MODERATE DEVIATIONS FOR FUNCTIONAL U-PROCESSES

PETER EICHELSBACHER

Abstract. The moderate deviations principle is shown for the partial sums processes built on U-empirical measures of Polish space valued random variables and on U-statistics of real valued kernel functions. It is proved that in the non-degenerate case the conditions for the time fixed principles suffice for the moderate deviations principle to carry over to the corresponding partial sums processes. Given a uniformly bounded VC subgraph class of functions, we obtain corresponding moderate deviations for time dependent U-processes. Among other techniques we use the decoupling technique for U-statistics due to de la Peña and Montgomery-Smith. We apply an improved version of a Bernstein-type inequality for degenerate U-statistics due to Arcones and Giné as well as some moment inequalities for U-statistics. Moreover we prove and use a Lévy-type maximal inequality for U-statistics.

1. Introduction and Statement of the results

For a sequence of $\mathbb{R}^d$-valued i. i. d. random variables X_i with a finite moment generating function Borovkov and Mogulskii [6] investigated the moderate deviation behaviour of the partial sums process

$$S_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} X_i, \quad t \in [0,1].$$

In an appropriate topology, Dembo and Zajic [9] proved a moderate deviation principle for the partial sums empirical process

$$L_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} \delta_{X_i}, \quad t \in [0,1].$$

Given a class of bounded functions $\mathcal{F}$, Dembo and Zajic [10] considered the moderate deviation principles for functional empirical processes

$$L_n(t,f) = \frac{1}{n} \sum_{i=1}^{[nt]} f(X_i), \quad t \in [0,1],\ f \in \mathcal{F}.$$

The aim of this paper is to extend the moderate deviation principles when passing from linear statistics to higher order statistics.

1991 *Mathematics Subject Classification.* 60F10 (primary), 60E15, 60F05.

Key words and phrases. Moderate deviations, partial sums, U-processes, VC-classes, decoupling inequality, maximal inequality for U-statistics, Bernstein-type inequality.

Let us recall the definition of the large deviations principle (LDP). A sequence of probability measures $\{\mu_n\}_{n\in\mathbb{N}}$ on a topological space $\mathcal{X}$ equipped with σ-field $\mathcal{B}$ is said to satisfy the LDP with speed $a_n \downarrow 0$ and good rate function $I(\cdot)$ if the level sets $\{x : I(x) \leq \alpha\}$ are compact for all $\alpha < \infty$ and for all $\Gamma \in \mathcal{B}$ the lower bound

$$\liminf_{n\to\infty} a_n \log \mu_n(\Gamma) \geq - \inf_{x\in \mathrm{int}(\Gamma)} I(x),$$

and the upper bound

$$\limsup_{n\to\infty} a_n \log \mu_n(\Gamma) \leq - \inf_{x\in \mathrm{cl}(\Gamma)} I(x)$$

hold, where $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ, respectively. We say that a sequence of random variables satisfies the LDP when the sequence of measures induced by these variables satisfies the LDP. Let $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ be a sequence satisfying

$$\lim_{n\to\infty} \frac{b_n}{n} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{n}{b_n^2} = 0. \tag{1.1}$$

If $a\,x \in \mathcal{X}$ for all $a > 0$ and $x \in \mathcal{X}$ then a sequence of random variables Z_n shall satisfy the moderate deviations principle (MDP) with speed $\frac{n}{b_n^2}$ and with good rate function $I(\cdot)$, if the sequence $\frac{n}{b_n} Z_n$ satisfies the LDP in $\mathcal{X}$ with the good rate function $I(\cdot)$ and with speed $\frac{n}{b_n^2}$.

Denote by $L_\infty([0,1],\mathbb{R}^d)$ the space of (equivalence classes modulo equality a.e. of) bounded measurable functions on $[0,1]$, equipped with the uniform topology. Consider the polygonal approximation of $S_n(\cdot)$, that is

$$\widetilde{S}_n(t) := S_n(t) + \left(t - \frac{[nt]}{n}\right) X_{[nt]+1}.$$

Note that $\widetilde{S}_n(\cdot)$ is continuous and it carries the same information as $S_n(\cdot)$. The MDP for $\widetilde{S}_n(\cdot)$ was established in $L_\infty([0,1],\mathbb{R}^d)$. By $\mathcal{AC}_0([0,1],\mathbb{R}^d)$ denote the subspace of absolutely continuous functions ϕ on $[0,1]$ with $\phi(0) = 0$. Let $\{X_n, n \in \mathbb{N}\}$ be a sequence of i. i. d. random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^d$, common law μ and $\mathbb{E}(X_1) = 0$. Borovkov and Mogulskii [6] considered (essentially) the MDP in the i.i.d. case under the condition that

$$\log \mathbb{E} \exp(\langle \theta, X_1 \rangle) < \infty \tag{1.2}$$

in some ball centered in the origin. $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product in $\mathbb{R}^d$ and $\|\cdot\|$ a norm in $\mathbb{R}^d$. The sequences $\frac{n}{b_n} \widetilde{S}_n(\cdot)$ satisfies the LDP in $L_\infty([0,1],\mathbb{R}^d)$, equipped with the uniform topology, with good rate function

$$I_\infty(\phi) = \int_0^1 \Lambda^*(\dot{\phi}) dt, \tag{1.3}$$

if $\phi \in \mathcal{AC}_0([0,1],\mathbb{R}^d)$ and $I_\infty(\phi) = \infty$ otherwise. Here Λ^* denotes the convex dual of $\Lambda(\theta) = \mathbb{E}(\langle \theta, X_1 \rangle^2)/2$ (see also [21, Theorem 1] and [5, Theorem 3.1]). The MDP for $\{S_n(\cdot), n \in \mathbb{N}\}$ follows from the following

Lemma 1.4. *Assume Condition (1.2) holds. Then* $\{S_n(\cdot), n \in \mathbb{N}\}$ *and* $\{\widetilde{S}_n(\cdot), n \in \mathbb{N}\}$ *have identical moderate deviation behaviour.*

Proof. For each $\eta > 0$ we obtain

$$\mathbb{P}\bigg(\sup_{0 \le t \le 1} \big\| \frac{n}{b_n}(S_n(t) - \widetilde{S}_n(t)) \big\| > \eta \bigg) \le \mathbb{P}\bigg(\sup_{0 \le t \le 1} ||X_{[nt]+1}|| > \eta\, b_n \bigg).$$

Now using the continuous analog of the Lévy–Ottaviani inequality (cf. [20, Lemma 6.2]) we obtain for $c > 0$ choosen such that $\mathbb{P}(||X_1|| > c) \le 1/2$ and for each $\lambda > 0$:

$$\begin{aligned}\mathbb{P}\bigg(\sup_{0 \le t \le 1} \big\| \frac{n}{b_n}(S_n(t) - \widetilde{S}_n(t)) \big\| > \eta \bigg) &\le 2\mathbb{P}(||X_1|| > \eta b_n - c) \\ &\le 2 \exp(-(\eta b_n - c)\lambda)\ \mathbb{E} \exp(\lambda ||X_1||).\end{aligned}$$

We conclude

$$\begin{aligned}&\frac{n}{b_n^2} \log \mathbb{P}\bigg(\sup_{0 \le t \le 1} \big\| \frac{n}{b_n}(S_n(t) - \widetilde{S}_n(t)) \big\| > \eta \bigg) \\ &\qquad \le \frac{n}{b_n^2} \log 2 - \frac{n\lambda(\eta b_n - c)}{b_n^2} + \frac{n}{b_n^2} \log \mathbb{E} \exp(\lambda ||X_1||).\end{aligned}$$

The MDP for $\{S_n(\cdot), n \in \mathbb{N}\}$ follows from [11, Theorem 4.2.13] by choosing $\lambda > 0$ as in (1.2), letting $n \to \infty$ and using the properties of the sequence b_n. $\square$

1.1. **Moderate deviations for partial sums U-statistics.** We will prove the MDP for *partial sums U-statistics*, i.e.

$$U_n^m(t, h) = \frac{1}{\binom{n}{m}} \sum_{C_m^{[nt]}} h(X_{i_1}, \ldots, X_{i_m}), \quad t \in [0, 1]$$

(we call $U_n^m(h) := U_n^m(1, h)$ a U-statistic). Here the X_i are i. i. d. random variables and h is a measurable, symmetric $\mathbb{R}^d$-valued function, called kernel function, where symmetric means that h is invariant under all permutations of its arguments. C_m^k with $k, m \in \mathbb{N}$ denotes the set $\{(i_1, \ldots, i_m) : 1 \le i_1 < \cdots < i_m \le k\}$. We will consider the non-degenerate case and will prove the MDP in the uniform topology. To be able to formulate our result for a non-degenerate partial sums U-statistic, we state the Hoeffding decompositon together with some notations. The operator $\pi_{k,m}^{\mu} = \pi_{k,m}$ acts on $\mu^{\otimes m}$ -integrable symmetric functions $h : S^m \to \mathbb{R}$ as follows:

$$\pi_{k,m} h(x_1, \ldots, x_k) = (\delta_{x_1} - \mu) \otimes \cdots \otimes (\delta_{x_k} - \mu) \otimes \mu^{\otimes(m-k)} h,$$

where $\nu_1 \otimes \cdots \otimes \nu_m h = \int \cdots \int h(x_1, \ldots, x_m) d\nu_1(x_1) \cdots d\nu_m(x_m)$ and

$$\mu^{\otimes(m-1)} h(x) = \int \cdots \int h(x_1, \ldots, x_{m-1}, x) d\mu(x_1) \cdots d\mu(x_{m-1}).$$

A function h is called *μ-canonical* or *completely degenerate* if $\int h(x_1, \ldots, x_m) d\mu(x_i) = 0$ for all $1 \le i \le m$. Note that $\pi_{k,m} h$ is a μ-canonical function of k variables. With this notation we can decompose a U-statistic into a sum of μ-canonical U-statistics of different orders. For all $\mu^{\otimes m}$-integrable functions $h: S^m \to \mathbb{R}$ the following relation holds true

$$U_n^m(h) = \sum_{k=0}^{m} \binom{m}{k} U_n^k(\pi_{k,m}h) \tag{1.5}$$

(cf. (1.1.9) in [8]). To prove the MDP in the non-degenerate case for $\{U_n(\cdot, h), n \in \mathbb{N}\}$, we will prove that $\{U_n(\cdot, \pi_{1,m}h), n \in \mathbb{N}\}$ satisfies a MDP and that the sequences $\{U_n^k(\cdot, \pi_{k,m}h)\, n \in \mathbb{N}\}$ do not contribute to the moderate upper and lower bounds for $2 \leq k \leq m$. Therefore we will use an improved Bernstein-type inequality; for bounded kernel functions h this typ of inequality was proved in [2]. Moreover we will prove and apply a Lévy-type maximal inequality for U-statistics. Consider

Condition 1.6 (Weak Cramér condition)**.** *For each $2 \leq k \leq m$ there exists at least one $\delta_k > 0$ such that*

$$\int_{S^k} \exp(\delta_k \|\pi_{k,m}h)\|^2)\, d\mu^{\otimes k} < \infty$$

and

$$\int_{S^k} \exp(\delta_k \|\pi_{k,m}h \circ \pi_\tau\|)\, d\mu^{\otimes k} < \infty$$

for every map $\tau: \{1, \dots, k\} \to \{1, \dots, k\}$, where $\pi_\tau: S^k \to S^k$ is defined by $\pi_\tau(s) = (s_{\tau(1)}, \dots, s_{\tau(k)})$ for every $s = (s_1, \dots, s_k) \in S^k$; and there exists at least one $\delta > 0$ such that

$$\int_S \exp(\delta \|\pi_{1,m}h\|)\, d\mu < \infty. \tag{1.7}$$

Theorem 1.8 (Moderate deviations of partial sums U-statistics)**.**
If Condition 1.6 is satisfied for a symmetric kernel function h, then $\{U_n^m(\cdot, h), n \in \mathbb{N}\}$ satisfies the MDP in $L_\infty([0,1], \mathbb{R}^d)$ with the good rate function $I_\infty(\cdot)$ given by (1.3), where $\Lambda(\theta) = \mathbb{E}(\langle \theta, \pi_{1,m}h(X_1)\rangle^2)/2$ and with speed $\frac{n}{b_n^2}$.

Remark 1.9. (a) Actually Theorem 1.8 is true under weaker moment assumptions for $2 \leq k \leq m$: Assume that for each $2 \leq k \leq m$

$$\mathbb{E}\,\|\pi_{k,m}h\|^{2\,l} \leq \frac{1}{2}\, l!\, H^{l-2} \mathbb{E}\,\|\pi_{k,m}h\|^4, \quad l = 2, 3, \dots,$$

and that there exists at least one $\delta > 0$ such that (1.7) holds true. Then the statement of the Theorem is true. The arguments are given in the proof of Lemma 2.1.

(b) We can easily adapt the result to a time intervall $[0, T]$. Applying Theorem 4.6.1 of [11] for $T \in \mathbb{N}$ yields a MDP for $\{U_n^m(\cdot, h), n \in \mathbb{N}\}$ in $\mathcal{AC}_0(\mathbb{R}_+, \mathbb{R}^d)$ equipped with the topology of uniform convergence on compact subsets of $\mathbb{R}_+$.

Example 1.10. Consider the *sample variance* U_n^{var}, which is a U-statistic of degree 2 with kernel function $h(x,y) = \frac{1}{2}(x-y)^2$. A simple calculation shows that

$$\pi_{1,2}h(x) = \frac{1}{2}\big((x-\mathbb{E}(X_1))^2 - \mathrm{Var}(X_1)\big),$$

where $\mathrm{Var}(X_i)$ denotes the variance of X_i under μ. It is well known, that we are in the non-degenerate regime, if the distribution μ satisfies the condition $\mathbb{E}(X_1 - \mathbb{E}(X_1))^4 > \mathrm{Var}(X_1)^2$. Let X_i be Gaussian random variables with mean a and variance σ^2. Then the condition is fulfilled since $\mathbb{E}(X_1 - \mathbb{E}(X_1))^4 = 3\sigma^4 > \sigma^4$. Moreover notice that $\pi_{1,2}h(X_1)$ is distributed like $\frac{1}{2}\sigma^2\, Y - \frac{1}{2}\sigma^2$, where Y is a χ_1^2-distributed random variable. Thus $\mathbb{E}\big(\frac{1}{2}\sigma^2\, Y - \frac{1}{2}\sigma^2\big) = 0$, it follows that

$$\Lambda(\theta) = \mathbb{E}\bigg(\theta^2\big(\frac{1}{2}\sigma^2\, Y - \frac{1}{2}\sigma^2\big)^2\bigg)/2 = \frac{\theta^2\,\sigma^4}{4}.$$

Thus we obtain $\Lambda^*(x) = \frac{x^2}{\sigma^4}$. Since Condition 1.6 is fulfilled (see [29, Theorem 2.1.1]) the sequence $\{U_n^{\mathrm{var}}(\cdot), n \in \mathbb{N}\}$ satisfies the MDP in $\mathcal{AC}_0(\mathbb{R}_+, \mathbb{R}^d)$, equipped with the topology of uniform convergence on compact subsets of $\mathbb{R}_+$, with the good rate function

$$I_{\mathrm{var}}^{\mathrm{gauss}}(\phi) = \frac{1}{\sigma^4}\int_0^\infty (\dot{\phi})^2\, dt.$$

If $\mathbb{P}(X_1 = 1) = p$ and $\mathbb{P}(X_1 = 0) = 1-p$ and if $0 < p < 1$ and $p \neq 1/2$, then we are in the non-degenerate case and the corresponding rate function is

$$I_{\mathrm{var}}^{\mathrm{bernoulli}}(\phi) = \frac{2}{p(1-p)(2p-1)^2}\int_0^\infty (\dot{\phi})^2\, dt.$$

Example 1.11. Note that for the kernel function $h(x,y) = xy$ (sample second moment), in the Gaussian case the rate function is $\frac{1}{2a^2\sigma^2}\int_0^\infty (\dot{\phi})^2\, dt$ and in the case of Bernoulli random variables we obtain the rate $\frac{1}{2p^3\big((1-p)^2+p^2\big)}\int_0^\infty (\dot{\phi})^2\, dt$.

Example 1.12. In the case of the *Wilcoxon one sample statistic* the kernel is given by $h(x,y) = 1_{\{x+y\leq 0\}}(x,y)$. Here $\pi_{1,2}h(x) = 1_{(0,\infty)}(x)$. For every continuous distribution function which is symmetric around 0, the rate is $\frac{1}{4}\int_0^\infty (\dot{\phi})^2\, dt$ (the result says that for the special class of distributions, the Wilcoxon statistic is asymptotically distributuion free). For an arbitrary continuous distribution function F with $F(0) \neq 1$, the rate is $\frac{1}{2(1-F(0))}\int_0^\infty (\dot{\phi})^2\, dt$.

The following Corollary does not exist in the literature and follows directly from Theorem 1.8:

Corollary 1.13. *If Condition 1.6 is satisfied for the symmetric kernel function h, then $\{U_n^m(h), n \in \mathbb{N}\}$ satisfies the MDP in $\mathbb{R}^d$ with the good rate function $\Lambda^*(\cdot)$, the convex dual of $\Lambda(\theta) = \mathbb{E}(\langle\theta, \pi_{1,m}h(X_1)\rangle^2)/2$ and with speed $\frac{n}{b_n^2}$.*

The large deviation principle (LDP) for $\{U_n^m(h), n \in \mathbb{N}\}$ is proved in [13].

1.2. **Moderate deviations for partial sums U-empirical measures.** To formulate the result on the empirical measure level we need some more notations. Let (S,d) be a Polish space with metric d. Fix $m \in \mathbb{N}$. Denote by $B(S^m, \mathbb{R}^d)$ the set of bounded Borel measurable functions, with $B'(S^m, \mathbb{R}^d)$ the algebraic dual of $B(S^m, \mathbb{R}^d)$ and by $C_b(S^m, \mathbb{R}^d)$ the set of bounded continuous functions.

Denote by $\mathcal{M}(S^m)$ and $\mathcal{M}_t(S^m)$, respectively, the set of Borel measures on the Polish space S^m which are signed and positive having total measure t, respectively. Unless explicitely stated otherwise, these spaces are equipped with the τ-topology.

Let $D_1[\mathbb{R}^d]$ denote the space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ equipped with the topology of uniform convergence on compact subsets of $\mathbb{R}_+$ and the corresponding Borel σ-field. Let $D_1[\mathcal{M}(S^m)]$ denote the space of càdlàg functions from $\mathbb{R}_+$ to $\mathcal{M}(S^m)$ equipped with the weakest topology such that the maps $y(\cdot) \mapsto \langle \varphi, y(\cdot)\rangle : D_1[\mathcal{M}(S^m)] \mapsto D_1[\mathbb{R}]$ are continuous, where $\varphi \in B(S^m, \mathbb{R})$ (and $\langle \varphi, y(\cdot)\rangle$ denotes the value of $\int \varphi \, d\, y(\cdot)$), and the smallest σ-field $\mathcal{B}$ such that these maps are mesurable.

Next let $D_2[[0,T], \mathbb{R}^d]$, $T > 0$, denote the Banach space of càdlàg functions $f : [0,T] \to \mathbb{R}^d$ whose norm $\sup_{t \in [0,T]} \|f(t)\|/(t+1)$ is finite. Let $D_2^{\mathrm{proj}}[\mathbb{R}^d]$ denote the space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ equipped with the projective limit topology of the system $(D_2[[0,T], \mathbb{R}^d], T \in \mathbb{R}_+)$ and let $D_2[\mathbb{R}^d]$ denote the Banach space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ whose norm $\sup_{t \in \mathbb{R}_+} \|f(t)\|/(t+1)^m$ is finite. All spaces are equipped with the corresponding Borel σ-field.

Let $D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ and $D_\beta[\mathbb{R}^d]$, respectively, denote the space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ corresponding to replacing $t+1$ by $\beta(t) = \sqrt{2(t \vee 1) \log\log(t \vee 3)}$ in the definition of the norms. Moreover define $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ as the space of càdlàg functions $y : \mathbb{R}_+ \to \mathcal{M}(S^m)$ such that $\langle \varphi, y(\cdot)\rangle \in D_2^{\mathrm{proj}}[\mathbb{R}^d]$ for every $\varphi \in B(S^m, \mathbb{R}^d)$, equipped with the weakest topology such that the maps $y(\cdot) \mapsto \langle \varphi, y(\cdot)\rangle : D_2^{\mathrm{proj}}[\mathcal{M}(S^m)] \mapsto D_2^{\mathrm{proj}}[\mathbb{R}^d]$ are continuous and the smallest σ-field, such that these maps are measurable. The space $D_2[\mathcal{M}(S^m)]$ and $D_2[B'(S^m, \mathbb{R}^d)]$ are defined similarly.

If $\{X_n, n \in \mathbb{N}\}$ is an i.i.d. sequence with state space S, Dembo and Zajic proved in [10, Theorem 1] the MDP for $\{L_n(\cdot), n \in \mathbb{N}\}$ in $D_1[\mathcal{M}(S)]$ as well as in $D_2[\mathcal{M}(S)]$ with a convex good rate function

$$J_\infty(\nu(\cdot)) = \int_0^\infty J(\dot\nu|\mu)\, dt,$$

if $\nu(\cdot) \in \mathcal{AC}_0[\mathcal{M}(S)]$ and $J_\infty(\nu(\cdot)) = +\infty$ otherwise. Here the functional $J(\cdot|\mu) : \mathcal{M}(S) \to [0,\infty]$ is defined by

$$J(\nu|\mu) = \frac{1}{2} \int_S \Big(\frac{d\nu}{d\mu}\Big)^2 \, d\mu$$

if $\nu(S) = 0$, $\nu \ll \mu$ and $J(\nu|\mu) = \infty$ otherwise (called the Fisher information). Using the convexity of $\mathbb{R} \ni x \mapsto x^2$, it follows that $J(\cdot|\mu)$ is convex. $\mathcal{AC}_0[\mathcal{M}(S)]$

is defined to be the set of all maps $\nu : \mathbb{R}_+ \to \mathcal{M}(S)$ with $\nu(0) = 0$ which are absolutely continuous with respect to the variation norm $||\cdot||_{\mathrm{var}}$ and possess a weak derivative for almost all t (the latter means that for almost every t the expression $\langle f, \nu(t+h) - \nu(t)\rangle/h$ converges to a limit $\langle f, \dot\nu(t)\rangle$ for every $f \in C_b(S, \mathbb{R})$). In [7] the moderate deviation behaviour of $L_n(\cdot)$ had already been considered in a weaker form.

Consider the measures $\tilde L_n^m : \Omega \to \mathcal{M}_1(S^m)$ with $n \geq m$, which are defined by

$$\tilde L_n^m = \frac{1}{n_{(m)}} \sum_{(i_1,\ldots,i_m)\in I(m,n)} \delta_{(X_{i_1},\ldots,X_{i_m})}, \tag{1.14}$$

where $n_{(m)} \equiv \prod_{k=0}^{m-1}(n-k)$ and where $I(m,n) \subset \{1,\ldots,n\}^m$ contains all m-tuples with pairwise different components. Due to their application, we call the measures $\{\tilde L_n^m\}_{n\geq m}$ the U-empirical measures of order m. We define the function $J_m(\cdot|\mu) : \mathcal{M}(S^m) \to [0,\infty]$ by

$$J_m(\nu|\mu) = \frac{1}{2}\int_S \Big(\frac{d\nu_1}{d\mu}\Big)^2 d\mu$$

if $\nu_1(S) = 0$, $\nu_1 \ll \mu$ and $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}$, and we define $J_m(\nu|\mu) = \infty$ otherwise. Using the convexity of $\mathbb{R} \ni x \mapsto x^2$, it follows that $J_m(\cdot|\mu)$ is convex, too. For every $n \geq m$ define the *partial sums U-empirical measure* by $L_n^m(\cdot) = (\tilde L_n^m - \mu^{\otimes m})(\cdot) : \Omega \to D_1[\mathcal{M}(S^m)]$ and denote by $M_n^m(\cdot) = \frac{n}{b_n} L_n^m(\cdot)$. In [16, Theorem 1.24] the MDP for $\{L_n^m(1), n \in \mathbb{N}\}$ (with rate function $J_m(\cdot|\mu)$) is proved for an arbitrary measurable space $(S^m, \mathcal{S}^{\otimes m})$ on a suitable subset of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$, endowed with a topology, which is stronger than the usual τ-topology and which makes maps $\nu \mapsto \int_{S^m} \varphi\, d\nu$ continuous even for certain unbounded φ taking values in a Banach space.

For technical reasons we have to consider functions $\varphi : S^m \to \mathbb{R}^d$ which are not symmetric. Otherwise for $\mathcal{S} \neq \{\emptyset, S\}$ and $m \geq 2$ we would not be able to separate the zero measure from the measure $(\delta_x^{\otimes m-1} \otimes \delta_y) - (\delta_y \otimes \delta_x^{m-1})$ with $x \in A \in \mathcal{S}$ and $y \in S \setminus A$ for example, hence the τ-topology would loose the Hausdorff property. Given φ and a nonempty subset A of $\{1,\ldots,m\}$, define $\varphi_A \in L_1(\mu^{\otimes|A|})$ by μ-integrating $\varphi(s_1,\ldots,s_m)$ with respect to every s_i with $i \in \{1,\ldots,m\}\setminus A$. By convention, $\varphi_\varnothing \equiv \int_{S^m} \varphi\, d\mu^{\otimes m} \in \mathbb{R}^d$. Furthermore, define $\tilde\varphi_A \in L_1(\mu^{\otimes|A|})$ by

$$\tilde\varphi_A\big((s_i)_{i\in A}\big) = \sum_{B\subset A} (-1)^{|A\setminus B|} \varphi_B\big((s_i)_{i\in B}\big), \tag{1.15}$$

for every nonempty $A \subset \{1,\ldots,m\}$, and let $\tilde\varphi_\varnothing = \varphi_\varnothing$. According to the inclusion–exclusion principle or the Möbius inversion formula,

$$\varphi(s_1,\ldots,s_m) = \sum_{A\subset\{1,\ldots,m\}} \tilde\varphi_A\big((s_i)_{i\in A}\big)$$

for $\mu^{\otimes m}$-almost all $(s_1,\dots,s_m)\in S^m$. Hence, for every $n\ge m$,

$$\int_{S^m}\varphi\,dL_n^m=\tilde\varphi_0+\sum_{c=0}^m\int_{S^c}\tilde\varphi_c\,dL_n^c \tag{1.16}$$

$\mathbb{P}$-almost surely, where, for every $c\in\{0,1,\dots,m\}$,

$$\tilde\varphi_c\equiv\sum_{\substack{A\subset\{1,\dots,m\}\\ |A|=c}}\tilde\varphi_A. \tag{1.17}$$

Note that every $\tilde\varphi_A$ with nonempty $A\subset\{1,\dots,m\}$ is completely μ-degenerate. For a symmetric φ, formula (1.16) is closely related to the Hoeffding decomposition of the corresponding U-statistic.

Theorem 1.18 (Moderate deviations of partial sums U-empirical measures)**.** *The sequence $\{L_n^m(\cdot),n\in\mathbb{N}\}$ satisfies the MDP in $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ with the good rate function*

$$J_\infty^m(\nu(\cdot))=\int_0^\infty J_m(\dot\nu|\mu)dt$$

for $\nu(\cdot)\in\mathcal{AC}_0[\mathcal{M}(S^m)]$ and $J_\infty^m(\nu(\cdot))=+\infty$ otherwise. The speed is $\frac{n}{b_n^2}$. Moreover the sequence $\{L_n^m(\cdot),n\in\mathbb{N}\}$ satisfies the MDP in $D_2[\mathcal{M}(S^m)]$ with the same good rate function and the same speed.

Remark 1.19. We can derive another representation for the rate function $I_\infty(\cdot)$ in Theorem 1.8:

$$I_\infty(\phi)=\inf\Big\{\int_0^1 J_m(\dot\nu|\mu)\,dt,\nu\in\mathcal{AC}_0[\mathcal{M}(S^m)]\cap K_\infty\text{ and }\int_{S^m}h\,d\nu(\cdot)=\phi(\cdot)\Big\}, \tag{1.20}$$

where $K_\infty:=\cup_{L\ge0}\{\nu(\cdot):\ J_\infty^m(\nu(\cdot))\le L\}$. Therefore Theorem 1.8 can be derived via the contraction principle ([11, Section 4.2]) from Theorem 1.18. A sketch of the proof of (1.20) is given after the proof of Theorem 1.18.

For the proof we will first establish the moderate principle for $U_n^m(\cdot,h)$ for bounded, asymmetric kernels h. Therefore we apply moment inequalities for U-statistics which can be deduced from decoupling and hypercontractive methods (cf. [8, Section 2.5-2.7]). Moreover we will use the MDP for $M_n^m(1)$, established in [16, Theorem 1.21] as well as a Lévy-type maximal inequality for U-statistics.

1.3. **Moderate deviations for functional U-processes.** Let $\mathcal{H}\subset B(S^m,\mathbb{R})$ be a class of functions such that $0\le h\le1$ for all $h\in\mathcal{H}$ and define the pseudo-metric $d_2(g,h)=\left(\int_{S^m}(g-h)^2\,d\mu^{\otimes m}\right)^{1/2}$. Let $l_\infty(\mathcal{H})$ be the Banach space of all bounded real functions on $\mathcal{H}$ with the supremum norm $||H||_{\mathcal{H}}=\sup_{h\in\mathcal{H}}|H(h)|$. Let $D_\beta[l_\infty(\mathcal{H})]$ denote the Banach space of càdlàg functions $H:\mathbb{R}_+\to l_\infty(\mathcal{H})$ such that

$$||H||_{\mathcal{H},\beta}:=\sup_{t\in\mathbb{R}+}\frac{||H_t||_{\mathcal{H}}}{\beta(t)^m}<\infty,$$

equipped with the Borel σ-field. Every signed measure $\nu \in \mathcal{M}(S^m)$ of finite variation corresponds to an element $E_\nu \in l_\infty(\mathcal{H})$ such that $E_\nu(h) = \int h\, d\nu$ for all $h \in \mathcal{H}$. We consider the random variables $E_{M_n^m(\cdot)}$ as elements of $D_\beta[l_\infty(\mathcal{H})]$. Throughout this paper we assume that the class $\mathcal{H}$ is countable or that the processes $\{\int_{S^m} h\, dM_n^m : h \in \mathcal{H}\}$ are separable, i.e. there exists a countable $\mathcal{F} \subset \mathcal{H}$ such that $\int_{S^m} h\, dM_n^m(\omega) \in \mathrm{cl}\big(\{\int_{S^m} f\, dM_n^m(\omega) : f \in \mathcal{F},\, d_2(h,f) < \varepsilon\}\big)$ for almost every ω, every $\varepsilon > 0$ and $h \in \mathcal{H}$.

To state the result we have to introduce some more notations. Given a pseudometric space (T,d), the ε-covering number $N(\varepsilon, T, d)$ is defined as

$$N(\varepsilon, T, d) = \min\{n \in \mathbb{N} : \text{ there exists a covering of } T \text{ by } n \text{ balls of } d\text{−radius } \leq \varepsilon\}.$$

The metric entropy of (T,d) is the function $\log N(\varepsilon, T, d)$. We define $N_2(\varepsilon, \mathcal{H}, \mu) := N(\varepsilon, \mathcal{H}, ||\cdot||_{L_2(\mu)})$. Some classes of functions satisfy a uniform bound in the entropy. A class of real functions $\mathcal{H}$ is Vapnic-Čhervonenkis (VC for short) subgraph class if the subgraphs of the functions in the class form a VC class of sets (subgraph of $h \in \mathcal{H}$: $\{(x,t) \in S^m \times \mathbb{R} : 0 \leq t \leq h(x_1, \ldots, x_m) \text{ or } h(x_1, \ldots, x_m) \leq t \leq 0\}$). For a definition of a VC class see for example [12]. Any finite-dimensional vector space of functions (e.g., polynomials of bounded degree on $\mathbb{R}^d$) is a VC subgraph class. Notice moreover, that if $\mathcal{C}$ is a VC class of sets and q a real function on $\mathcal{C}$, then the class $\{1_C/q(C) : C \in \mathcal{C}\}$ corresponding to a weighted empirical process is a VC subgraph class. The envelope H of $\mathcal{H}$ is defined as $\sup_{h \in \mathcal{H}} |h|$. It is well known ([25, Proposition II 2.5]) that if $\mathcal{H}$ is a VC subgraph class then there are finite constants A and v such that, for each probability measure μ with $\mu^{\otimes m}(H^2) < \infty$,

$$N_2(\varepsilon, \mathcal{H}, \mu) \leq A(\mu^{\otimes m}(H^2)^{1/2}/\varepsilon)^v.$$

Consider the following different typs of classes $\mathcal{H}$:

Condition 1.21. *Let $\mathcal{H}$ be a measurable class of functions $h : S^m \to \mathbb{R}$ satisfying:*

(a) *$\mathcal{H}$ is uniformly bounded.*
(b) *There are $A > 0$ and $v < \infty$ such that $N_2(\varepsilon, \mathcal{H}, \mu) \leq (A/\varepsilon)^v$ for all probability measures μ.*

Let $\{\varepsilon_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d. random variables defined on $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, \mathbb{P})$ with $\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$, independent of the underlying sequence $\{X_i\}_{i \in \mathbb{N}}$. Let $\{X_i^k : i \in \mathbb{N}\}_{k=1}^m$ be i.i.d. copies of $\{X_i : i \in \mathbb{N}\}$ and $\{\varepsilon_i^k : i \in \mathbb{N}\}_{k=1}^m$ be independent copies of $\{\varepsilon_i : i \in \mathbb{N}\}$, independent of $\{X_i^k : i \in \mathbb{N}\}_{k=1}^m$.

Condition 1.22. *Let $\mathcal{H}$ be a measurable class of functions $h : S^m \to \mathbb{R}$ satisfying:*

(a) *$\mathcal{H}$ is uniformly bounded.*
(b) *There is a finite constant c such that*

$$\mathbb{E}\Big(\big\| n^{-m/2} \sum_{I_{m,n}} \varepsilon_{i_1}^1 \varepsilon_{i_2}^2 h(X_{i_1}^1, \ldots, X_{i_m}^m) \big\|_{\mathcal{H}} \Big) \leq c. \tag{1.23}$$

Notice that uniformly bounded VC subgraph classes $\mathcal{H}$ satisfy Condition 1.21 as well as Condition 1.22 (for Condition 1.22, see [2, Proposition 2.6]). We introduce the following quantity:

$$H_1(n,\eta) := \mathbb{E}\Big\| \sum_{i=1}^n \varepsilon_i \pi_{1,m} h(X_i) \Big\|_{\mathcal{H}_{\eta,1}},$$

where

$$\mathcal{H}_{\eta,1} := \{\pi_{1,m} f - \pi_{1,m} g : \; f,g \in \mathcal{H}; \; d_2(\pi_{1,m} f, \pi_{1,m} g) \le \eta\}.$$

Theorem 1.24 (Moderate deviations for functional U-processes)**.**
Assume that the class $\mathcal{H}$ satisfies one of

(a) *Let $\mathcal{H}$ be a class of symmetric functions satisfying Condition 1.21.*
(b) *Assume that $\mathcal{H}$ satisfies Condition 1.22. Moreover assume that $\{\pi_{1,m} h : h \in \mathcal{H}\}$ is totally bounded with respect to d_2 and satisfies*

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{H_1(n,\eta)}{b_n} = 0. \tag{1.25}$$

Then $\{E_{L_n^m(\cdot)}, n \in \mathbb{N}\}$ satisfies the MDP in $D_\beta[l_\infty(\mathcal{H})]$ with speed n/b_n^2 and with the good rate function

$$J_{\mathcal{H}}^m(H(\cdot)) = \inf\Big\{ \int_0^\infty J_m(\dot{\nu}|\mu)\, dt : \; \nu(\cdot) \in \mathcal{AC}_0[\mathcal{M}(S^m)] \text{ and } E_{\nu(\cdot)} = H(\cdot) \Big\}.$$

An important and well studied class is the class of uniformly bounded VC subgraph classes of symmetric functions on S^m with values in $\mathbb{R}$.

Corollary 1.26. *If $\mathcal{H}$ is a uniformly bounded VC subgraph class of symmetric functions, then in both cases of Theorem 1.24 the assumptions on $\mathcal{H}$ are fullfilled.*

Proof. We have already mentioned that a uniformly bounded VC subgraph class of functions satisfies Condition 1.21 as well as Condition 1.22. The proof of Corollary 5.7 in [2] shows that $\{\pi_{1,m} h : \; h \in \mathcal{H}\}$ is a measurable uniformly bounded VC subgraph class. Thus the conditions for $\{\pi_{1,m} h : \; h \in \mathcal{H}\}$ hold. $\square$

Example 1.27. There are several examples of classes $\mathcal{H}$ in the literature satisfying the assumptions of Theorem 1.24. We only mention, that the simplicial depth process, empirical distribution functions of U-statistic structure and a class of uniform Hölder functions come into play. A lot of these examples can be found in [2].

If $\mathcal{H}$ is a uniformly bounded VC subgraph class of symmetric functions, the MDP for $\{E_{L_n^m(1)}, n \in \mathbb{N}\}$ was established in [14]. For the proof of Theorem 1.24 we will apply again the maximal inequality as well as the contraction principle.

2. Decoupling inequalities and consequences

One key for the proofs is a Bernstein type inequality for μ-degenerate U-statistics of degree r with $1 \leq r \leq m$. For bounded $\mathbb{R}$-valued kernel functions, the proof is given in [2, Proposition 2.3], for bounded kernel functions with values in a Banach space of type 2 a Bernstein type inequality is given in [8, Theorem 8.15, Corollary 8.1.5]. In [16] the following Bernstein type inequality for unbounded kernel functions, whose squared norm satisfies the weak Cramér condition, is proved. For the sake of completeness we will give a sketch of the proof.

Lemma 2.1 (Bernstein-type inequality)**.** *Consider a symmetric and completely μ-degenerate kernel function $\varphi : S^r \to \mathbb{R}^d$ with $1 \leq r \leq m$. Assume that there exists an $\alpha > 0$ such that*

$$a \equiv \int_{S^r} \exp\bigl(\alpha\|\varphi\|^2\bigr)\, d\mu^{\otimes r} < \infty. \tag{2.2}$$

Define $\sigma^2 = \mathbb{E}\bigl[\|\varphi(X_1,\dots,X_r)\|^2\bigr]$. If $\mathbb{E}[\varphi(X_1,\dots,X_r)] = 0$, then there exist constants c_1, c_2, c_3, depending only on r, and a constant H depending only on μ and φ such that

$$\mathbb{P}\bigl(\|n^{r/2}U_n^r(\varphi)\| \geq x\bigr) \leq c_1 \exp\left(-\frac{c_2 x^{2/r}}{\sigma^{2/r} + (c_3 H x^{1/r} n^{-1/2})^{2/(r+1)}}\right) \tag{2.3}$$

for all $0 < x < c\sigma^{r+1} n^{r/2} (H^{r/2})^{-1}$ and all $n \geq r$, where c is a constant, depending only on r.

Proof. Using Chebychev's inequality we obtain for every $t > 0$

$$\begin{aligned}
&\mathbb{P}\bigl(\|n^{r/2}U_n^r(\varphi)\| \geq x\bigr)\\
&\quad\leq \exp\bigl(-tx^{2/(r+1)}\bigr)\,\mathbb{E}\bigl[\exp\bigl(t\|n^{r/2}U_n^r(\varphi)\|^{2/(r+1)}\bigr)\bigr]\\
&\quad\leq \exp\bigl(-tx^{2/(r+1)}\bigr)\\
&\qquad\times \mathbb{E}\left[\exp\left((r^r r!)^{2/(r+1)} t \Bigl\|n^{-r/2} \sum_{1\leq i_1<\cdots<i_r\leq n} \varphi(X_{i_1},\dots,X_{i_r})\Bigr\|^{2/(r+1)}\right)\right],
\end{aligned}$$

since $\binom{n}{r}^{-1} \leq n^{-r} r^r r!$. Now we apply Borell's inequality (a Khinchin-type inequality for the Rademacher chaos, cf. [8, (2.6.5)]), the decoupling technique [8, Theorem 2.5.4], a symmetrization lemma [8, Lemma 2.4.5] as well as Hoffdings formula [8, (1.1.16)]. We get as in the proof of [8, Theorem 8.1.2]:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left((r^r r!)^{2/(r+1)} t \Bigl\|n^{-r/2} \sum_{1\leq i_1<\cdots<i_r\leq n} \varphi(X_{i_1},\dots,X_{i_r})\Bigr\|^{2/(r+1)}\right)\right]\\
&\leq c_1 \exp\bigl(c_2\sigma^2 t^{r+1}\bigr)\,\mathbb{E}\left[\exp\left(2r\,c_2 t^{r+1} n^{-1} \sum_{i=1}^{\lfloor n/r\rfloor} \bigl(\|\varphi(X_{(i-1)r+1},\dots,X_{ir})\|^2 - \sigma^2\bigr)\right)\right].
\end{aligned}$$

Since $a < \infty$ by (2.2), we obtain for each $l \geq 2$ the inequality $\int_{S^r} \|\varphi\|^{2l}\, d\mu^{\otimes r} \leq \frac{1}{2} l!\, H^{l-2} \int_{S^r} \|\varphi\|^4\, d\mu^{\otimes r}$ with $H \equiv \sup_{l\geq 3} (2a)^{1/(l-2)} (\alpha^l \int_{S^r} \|\varphi\|^4\, d\mu^{\otimes r})^{-1/(l-2)} < \infty$, cf. [29, Remark 3.6.1]. Thus we can apply [28, Theorem 2.1], especially formula [28, (2.7)]: We obtain

$$\mathbb{E}\Big[\exp\Big(h\sum_{i=1}^{\lfloor n/r\rfloor}\big(\|\varphi(X_{(i-1)r+1},\dots,X_{ir})\|^2-\sigma^2\big)\Big)\Big]\le\exp\left(\frac{h^2B_n^2}{2(1-hH)}\right)$$

for every $h\in[0,1/(H))$, where $B_n^2=\lfloor n/r\rfloor\int_{S^r}\|\varphi\|^4\,d\mu^{\otimes r}$. Considering $h=2r\,c_2t^{r+1}/n$ we obtain

$$\begin{aligned}\mathbb{E}\Big[\exp\Big((r^r r!)^{2/(r+1)}t\,\Big\|n^{-r/2}\sum_{1\le i_1<\cdots<i_r\le n}\varphi(X_{i_1},\dots,X_{i_r})\Big\|^{2/(r+1)}\Big)\Big]&\\ \le c_1\exp\Big(c_2t^{r+1}\sigma^2+\frac{2r\,c_2^2t^{2(r+1)}\int_{S^r}\|\varphi\|^4\,d\mu^{\otimes r}}{n-2r\,c_2t^{r+1}H}\Big).&\end{aligned}$$

for all $t^{r+1}\le n(2r\,c_2H)^{-1}$. Notice that the r-dependent constants change from step to step. Thus we get estimates similar to [2, (2.7),(2.8)] and the result follows by adopting the calculations of the proof of [2, Proposition 2.3(c), page 1503]. $\square$

Another key is a Lévy–type maximal inequality for the tails of U-statistics. Although there are some places in the literature, where the decoupling method is used to prove some maximal estimations for non-linear statistics (see for example [4, Lemma 2.4], where maximal inequalities for moments are obtained, [19, Lemma 3.3], where maximal tail estimations are proved for U-statistics with decoupled entries; see also [18, Chapter 2 and 3]), the following result does not exist in the literature. The proof was suggested by E. Giné in a private communication.

Lemma 2.4 (Lévy–type maximal inequality, $m=2$). *Let $\{X_n,n\in\mathbb{N}\}$ be a sequence of independent identically distributed random variables with values in a measurable space $(S,\mathcal{S})$ and suppose that the kernel function $h:\ S^2\to\mathbb{R}^d$ is a symmetric measurable function. Then there exist universal constants e_1,e_2 and f_1,f_2 (finite, positive and depending only on m) such that*

$$\begin{aligned}&\mathbb{P}\Big(\max_{k\le n}\big\|\sum_{I(2,k)}h(X_i,X_j)\big\|>t\Big)\\ &\le e_1\mathbb{P}\Big(\big\|\sum_{I(2,n)}h(X_i,X_j)\big\|>e_2\,t\Big)+f_1\mathbb{P}\Big(\big\|\sum_{i=1}^n h(X_i,X_i)\big\|>f_2\,t\Big).\end{aligned}\tag{2.5}$$

Since the diagonal terms are coming into play, it is more complicated to formulate this type of inequality for an arbitrary degree m. For $c=1,\dots,m$ let us define the symmetric function $h_{mc}:\ S^c\to\mathbb{R}^d$ by

$$h_{mc}(x_1,\dots,x_c):=\sum_{\substack{\nu_1+\cdots+\nu_c=m\\ \nu_j\ge1}}\frac{m!}{\nu_1!\cdots\nu_c!}h(x_1^{\nu_1},\dots,x_c^{\nu_c}),$$

where $x_j^{\nu_j} = (x_j, \dots, x_j)$ is a vector where all ν_j coordinates are equal to x_j. For $n \geq m$ we get the following identity:

$$\sum_{i_1,\dots,i_m=1}^{n} h(X_{i_1}, \dots, X_{i_m}) = \sum_{c=1}^{m} \sum_{I(c,n)} h_{mc}(X_{i_1}, \dots, X_{i_c}) \tag{2.6}$$

(see (1.3.3) in [8]).

Lemma 2.7 (Lévy–type maximal inequality)**.** *Let $\{X_n, n \in \mathbb{N}\}$ be a sequence of independent identically distributed random variables with values in a measurable space $(S, \mathcal{S})$ and suppose that the kernel function h is a symmetric measurable $\mathbb{R}^d$-valued function. Then there exist universal constants $e_c(m)$ and $f_c(m)$, $c = 1, \dots, m$ (finite, positive and depending only on m) such that*

$$\begin{aligned}\mathbb{P}\Big(\max_{k\leq n} \Big\| \sum_{I(m,k)} h(X_{i_1}, \dots, X_{i_m})\Big\| > t\Big) \\ \leq \sum_{c=1}^{m} e_c(m)\, \mathbb{P}\Big(\Big\| \sum_{I(c,n)} h_{mc}(X_{i_1}, \dots, X_{i_c})\Big\| > f_c(m)\, t\Big).\end{aligned} \tag{2.8}$$

A decoupling inequality for the tail probabilities of multivariate U-statistics is the key for the proof of Lemma 2.4 and 2.7. We state here the pertinent result from de la Peña and Montgomery-Smith ([23, Theorem 1]):

Proposition 2.9 (Decoupling for tails)**.** *For natural numbers $n \geq m$, let $\{X_n, n \in \mathbb{N}\}$ be a sequence of independent random variables with values in a measurable space $(S, \mathcal{S})$, and let $\{X_n^k, n \in \mathbb{N}\}$, $k = 1, \dots, m$, be m independent copies of this sequence. Let E be a separable Banach space and, for each $(i_1, \dots, i_m) \in I(n,m)$, let $h_{i_1,\dots,i_m} : S^m \to E$ be measurable functions. Then there are constants $c_m \in (0, \infty)$, depending only on m, such that for all $t > 0$*

$$\begin{aligned}\mathbb{P}\Big(\Big\| \sum_{I(m,n)} h_{i_1,\dots,i_m}(X_{i_1}, \dots, X_{i_m})\Big\|_E > t\Big) \\ \leq c_m \mathbb{P}\Big(c_m \Big\| \sum_{I(m,n)} h_{i_1,\dots,i_m}(X_{i_1}^1, \dots, X_{i_m}^m)\Big\|_E > t\Big).\end{aligned} \tag{2.10}$$

If, moreover, the functions $h_{i_1,\dots,i_m}$ are symmetric in the sence that, for all $x_1, \dots, x_m \in S$ and all permutations π of $\{1, \dots, m\}$,

$$h_{i_1,\dots,i_m}(x_1, \dots, x_m) = h_{i_{\pi(1)},\dots i_{\pi(m)}}(x_{\pi(1)}, \dots x_{\pi(m)}),$$

then the reverse inequality holds true that is, there are constants $d_m \in (0, \infty)$ depending only on m, such that for all $t > 0$

$$\begin{aligned}\mathbb{P}\Big(\Big\| \sum_{I(m,n)} h_{i_1,\dots,i_m}(X_{i_1}^1, \dots, X_{i_m}^m)\Big\|_E > t\Big) \\ \leq d_m \mathbb{P}\Big(d_m \Big\| \sum_{I(m,n)} h_{i_1,\dots,i_m}(X_{i_1}, \dots, X_{i_m})\Big\|_E > t\Big).\end{aligned} \tag{2.11}$$

Moreover for the proof of Lemma 2.4 and 2.7 we use an extension of the classical Lévy inequality for sums of independent symmetric random vectors to sums of not necessarily symmetric, but i.i.d. random vectors. The proof has been found recently in [18, Section 2.4]. Note that the crucial points for this extension consists of [22, Theorem 1] and of a simple consequence of the Lévy-Ottaviani inequality, that is: let $\{X_i\}_{i \in \mathbb{N}}$ be independent random vectors. Then for all $t \geq 0$,

$$\mathbb{P}\Big(\max_{k \leq n} \| \sum_{i=1}^k X_i \| > t \Big) \leq 3 \max_{k \leq n} \mathbb{P}\Big(\| \sum_{i=1}^k X_i \| > 1/3 \, t \Big).$$

Proposition 2.12 (Lévy's inequality for asymmetric random vectors). *There exist universal constants c_1 and c_2 such that if $\{X_i\}_{i \in \mathbb{N}}$ are i.i.d. E-valued random variables, where $(E, \| \cdot \|_E)$ is a separable Banach space, then, for $1 \leq k \leq n$,*

$$\mathbb{P}\Big(\max_{k \leq n} \| \sum_{i=1}^k X_i \|_E > t \Big) \leq c_1 \, \mathbb{P}\Big(\| \sum_{k=1}^n X_i \|_E > c_2 \, t \Big).$$

We can choose $c_1 = 9$ and $c_2 = 1/30$.

Propositions 2.9 and 2.12 can be extended to Banach spaces E not necessarily separable, a situation that arises with U-processes. For details a good reference is [18] and references therein. Thus we obtain (details are left to the reader):

Corollary 2.13. *Like Proposition 2.9 and Proposition 2.12, Lemma 2.7 has also a version for U-processes indexed by arbitrary families of kernels.*

Proof of Lemma 2.4 and 2.7. For each $k \in \mathbb{N}$ define new functions h_k taking values in $(l_\infty^n, \| \cdot \|_\infty)$, the space $(\mathbb{R}^d)^n$ endowed with the supremum norm, as follows:

$$h_k := (\underbrace{0, \dots, 0}_{k \text{ times}}, h, h, \dots, h).$$

Thus it is obvious that

$$\max_{k \leq n} \| \sum_{I(m,k)} h(X_{i_1}, \dots, X_{i_m}) \| = \| \sum_{I(m,n)} h_{i_1 \vee i_2 \cdots \vee i_m}(X_{i_1}, \dots, X_{i_m}) \|_\infty . \tag{2.14}$$

Applying first (2.14), than Proposition 2.9 for $E = l_\infty^n$ and again (2.14) it follows that

$$\begin{aligned} &\mathbb{P}\Big(\max_{k \leq n} \| \sum_{I(m,k)} h(X_{i_1}, \dots, X_{i_m}) \| > t \Big) \\ &\leq c_m \, \mathbb{P}\Big(c_m \| \sum_{I(m,n)} h_{i_1 \vee i_2 \vee \cdots \vee i_m}(X_{i_1}^1, \dots, X_{i_m}^m) \| > t \Big) \qquad (2.15) \\ &= c_m \, \mathbb{P}\Big(c_m \max_{k \leq n} \| \sum_{I(m,k)} h(X_{i_1}^1, \dots, X_{i_m}^m) \| > t \Big). \end{aligned}$$

For clarify of exposition we present only the proof of Lemma 2.4. This is notationally much less involved than the general case and already contains the main idea. The general case is proved by recurrence. Remark that

$$\begin{aligned} &\mathbb{P}\Big(\max_{k\le n} \big\|\sum_{I(2,k)} h(X_i^1,X_j^2)\big\| > t\Big) \\ &\le \mathbb{P}\Big(\max_{k\le n} \big\|\sum_{i,j=1}^{k} h(X_i^1,X_j^2)\big\| > t/2\Big) + \mathbb{P}\Big(\max_{k\le n}\big\|\sum_{i=1}^{k} h(X_i^1,X_i^2)\big\| > t/2\Big). \end{aligned} \tag{2.16}$$

The result can now be obtained by conditionally applying Lévy's inequality (Proposition 2.12) 2 times: Let s_l be a defined as $\sum_{i=1}^k h(X_i^1,X_l^2)$, $1\le l\le k$. We denote by $\mathbb{P}_2$ the conditional probability given $\{X_i^1, i\in\mathbb{N}\}$. Now Lévy's inequality applied to the conditionally independent and identically distributed random variables s_l gives

$$\begin{aligned} &\mathbb{P}\Big(\max_{k\le n} \big\|\sum_{i,j=1}^{k} h(X_i^1,X_j^2)\big\| > t/2\Big) = \mathbb{E}\,\mathbb{P}_2\Big(\max_{k\le n}\big\|\sum_{l=1}^{k} s_l\big\| > t/2\Big) \\ &\quad\le 9\,\mathbb{E}\,\mathbb{P}_2\Big(30\,\big\|\sum_{l=1}^{n} s_l\big\| > t/2\Big) = 9\,\mathbb{P}\Big(\max_{k\le n}\big\|30\sum_{i=1}^{k}\sum_{j=1}^{n} h(X_i^1,X_j^2)\big\| > t/2\Big). \end{aligned} \tag{2.17}$$

The inequality follows now by iteration and by (2.11) of Proposition 2.9; moreover we apply Proposition 2.12 as well as (2.11) of Proposition 2.9 for the second term in (2.16). The constants are $e_1 = c_2\,d_2\,9^2$ and $e_2 = (4\,c_2\,d_2\,30^2)^{-1}$ as well as $f_1 = c_2(9+9^2)$ and $f_2 = (4\,c_2\,30^2)^{-1}$. Using the identity (2.6) we obtain the result in decomposing these statistics again and again and conditionally applying Lévy's inequality. The constants $e_c(m)$ and $f_c(m)$ can be calculated explicitely as a function of the degree m. We omit this calculation, since for our application it is enough to know, that the constants depend only on m. □

Remark 2.18. It is a surprise that we have to care about all diagonal terms in the maximal inequality for U-statistics, see (2.8). For a symmetrized kernel function $\varepsilon_{i_1}^1\,\varepsilon_{i_2}^2\cdots\varepsilon_{i_m}^m\,h(X_{i_1}^1,\dots,X_{i_m}^m)$ one can neglect the diogonal terms. The following arguments were given in the proof of Lemma 3.3 in [19]. Let

$$I(m,n,k,l) := \{(i_1,\dots,i_m)\in I(m,n) :\ i_j\le k \text{ for } j\le l\}.$$

and set s_i be a vector in l_∞^n with coordinates all possible sums

$$\sum_{i_j\ne i,(i_1,\dots,i_{m-1},i)\in I(m,n,k,m-1)} \varepsilon_{i_1}^1\cdots\varepsilon_{i_m}^m\,h(X_{i_1}^1,\dots,X_{i_{m-1}}^{m-1},X_i^m)$$

with $1\le k\le n$. $\mathbb{P}_m$ denotes the conditional probability given $\{X_j^r, r\ne m, j\in\mathbb{N}\}$. For the symmetrized kernel the Lévy type inequality for U-statistics now can be obtained by conditionally applying Lévy's inequality for the symmetric random vectors s_i ([20, Proposition 2.3]) m times:

$$\mathbb{P}\Big(\max_{k\le n} \| \sum_{I(m,n,k,m)} \varepsilon^1_{i_1}\cdots\varepsilon^m_{i_m} h(X^1_{i_1},\dots,X^m_{i_m})\| > t\Big)$$
$$= \mathbb{E}\mathbb{P}_m\left(\max_{k\le n} \|\sum_{i=1}^k s_i\|_\infty > t\right) \le 2\,\mathbb{E}\,\mathbb{P}_m\left(\|\sum_{i=1}^n s_i\|_\infty > t\right)$$
$$\le 2\,\mathbb{P}\Big(\max_{k\le n} \| \sum_{I(m,n,k,m-1)} \varepsilon^1_{i_1}\cdots\varepsilon^m_{i_m}\, h(X^1_{i_1},\dots,X^m_{i_m})\| > t\Big).$$

The maximal inequality follows by iteration and by Proposition 2.9. But notice that inequalities comparing tail probabilities for the random variables $\sum_{I(m,n)} h(X_{i_1},\dots,X_{i_m})$ and $\sum_{I(m,n)} \varepsilon_{i_1}\cdots\varepsilon_{i_m} h(X_{i_1},\dots,X_{i_m})$ (which are referred to as randomization inequalities) are obtained only one sided: one obtains by Theorem 1.5, Chapter 3, in [18]:

$$\mathbb{P}\Big(\| \sum_{I(m,n)} \varepsilon^1_{i_1}\cdots\varepsilon^m_{i_m}\, h(X^1_{i_1},\dots,X^m_{i_m})\| > t\Big)$$
$$\le c(m)\mathbb{P}\Big(\| \sum_{I(m,n)} h(X_{i_1},\dots,X_{i_m})\| > c(m)t\Big)$$

for some constants $c(m)$ depending only on m. But there is no converse to this inequality in general, even for $m = 1$: see the counterexample in Chapter 3 in [18] (after Theorem 1.5). Thus we have to take into consideration the diagonal terms.

We will use the following moment inequality for U-statistics established with decoupling and hypercontractive methods ([8, Theorem 2.7.1, Corollary 2.7.1]).

Proposition 2.19. *Let E be a Banach space of type p, $1 \le p \le 2$ with norm $\|\cdot\|_E$. Let $\{X_n, n\in\mathbb{N}\}$ be a sequence of independent random variables with identical distributions and suppose that the kernel function h is a symmetric E-valued function with rank m and $\mathbb{E}\|h\|^q_E < \infty, q\ge 1$. Then*

$$\mathbb{E}\| \sum_{I(m,n)} h(X_{i_1},\dots,X_{i_m})\|^q_E$$
$$\le (6\,A^{1/p}\,m^3)^{mq}(1\vee(q-1)^{m\,q/2})\,(n_{(m)})^{1\vee(q/p)}\mathbb{E}\|h(X_1,\dots,X_m)\|^q_E, \tag{2.20}$$

where A is the type p constant, depending only on E and p.

We will use the following two typs of Bernstein-type inequalities for U-processes. The first one is proved with decoupling techniques in [3, Theorem 3.2].

Proposition 2.21 (Bernstein inequality for U-processes, type 1)**.**
Let $\mathcal{H}$ be a measurable class of μ-canonical functions $h: S^m \to \mathbb{R}$ satisfying:

(a) *$\mathcal{H}$ is uniformly bounded by 1.*

(b) *There is a Lebesgue integrable function* $\lambda: (0,\infty) \to [0,\infty)$ *such that for each probability measure* ν

$$(\log N_2(\varepsilon, \mathcal{H}, \nu))^{m/2} \leq \lambda(\varepsilon), \quad \varepsilon > 0.$$

Then there are constants c *and* c'*, depending only on* m *such that for all* $t > 0$ *and* $n \geq m$

$$\mathbb{P}\Big(\big\| n^{-m/2} \sum_{I(m,n)} h(X_{i_1}, \dots, X_{i_m}) \big\|_{\mathcal{H}} \geq t \Big) \leq c \, \exp(-c' \, t^{2/m}). \tag{2.22}$$

The second is a modification of [1, Proposition 4].

Lemma 2.23 (Bernstein inequality for U-processes, type 2)**.**
Let $\mathcal{H}$ *be a measurable class of* μ*-canonical functions* $h: S^m \to \mathbb{R}$ *satisfying Condition 1.22 with* $\sup_{h \in \mathcal{H}} |h(X)| \leq 1$*. Then there are constants* c *and* c'*, depending only on* m *such that for all* $t > 0$ *and* $n \geq m$ (2.22) *holds.*

Proof. We combine some arguments of the proofs of [1, Proposition 1 and 4] with some of the proof of [3, Theroem 3.2]. We only give a sketch of the proof. Decoupling and symmetrization gives

$$\begin{aligned} &\mathbb{E}\Big[\exp\Big(t \, \big\| n^{-m/2} \sum_{I(m,n)} h(X_{i_1}, \dots, X_{i_m}) \big\|_{\mathcal{H}}^{1/m} \Big) \Big] \\ &\quad \leq \mathbb{E}\Big[\exp\Big(2^2 m^m t \, \big\| n^{-m/2} \sum_{I(m,n)} \varepsilon_{i_1}^1 \varepsilon_{i_2}^2 \, h(X_{i_1}^1, \dots, X_{i_m}^m) \big\|_{\mathcal{H}}^{1/m} \Big) \Big]. \end{aligned}$$

For $X = \sum_{I(m,n)} a_{i_1,\dots,i_m} \, \varepsilon_{i_1} \cdots \varepsilon_{i_m}$ (with symmetric $a_{i_1,\dots,i_m}$, which are elements in a Banach space $(B, \|\cdot\|)$), there exists constants c_1, c_2, depending only on m, such that

$$\mathbb{E}\big(\exp(\|X\|^{1/m})\big) \leq c_1 \exp\big(c_2 \mathbb{E}(\|X\|^2)^{1/m}\big)$$

(cf. [2, (2.4)], the proof used the Khinchin-type inequality for the Rademacher chaos, cf. [2, Proposition 2.2]). Thus we obtain that the right hand side of the last expression is bounded by

$$c_1 \, \mathbb{E}\Big[\exp\Big(c_2 2^2 m^m t^2 \Big(\mathbb{E}_\varepsilon \big\| n^{-m/2} \sum_{I(m,n)} \varepsilon_{i_1}^1 \varepsilon_{i_2}^2 \, h(X_{i_1}^1, \dots, X_{i_m}^m) \big\|_{\mathcal{H}}^2 \Big)^{1/m} \Big) \Big].$$

For a Rademacher chaos K of order 2, using $\|X\|_p \leq (p-1)^{m/2} \|X\|_2$ for all $p \geq 2$ (cf. [2, Proposition 2.2]) and Hölder's inequality we obtain $\|K\|_2 \leq 2^3 \|K\|_1$ (cf. [1, Proposition 4]). Thus by our hypothesis the expectation $\mathbb{E}_\varepsilon$ is uniformly bounded in n. The result follows by Markov's inequality applied to the function $\exp\big(\lambda \| \cdot \|_{\mathcal{H}}^{1/m}\big)$ $\square$

3. Proofs of the moderate deviations results

Proof of Theorem 1.8. We will check that the linear term in the Hoeffding decomposition (1.16) is the leading term for the moderate deviations behaviour. Therefore we will prove that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} \sup_{t\in[0,1]} \Big\| \sum_{k=2}^{m} \binom{m}{k} U_n^k(t, \pi_{k,m}h) \Big\| \geq \delta \right) = -\infty \tag{3.1}$$

for all $\delta > 0$. Applying [11, Theorem 4.2.13] we see that the sequences $\{U_n^m(\cdot, h), n \in \mathbb{N}\}$ and $\{U_n^1(\cdot, \pi_{1,m}h), n \in \mathbb{N}\}$ have identical moderate deviations behaviour. Notice that the linear term $U_n^1(\cdot, \pi_{1,m}h)$ in the Hoeffding decomposition is of the form

$$U_n^1(\cdot, \pi_{1,m}h) = m \frac{1}{n} \sum_{i=1}^{[n\cdot]} \pi_{1,m}h(X_i).$$

Therefore the result follows from the MDP for linear statistics. By definition the values of $U_n^k(\cdot, \pi_{k,m}h)$ are constant on each intervall $[i/n, (i+1)/n)$ for $i \in \{0, \dots, n\}$ and $U_n^k(t, \pi_{k,m}h) = U_n^k(i/n, \pi_{k,m}h)$ for all $t \in [i/n, (i+1)/n)$. Lemma 2.7 yields

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in[0,1]} \Big\| \frac{n}{b_n} U_n^k(t, \pi_{k,m}h) \Big\| \geq \delta \Big) &\leq \mathbb{P}\left(\max_{i\in\{0,\dots,n\}} \frac{n}{b_n} U_n^k(i/n, \|\pi_{k,m}h\|) \geq \delta \right) \\ &\leq \sum_{c=1}^{k} e_c(k) \mathbb{P}\left(\frac{n}{b_n\, n_{(k)}} \Big\| \sum_{I(c,n)} (\pi_{k,m}h)_{kc}(X_{i_1}, \dots, X_{i_c}) \Big\| \geq f_c(k)\, \delta \right).\end{aligned} \tag{3.2}$$

For each fixed $2 \leq k \leq m$ we get for the k-th summand with Lemma 2.1 for sufficiently large n:

$$\begin{aligned}&\mathbb{P}\Big(\Big\| \frac{n}{b_n} U_n^k(1, \pi_{k,m}h) \Big\| \geq f_k(k)\, \delta \Big) = \\ &\qquad \mathbb{P}\Big(\Big\| n^{k/2} U_n^k(1, \pi_{k,m}h) \Big\| \geq \frac{f_k(k)\, b_n \delta n^{k/2}}{n} \Big) \\ &\qquad \leq c_1(k) \exp\Big(-c_2(k) \Big(\frac{\delta^{2/k} b_n^{2/k} n^{1-2/k}}{\sigma^{2/k} + (\delta^{1/k} H c_3^k)^{2/(k+1)} (b_n/n)^{2/(k(k+1))}} \Big) \Big),\end{aligned} \tag{3.3}$$

where $c_i(k), i = 1, 2$, are some constants depending only on k, which change from step to step. For n sufficiently large this implies

$$\begin{aligned}&\frac{n}{b_n^2} \log \mathbb{P}\Big(\Big\| \frac{n}{b_n} U_n^k(1, \pi_{k,m}h) \Big\| \geq f_k(k)\, \delta \Big) \\ &\quad \leq \frac{n}{b_n^2} \log c_1(k) - c_2(k) \Big(c_1(k) \Big(\frac{n}{b_n} \Big)^{2-2/k} \Big) \Big(\sigma^{2/r} + c_2(k) \Big(\frac{b_n}{n} \Big)^{2/(k(k+1))} \Big)^{-1}.\end{aligned} \tag{3.4}$$

The right hand side decreases in n to $-\infty$ by the assumptions for $\{b_n\}_{n\in\mathbb{N}}$. For $c = 1,\dots,k-1$ we obtain that

$$\mathbb{P}\left(\frac{n}{b_n\, n_{(k)}}\Big\|\sum_{I(c,n)}(\pi_{k,m}h)_{kc}(X_{i_1},\dots,X_{i_c})\Big\| \geq f_c(k)\,\delta\right)$$
$$\leq \mathbb{P}\bigg(\big\|n^{c/2}U_n^c(1,(\pi_{k,m}h)_{kc})\big\| \geq \frac{f_c(k)\,\delta\, b_n\, n^{c/2}}{n}\bigg),$$

and therefore by the same calculations (as in (3.3) and (3.4)) (3.1) follows from Lemma 2.1. Notice that we apply the Bernstein type inequality [2, Proposition 2.3(b)] (more precisely the unbounded version) to see that the summands for $c = 1,\dots,k-1$ in (3.2) do not contribute to the MDP behaviour. Since

$$\left\{\sup_{t\in[0,1]}\Big\|\frac{n}{b_n}\sum_{k=2}^m\binom{m}{k}U_n^k(t,\pi_{k,m}h)\Big\|\geq\delta\right\}$$
$$\subset\bigcup_{k=2}^m\left\{\frac{n}{b_n}\sup_{t\in[0,1]}\Big|\Big|\binom{m}{k}U_n^k(t,\pi_{k,m}h)\Big|\Big|\geq\frac{\delta}{m}\right\},$$

Lemma 1.2.15 in [11] yields (3.1). □

Proof of Theorem 1.18. We follow the lines of the proof of Theorem 1(b) in [10]. First we will check for fixed $d\in\mathbb{N}$ and $f\in B(S^m,\mathbb{R}^d)$ that the sequence $\{\langle f, L_n^m(\cdot)\rangle, n\in\mathbb{N}\}$ satisfies a MDP in $D[\mathbb{R}^d]$ with a convex good rate function

$$I_f(\phi)=\int_0^\infty\Lambda^*(\dot\phi)\,dt$$

for $\phi\in\mathcal{AC}_0([0,1],\mathbb{R}^d)$ and $I_f(\phi)=+\infty$ otherwise. Here $\Lambda^*(\theta)=\sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda,\theta\rangle-\frac{1}{2}\int(\langle\lambda,\widetilde{f}_1-\int\widetilde{f}_1\,d\mu\rangle)^2\,d\mu\}$, where $\widetilde{f}_1$ denotes the completely μ-degenerate function for f as in (1.17). First we check the MDP on the space $D([0,1],\mathbb{R}^d)$, the space of càdlàg functions from $[0,1]$ to $\mathbb{R}^d$, equipped with the topology of uniform convergence. Again we check that the non-linear terms of the decomposition (1.16) do not contribute to the MDP. The MDP then follows from [6] (see also [21] and [5]). It suffices to prove that for every $\eta>0$

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\bigg(\frac{n}{b_n}\sup_{t\in[0,1]}\Big\|\int_{S^a}\tilde{f}_a\,dL_n^a(t)\Big\|\geq\eta\bigg)=-\infty \tag{3.5}$$

for every $a\in\{2,\dots,m\}$, where $\widetilde{f}_a$ denotes the completely μ-degenerate function for f as in (1.17). Define $k_n=b_n^2/n$. Due to (1.1), we may assume that $k_n\geq 2$ for all $n\in\mathbb{N}$. Using Chebychev's inequality for the first step and Lemma 2.19 for the second step we obtain the estimate

$$\begin{aligned}
&\mathbb{P}\Big(\frac{n}{b_n}\Big\|\int_{S^a}\tilde f_a\,dL_n^a(1)\Big\|\geq\eta\Big)\\
&\quad\leq\Big(\frac{n}{b_n\,\eta\,n_{(a)}}\Big)^{k_n}\mathbb{E}\Big[\Big\|\sum_{(i_1,\dots,i_a)\in I(a,n)}\tilde f_a(X_{i_1},\dots,X_{i_a})\Big\|^{k_n}\Big]\\
&\quad\leq\Big(\frac{n}{b_n\,\eta\,\sqrt{n_{(a)}}}\Big)^{k_n}\big(6a^3\sqrt{const(k_n-1)}\big)^{ak_n}\int_{S^a}\|\tilde f_a\|^{k_n}\,d\mu^{\otimes a}.
\end{aligned}$$

Moreover with the notation of (2.6) we get for every $c=1,\dots,a-1$:

$$\mathbb{P}\Big(\frac{n}{b_n\,n_{(a)}}\Big\|\sum_{I(c,n)}(\tilde f_a)_{ac}(X_{i_1},\dots,X_{i_c})\Big\|\geq\eta\Big)\leq\mathbb{P}\Big(\frac{n}{b_n}\Big\|\int_{S^c}(\tilde f_a)_{ac}\,dL_n^c(1)\Big\|\geq\eta\Big),$$

which can be bounded via Chebychev's inequality and Lemma 2.19 as above. Using Lemma 2.7 this implies (3.5), since f is bounded. The result can easily be adapted to cover the time intervall $[0,T]$ instead of $[0,1]$, then we apply the projective limit approach Theorem 4.6.1 of [11] for $T\in\mathbb{N}$ and the MDP for the sequence $\{\langle f,L_n^m(\cdot)\rangle, n\in\mathbb{N}\}$ follows for $f\in B(S^m,\mathbb{R}^d)$, $d\in\mathbb{N}$ fixed (for the form of the rate function, see [11, (5.1.11)]). The MDP for $\{\langle f,L_n^m(\cdot)\rangle, n\in\mathbb{N}\}$ with the same good rate function extends to the space $D_\beta^{\text{proj}}[\mathbb{R}^d]$ (and hence also to $D_2^{\text{proj}}[\mathbb{R}^d]$): First of all note that almost surely $\langle f,L_n^m(\cdot)\rangle\in D_\beta[\mathbb{R}^d]\subset D_\beta^{\text{proj}}[\mathbb{R}^d]$ (this is the LIL for non-degenerate U-statistics, see for example [3, Corollary 3.5]). Moreover we will show that $\mathcal{D}_{I_f}:=\{\phi:\ I_f(\phi)<\infty\}\subset D_\beta[\mathbb{R}^d]\subset D_\beta^{\text{proj}}[\mathbb{R}^d]$. Let ϕ be choosen such that $I_f(\phi)\leq L$, $L>0$, it follows that $\phi(t)\leq\sqrt{LMt}$ for every $t\in\mathbb{R}_+$, where M is a constant such that $\Lambda^*(y)\geq\|y\|^2/M$. Thus $\phi\in D_\beta[\mathbb{R}^d]$. We apply the contraction principle [11, Theorem 4.2.23] (see the proof of [10, Theorem 1(b)]) to the measurable extension to $D_1[\mathbb{R}^d]$ of the identity map on $D_\beta^{\text{proj}}[\mathbb{R}^d]$. Therefore notice that the topology on $D_\beta^{\text{proj}}[\mathbb{R}^d]$ is generated by the familiy of metrics $d_T(y,z):=\sup_{t\in[0,T]}\|y(t)-z(t)\|/(\beta(t))$, $y,z\in D_\beta^{\text{proj}}[\mathbb{R}^d]$ with $T>0$. This family is separating. Using the concept of exponential equivalence in completely regular topological spaces ([17, Theroem 1.5]) we have to verify that for every $\eta>0$ and every $T>0$

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\Big(\frac{n}{b_n}\sup_{t\in[0,T]}\Big\|\int_{S^a}\tilde f_a\,dL_n^a(t)\Big\|/(\beta(t))\geq\eta\Big)=-\infty \tag{3.6}$$

for every $a\in\{2,\dots,m\}$. This follows by Chebychev's inequality, Lemma 2.7 and Lemma 2.19 as above. Applying the contraction principle [11, Theorem 4.2.23], Etemadi's maximal inequality and Hoeffdings's inequality, we get the MDP for $\{\int_S\widetilde{f}_1\,dL_n(\cdot), n\in\mathbb{N}\}$ in $D_\beta^1[\mathbb{R}^d]$ as in the proof of [10, Theorem 1(b)]. Here $D_\beta^1[\mathbb{R}^d]$ denotes the space of all càdlàg functions $f:\ \mathbb{R}_+\to\mathbb{R}^d$ such that $\sup_{t\in\mathbb{R}_+}\frac{\|f(t)\|}{\beta(t)}<\infty$.

To strengthen the MDP $\{\langle f,L_n^m(\cdot)\rangle, n\in\mathbb{N}\}$ to the topology induced by the norm $\|f\|_\beta:=\sup_{t\in\mathbb{R}_+}\|f(t)\|/(\beta(t)^m)$ and hence also to $D_2[\mathbb{R}^d]$, we apply the contraction principle [11, Theorem 4.2.23], by which it suffices to prove exponential equivalence in the space $D_\beta[\mathbb{R}^d]$. Let F denote a measurable extension to $D_1[\mathbb{R}^d]$ of the identity map on $D_\beta[\mathbb{R}^d]$ and consider its continuous approximations $F_l : D_1[\mathbb{R}^d] \to D_\beta[\mathbb{R}^d]$ such that $F_l(y)(t) = y(t)1_{[0,l)}(t)$. We have already checked that $\mathbb{P}\big(\frac{n}{b_n}\langle f, L_n^m(\cdot)\rangle \in D_\beta[\mathbb{R}^d]\big) = 1$ and since $\phi(t) \le \sqrt{LMt}$ for all ϕ with $I_f(\phi) \le L$, we obtain for every $L < \infty$

$$\lim_{l\to\infty} \sup_{\phi:\ I_f(\phi)\le L} \|F_l(\phi) - F(\phi)\|_\beta = 0.$$

Thus the MDP extends to the Banach space $D_\beta[\mathbb{R}^d]$ and hence also to $D_2[\mathbb{R}^d]$ if we show that for every $\eta > 0$ and for each $a \in \{1, \dots, m\}$

$$\lim_{l\to\infty} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\sup_{t\ge l} \frac{n}{b_n} \Big\| \int_{S^a} \tilde f_a \, dL_n^a(t) \Big\|_\beta \ge \eta \Big) = -\infty.$$

Notice that for a symmetric and completely μ-degenerate function $g : S^r \to \mathbb{R}^d$ with $\|g\| \le c$ we get by [2, Proposition 2.3(d)]:

$$\mathbb{P}\bigg(\Big\| n^{-r/2} \sum_{(i_1,\dots,i_r)\in I(r,n)} g(X_{i_1}, \dots, X_{i_r}) \Big\| \ge x \bigg) \le c_1 \exp\big(-c_2(x/c)^{2/r}\big) \tag{3.7}$$

for all $x > 0$ and all $n \ge r$, where the constants c_i depend only on r. Lemma 2.7 and (3.7) yield

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\ge l} \frac{n}{b_n} \Big\| \int_{S^a} \tilde f_a \, dL_n^a(t) \Big\|_\beta \ge \eta \Big) \\
&\le \sum_{k=0}^\infty \mathbb{P}\bigg(\max_{nl2^k \le j \le nl2^{k+1}} \Big\| \sum_{(i_1,\dots,i_a)\in I(a,j)} \tilde f_a(X_{i_1}, \dots, X_{i_a}) \Big\| \ge \frac{\eta\, n_{(a)}\, b_n\, \beta(l2^k)^m}{n} \bigg) \\
&\le \sum_{k=0}^\infty \sum_{c=1}^a e_c(a) \mathbb{P}\bigg(\Big\| \sum_{\substack{(i_1,\dots,i_c) \\ \in I(c,l\,2^{k+1})}} (\tilde f_a)_{ac}(X_{i_1}, \dots, X_{i_c}) \Big\| \ge \frac{f_c(a)\, \eta\, n_{(a)}\, b_n\, \beta(l2^k)^m}{n} \bigg) \\
&\le \sum_{k=0}^\infty \sum_{c=1}^a e_c(a) \mathbb{P}\bigg(\frac{1}{(n\,l\,2^{k+1})^{c/2}} \Big\| \sum_{(i_1,\dots,i_c)\in I(c,l\,2^{k+1})} (\tilde f_a)_{ac}(X_{i_1}, \dots, X_{i_c}) \Big\| \\
&\qquad\qquad \ge \frac{\eta\, f_c(a) n_{(a)}\, b_n\, \beta(l2^k)^m}{n\,(n\,l\,2^{k+1})^{c/2}} \bigg) \\
&\le \sum_{k=0}^\infty \sum_{c=1}^a \tilde e_c(a) \exp\bigg(-\eta\, \tilde f_c(a) \frac{(b_n\, n_{(a)})^{2/c}}{n^{1+2/c}} \log(\log l + k \log 2) \bigg).
\end{aligned} \tag{3.8}$$

Therefore

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\ge l} \frac{n}{b_n} \Big\| \int_{S^a} \tilde f_a \, dL_n^a(t) \Big\|_\beta \ge \eta \Big) \\
&\sum_{k=0}^\infty a\, c_1(a) \exp\bigg(-\eta\, c_2(a)\, \frac{(b_n\, n_{(a)})^{2/a}}{n^3} \log(\log l + k\, \log 2) \bigg).
\end{aligned}$$

Without loss of generality $\eta\, c_2(a) = 1$. We obtain

$$\frac{n}{b_n^2} \log \mathbb{P}\Big(\sup_{t \geq l} \frac{n}{b_n} \Big\| \int_{S^a} \tilde{f}_a \, dL_n^a(t) \Big\|_\beta \geq \eta \Big)$$
$$\leq \frac{n}{b_n^2} \log a\, c_1(a) - a \, \frac{n}{b_n^2} \, \frac{(b_n \, n_{(a)})^{2/a}}{n^3} \, \log \, \log l + \frac{n}{b_n^2} \,\text{const}\,.$$

For $a = 1$, considering $n \to \infty$ and then $l \to \infty$ yields the result. Since for $a \in \{2, \dots, m\}$ the right hand side converges to $-\infty$ for $n \to \infty$, independent of l, we actually have proved that all terms of order $a \geq 2$ do not contribute to the MDP in $D_\beta[\mathbb{R}^d]$ and therefore the result in $D_\beta[\mathbb{R}^d]$ is proved.

We equip $D_2[B'(S^m, \mathbb{R})]$, the space of all càdlàg functions from $\mathbb{R}_+$ to $B'(S^m, \mathbb{R})$ with the weakest topology such that the maps $y(\cdot) \mapsto \langle \phi, y(\cdot) \rangle : \; D_2[B'(S^m, \mathbb{R})] \mapsto D_2[\mathbb{R}]$ are continuous and with the minimal σ-field $\mathcal{B}_2$ for which these maps are measurable, where $\phi \in B(S^m, \mathbb{R})$. Moreover choose $\mathcal{B}_2$ such that $D_2[\mathcal{M}(S^m)] \in \mathcal{B}_2$. We apply the projective limit approach for càdlàg function spaces, introduced in [10, Theorem A.1]: we obtain the MDP for $\{L_n^m(\cdot), n \in \mathbb{N}\}$ in the space $D_2[B'(S^m, \mathbb{R})]$. The corresponding good rate function has the form

$$\widetilde{J}_\infty^m(\nu(\cdot)) = \sup_{m \in \mathbb{N}, 0 < t_1 < \cdots < t_m < \infty} \sum_{i=1}^m (t_i - t_{i-1}) J_m \left(\frac{\nu(t_i) - \nu(t_{i-1})}{t_i - t_{i-1}} \middle| \mu \right). \tag{3.9}$$

Since $\{L_n^m(1), n \in \mathbb{N}\}$ satisfies the MDP in $B'(S^m, \mathbb{R}^d)$ with the convex rate function $J_m(\cdot|\mu)$ (cf. [16, Theorem 1.21]), this representation follows from Lemma A.3 in [10]. Moreover since both $\{M_n^m(\cdot), n \in \mathbb{N}\}$ and $\{\nu(\cdot) : \; \widetilde{J}_\infty^m(\nu(\cdot)) < \infty\}$ are subsets of $D_2[\mathcal{M}(S^m)]$ the MDP also holds in this space by Lemma 4.1.5 in [11]. Next we define

$$K_m(\nu|\mu^{\otimes m}) = \frac{1}{2} \int_{S^m} \left(\frac{d\nu}{d\mu^{\otimes m}} \right)^2 d\mu^{\otimes m},$$

if $\nu \ll \mu^{\otimes m}$ and $\nu(S^m) = 0$ and defining $K_m(\nu|\mu^{\otimes m}) = \infty$ otherwise. Let $\nu \in \mathcal{M}(S^m)$ satisfies $J_m(\nu|\mu) < \infty$. Then the representation $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \widetilde{\nu} \otimes \mu^{\otimes m-i}$ implies that

$$\frac{d\,\nu}{d\,\mu^{\otimes m}}(s_1, \dots, s_m) = \sum_{i=1}^m \frac{d\,\widetilde{\nu}}{d\,\mu}(s_i) \quad \mu^{\otimes m} - a.s.$$

Using the definition of $K_m(\cdot|\mu^{\otimes m})$, $\widetilde{\nu}(S) = 0$ and the definition of $J_m(\cdot|\mu)$, it follows that

$$K_m(\nu|\mu^{\otimes m}) = m \, J_m(\nu|\mu).$$

Since $K_m(\nu|\mu^{\otimes m}) \geq \|\nu\|_{\text{var}}/2$ we obtain $J_m(\nu|\mu) \geq \|\nu\|_{\text{var}}/2m$. Using again the convexity of $J_m(\cdot|\mu)$ we can apply Lemma A.6 in [10] to get $\widetilde{J}_\infty^m(\nu(\cdot)) = J_\infty^m(\nu(\cdot))$ for all relevant paths. Remark that the result in $D_2^{\text{proj}}[\mathcal{M}(S^m)]$ as well as in $D_1[\mathcal{M}(S^m)]$ follow by the same arguments. $\square$

Remark 3.10. In $D_\beta[\mathbb{R}^d]$ and in $D_2[\mathbb{R}^d]$, in the case $m \geq 2$, the result does not follow if the norm would be $\sup_{t\in\mathbb{R}_+} \|f(t)\|/(\beta(t))$. We have to replace β by β^m, otherwise in (3.8) the series in k does not converge. Moreover notice, that the results of Lemma 2.1 and of [4, Proposition 2.3] show that the tail behaviour of the normalized U-statistics for completely μ-degenerate kernel functions are of the order of the tail of the limiting processes in the regime of interest. This opitmality property may be a hint, why we have to restrict ourselves to a less general situation with respect to the topology. The calculations in (3.8) show that

$$\lim_{l\to\infty} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\sup_{t\geq l} \big(\frac{n}{b_n}\big)^a \Big\| \int_{S^a} \tilde{f}_a \, dL_n^a(t) \Big\| / (\beta(t))^a \geq \eta \Big) = -\infty$$

for all $1 \leq a \leq m$. This follows immediately. The scaling $\big(\frac{n}{b_n\,\beta(t)}\big)^a$ is the right one for the degenerate case. Our calculations show that for every rank a we have to choose the rank-independent norm $\|\cdot\|_\beta$, introduced above.

Remark 3.11. What is the leading term in the MDP of the sequence $\{L_n^m(\cdot), n \in \mathbb{N}\}$? Consider

$$L_n^{m,1} := \sum_{i=1}^m \mu^{\otimes i-1} \otimes L_n \otimes \mu^{\otimes m-i}, \tag{3.12}$$

which is the first term in the "Hoeffding decompositon" of L_n^m (see [16] for details). We check that $\{L_n^{m,1}(\cdot), n \in \mathbb{N}\}$ and $\{L_n^m(\cdot), n \in \mathbb{N}\}$ have identical moderate behaviour on $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ (and on $D_1[\mathcal{M}(S^m)]$ as well). Therefore we will use the concept of exponential equivalence in completely regular topological spaces introduced in [17]. The weak topology on $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ is generated by the family of pseudometrics

$$d_{2,\varphi,T}(y,z) := \sup_{t\in[0,T]} \Big\| \int_{S^m} f \, dy(t) - \int_{S^m} f \, dz(t) \Big\| / (t+1), \qquad y, z \in D_2^{\mathrm{proj}}[\mathcal{M}(S^m)],$$

with $f \in B(S^m, \mathbb{R}^d)$ and $T > 0$. This family is separating. The family of separating pseudometrics which generates the weak topology on $D_1[\mathcal{M}(S^m)]$ is given by

$$d_{1,\varphi,T}(y,z) := \sup_{t\in[0,T]} \Big\| \int_{S^m} f \, dy(t) - \int_{S^m} f \, dz(t) \Big\|, \qquad y, z \in D_1[\mathcal{M}(S^m)],$$

with the same f and T. Note that

$$\int_{S^m} f \, dL_n^{m,1}(t) - \int_{S^m} f \, dL_n^m(t) = \sum_{a=2}^m \int_{S^a} \tilde{f}_a \, dL_n^a(t)$$

(for details see [16]). With [17, Theorem 1.5] the exponential equivalence is proved, provided we show that for every $f \in B(S^m, \mathbb{R}^d)$ and for every $T > 0$ and for every $\eta > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\frac{n}{b_n} \sup_{t\in[0,T]} \Big\| \int_{S^a} \tilde{f}_a \, dL_n^a(t) \Big\| / (t+1) \geq \eta \Big) = -\infty$$

for every $a \in \{2, \dots, m\}$. But this follows from the calculations in the proof of Theorem 1.18.

Remark 3.13. To generalize the MDP $\{\langle f, L_n^m(\cdot)\rangle, n \in \mathbb{N}\}$ to the topology induced by the norm $\|f\|_\beta := \sup_{t\in\mathbb{R}_+} \|f(t)\|/(\beta(t)^m)$ we can alternatively apply the inverse contraction principle [11, Theorem 4.2.6, Corollary 4.2.7], by which it suffices to prove exponential tightness in the space $C_\beta[\mathbb{R}^d]$, the closed subset of all continuous elements in $D_\beta[\mathbb{R}^d]$. Therefore we consider the polygonal approximation of $L_n^m(\cdot)$. Notice that $C_\beta[\mathbb{R}^d]$ can be identified with the Polish space of continuous functions on the extended real line, equipped with the supremum norm. In particular, $C_\beta[\mathbb{R}^d]$ is a Polish space. The exponential tightness follows now with [24, Theorem 3, Lemma 4, Lemma 5] together with the calculations in (3.8).

Proof of Remark 1.19. We only give a sketch of the proof of (1.20). A similar proof was already given in [15, Corollary 2.7]. We have to check

$$\inf\Big\{\int_0^1 J_m(\dot\nu|\mu)\,dt, \nu \in \mathcal{AC}_0[\mathcal{M}(S^m)] \cap K_\infty \text{ and } \int_{S^m} h\,d\nu(\cdot) = \phi(\cdot)\Big\} = I_\infty(\phi). \tag{3.14}$$

Assume that the right hand side is finite. Using $\widetilde{J}_\infty^m(\nu(\cdot)) = J_\infty^m(\nu(\cdot))$ and the identity (3.9) we obtain that the left hand side is greater than or equal to

$$\inf\Big\{\sum_{i=1}^k (t_i - t_{i-1}) J_m\Big(\frac{\nu(t_i) - \nu(t_{i-1})}{t_i - t_{i-1}}\Big|\mu\Big); \int_{S^m} h\,d\nu(t_i) = \phi(t_i), 1 \le i \le k\Big\}$$

for arbitrary $0 = t_0 < t_1 < \cdots < t_k \le 1$. Consider the case $k = 1$ and $t_1 = 1$, than we obtain

$$\inf\Big\{J_m(\nu|\mu); \int_{S^m} h\,d\nu(1) = \phi(1), \quad \nu(1) \in K_{1,\infty}\Big\} = \Lambda^*(\phi(1)),$$

where $K_{1,\infty} = \cup_{L\ge 0}\{\varrho \in \mathcal{M}(S^m) : J_m(\varrho|\mu) \le L\}$ and $\Lambda^*(\cdot)$ is the convex dual of $\Lambda(\cdot)$ defined in the statement of Theorem 1.8. This time-fixed variational identity follows from [16, Lemma 3.19]. A simple argument, using the linearity of $\nu \mapsto \int_{S^m} h\,d\nu$ and the convexity of $J_m(\cdot|\mu)$ yields

$$\sum_{i=1}^k (t_i - t_{i-1}) J_m\Big(\frac{\nu(t_i) - \nu(t_{i-1})}{t_i - t_{i-1}}\Big|\mu\Big) = \sum_{i=1}^k (t_i - t_{i-1}) \Lambda^*\Big(\frac{\phi(t_i) - \phi(t_{i-1})}{t_i - t_{i-1}}\Big).$$

Since this is true for every $(t_0, \dots, t_k)$ we get by Lemma 5.1.6 in [11] that the right hand side in (3.14) is smaller than the left hand side. In the case $\int_0^1 \Lambda^*(\dot\phi)\,dt = \infty$ it follows easily that $\{\nu(\cdot) : \nu(\cdot) \in K_\infty \quad \text{and} \quad \int h\,d\nu = \phi(\cdot)\} = \emptyset$.

For fixed $t \in [0,1]$ the infimum of $J_m(\nu(t)|\mu)$ with $\nu(t) \in K_{1,\infty}$ can be calculated for all $\nu(t) = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1(t) \otimes \mu^{\otimes m-i}$. The symmetry of h yields

$$m \int_{S_m} h\,d\nu(t) = m \int_S \mu^{\otimes m-1} h\,d\nu_1(t) = m \int_S \tilde h_1\,d\nu_1(t).$$

Thus the other inequality, which is more involed, follows step by step as in the proof of [15, Corollary 2.7]. $\square$

Proof of Theorem 1.24. First we will check that in our scaling the non-linear terms of the Hoeffding decomposition do not contribute to the MDP. Denote by $R_n^m(t,h) = \frac{n}{b_n}\sum_{a=2}^m \binom{m}{k} U_n^a(t,\pi_{a,m}h)$. It suffices to prove that every $\eta > 0$

$$\lim_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\big(\|E_{R_n^m}\|_{\mathcal{H},\beta} > \eta\big) = -\infty. \tag{3.15}$$

Assume that $\mathcal{H}$ satisfies Condition 1.21. Note that for any probability measure ν on S^a the quantity $||\pi_{a,m}h||_{L_2(\nu)}$ is dominated by the sum of 2^a L_2-distances. Thus the condition $N_2(\varepsilon,\mathcal{H},\mu) \leq (A/\varepsilon)^v$ (for all probability measures μ) implies that the classes of canonical functions $\{\pi_{a,m}h : h \in \mathcal{H}\}$ satisfy the second part of the conditions of Proposition 2.21. With (2.22) and Corollary 2.13 we get for each $2 \leq a \leq m$ and every $l \geq e$

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\geq l} \frac{n}{b_n}\big\|U_n^a(t,\pi_{a,m}h)\big\|_{\mathcal{H}}/(\beta(t)^m) \geq \eta\right)\\
&\qquad \leq a\, c_1(a) \sum_{k=0}^{\infty} \exp\left(-\eta\, c_2(a) \frac{(b_n\, n_{(a)})^{2/a}}{n^3} \log(\log l + k \log 2\,)\right).
\end{aligned}$$

Since the calculations are easy for a time intervall $[0,l]$, we are done via the principle of the largest term [11, Lemma 1.2.15]. If $\mathcal{H}$ satisfies Condition 1.22, we apply Lemma 2.23 and Corollary 2.13 to see (3.15) via exactly the same calculations.
If $\mathcal{H}$ satisfies Condition 1.21 we can use the proof of Corollary 5.7 and the following remark in [2], using the uniform boundedness of $\mathcal{H}$, to get the following remarkable fact: all projections $\pi_{a,m}\mathcal{H} = \{\pi_{a,m}h : h \in \mathcal{H}\}$, $a = 1,\dots,m$, satisfy the Central Limit Theorem in $l_\infty(\mathcal{H})$. Especially this implies using [20, Theorem 14.6] that the class $\{\pi_{1,m}h : h \in \mathcal{H}\}$ satisfies the sufficient (and necessary) conditions of Theorem 2 in [27]: the class is totally bounded and (1.25) holds. Notice that $\{\pi_{1,m}h : h \in \mathcal{H}\}$ is uniformly bounded. For fixed $k \in \mathbb{N}$ let $\mathcal{F}_k \in \mathcal{H}$ be finite $1/k$-nets of $\mathcal{H}$, i.e. $\sup_{h\in\mathcal{H}} \min_{f\in\mathcal{F}_k} d_2(h,f) \leq 1/k$. Define $f_k(h)$ via $d_2(h,f_k(h)) = \min_{f\in\mathcal{F}_k} d_2(h,f)$ with tie-breaking such that $h \mapsto f_k(h)$ is Borel measurable in $(\mathcal{H},d_2)$. Moreover denote $E_{k,\nu}(\cdot) = \langle f_k(\cdot),\nu\rangle \in l_\infty(\mathcal{H})$. Remember the notation $L_n^{m,1}$ defined in (3.12). Adopting the arguments of [10, Theorem 2(b)], which uses [26, Theorem 3.5] we obtain

$$\lim_{k\to\infty} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\big\|E_{\frac{n}{b_n}L_n^{m,1}} - E_{k,\frac{n}{b_n}L_n^{m,1}}\big\|_{\mathcal{H},\beta} > \eta\right) = -\infty.$$

For the last step of the proof we adapt the proof of [10, Theorem 2(b)] using some properties of the rate function $J_m(\cdot|\mu)$. For the sake of completeness we give the arguments. Let $\mathcal{F}_k = \{f_1,\dots,f_d\}$ for some $d = d_k$ and denote $f = (f_1,\dots,f_d)$. By Theorem 1.18, applying the contraction principle, we know that

$\langle f, M_n^m(\cdot)\rangle$ satisfies the MDP in $D_2[\mathbb{R}^d]$ with the good rate function $I_f(\phi(\cdot)) = \inf_{\{\nu(\cdot)\,:\;\phi(\cdot)=\langle f,\nu(\cdot)\rangle\}} J_\infty^m(\nu(\cdot))$. Denote by

$$K_L := \{\nu(\cdot)\,:\; J_\infty^m(\nu(\cdot)) \le L\} \subset D_2[\mathcal{M}(S^m)].$$

With (3.9) it follows that $t\, J_m\big(\frac{\nu(t)}{t}\big|\mu\big) \le J_\infty^m(\nu(\cdot))$ for every $t > 0$. For each $\nu(\cdot) \in K_L$ and every $t > 0$ there exists a density $d\nu(t)/d\,\mu^{\otimes m}$ such that

$$\frac{1}{2\,t}\int_{S^m}\left(\frac{d\nu(t)}{d\mu^{\otimes m}}\right)^2 d\mu^{\otimes m} = tK_m\left(\frac{\nu(t)}{t}\middle|\mu^{\otimes m}\right) = t\,m\,J_m\left(\frac{\nu(t)}{t}\middle|\mu\right) \le L\ m.$$

Thus $\phi(t) \le \sqrt{2t\,L\,m}$ for $I_f(\phi(\cdot)) < \infty$ and is follows that $\{\phi(\cdot)\,:\; I_f(\phi(\cdot)) < \infty\} \subset D_\beta[\mathbb{R}^d]$. We have seen in the proof of Theorem 1.18 that $\langle f, M_n^m(\cdot)\rangle$ satisfies the MDP in $D_\beta[\mathbb{R}^d]$. The uniqueness of the rate function (cf. [11, Lemma 4.1.4]) and [11, Lemma 4.1.5] imply that the rate function is $I_f(\cdot)$. Identifying $E_{k,\nu(\cdot)}$ with $\langle f, \nu(\cdot)\rangle$ the MDP in $D_\beta[l_\infty(\mathcal{H})]$ with the good rate

$$J_{\mathcal{H},k}^m(H(\cdot)) = \inf\left\{J_\infty^m(\nu(\cdot))\,:\; \nu(\cdot) \in \mathcal{AC}_0[\mathcal{M}(S^m)] \text{ and } E_{k,\nu(\cdot)} = H(\cdot)\right\}$$

follows via the contraction principle [11, Theorem 4.2.23]: the mapping $\nu(\cdot) \mapsto E_{k,\nu(\cdot)}\,:\; K_L \to D_\beta[l_\infty(\mathcal{H})]$ is continuous. Define
$\mathcal{H}_\eta := \{h - g\,:\; h, g \in \mathcal{H} \text{ and } d_2(h,g) \le \eta\}$. For $\nu(\cdot) \in K_L$ it follows that

$$\|\nu(t)\|_{\mathcal{H}_{1/k}} \le \sup_{\{h\,:\;\int_{S^m} h^2\, d\mu^{\otimes m} \le 1/k^2, \|h\|\le 1\}} \left\|\int_{S^m} h\frac{d\nu(t)}{d\mu^{\otimes m}}\, d\mu^{\otimes m}\right\| \le c(m)\frac{\sqrt{2L\,t\,m}}{k},$$

where $c(m)$ is a constant, depending only on m. Using the fact that $h - f_k(h) \in \mathcal{H}_{1/k}$ for every $h \in \mathcal{H}$ implies that $\|E_\nu - E_{k,\nu}\|_{\mathcal{H}} \le \|\nu\|_{\mathcal{H}_{1/k}}$ for every $\nu \in \mathcal{M}(S^m)$, we obtain that $\|E_\nu - E_{k,\nu}\|_{\mathcal{H},\beta} \to 0$ as $k \to \infty$, uniformly over K_L. □

Acknowledgement. I would like to thank Prof. E. Giné for discussions concerning Lemma 2.7.

References

1. M. A. Arcones, *A Bernstein-type inequality for U-statistics and U-processes*, Statist. Probab. Lett. **22** (1995), 239–247.
2. M. A. Arcones and E. Giné, *Limit theorems for U-processes*, Ann. Probab. **21** (1993), 1494–1542.
3. ———, *U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters*, Stochastic Process. Appl. **52** (1994), 17–38.
4. ———, *On the law of the iterated logarithm for canonical U-statistics and processes*, Stochastic Process. Appl. **58** (1995), 217–245.
5. P. Baldi, *Large deviations and stochastic homogenization*, Ann. Mat. Pura Appl. **151** (1988), 161–177.
6. A. A. Borovkov and A. A. Mogulskii, *Probabilities of large deviations in topological spaces I*, Siberian Math. J. **19** (1978), 697–709.
7. ———, *Probabilities of large deviations in topological spaces II*, Siberian Math. J. **21** (1980), 653–664.

8. Yu. V. Borovskikh, *U-Statistics in Banach Spaces*, VSP, Utrecht, 1996.
9. A. Dembo and T. Zajic, *Large deviations: From emprical mean and measure to partial sums process*, Stochastic Process. Appl. **57** (1995), 191–224.
10. ———, *Uniform large and moderate deviations for functional empirical processes*, Stochastic Process. Appl. **67** (1997), 195–211.
11. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, MA, 1993.
12. R. M. Dudley, *A course on empirical processes*, École d'Eté de Probabilités de Saint-Flour XII. Lecture Notes in Math. 1097 (A. Dold and B. Eckmann, eds.), Springer, New York, 1984, pp. 1–142.
13. P. Eichelsbacher, *Large deviations for products of empirical probability measures in the τ-topology*, J. Theoret. Probab. (1997), (to appear).
14. ———, *Moderate and large deviations for U-processes*, Stochastic Process. Appl. (1997), (to appear).
15. P. Eichelsbacher and M. Löwe, *Large deviations for partial sums U-processes*, Theory of Probab. and Its Appl. (1997), (to appear).
16. P. Eichelsbacher and U. Schmock, *Large and moderate deviations of products of empirical measures and U-empirical measures in strong topologies*, SFB 343 Diskrete Strukturen in der Mathematik, University of Bielefeld, preprint 97-082, 1997.
17. ———, *Exponential approximations in completely regular topological spaces and extensions of Sanov's theorem*, submitted, 1996.
18. E. Giné and V. de la Peña, *Decoupling: from dependence to independence*, Springer-Verlag, 1997.
19. E. Giné and C.-H. Zhang, *On integrability in the LIL for degenerate U-statistics*, J. Theoret. Probab. **9** (1996), no. 2, 385–412.
20. M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.
21. A. A. Mogulskii, *Large deviations for trajectories of multidimensional random walks*, Theory Probab. Appl. **21** (1976), 300–315.
22. S. J. Montgomery-Smith, *Comparison of sums of independent identically distributed random vectors*, Probab. Math. Statist. **14** (1994), no. 2, 281–285.
23. S. J. Montgomery-Smith and V. de la Peña, *Decoupling inequalities for the tail probabilities of multivariate U-statistics*, Ann. Probab. **23** (1995), no. 2, 806–816.
24. N. O'Connell, *Stronger topologies for sample path large deviations in euclidean space*, preprint, 1996.
25. D. Pollard, *Convergence of Stochastic Processes*, Spinger, New York, 1984.
26. M. Talagrand, *Sharper bounds for gaussian and empirical processes*, Ann. Probab. **22** (1994), no. 1, 28–76.
27. L. Wu, *Large deviations, moderate deviations and LIL for empirical processes*, Ann. Probab. **22** (1994), no. 1, 17–27.
28. V. V. Yurinsky, *Exponential inequalities for sums of random vectors*, J. Mult. Anal. **6** (1976), no. 4, 473–499.
29. ———, *Sums and Gaussian vectors*, Lecture Notes in Mathematics, vol. 1617, Springer-Verlag, Berlin, 1995.

(Peter Eichelsbacher) Fakultät für Mathematik, Universität Bielefeld, Universitätsstrasse 1, D-33501 Bielefeld, Germany

E-mail address, Peter Eichelsbacher: `peter@mathematik.uni-bielefeld.de`