the form $\prod_{s=1}^{j+1} B_{\ell_s}$ (coefficients omitted) such that $\sum_{s=1}^{j+1} \ell_s = j(j+1)$. Note that by (10.3) each monomial M in $B_\ell(x_1, \ldots, x_\ell)$ satisfies that

$$\deg_{x_1}(M) + \deg_{x_2}(M) + \cdots + \deg_{x_{2j}}(M) \geq 1 \quad \text{for } \ell \geq 1,$$
$$\deg_{x_1}(M) + 2\deg_{x_2}(M) + \cdots + (2j)\deg_{x_{2j}}(M) = \ell.$$

Thus, each monomial $a^{(j)}_{t_1, \ldots, t_{2j}} x_1^{t_1} \cdots x_{2j}^{t_{2j}}$ in $L_j(x_1, \ldots, x_{2j})$ satisfies that

$$t_1 + t_2 + \cdots + t_{2j} \geq \left| \{ s \in [j+1] : \ell_s \geq 1 \} \right| \geq j,$$
$$t_1 + 2t_2 + \cdots + (2j)t_{2j} = \sum_{s=1}^{j+1} \ell_s = j(j+1).$$

Now if we take $x_1 = 0$, then $t_1 = 0$ holds for all remaining terms, and (10.7) follows.

We further define in the following way a sequence of univariate polynomials which will be essential in our construction in Lemma 8.8, by setting

$$P_j(x) := L_j(0, 1, x, x^2, x^3, \ldots, x^{2j-2}).$$

Firstly, we present a lemma on the properties of $P_j(x)$.

Lemma 10.2. $P_j(x)$ *is a polynomial of degree at most $j(j-1)$ with only even-degree terms and if we write*

$$P_j(x) = \sum_{\ell=0}^{j(j-1)/2} b^{(j)}_{2\ell} x^{2\ell},$$

we have $b^{(j)}_0 = a^{(j)}_{j(j+1)/2, 0, \ldots, 0} \geq 2$ for any $j \geq 2$, $j \in \mathbb{N}_+$.

Proof. Note that by applying (10.6) we obtain that

$$P_j(x) = L_j(0, 1, x, \ldots, x^{2j-2}) = \sum a^{(j)}_{t_2, \ldots, t_{2j}} x^{t_3 + 2t_4 + \cdots + (2j-2)t_{2j}}, \qquad (10.8)$$

where the sum is taken over

$$t_2 + t_3 + \cdots + t_{2j} \geq j,$$
$$2t_2 + 3t_3 + \cdots + (2j)t_{2j} = j(j+1).$$

The degree of each term in (10.8) is

$$t_3 + 2t_4 + \cdots + (2j-2)t_{2j}$$
$$= (2t_2 + 3t_3 + \cdots + (2j)t_{2j}) - 2(t_2 + t_3 + \cdots + t_{2j})$$
$$= j(j+1) - 2(t_2 + t_3 + \cdots + t_{2j}).$$

This is even and no greater than $j(j-1)$ since $t_2 + t_3 + \cdots + t_{2j} \geq j$.

Then we show the constant term $b^{(j)}_0 \geq 2$. Consider a standard normal random variable $\xi \sim \mathcal{N}(0,1)$. Then $\kappa_j(\xi) = 0$ for all $j \geq 1, j \neq 2$, and $\kappa_2(\xi) = 1$, which is straightforward by checking that the moment generating function of ξ is $\exp(t^2/2)$. By Lemma 10.1, we have

$$b^{(j)}_0 = P_j(0) = L_j(0, 1, 0, \ldots, 0)$$
$$= L_j(\kappa_1(\xi), \kappa_2(\xi), \ldots, \kappa_{2j}(\xi))$$
$$= H_j(\mu_0(\xi), \mu_1(\xi), \ldots, \mu_{2j}(\xi)) > 0.$$

Since $\mu_{2\ell}(\xi) = (2\ell - 1)!!$ and $\mu_{2\ell-1}(\xi) = 0$ are integers for $\ell \in \mathbb{N}_+$, $b^{(j)}_0$ is also an integer. Checking Leibniz formula of the determinant for the Hankel matrix H_j [26], we observe

that there is an even number of terms and that each term is odd. In specific, the determinant for the Hankel matrix is given by

$$b_0^{(j)} = H_j(\mu_0(\xi), \mu_1(\xi), \ldots, \mu_{2j}(\xi)) = \sum_{\tau \in S_j} \text{sgn}(\tau) \prod_{i=1}^{j} \mu_{\tau(i)+i-2}(\xi),$$

where by abuse of notation sgn is the sign function of permutations in the j-th permutation group S_j, which returns $+1$ and -1 for even and odd permutations, respectively. Since $\mu_{2\ell}(\xi) = (2\ell-1)!!$ and $\mu_{2\ell-1}(\xi) = 0$ for all $\ell \in \mathbb{N}_+$, we have

$$\text{sgn}(\tau) \prod_{i=1}^{j} \mu_{\tau(i)+i-2}(\xi) \begin{cases} \text{is odd} & \text{if } \tau(i)+i \text{ is even } \forall i = 1, \ldots, j \\ = 0 & \text{otherwise} \end{cases}.$$

Noting that the number of permutations τ that satisfies $\tau(i) + i$ is even for all $i = 1, \ldots, j$ is $(j!)^2$, which is even when $j \geq 2$, we conclude that $b_0^{(j)}$ is even, and thus, it should be at least 2. $\qquad \square$

As we have explained at the beginning of this section, we would like to construct a 'moment' sequence such that the corresponding Hankel kernel is positive definite. The following lemma offers one single step in the construction.

Lemma 10.3. *Suppose there is some constant C such that $|\mu_\ell| \leq C$ for $\ell = 1, \ldots, 2j+1$ and $H_j(\mu_0, \ldots, \mu_{2j}) \geq 1$. Then there exists C' only depending on j and C such that*

$$H_{j+1}(\mu_0, \ldots, \mu_{2j}, \mu_{2j+1}, C') \geq 1.$$

Proof. Let $C' = (j+1)(j+1)!C^{j+2} + 1$. Then by the Laplace expansion [26] of the determinant, we have

$$H_{j+1}(\mu_0, \ldots, \mu_{2j}, \mu_{2j+1}, C') = C'H_j(\mu_0, \ldots, \mu_{2j}) + \sum_{\ell=0}^{j} (-1)^{j+1+\ell} \mu_{j+1+\ell} A_{j+2,\ell+1}$$

$$\geq C' - (j+1)C \cdot (j+1)!C^{j+1} \geq 1,$$

where $A_{j+2,\ell+1}$ is the determinant of the $(j+1) \times (j+1)$ submatrix obtained by deleting the $(j+2)$-th row and $(\ell+1)$-th column of

$$A = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_{j+1} \\ \mu_1 & \mu_2 & \cdots & \mu_{j+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{j+1} & \mu_{j+2} & \cdots & C' \end{pmatrix}. \qquad \square$$

Now we prove Lemma 8.8.

Proof of Lemma 8.8. The key of the proof will be to use Lemma 10.1. To do so we need to postulate an infinite sequence that will be our candidates for of potential moments and check that the conditions of Lemma 10.1 hold. We remark that as we already know what we want the first $k+1$ cumulants to be, we already know what the candidates are for the first $k+1$ moments; and we only to find adequate proposal for the $(k+2)$-th moment onward. We will do so by iteratively using Lemma 10.3.

In this goal, we remark that since by Lemma 10.2 we know that $b_0^{(j)} \geq 2$. Therefore, we can choose a small enough constant $0 < C_p < 1$ only depending on $k = \lceil p \rceil$ such that

$$b_0^{(j)} - \sum_{\ell=1}^{j(j-1)/2} \sum_{\substack{2t_2+2t_3+\cdots+2t_{2j}=j(j+1)-2\ell \\ 2t_2+3t_3+\cdots+2jt_{2j}=j(j+1)}} |a_{t_2,\ldots,t_{2j}}^{(j)}| C_p^{2\ell} \geq 1, \tag{10.9}$$

for any integer $j = 1, \ldots, \lceil k/2 \rceil$. Given an index set I_n, if $u_j^{(n)} = 0$ for all $j = 1, \ldots, k-1$, let $\xi^{(n)} \sim \mathcal{N}(0,1)$ and $q_n \gg |I_n|$. Then q_n and $\xi^{(n)}$ satisfy all the requirements since $\kappa_j(\xi^{(n)}) = 0$ for all $j \in \mathbb{N}_+, j \neq 2$ and $\kappa_2(\xi^{(n)}) = 1$, which is straightforward by checking that the momemt generating function of $\xi^{(n)}$ is $\exp(t^2/2)$.

Otherwise, let

$$q_n := \left\lfloor \min_{1 \leq j \leq k-1, u_j^{(n)} \neq 0} \left\{ C_p^2 |u_j^{(n)}|^{-2/j} \right\} \right\rfloor, \tag{10.10}$$

where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. Since by assumption, for any $j = 1, \ldots, k-1$, $u_j^{(n)} \to 0$ as $n \to \infty$, then we know that there exists $N > 0$ such that (i) $q_n \geq 1$ for any $n > N$ and (ii) $q_n \to \infty$ as $n \to \infty$. We note that by definition $\min_{1 \leq j \leq k-1, u_j^{(n)} \neq 0} \left\{ C_p^2 |u_j^{(n)}|^{-2/j} \right\} < q_n + 1$, which implies

$$\max_{1 \leq j \leq k-1} \left\{ q_n^{j/2} |u_j^{(n)}| \right\} > C_p^j \left(q_n/(q_n+1) \right)^{j/2} > C_p^p / 2^{p/2}. \tag{10.11}$$

On the other hand, (10.10) also implies that $C_p^2 |u_j^{(n)}|^{-2/j} \geq q_n$. Thus, $q_n^{j/2} |u_j^{(n)}| \leq C_p^j$. Now let $\widetilde{\kappa}_{j+2} := q_n^{j/2} u_j^{(n)}$. We remark that $|\widetilde{\kappa}_{j+2}| \leq C_p^j$ and $|\widetilde{\kappa}_{j+2}| \geq C_p^p / 2^{p/2}$. We write $\widetilde{\mu}_{j+2} := B_{j+2}(0, \widetilde{\kappa}_2, \ldots, \widetilde{\kappa}_{j+2})$ for $j = 1, \ldots, k-1$. Those will be our candidates for the first $k+1$ moments. Moreover, if k is odd, we also propose a candidate for $(k+2)$-th moment by setting $\widetilde{\mu}_{k+2} := 0$.

For $j = 1, \ldots, \lceil k/2 \rceil$ by (10.6) we have

$$H_j(1, 0, \widetilde{\mu}_2, \widetilde{\mu}_3, \ldots, \widetilde{\mu}_{2j}) = L_j(0, \widetilde{\kappa}_2, \widetilde{\kappa}_3, \ldots, \widetilde{\kappa}_{2j})$$

$$= \sum_{\substack{2t_2 + 3t_3 + \cdots + 2jt_{2j} = j(j+1)}} a_{t_2, \ldots, t_{2j}}^{(j)} \widetilde{\kappa}_2^{t_2} \cdots \widetilde{\kappa}_{2j}^{t_{2j}}$$

$$= \sum_{\ell=0}^{j(j-1)/2} \sum_{\substack{2t_2 + 2t_3 + \cdots + 2t_{2j} = j(j+1) - 2\ell \\ 2t_2 + 3t_3 + \cdots + 2jt_{2j} = j(j+1)}} a_{t_2, \ldots, t_{2j}}^{(j)} \widetilde{\kappa}_2^{t_2} \cdots \widetilde{\kappa}_{2j}^{t_{2j}}$$

$$\overset{(a)}{\geq} b_0^{(j)} - \sum_{\ell=1}^{j(j-1)/2} \sum_{\substack{2t_2 + 2t_3 + \cdots + 2t_{2j} = j(j+1) - 2\ell \\ 2t_2 + 3t_3 + \cdots + 2jt_{2j} = j(j+1)}} \left| a_{t_2, \ldots, t_{2j}}^{(j)} \widetilde{\kappa}_2^{t_2} \cdots \widetilde{\kappa}_{2j}^{t_{2j}} \right|$$

$$\overset{(b)}{\geq} b_0^{(j)} - \sum_{\ell=1}^{j(j-1)/2} \sum_{\substack{2t_2 + 2t_3 + \cdots + 2t_{2j} = j(j+1) - 2\ell \\ 2t_2 + 3t_3 + \cdots + 2jt_{2j} = j(j+1)}} \left| a_{t_2, \ldots, t_{2j}}^{(j)} \right| C_p^{2\ell} \overset{(c)}{\geq} 1,$$

where to get (a) we used the definition of $b_0^{(j)}$, and where to obtain (b) we used the fact that $|\widetilde{\kappa}_{j+2}| \leq C_p^j$, and where to get (c) we used (10.9). Moreover, as $|\widetilde{\kappa}_{j+2}| \leq C_p^j$, then we know that there exists some constant C_p' such that $|\widetilde{\mu}_{j+2}| = |B_{j+2}(0, \widetilde{\kappa}_2, \ldots, \widetilde{\kappa}_{j+2})| \leq C_p'$ for any integer $j = 1, \ldots, 2\lceil k/2 \rceil - 1$. Therefore, by Lemma 10.3, there exists C_p'' depending on $k = \lceil p \rceil$ and C_p' such that

$$H_{\lceil k/2 \rceil + 1}(1, 0, \widetilde{\mu}_2, \ldots, \widetilde{\mu}_{2\lceil k/2 \rceil + 1}, C_p'') \geq 1.$$

Let $\widetilde{\mu}_{2\lceil k/2 \rceil + 2} := C_p''$. Applying Lemma 10.3 repeatedly, we get a sequence $(\widetilde{\mu}_j)_{j \geq 1}$ such that $\widetilde{\mu}_0 = 1$ and $H_j(\widetilde{\mu}_0, \widetilde{\mu}_1, \ldots, \widetilde{\mu}_{2j}) \geq 1 > 0$ for any $j \in \mathbb{N}_+$. The sequence $(\widetilde{\mu}_j)$ is then our candidate for the moments and we remark that they satisfy the conditions of Lemma 10.1. Therefore, by Lemma 10.1, we conclude that there exists $\xi^{(n)}$ such that $\mu_j(\xi^{(n)}) = \widetilde{\mu}_j$ for any $j \in \mathbb{N}_+$. As the first $k+1$ moments uniquely define the first $k+1$ cumulants of a random variable we have $\kappa_{j+2}(\xi^{(n)}) = \widetilde{\kappa}_{j+2} = q_n^{j/2} u_j^{(n)}$ for all $j = 1, \ldots, k-1$. Thus, the q_n and $\xi^{(n)}$ that we have constructed meet the requirements of Lemmas 8.8a and 8.8b.

Moreover, (10.11) implies that Lemma 8.8c is also satisfied. Lastly, to show Lemma 8.8d we note that

$$
\begin{aligned}
\mathbb{E}[|\xi^{(n)}|^{p+2}] = \|\xi^{(n)}\|_{p+2}^{p+2} &\overset{(*)}{\leq} \|\xi^{(n)}\|_{2\lceil k/2\rceil+2}^{p+2} \\
&= \big(\mu_{2\lceil k/2\rceil+2}(\xi^{(n)})\big)^{(p+2)/(2\lceil k/2\rceil+2)} \\
&\leq (C_p'')^{(p+2)/(2\lceil k/2\rceil+2)}.
\end{aligned}
$$

Here $(*)$ is due to the fact that $k = \lceil p\rceil \geq p$. $\hfill\square$

11 Proofs of other results

In this section, we provide the proofs of all the other results.

11.1 Proof of Proposition 3.3

For ease of notation, in this section we will drop the dependence on n in our notation and write W, $N(\cdot)$, σ, X_i, I and $R_{j,\omega}$ for respectively W_n, $N_n(\cdot)$, σ_n, $X_i^{(n)}$, I_n and $R_{j,\omega,n}$.

Before we prove the bounds for $R_{k,\omega}$, we note that $R_{k,\omega}$ can be defined without assuming local dependence [LD*]. Thus, we first aim to generalize this concept, which makes the result derived in Proposition 11.1 also applicable in general dependent situations. Let $(X_i)_{i\in I}$ be a class of mean zero random variables indexed by I. For any graph G (not necessarily the dependency graph) with the vertex set I and a subset $J \subseteq I$, we define $N(J)$ to be vertex set of the neighborhood of J. As in Section 9, we assume $\mathrm{Var}\big(\sum_{i\in I} X_i\big) = 1$, without loss of generality. Let $W = \sum_{i\in I} X_i$.

We extend the notation of $\mathcal{R}$-sums defined in (9.9) to this general setting. Given an integer $k \in \mathbb{N}_+$ such that $k \geq 2$, for any $t_{1:k} \in \mathbb{Z}$ such that $|t_j| \leq j-1$ for any $j \in [k]$, let $z = \big|\{j : t_j > 0\}\big|$. If $z \geq 1$, we write $\{j : t_j > 0\} = \{q_1, \dots, q_z\}$, where the sequence $2 \leq q_1 < \cdots < q_z \leq k$ is taken to be increasing. We further let $q_0 := 1$ and $q_{z+1} := k+1$. Then we could still define the $\mathcal{R}$-sums by

$$
\mathcal{R}_\omega[t_1, t_2, \dots, t_k] :=
$$
$$
\sum_{i_1\in N_1}\sum_{i_2\in N_2}\cdots\sum_{i_{k-1}\in N_{k-1}} [q_1 - q_0, \dots, q_{z+1} - q_z] \rhd \Big(|X_{i_1}|, \dots, |X_{i_{k-1}}|, \big(\sum_{i_k\in N_k} |X_{i_k}|\big)^\omega \Big),
$$

where $N_1 := I$, and for $2 \leq j \leq k$

$$
N_j := \begin{cases} N(i_{1:|t_j|}) = N(i_1, \dots, i_{|t_j|}) & \text{if } t_j \neq 0 \\ \emptyset & \text{if } t_j = 0 \end{cases}.
$$

Now the remainder term $R_{k,\omega}$ is defined as

$$
R_{k,\omega} := \sum_{(\ell,\eta_{1:\ell})\in C^*(k+2)}\sum_{i_1\in N_1'}\sum_{i_2\in N_2'}\cdots\sum_{i_{k+1}\in N_{k+1}'} \tag{11.1}
$$
$$
[\eta_1, \dots, \eta_\ell] \rhd \Big(|X_{i_1}|, \dots, |X_{i_{k+1}}|, \big(\sum_{i_{k+2}\in N_{k+2}'} |X_{i_{k+2}}|\big)^\omega \Big)
$$
$$
= \sum_{t_{1:(k+2)}\in\mathcal{M}_{1,k+2}} \mathcal{R}_\omega[t_1, t_2, \dots, t_{k+2}],
$$

where $N_1' := I$ and $N_j' := N(i_{1:(j-1)})$ for $j \geq 2$. $C^*(k+2)$ and $\mathcal{M}_{1,k+2}$ are given by

$$
C^*(k+2) = \big\{\ell, \eta_{1:\ell} \in \mathbb{N}_+ : \eta_j \geq 2 \ \forall j \in [\ell-1], \ \sum_{j=1}^\ell \eta_j = k+2\big\},
$$

and
$$\mathcal{M}_{1,k+2} := \left\{ t_{1:(k+2)} : \ t_{j+1} = \pm j \ \& \ t_j \wedge t_{j+1} < 0 \ \forall 1 \le j \le k+1 \right\}.$$

Note that the expressions of $\mathcal{R}$-sums and $R_{k,\omega}$ have the same forms as those in Section 9.2, but here we do not impose the assumption of the local dependence of $(X_i)_{i\in I}$ anymore as $N(i_{1:q})$'s are defined directly from the graph structure we constructed on I. The main goal of this section is to prove the following proposition.

Proposition 11.1. *Fix $k \in \mathbb{N}_+$ such that $k \ge 2$ and real number $\omega \in (0,1]$. Let $N(J)$ be defined as above and suppose the cardinality of $N(J)$ is upper-bounded by M for any $|J| \le k$. Then there exists a constant $C_{k+\omega}$ only depending on $k + \omega$ such that*

$$\mathcal{R}_\omega[t_1, t_2, \dots, t_k] \le C_{k+\omega} M^{k-2+\omega} \sum_{i\in I} \mathbb{E}[|X_i|^{k-1+\omega}].$$

Before proving Proposition 11.1, we need the following two lemmas. Lemma 11.2 helps us change the order of summation in $\mathcal{R}_\omega[t_1, \dots, t_k]$ and Lemma 11.3 is a generalized version of Young's inequality, which allows us to bound the expectations of products by sums of moments.

Lemma 11.2. *Fix $k \in \mathbb{N}_+$ such that $k \ge 2$. For any $J \subseteq I$, let $N(J)$ be defined as above. Suppose $(i_1, \dots, i_k)$ is a tuple such that $i_1 \in I$, $i_2 \in N(i_1), \dots, i_k \in N(i_{1:(k-1)})$. Then for any $1 \le h \le k$, there exists a permutation π on $[k]$ such that $\pi(1) = h$, $i_{\pi(1)} \in I$, $i_{\pi(2)} \in N(i_{\pi(1)}), \dots, i_{\pi(k)} \in N(i_{\pi(1)}, \dots, i_{\pi(k-1)})$.*

Proof. We perform induction on k.

Firstly, suppose that $k = 2$, then we remark that $i_2 \in N(i_1) \Leftrightarrow i_1 \in N(i_2)$. For $h = 1$, we can choose π to be the identity and the desired identity holds. For $h = 2$, we let $\pi(1) := 2$ and $\pi(2) := 1$ and remark than once again the desired result holds.

Suppose that the proposition is true for $2, \dots, k-1$. We want to prove that it holds for k. If $h < k$, consider the tuple $(i_1, \dots, i_h)$. By inductive hypothesis, there is a permutation $\widetilde{\pi}$ on $\{1, 2, \dots, h\}$ such that $\widetilde{\pi}(1) = h$, $i_{\widetilde{\pi}(2)} \in N(i_{\widetilde{\pi}(1)}), \dots, i_{\widetilde{\pi}(h)} \in N(i_{\widetilde{\pi}(1)}, \dots, i_{\widetilde{\pi}(q-1)})$. Define

$$\pi(j) := \begin{cases} \widetilde{\pi}(j) & \text{if } 1 \le j \le h \\ j & \text{if } h+1 \le j \le k \end{cases}.$$

Then π satisfies the requirements in the lemma.

Now suppose $h = k$. $i_k \in N(i_{1:(k-1)})$ indicates that i_k is a neighbor of $\{i_1, \dots, i_{k-1}\}$. Then there exists $1 \le \ell \le k-1$ such that there is an edge between i_k and i_ℓ in the graph $G = (I, E)$. Thus, $i_h \in N(i_\ell)$.

By inductive hypothesis, there is a permutation $\widetilde{\pi}$ on $[\ell]$ such that $\widetilde{\pi}(1) = \ell$, $i_{\widetilde{\pi}(2)} \in N(i_{\widetilde{\pi}(1)}), \dots, i_{\widetilde{\pi}(\ell)} \in N(i_{\widetilde{\pi}(1)}, \dots, i_{\widetilde{\pi}(\ell-1)})$.

Define

$$\pi(j) := \begin{cases} k & \text{if } j = 1 \\ \widetilde{\pi}(j-1) & \text{if } 2 \le j \le \ell+1 \\ j-1 & \text{if } \ell+2 \le j \le k \end{cases}.$$

Then $\pi(1) = h = k$. Moreover, we have $i_{\pi(2)} = i_\ell \in N(i_k) = N(i_{\pi(1)})$, and note that for all $j = 3, \dots, \ell$ we have $i_{\pi(j+1)} = i_{\widetilde{\pi}(j)} \in N(i_{\widetilde{\pi}(1)}, \dots, i_{\widetilde{\pi}(j-1)}) = N(i_{\pi(1)}, \dots, i_{\pi(j)})$. Finally, for all $j \ge \ell+1$ we have $i_{\pi(j+1)} = i_j \in N(i_{1:(j-1)}) \subseteq N(i_1, \dots, i_{j-1}, i_k) = N(i_{\pi(1)}, \dots, i_{\pi(j)})$. Thus, the lemma holds for k as well. By induction, the proof is complete. $\square$

Also, we need a generalization of Young's inequality.

Lemma 11.3. *Given* $t \in \mathbb{N}_+$, *let* $Y_1, \ldots, Y_t$ *be a sequence of random variables, and real numbers* $p_1, \ldots, p_t > 1$ *satisfy that* $1/p_1 + \cdots + 1/p_t = 1$. *Then for any* $(\ell, \eta_{1:\ell}) \in C(t) := \{\ell, \eta_{1:\ell} \in \mathbb{N}_+ : \sum_{j=1}^{\ell} \eta_j = t\}$, *we have that*

$$[\eta_1, \ldots, \eta_\ell] \triangleright (|Y_1|, \ldots, |Y_t|) \leq \frac{1}{p_1} \mathbb{E}[|Y_1|^{p_1}] + \cdots + \frac{1}{p_t} \mathbb{E}[|Y_t|^{p_t}]. \tag{11.2}$$

Proof. First, we prove

$$\mathbb{E}[|Y_1 \cdots Y_t|] \leq \frac{1}{p_1} \mathbb{E}[|Y_1|^{p_1}] + \cdots + \frac{1}{p_t} \mathbb{E}[|Y_t|^{p_t}], \tag{11.3}$$

$$\mathbb{E}[|Y_1|] \cdots \mathbb{E}[|Y_t|] \leq \frac{1}{p_1} \mathbb{E}[|Y_1|^{p_1}] + \cdots + \frac{1}{p_t} \mathbb{E}[|Y_t|^{p_t}]. \tag{11.4}$$

In this goal, note that Young's inequality is stated as follows: For any $a_1, \ldots, a_t \geq 0$, and $p_1, \ldots, p_t > 1$ such that $1/p_1 + \cdots + 1/p_t = 1$, we have

$$a_1 \cdots a_t \leq \frac{1}{p_1} a_1^{p_1} + \cdots + \frac{1}{p_t} a_t^{p_t}.$$

Thus, by Young's inequality we know that

$$|Y_1 \cdots Y_t| \leq \frac{1}{p_1} |Y_1|^{p_1} + \cdots + \frac{1}{p_t} |Y_t|^{p_t}.$$

Taking the expectation, we have

$$\mathbb{E}[|Y_1 \cdots Y_t|] \leq \frac{1}{p_1} \mathbb{E}[|Y_1|^{p_1}] + \cdots + \frac{1}{p_t} \mathbb{E}[|Y_t|^{p_t}].$$

Again by Young's inequality, we obtain that

$$\mathbb{E}[|Y_1|] \cdots \mathbb{E}[|Y_t|] \leq \frac{1}{p_1} \mathbb{E}[|Y_1|]^{p_1} + \cdots + \frac{1}{p_t} \mathbb{E}[|Y_t|]^{p_t}.$$

By Jensen's inequality, $\mathbb{E}[|Y_i|]^{p_i} \leq \mathbb{E}[|Y_i|^{p_i}]$ for $i \in [t]$. This implies that

$$\mathbb{E}[|Y_1|] \cdots \mathbb{E}[|Y_t|] \leq \frac{1}{p_1} \mathbb{E}[|Y_1|^{p_1}] + \cdots + \frac{1}{p_t} \mathbb{E}[|Y_t|^{p_t}].$$

Finally, we prove (11.2). Let $1/q_j := \sum_{i=\eta_{j-1}+1}^{\eta_j} 1/p_i$ for $1 \leq j \leq k$.

$$\begin{aligned}
&[\eta_1, \ldots, \eta_\ell] \triangleright (|Y_1|, \ldots, |Y_k|) \\
&= \mathbb{E}\big[|Y_1 \cdots Y_{\eta_1}|\big] \, \mathbb{E}\big[|Y_{\eta_1+1} \cdots Y_{\eta_2}|\big] \, \cdots \, \mathbb{E}\big[|Y_{\eta_1+\cdots+\eta_{\ell-1}+1} \cdots Y_k|\big] \\
&\overset{(11.4)}{\leq} \frac{1}{q_1} \mathbb{E}\big[|Y_1 \cdots Y_{\eta_1}|^{q_1}\big] + \cdots + \frac{1}{q_k} \mathbb{E}\big[|Y_{\eta_1+\cdots+\eta_{\ell-1}+1} \cdots Y_k|^{q_k}\big] \\
&\overset{(11.3)}{\leq} \frac{1}{p_1} \mathbb{E}[|Y_1|^{p_1}] + \cdots + \frac{1}{p_{\eta_1}} \mathbb{E}[|Y_{\eta_1}|^{p_{\eta_1}}] + \cdots \\
&\quad + \frac{1}{p_{\eta_1+\cdots+\eta_{\ell-1}+1}} \mathbb{E}[|Y_{k+1-u_\ell}|^{p_{\eta_1+\cdots+\eta_{\ell-1}+1}}] + \cdots + \frac{1}{p_k} \mathbb{E}[|Y_k|^{p_k}]. \qquad \square
\end{aligned}$$

Now we are ready to prove Proposition 11.1.

Proof of Proposition 11.1. By (9.11), we only need to prove that the following inequality holds for any $k \in \mathbb{N}_+$:

$$\mathcal{R}_\omega[0, \pm 1, \ldots, \pm k] \lesssim M^{k-1+\omega} \sum_{i \in I} \mathbb{E}[|X_i|^{k+\omega}].$$

Once again we write $z := \big|\{j : t_j > 0\}\big|$. If $z \geq 1$, we write $\{j : t_j > 0\} = \{q_1, \ldots, q_z\}$, where $2 \leq q_1 < \cdots < q_z \leq k$ is increasing. Further let $q_0 := 1$ and $q_{z+1} := k + 1$.

Noticing that

$$\underbrace{\frac{1}{k+\omega} + \cdots + \frac{1}{k+\omega}}_{k \text{ times}} + \frac{\omega}{k+\omega} = 1,$$

we apply Lemma 11.3 and obtain that

$$[q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright \left(|X_{i_1}|, \ldots, |X_{i_k}|, \left(\frac{1}{M} \sum_{i_{k+1} \in N(i_{1:k})} |X_{i_{k+1}}| \right)^{\omega} \right) \qquad (11.5)$$

$$\lesssim \mathbb{E}[|X_{i_1}|^{k+\omega}] + \cdots + \mathbb{E}[|X_{i_k}|^{k+\omega}] + \mathbb{E}\left[\left(\frac{1}{M} \sum_{i_{k+1} \in N(i_{1:k})} |X_{i_{k+1}}| \right)^{k+\omega} \right].$$

Now by Jensen's inequality and the fact that $\big|N(i_{1:k})\big| \leq M$, we get that

$$\mathbb{E}\left[\left(\frac{1}{M} \sum_{i_{k+1} \in N(i_{1:k})} |X_{i_{k+1}}| \right)^{k+\omega} \right] \leq \frac{1}{M} \sum_{i_{k+1} \in N(i_{1:k})} \mathbb{E}[|X_{i_{k+1}}|^{k+\omega}].$$

Moreover, we remark that

$$M^{\omega}[q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright \left(|X_{i_1}|, \ldots, |X_{i_k}|, \left(\frac{1}{M} \sum_{i_{k+1} \in N(i_{1:k})} |X_{i_{k+1}}| \right)^{\omega} \right)$$

$$= [q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright \left(|X_{i_1}|, \ldots, |X_{i_k}|, \left(\sum_{i_{k+1} \in N(i_{1:k})} |X_{i_{k+1}}| \right)^{\omega} \right) \qquad (11.6)$$

Thus, this implies that

$$\mathcal{R}_{\omega}[0, \pm 1, \ldots, \pm k] \qquad (11.7)$$

$$= \sum_{i_1 \in I} \cdots \sum_{i_k \in N(i_{1:(k-1)})} [q_1 - q_0, \ldots, q_{z+1} - q_z] \triangleright \left(|X_{i_1}|, \ldots, |X_{i_k}|, \left(\sum_{i_{k+1} \in N(i_{1:k})} |X_{i_{k+1}}| \right)^{\omega} \right)$$

$$\lesssim M^{\omega} \sum_{i_1 \in I} \cdots \sum_{i_k \in N(i_{1:(k-1)})} \left(\mathbb{E}[|X_{i_1}|^{k+\omega}] + \cdots + \mathbb{E}[|X_{i_k}|^{k+\omega}] + \frac{1}{M} \sum_{i_{k+1} \in N(i_{1:k})} \mathbb{E}[|X_{i_{k+1}}|^{k+\omega}] \right).$$

Since the cardinality of $N(i_1), \ldots, N(i_{1:k})$ are bounded by M, for $j = 1$ we have

$$\sum_{i_1 \in I} \sum_{i_2 \in N(i_1)} \cdots \sum_{i_k \in N(i_{1:(k-1)})} \mathbb{E}[|X_{i_j}|^{k+\omega}] \leq M^{k-1} \sum_{i \in I} \mathbb{E}[|X_i|^{k+\omega}]. \qquad (11.8)$$

Now we bound

$$\sum_{i_1 \in I} \sum_{i_2 \in N(i_1)} \cdots \sum_{i_k \in N(i_{1:(k-1)})} \mathbb{E}[|X_{i_j}|^{k+\omega}],$$

where $j = 2, \ldots, k$.

By Lemma 11.2, for any tuple $(i_1, \ldots, i_k)$ in the summation, there exists a permutation π such that $\pi(1) = j$, $i_{\pi(2)} \in N\big(i_{\pi(1)}\big), \ldots, i_{\pi(k)} \in N\big(i_{\pi(1)}, \ldots, i_{\pi(k-1)}\big)$. Let ϕ_j be a map that sends $(i_1, \ldots, i_k)$ to $\big(i_{\pi(1)}, \ldots, i_{\pi(k)}\big)$. Then no more than $(k-1)!$ tuples are mapped to the same destination since $(i_1, \ldots, i_k)$ is a permutation of $\big(i_{\pi(1)}, \ldots, i_{\pi(k)}\big)$ and i_j is fixed to be $i_{\pi(1)}$. Thus, we obtain that

$$\sum_{i_1 \in I} \sum_{i_2 \in N(i_1)} \cdots \sum_{i_k \in N(i_{1:(k-1)})} \mathbb{E}[|X_{i_j}|^{k+\omega}]$$

$$
\begin{aligned}
\leq &(k-1)! \sum_{\pi:\pi(1)=j} \sum_{i_{\pi(1)} \in I} \sum_{i_{\pi(2)} \in N(i_{\pi(1)})} \cdots \sum_{i_{\pi(k)} \in N(i_{\pi(1)},\ldots,i_{\pi(k-1)})} \mathbb{E}[|X_{i_{\pi(1)}}|^{k+\omega}] \\
\leq &(k-1)! \sum_{\pi:\pi(1)=j} \sum_{i_1 \in I} \sum_{i_2 \in N(i_1)} \cdots \sum_{i_k \in N(i_{1:(k-1)})} \mathbb{E}[|X_{i_j}|^{k+\omega}] \\
\leq &((k-1)!)^2 M^{k-1} \sum_{i \in I} \mathbb{E}[|X_i|^{k+\omega}] \lesssim M^{k-1} \sum_{i \in I} \mathbb{E}[|X_i|^{k+\omega}]. \qquad (11.9)
\end{aligned}
$$

Similarly,

$$
\sum_{i_1 \in I} \sum_{i_2 \in N(i_1)} \cdots \sum_{i_{k+1} \in N(i_{1:k})} \mathbb{E}[|X_{i_{k+1}}|^{k+\omega}] \lesssim M^k \sum_{i \in I} \mathbb{E}[|X_i|^{k+\omega}]. \qquad (11.10)
$$

Substituting (11.8), (11.9), and (11.10) into (11.7), we conclude

$$
\begin{aligned}
\mathcal{R}_\omega[t_1, t_2, \ldots, t_k] \leq &\mathcal{R}_\omega\big[0, \operatorname{sgn}(t_2), 2 \cdot \operatorname{sgn}(t_3), \ldots, (k-1)\operatorname{sgn}(t_{k-1})\big] \\
\lesssim &M^{k-2+\omega} \sum_{i \in I} \mathbb{E}[|X_i|^{k-1+\omega}]. \qquad \square
\end{aligned}
$$

Proof of Proposition 3.3. By Proposition 11.1, we have

$$
R_{k,\omega} \overset{(11.1)}{=} \sum_{t_{1:(k+2)} \in \mathcal{M}_{1,k+2}} \mathcal{R}_\omega[t_1, t_2, \ldots, t_{k+2}] \lesssim \sum_{t_{1:(k+2)} \in \mathcal{M}_{1,k+2}} M^{k+\omega} \sum_{i \in I} \mathbb{E}[|X_i|^{k+1+\omega}].
$$

Noting that $|\mathcal{M}_{1,k+2}| < 2^{k+1}$ [22], we conclude that

$$
R_{k,\omega} \lesssim M^{k+\omega} \sum_{i \in I} \mathbb{E}[|X_i|^{k+1+\omega}]. \qquad \square
$$

11.2   Proof of Theorem 3.4

The proof of Theorem 3.4 relies on Theorem 3.1 and Proposition 3.3.

Proof of Theorem 3.4. Let $k := \lceil p \rceil$. Then $p = k + \omega - 1$. Without loss of generality, we assume $\sigma_n = 1$. By Proposition 3.3,

$$
R_{j,\omega,n} \lesssim M_n^{j+\omega} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{j+1+\omega}\big].
$$

If we let $q_1 = (k-1)/(k-j)$ and $q_2 = (k-1)/(j-1)$, then $1/q_1 + 1/q_2 = 1$ and $(2+\omega)/q_1 + (k+1+\omega)/q_2 = j+1+\omega$. Thus,

$$
\big|X_i^{(n)}\big|^{j+1+\omega} = \big|X_i^{(n)}\big|^{(2+\omega)/q_1} \cdot \big|X_i^{(n)}\big|^{(k+1+\omega)/q_2}.
$$

By Hölder's inequality,

$$
\begin{aligned}
&M_n^{j+\omega} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{j+1+\omega}\big] \\
\leq &\Big(M_n^{1+\omega} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\Big)^{1/q_1} \Big(M_n^{k+\omega} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\Big)^{1/q_2} \\
= &\Big(M_n^{1+\omega} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\Big)^{(k-j)/(k-1)} \Big(M_n^{k+\omega} \sum_{i \in I_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\Big)^{(j-1)/(k-1)}.
\end{aligned}
$$

Since

$$
\frac{\omega(k-j)}{(k-1)(j+\omega-1)} + \frac{(j-1)(k+\omega-1)}{(k-1)(j+\omega-1)} = 1,
$$

by Young's inequality (See Lemma 11.3 for details), we get

$$\left(M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\right)^{\frac{k-j}{(k-1)(j+\omega-1)}}\left(M_n^{k+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\right)^{\frac{j-1}{(k-1)(j+\omega-1)}}$$

$$\leq\frac{\omega(k-j)}{(k-1)(j+\omega-1)}\left(M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\right)^{1/\omega}$$

$$+\frac{(j-1)(k+\omega-1)}{(k-1)(j+\omega-1)}\left(M_n^{k+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\right)^{1/(k+\omega-1)}.$$

Thus, we have

$$R_{j,\omega,n}^{1/(j+\omega-1)}\lesssim\left(M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\right)^{1/\omega}+\left(M_n^{k+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\right)^{1/(k+\omega-1)}.$$

Similarly, we derive that

$$R_{j,1,n}^{1/j}\lesssim\left(M_n^{j+1}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{j+2}\big]\right)^{1/j}$$

$$\leq\left(M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\right)^{\frac{k+\omega-j-1}{kj}}\left(M_n^{k+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\right)^{\frac{j-\omega}{(k-1)j}}$$

$$\lesssim\left(M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{2+\omega}\big]\right)^{1/\omega}+\left(M_n^{k+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{k+1+\omega}\big]\right)^{1/(k+\omega-1)}.$$

Since by assumption $M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{\omega+2}\big]\to 0$ and $M_n^{k+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big]\to 0$ as $n\to\infty$, we have that $R_{j,1,n}\to 0$ as $n\to\infty$. Therefore, by Theorem 3.1 and noting the fact that $p=k+\omega-1$, we conclude

$$\mathcal{W}_p(\mathcal{L}(W_n),\mathcal{N}(0,1))$$

$$\leq C_p\left(\left(M_n^{1+\omega}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{\omega+2}\big]\right)^{1/\omega}+\left(M_n^{p+1}\sum_{i\in I_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big]\right)^{1/p}\right),$$

where C_p only depends on p. $\qquad\square$

11.3 Proofs of Corollaries 4.2 and 5.2

Proof of Corollary 4.2. Define the graph (T_n,E_n) to be such that there is an edge between $i_1,i_2\in T_n$ if and only if $\|i_1-i_2\|\leq m$. From the definition of the m-dependent random field, $\big(X_i^{(n)}\big)_{i\in T_n}$ satisfies [LD*]. We will therefore apply Theorem 3.4 to obtain the desired result. We remark that $j\in N_n\big(i_{1:(\lceil p\rceil+1)}\big)$ only if there is $\ell\in[\lceil p\rceil+1]$ such that $\|i_\ell-j\|\leq m$, which directly implies that $\big|N_n\big(i_{1:(\lceil p\rceil+1)}\big)\big|\leq(2m+1)^d(\lceil p\rceil+1)$.

Moreover, by Hölder's inequality, we have

$$\sum_{i\in T_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{\omega+2}\big]\leq\left(\sum_{i\in T_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^2\big]\right)^{(p-\omega)/p}\left(\sum_{i\in T_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big]\right)^{\omega/p}$$

$$\overset{(a)}{\leq}M^{(p-\omega)/p}\sigma^{2(p-\omega)/p}\left(\sum_{i\in T_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big]\right)^{\omega/p}.$$

Here (a) is due to the non-degeneracy condition. And this directly implies that

$$m^{(1+\omega)d/\omega}\left(\sigma_n^{-(\omega+2)}\sum_{i\in T_n}\mathbb{E}\big[\big|X_i^{(n)}\big|^{\omega+2}\big]\right)^{1/\omega}$$

$$\leq m^{\frac{(1+\omega)d}{\omega}} M^{\frac{p-\omega}{p\omega}} \left(\sigma_n^{-(p+2)} \sum_{i \in T_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big] \right)^{1/p} \to 0 \quad \text{as } n \to \infty.$$

Therefore, by Theorem 3.4, there exists $C_{p,d} > 0$ such that for n large enough we have

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq C_{p,d} m^{\frac{(1+\omega)d}{\omega}} M^{\frac{p-\omega}{p\omega}} \sigma_n^{-\frac{p+2}{p}} \left(\sum_{i \in T_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big] \right)^{1/p}.$$

Moreover, if $\big(X_i^{(n)}\big)$ is in addition assumed to be stationary, then by assumption there is a constant K such that $\liminf_{n \to \infty} \sigma_n^2/|T_n| \geq K$. Therefore, we get that

$$\sigma_n^{-(p+2)} \sum_{i \in T_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big] \asymp |T_n|^{-(p+2)/2} \cdot |T_n| = |T_n|^{-p/2} \to 0,$$

and

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) \leq C_{p,d} m^{\frac{(1+\omega)d}{\omega}} M^{\frac{p-\omega}{p\omega}} \sigma_n^{-\frac{p+2}{p}} \left(\sum_{i \in T_n} \mathbb{E}\big[\big|X_i^{(n)}\big|^{p+2}\big] \right)^{1/p} = \mathcal{O}(|T_n|^{-1/2}). \quad \square$$

Proof of Corollary 5.2. Consider the index set $I_n = \{\boldsymbol{i} = (i_1, \ldots, i_m) : 1 \leq i_1 \leq \cdots \leq i_m \leq n\} \subseteq \mathbb{Z}^m$. For each $\boldsymbol{i} \in I_n$, let $\xi_{\boldsymbol{i}} := h(X_{i_1}, \ldots, X_{i_m})$. Then $W_n = \sigma_n^{-1} \sum_{\boldsymbol{i} \in I} \xi_{\boldsymbol{i}}$. Let (I_n, E_n) be the graph such that there is an edge between $\boldsymbol{i}, \boldsymbol{j} \in I_n$ if and only if $\{i_1, \ldots, i_m\} \cap \{j_1, \ldots, j_m\} \neq \emptyset$.

Then we remark that the conditions [LD*] holds. Moreover, note that $\boldsymbol{j}$ is in $N_n(\boldsymbol{i}_{1:(\lceil p \rceil + 1)})$ only if there is $\ell \in [\lceil p \rceil + 1]$ and $k_1, k_2 \in [m]$ such that $j_{k_1} = (\boldsymbol{i}_\ell)_{k_2}$, where $(\boldsymbol{i}_\ell)_{k_2}$ denotes the k_2-th component of the vector $\boldsymbol{i}_\ell$. This directly implies that the cardinality of the dependency neighborhoods are bounded by $n^m - \big(n - m(\lceil p \rceil + 1)\big)^m \asymp n^{m-1}$. Moreover, the non-degeneracy condition of the U-statistic implies that $\sigma_n^2 \asymp n^{2m-1}$ [14]. Applying Theorem 3.4, we get that

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1))$$
$$\lesssim \left(n^m (n^{m-1})^{1+\omega} \frac{1}{\sigma_n^{\omega+2}} \mathbb{E}\big[\big|h(X_1, \ldots, X_m)\big|^{\omega+2}\big] \right)^{1/\omega}$$
$$\quad + \left(n^m (n^{m-1})^{p+1} \frac{1}{\sigma_n^{p+2}} \mathbb{E}\big[\big|h(X_1, \ldots, X_m)\big|^{p+2}\big] \right)^{1/p}$$
$$\lesssim n^{-1/2} \left(\mathbb{E}\big[\big|h(X_1, \ldots, X_m)\big|^{\omega+2}\big] \right)^{1/\omega} + n^{-1/2} \left(\mathbb{E}\big[\big|h(X_1, \ldots, X_m)\big|^{p+2}\big] \right)^{1/p}$$
$$\leq n^{-1/2} \big\|h(X_1, \ldots, X_m)\big\|_{p+2}^{(\omega+2)/\omega} + n^{-1/2} \big\|h(X_1, \ldots, X_m)\big\|_{p+2}^{(p+2)/p}.$$

By the moment condition, $\big\|h(X_1, \ldots, X_m)\big\|_{p+2} < \infty$. Thus, we conclude

$$\mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) = \mathcal{O}(n^{-1/2}). \quad \square$$

11.4 Proof of Theorem 6.1

Proof. For ease of notation we write $\omega_p := \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1))$. Choose $\rho \in (0,1)$. Then remark that for all $\epsilon > 0$ there is $G \sim \mathcal{N}(0,1)$ such that $\|G - W_n\|_p \leq \mathcal{W}_p(\mathcal{L}(W_n), \mathcal{N}(0,1)) + \epsilon$. Therefore, by the union bound we have

$$\begin{aligned}
\mathbb{P}\big(W_n \geq t\big) &= \mathbb{P}\big(W_n - G + G \geq t\big) \\
&\leq \mathbb{P}\big(W_n - G \geq (1-\rho)t\big) + \mathbb{P}\big(G \geq \rho t\big) \\
&\stackrel{(a)}{\leq} \Phi^c(\rho t) + \frac{\|W_n - G\|_p^p}{((1-\rho)t)^p}
\end{aligned}$$

$$\leq \Phi^c(\rho t) + \frac{(\omega_p + \epsilon)^p}{((1-\rho)t)^p}$$

where to obtain (a) we have used Markov's inequality. Now as this holds for any arbitrary choice of $\epsilon > 0$ we conclude that

$$\mathbb{P}(W_n \geq t) \leq \Phi^c(\rho t) + \frac{\omega_p^p}{((1-\rho)t)^p}.$$

Define the function $g_t : x \mapsto (1-x)^{p+1} e^{-\frac{(xt)^2}{2}}$, then we can remark that $g_t : [0,1] \to [0,1]$ is a bijection. Choose $\rho_t^* := g_t^{-1}\left(\frac{\sqrt{2\pi}p\omega_p^p}{t^{p+1}}\right)$. Moreover, we obtain that

$$\mathbb{P}(W_n \geq t) \leq \Phi^c(\rho_t^* t) + \varphi(\rho_t^* t)(1-\rho_t^*)t \frac{\omega_p^p}{t^{p+1}(1-\rho_t^*)^{p+1}\varphi(\rho_t^* t)} \tag{11.11}$$

$$\overset{(a)}{\leq} \Phi^c(t) + (1-\rho_t^*)t\varphi(\rho_t^* t)\left(1 + \frac{1}{p}\right)$$

$$\leq \Phi^c(t) + \frac{p^{\frac{1}{p+1}}\omega_p^{1-\frac{1}{p+1}}}{t} \varphi\left(\rho_t^* t\left(1 - \frac{1}{p+1}\right)\right)\left(1 + \frac{1}{p}\right)$$

where to obtain (a) we used the fact that $\Phi^c(\rho_t^* t) \leq \Phi^c(t) + (1-\rho_t^*)t \sup_{x \in [\rho_t^* t, t]} \varphi(x)$.

Suppose that $t \geq 1$ and satisfies $1 - \frac{\sqrt{2\beta \log t}}{t} \leq 1$. Define

$$x := \frac{\sqrt{2\beta \log t}}{t},$$

we notice that $x \in [0,1]$. We remark that if

$$\omega_p \leq (\sqrt{2\pi}p)^{\frac{1}{p+1}}\left(1 - \frac{\sqrt{2\beta \log t}}{t}\right)t^{1-\frac{\beta}{p+1}},$$

then we have $g_t^{-1}(x) \geq \frac{\sqrt{2\pi}p\omega_p^p}{t^{p+1}}$. Therefore as $g_t^{-1}(\cdot)$ is a decreasing function we have that $x \leq \rho_t^*$ which implies that

$$\mathbb{P}(W_n \geq t) \leq \Phi^c(t) + \frac{1}{t^{1+\beta\left(1-\frac{1}{p+1}\right)}} p^{\frac{1}{p+1}}\omega_p^{1-\frac{1}{p+1}}\left(1 + \frac{1}{p}\right).$$

Moreover, similarly we can remark that

$$\mathbb{P}(G \geq (1+\rho)t) \leq \mathbb{P}(W_n \geq t) + \mathbb{P}(G - W_n \geq \rho t)$$

$$\leq \mathbb{P}(W_n \geq t) + \frac{(\omega_p + \epsilon)^p}{\rho^p t^p}.$$

Therefore, as this holds for any arbitrary $\epsilon > 0$ we obtain that

$$\Phi^c((1+\rho)t) \leq \mathbb{P}(W_n \geq t) + \frac{\omega_p^p}{\rho^p t^p}.$$

Moreover, we can definite $\widetilde{g}_t : x \mapsto e^{-(1+x)^2 t^2} x^{p+1}$ then choose $\widetilde{\rho}_t^* := \widetilde{g}_t^{-1}\left(\frac{\sqrt{2\pi}p\omega_p^p}{t^{p+1}}\right)$. We similarly obtain that

$$\mathbb{P}(W_n \geq t) \geq \Phi^c(t) - \frac{p^{\frac{1}{p+1}}\omega_p^{1-\frac{1}{p+1}}}{t}\varphi\left(t\left(1 - \frac{1}{p+1}\right)\right)\left(1 + \frac{1}{p}\right). \qquad \square$$

References

[1] Ralph Palmer Agnew, *Estimates for global central limit theorems*, The Annals of Mathematical Statistics **28** (1957), no. 1, 26–42. MR0084227

[2] Naum I Akhiezer, *The classical moment problem and some related questions in analysis*, SIAM, 2020. MR4191205

[3] Pierre Baldi and Yosef Rinott, *On normal approximations of distributions in terms of dependency graphs*, The Annals of Probability **17** (1989), no. 4, 1646–1650. MR1048950

[4] Andrew D Barbour, *Asymptotic expansions based on smooth functions in the central limit theorem*, Probability Theory and Related Fields **72** (1986), no. 2, 289–303. MR0836279

[5] Andrew D Barbour, Michal Karoński, and Andrzej Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, Journal of Combinatorial Theory, Series B **47** (1989), no. 2, 125–145. MR1047781

[6] P Bártfai, *über die entfernung der irrfahrtswege*, Studia Sci. Math. Hungar **5** (1970), 41–49. MR0275499

[7] Christian Berg, *Indeterminate moment problems and the theory of entire functions*, Journal of Computational and Applied Mathematics **65** (1995), no. 1-3, 27–55. MR1379118

[8] Sergey G Bobkov, *Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances*, Probability Theory and Related Fields **170** (2018), no. 1, 229–262. MR3748324

[9] Erwin Bolthausen, *On the central limit theorem for stationary mixing random fields*, The Annals of Probability (1982), 1047–1050. MR0672305

[10] Thomas Bonis, *Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem*, Probability Theory and Related Fields **178** (2020), no. 3, 827–860. MR4168389

[11] Sourav Chatterjee, Jason Fulman, and Adrian Röllin, *Exponential approximation by Stein's method and spectral graph theory*, ALEA Lat. Am. J. Probab. Math. Stat **8** (2011), no. 1, 197–223. MR2802856

[12] Louis HY Chen, *Poisson approximation for dependent trials*, The Annals of Probability **3** (1975), no. 3, 534–545. MR0428387

[13] Louis HY Chen and Qi-Man Shao, *Normal approximation under local dependence*, The Annals of Probability **32** (2004), no. 3, 1985–2028. MR2073183

[14] Louis HY Chen and Qi-Man Shao, *Normal approximation for nonlinear statistics using a concentration inequality approach*, Bernoulli **13** (2007), no. 2, 581–599. MR2331265

[15] Carl-Gustav Esseen, *On mean central limit theorems*, Trans. Roy. Inst. Tech., Stockholm **2** (1958), no. 120, 1–30. MR0097111

[16] Xiao Fang, *Wasserstein-2 bounds in normal approximation under local dependence*, Electronic Journal of Probability **24** (2019), 1–14. MR3940765

[17] Xiao Fang and Yuta Koike, *From p-Wasserstein bounds to moderate deviations*, arXiv preprint arXiv:2205.13307 (2022). MR4609449

[18] William Feller, *The fundamental limit theorems in probability*, Bulletin of the American Mathematical Society **51** (1945), no. 11, 800–832. MR0013252

[19] Valentin Féray, Pierre-Loïc Méliot, and Ashkan Nikeghbali, *Mod-φ convergence: Normality zones and precise deviations*, Springer, 2016. MR3585777

[20] George T Gilbert, *Positive definite matrices and Sylvester's criterion*, The American Mathematical Monthly **98** (1991), no. 1, 44–46. MR1083614

[21] Lothar Heinrich, *A method for the derivation of limit theorems for sums of m-dependent random variables*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **60** (1982), no. 4, 501–515. MR0665742

[22] Silvia Heubach and Toufik Mansour, *Combinatorics of compositions and words*, Chapman and Hall/CRC, 2009. MR2531482

[23] Wassily Hoeffding, *A class of statistics with asymptotically normal distribution*, The Annals of Mathematical Statistics **19** (1948), no. 3, 293 – 325. MR0026294

[24] Wassily Hoeffding and Herbert Robbins, *The central limit theorem for dependent random variables*, Duke Mathematical Journal **15** (1948), no. 3, 773–780. MR0026771

[25] Svante Janson, *Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs*, The Annals of Probability (1988), 305–312. MR0920273

[26] Serge Lang, *Introduction to linear algebra*, Springer Science & Business Media, 2012.

[27] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati, *Stein's method, logarithmic Sobolev and transport inequalities*, Geometric and Functional Analysis **25** (2015), no. 1, 256–306. MR3320893

[28] Paul Lévy, *Propriétés asymptotiques des sommes de variables indépendantes ou enchaînées*, Journal des mathématiques pures et appliquées. Series **9** (1935), no. 14, 4.

[29] Christophe Ley, Gesine Reinert, and Yvik Swan, *Stein's method for comparison of univariate distributions*, Probability Surveys **14** (2017), 1–52. MR3595350

[30] Song-Hao Liu and Zhuo-Song Zhang, *Cramér-type moderate deviations under local dependence*, arXiv preprint arXiv:2112.10946 (2021).

[31] Eugene Lukacs, *Characteristic functions*, A Charles Griffin Book, Griffin, 1970. MR0346874

[32] Felix Otto and Cédric Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, Journal of Functional Analysis **173** (2000), no. 2, 361–400. MR1760620

[33] Erol A Peköz and Adrian Röllin, *New rates for exponential approximation and the theorems of Rényi and Yaglom*, The Annals of Probability **39** (2011), no. 2, 587–608. MR2789507

[34] MB Petrovskaya and Aleksandr Mikhailovich Leontovich, *The central limit theorem for a sequence of random variables with a slowly growing number of dependences*, Theory of Probability & Its Applications **27** (1983), no. 4, 815–825. MR0681466

[35] Emmanuel Rio, *Distances minimales et distances idéales*, Comptes Rendus de l'Académie des Sciences-Series I-Mathematics **326** (1998), no. 9, 1127–1130. MR1647215

[36] Emmanuel Rio, *Upper bounds for minimal distances in the central limit theorem*, Annales de l'IHP Probabilités et statistiques **45** (2009), no. 3, 802–817. MR2548505

[37] Murray Rosenblatt, *A central limit theorem and a strong mixing condition*, Proceedings of the national Academy of Sciences **42** (1956), no. 1, 43–47. MR0074711

[38] Nathan Ross, *Fundamentals of Stein's method*, Probability Surveys **8** (2011), 210–293. MR2861132

[39] Aleksandr Ivanovich Sakhanenko, *Estimates in an invariance principle*, Matematicheskie Trudy **5** (1985), 27–44. MR0821751

[40] Peter J Smith, *A recursive formulation of the old problem of obtaining moments from cumulants and vice versa*, The American Statistician **49** (1995), no. 2, 217–218. MR1347727

[41] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, in: Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory, vol. 6, University of California Press, 1972, pp. 583–603. MR0402873

[42] Michel Talagrand, *Transportation cost for Gaussian and other product measures*, Geometric & Functional Analysis GAFA **6** (1996), no. 3, 587–600. MR1392331

[43] Jacob David Tamarkin, James Alexander Shohat, and JD Tamarkin, *The problem of moments*, American Mathematical Society, 1943. MR0008438

[44] Cédric Villani, *Optimal transport: Old and new*, vol. 338, Springer, 2009. MR2459454

[45] Vladimir Mikhailovich Zolotarev, *Probability metrics*, Theory of Probability & Its Applications **28** (1984), no. 2, 278–302. MR0700210

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System: https://vtex.lt/services/ejms-peer-review/
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: https://imstat.org/shop/donation/