


A note on concentration inequality for vector-valued martingales with weak exponential-type tails

Chris Junchi Li* 

Abstract

We present novel martingale concentration inequalities for martingale differences with finite Orlicz- ψ_α norms. Such martingale differences with weak exponential-type tails scatters in many statistical applications and can be heavier than sub-exponential distributions. In the case of one dimension, we prove in general that for a sequence of scalar-valued supermartingale difference, the tail bound depends solely on the sum of squared Orlicz- ψ_α norms instead of the maximal Orlicz- ψ_α norm, generalizing the results of [Lesigne & Volný, 2001] and [Fan et al., 2012]. In the multidimensional case, using a dimension reduction lemma proposed by [Kallenberg & Sztencel, 1991] we show that essentially the same concentration tail bound holds for vector-valued martingale difference sequences.

Keywords: Concentration inequalities ; Vector-valued martingales ; Weak exponential-type tails ; Dimension reduction lemma.

MSC2020 subject classifications: 62H05.

Submitted to ECP on , final version accepted on .

1 Introduction

This note concerns the following problem: let $\mathbf{u}_1, \dots, \mathbf{u}_N \in \mathbb{R}^d$ be a vector-martingale difference sequence that take place on the d -dimensional Euclidean space \mathbb{R}^d , where $\mathbb{E}[\mathbf{u}_i \mid \mathcal{F}_{i-1}] = \mathbf{0}$. Assume that \mathbf{u}_i has the following weak exponential-type tail condition: for some $\alpha > 0$ and all $i = 1, \dots, N$ we have for some scalar $K_i > 0$

$$\mathbb{E} \exp \left(\left\| \frac{\mathbf{u}_i}{K_i} \right\|^\alpha \right) \leq 2, \quad (1.1)$$

and hence by Markov's inequality their tails satisfy for each $n = 1, \dots, N$

$$\mathbb{P} \left(\left\| \frac{\mathbf{u}_i}{K_i} \right\| \geq z \right) \leq \exp(-z^\alpha) \mathbb{E} \exp \left(\left\| \frac{\mathbf{u}_i}{K_i} \right\|^\alpha \right) \leq 2 \exp(-z^\alpha),$$

then what can we conclude about the tail probability of the random variable $\left\| \sum_{n=1}^N \mathbf{u}_i \right\|$? Note for $\alpha < 1$ under the condition (1.1), the moment generating functions $\mathbb{E} \exp(t \|\mathbf{u}_i\|)$ are in general not available, and hence the classical analysis using moment generating functions do not work through and hence new analytical tools are in demand.

*University of California, Berkeley, United States of America. E-mail: junchili@cs.berkeley.edu

Our result makes several contributions upon the previous works. First, we conclude that in the one-dimensional case where one denotes $\mathbf{u}_i = u_i$, a one-sided maximal inequality can be concluded that, roughly,

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u_i \geq z \right) \leq \mathcal{L}_\alpha \left(\frac{\sum_{i=1}^N K_i^2}{z^2} \right) \exp \left\{ - \left(C \cdot \frac{z^2}{\sum_{i=1}^N K_i^2} \right)^{\frac{\alpha}{\alpha+2}} \right\} \quad (1.2)$$

where the factor $\mathcal{L}_\alpha(y)$ is solely dependent on y for any fixed $\alpha > 0$ and grows linearly in y , and $C < 100$ is a positive numerical constant. In above and the following, we allow the numerical constant C to change from paragraph to paragraph. This generalizes the bound of [Lesigne & Voln, 2001] and [Fan et al., 2012], where both groups of authors only consider the case $K_1 = \dots = K_N$ in the independent and martingale difference sequence cases, separately. See also the more recent paper [Fan et al., 2017] for similar concentration under a slightly weaker condition. In fact, we also know that the inequality (1.2) is optimal in the sense that it *cannot* be further improved for a class of martingale difference sequences that satisfy the exponential moment condition (1.1).

Secondly for the general dimension case, applying (1.2) as well as a dimension-reduction argument for vector martingales [Kallenberg & Sztencel, 1991, Hayes, 2005, Lee et al., 2016] allows us to conclude a one-sided bound on its Euclidean norm: under (1.1) we have

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \left\| \sum_{i=1}^n \mathbf{u}_i \right\| \geq z \right) \leq \mathcal{L}'_\alpha \left(\frac{\sum_{i=1}^N K_i^2}{z^2} \right) \exp \left\{ - \left(C \cdot \frac{z^2}{\sum_{i=1}^N K_i^2} \right)^{\frac{\alpha}{\alpha+2}} \right\} \quad (1.3)$$

where analogously, the factor $\mathcal{L}'_\alpha(y)$ is solely dependent on y for any fixed $\alpha > 0$ and grows linearly in y , and $C < 100$ is a positive numerical constant. To our best knowledge, this provides a first concentration result for vector-valued martingales with unbounded martingale differences under the weak exponential-type condition (1.1).

Concentration results of (1.2) and (1.3) potentially see many applications in probability and statistics, including the rate of convergence of martingales, the consistency of nonparametric regression estimation with errors of martingale difference sequence (see [Laib, 1999]), as well as online stochastic gradient algorithms for parameters estimation in linear models and PCA [Li et al., 2018].

2 Orlicz space and Orlicz norm

In this subsection, we briefly revisit the properties of Orlicz space and its ψ -norm that are mostly relevant. Readers who are interested in an exposure of Orlicz space from a Banach space point of view are referred to [Ledoux & Talagrand, 2013].

Let \mathbb{R}_+ be the set of nonnegative real numbers. Consider the Orlicz space of \mathbb{R}^d -valued random vector \mathbf{X} which lives in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}\psi(\|\mathbf{X}\|/K) < \infty$ some $K > 0$. Let $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a nondecreasing convex function with $\psi(0) = 0$ and $\lim_{x \rightarrow \infty} \psi(x) = \infty$, and equip the Orlicz space with the norm

$$\|\mathbf{X}\|_\psi := \inf \left\{ K > 0 : \mathbb{E}\psi \left(\frac{\|\mathbf{X}\|}{K} \right) \leq 1 \right\}.$$

One calls $\|\cdot\|_\psi$ the Orlicz- ψ norm. In special, random vector \mathbf{X} has an Orlicz- ψ norm defined as Orlicz- ψ norm of $\|\mathbf{X}\|$ as a scalar-valued random variable.

In this note, we are interested in the exponential-tailed distributions that corresponds to a family of ψ functions: $\psi_\alpha(x) = \exp(x^\alpha) - 1$, $\alpha \in (0, \infty)$, in which case the corre-

sponding Orlicz space is the collection of random variables with exponential moments $\mathbb{E} \exp \{ \|X/K\|^\alpha \} \leq 2$.¹

3 One dimensional result

We state our first main result that concludes the right-tailed bound (1.2) under a slightly more general condition that u_1, \dots, u_N forms a supermartingale difference sequence.

Theorem 3.1. *Let $\alpha \in (0, \infty)$ be given. Assume that $(u_i : i \geq 1)$ is a sequence of supermartingale differences with respect to \mathcal{F}_i , i.e. $\mathbb{E}[u_i | \mathcal{F}_{i-1}] \leq 0$, and it satisfies $\|u_i\|_{\psi_\alpha} < \infty$ for each $i = 1, \dots, N$. Then for an arbitrary $N \geq 1$ and $z > 0$,*

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u_i \geq z \right) \leq \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{64 \sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2}{z^2} \right] \exp \left\{ - \left(\frac{z^2}{32 \sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2} \right)^{\frac{\alpha}{\alpha+2}} \right\} \quad (3.1)$$

Remark We make several remarks on Theorem 3.1, as follows.

- (i) By replacing $\sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2$ by a larger value $N \max_{1 \leq i \leq N} \|u_i\|_{\psi_\alpha}^2$ in (3.1) of Theorem 3.1, one may rediscover essentially Theorem 2.1 in [Fan et al., 2012] which includes bound (1.1) of [Lesigne & Voln y, 2001] as a special case $\alpha = 1$.² In summary, Theorem 2.1 of [Fan et al., 2012] would provide a bound that depends on the maximum of $N\|u_i\|_{\psi_\alpha}$, while our new bound sharpens such bound of [Fan et al., 2012] and depends only on the Orlicz- ψ_α norm of the martingale differences $\|u_i\|_{\psi_\alpha}$ in terms of their squared sum. It turns out that the sharpened bound is more desirable to obtain useful upper bounds in many statistical applications.
- (ii) Theorem 2.1 in [Fan et al., 2012] is *optimal* in the sense that a counterexample that has the right hand of (3.1) as the lower bound (up to a constant factor in the exponent), and forbids the existence of a sharper bound for the martingale difference sequence class. Since our result generalizes their Theorem 2.1, one may apply the same counterexample and conclude the optimality of our bound. See more in the next paragraph.

Optimality of our result To claim optimality we note that (3.1) implies, for the special case $z = N$ and each $\|u_i\|_{\psi_\alpha} \leq 1$,

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u_i \geq N \right) \leq \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{64}{N} \right] \exp \left\{ - \left(\frac{N}{32} \right)^{\frac{\alpha}{\alpha+2}} \right\}, \quad (3.2)$$

which is $\mathcal{O} \left(\exp \left\{ -CN^{\frac{\alpha}{\alpha+2}} \right\} \right)$ as $N \rightarrow \infty$ for some $C \leq 1/32$. In the mean time, [Fan et al., 2012] generalizes the counterexample in [Lesigne & Voln y, 2001] where, in our terminology of ψ_α -norm, Theorem 2.1 of [Fan et al., 2012] provides for each $\alpha \in (0, \infty)$ an ergodic

¹Rigorously speaking, when $\alpha \in (0, 1)$ $\psi_\alpha(x)$ is *not* convex when x is in a neighborhood of 0. In this case, one can let the ψ function be

$$\psi(x) = \begin{cases} \exp(x^\alpha) - 1 & x \geq x_\alpha \\ \text{linear} & x \in [0, x_\alpha) \end{cases}$$

for some $x_\alpha > 0$ large enough, so that the function satisfies the condition. We choose not to adopt this definition of ψ_α simply for clarity of presentation.

²The work [Fan et al., 2012] assumes a slightly more general condition $\mathbb{E} \exp \{ |u_i/K|^\alpha \} \leq C_1$. Nevertheless, our result does not lose any generality in general, since C_1 (when greater than or equal to 2) can be absorbed into the Orlicz- ψ_α norm as a polylogarithmic factor.

sequence of martingale differences u_1^*, \dots, u_N^* and a sequence of positives x_1, \dots, x_N such that for all N sufficiently large,

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u_i^* \geq N \right) \geq \exp \left\{ -3N^{\frac{\alpha}{\alpha+2}} \right\}$$

Comparing the last equation with (3.2), we conclude the optimality of our result.

Comparison with conditional weak exponential-type conditions If we pose the additional assumption that u_i 's satisfy (1.1) in the conditional sense, the martingale concentration inequality can be further improved. Taking the example where $d = 1$ and $\alpha = 2$, if one poses a slightly stronger condition

$$\mathbb{E} \exp \left(\left| \frac{u_i}{K_i} \right|^2 \middle| \mathcal{F}_{i-1} \right) \leq 2, \quad (3.3)$$

i.e. the martingale differences are scalar-valued and conditionally subgaussian random variables, and one may conclude from the Hoeffding's concentration inequality [Wainwright, 2019]

$$\mathbb{P} \left(\left| \sum_{i=1}^N u_i \right| \geq z \right) \leq 2 \exp \left(-C \cdot \frac{z^2}{\sum_{i=1}^N K_i^2} \right). \quad (3.4)$$

Similar bound can be derived for sub-exponential variables. Observe that the power of the $z^2/(\sum_{i=1}^N K_i^2)$ term in the exponent of (3.4) is 1, and instead, our bound in (1.2) has an exponent of $1/3$ and is hence worse. Fortunately, to obtain an error probability $\leq \delta$ both inequalities give a cut-off point $z_\delta \sim \left(\sum_{i=1}^N K_i^2 \right)^{1/2}$ up to a different polylogarithmic factor of $1/\delta$, and these two cut-off points are equivalent if these factors are ignored.

4 Proof of Theorem 3.1

To prove our main result for the one-dimensional case, Theorem 3.1, we will use a maxima version of the classical Azuma-Hoeffding's inequality proposed by [Laib, 1999] for bounded martingale differences, and then apply an argument of [Lesigne & Volný, 2001] and [Fan et al., 2012] to truncate the tail and analyze the bounded and unbounded pieces separately.

- (i) First of all, for the sake of simplicity and with no loss of generality, throughout the following proof of Theorem 3.1 we shall pose the following extra condition

$$\sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2 = 1. \quad (4.1)$$

In other words, under the additional (4.1) condition proving (3.1) reduces to showing

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u_i \geq z \right) \leq \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{64}{z^2} \right] \exp \left\{ - \left(\frac{z^2}{32} \right)^{\frac{\alpha}{\alpha+2}} \right\}. \quad (4.2)$$

This can be made more clear from the following rescaling argument: one can put in the left of (4.2) $u_i / \left(\sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2 \right)^{1/2}$ in the place of u_i , and $z / \left(\sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2 \right)^{1/2}$ in the place of z , the left hand of (3.1) is just

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n \frac{u_i}{\left(\sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2 \right)^{1/2}} \geq \frac{z}{\left(\sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2 \right)^{1/2}} \right)$$

which, by (4.2), is upper-bounded by

$$\leq \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{64 \sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2}{z^2} \right] \exp \left\{ - \left(\frac{z^2}{32 \sum_{i=1}^N \|u_i\|_{\psi_\alpha}^2} \right)^{\frac{\alpha}{\alpha+2}} \right\},$$

proving (3.1).

- (ii) We apply a truncating argument used in [Lesigne & Volný, 2001] and later in [Fan et al., 2012]. Let $\mathcal{M} > 0$ be arbitrary, and we define

$$u'_i = u_i 1_{\{|u_i| \leq \mathcal{M}\|u_i\|_{\psi_\alpha}\}} - \mathbb{E}(u_i 1_{\{|u_i| \leq \mathcal{M}\|u_i\|_{\psi_\alpha}\}} \mid \mathcal{F}_{i-1}), \quad (4.3)$$

$$u''_i = u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} - \mathbb{E}(u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} \mid \mathcal{F}_{i-1}), \quad (4.4)$$

$$T'_n = \sum_{i=1}^n u'_i, \quad T''_n = \sum_{i=1}^n u''_i, \quad T'''_n = \sum_{i=1}^n \mathbb{E}(u_i \mid \mathcal{F}_{i-1}).$$

Since u_i is \mathcal{F}_i -measurable, u'_i and u''_i are two martingale difference sequences with respect to \mathcal{F}_i , and let T_n be defined as

$$T_n = \sum_{i=1}^n u_i \quad \text{and hence} \quad T_n = T'_n + T''_n + T'''_n. \quad (4.5)$$

Since u_i are supermartingale differences we have that T'''_n is \mathcal{F}_{n-1} -measurable with $T'''_n \leq T'''_0 = 0$, a.s., and hence for any $z > 0$,

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq n \leq N} T_n \geq 2z \right) &\leq \mathbb{P} \left(\max_{1 \leq n \leq N} T'_n + T'''_n \geq z \right) + \mathbb{P} \left(\max_{1 \leq n \leq N} T''_n \geq z \right) \\ &\leq \mathbb{P} \left(\max_{1 \leq n \leq N} T'_n \geq z \right) + \mathbb{P} \left(\max_{1 \leq n \leq N} T''_n \geq z \right) \end{aligned} \quad (4.6)$$

In the following, we analyze the tail bounds for T'_n and T''_n separately [Lesigne & Volný, 2001, Fan et al., 2012].

- (iii) To obtain the first bound, we recap Laib's inequality as follows:

Lemma 4.1. [Laib, 1999] Let $(w_i : 1 \leq i \leq N)$ be a real-valued martingale difference sequence with respect to some filtration \mathcal{F}_i , i.e. $\mathbb{E}[w_i \mid \mathcal{F}_{i-1}] = 0$, a.s., and the essential norm $\|w_i\|_\infty$ is finite. Then for an arbitrary $N \geq 1$ and $z > 0$,

$$\mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n w_i \geq z \right) \leq \exp \left\{ - \frac{z^2}{2 \sum_{i=1}^N \|w_i\|_\infty^2} \right\}. \quad (4.7)$$

(4.7) generalizes the folklore Azuma-Hoeffding's inequality, where the latter can be concluded from

$$\max_{n \leq N} \sum_{i=1}^n w_i \geq \sum_{i=1}^N w_i.$$

The proof of Lemma 4.1 is given in [Laib, 1999]. Recall our extra condition (4.1), then from the definition of u'_i in (4.3) that $|u'_i| \leq 2\mathcal{M}\|u_i\|_{\psi_\alpha}$, the desired bound follows immediately from Laib's inequality in Lemma 4.1 by setting $w_i = u'_i$:

$$\mathbb{P} \left(\max_{1 \leq n \leq N} T'_n \geq z \right) = \mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u'_i \geq z \right) \leq \exp \left\{ - \frac{z^2}{8\mathcal{M}^2} \right\} \quad (4.8)$$

To obtain the tail bound of T_n'' we only need to show

$$\mathbb{E}(u_i'')^2 \leq (6\mathcal{M}^2 + 8\mathcal{B}^2)\|u_i\|_{\psi_\alpha}^2 \exp\{-\mathcal{M}^\alpha\}, \quad (4.9)$$

where

$$\mathcal{B} := \left(\frac{3}{\alpha}\right)^{\frac{1}{\alpha}}, \quad (4.10)$$

from which, Doob's martingale inequality [Durrett, 2010, §5] implies immediately that

$$\mathbb{P}\left(\max_{1 \leq n \leq N} T_n'' \geq z\right) \leq \frac{1}{z^2} \sum_{i=1}^N \mathbb{E}(u_i'')^2 \leq \frac{6\mathcal{M}^2 + 8\mathcal{B}^2}{z^2} \exp\{-\mathcal{M}^\alpha\}. \quad (4.11)$$

To prove (4.9), first recall from the definition of u_i'' in (4.4) that

$$u_i'' = u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} - \mathbb{E}(u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} \mid \mathcal{F}_{i-1}).$$

Recall from the property of conditional expectation [Durrett, 2010] that for any random variable W and a σ -algebra $\mathcal{G} \subset \mathcal{F}$

$$\mathbb{E}[W - \mathbb{E}(W \mid \mathcal{G})]^2 = \mathbb{E}W^2 - \mathbb{E}[\mathbb{E}(W \mid \mathcal{G})]^2 \leq \mathbb{E}W^2 = \int_0^\infty 2y\mathbb{P}(|W| > y)dy$$

where the last equality is due to the second-moment formula for nonnegative random variable $|W|$ [Durrett, 2010]. Plugging in $W = u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}}$ and $\mathcal{G} = \mathcal{F}_{i-1}$ we have

$$\begin{aligned} \mathbb{E}(u_i'')^2 &= \mathbb{E}[u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} - \mathbb{E}(u_i 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} \mid \mathcal{F}_{i-1})]^2 \\ &\leq \int_0^\infty 2y\mathbb{P}(|u_i| 1_{\{|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}\}} > y)dy \\ &= \int_0^{\mathcal{M}\|u_i\|_{\psi_\alpha}} 2ydy \cdot \mathbb{P}(|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}) + \int_{\mathcal{M}\|u_i\|_{\psi_\alpha}}^\infty 2y\mathbb{P}(|u_i| > y)dy \\ &= \mathcal{M}^2\|u_i\|_{\psi_\alpha}^2 \mathbb{P}(|u_i| > \mathcal{M}\|u_i\|_{\psi_\alpha}) + \int_{\mathcal{M}}^\infty 2t\|u_i\|_{\psi_\alpha} \mathbb{P}(|u_i| > t\|u_i\|_{\psi_\alpha}) \|u_i\|_{\psi_\alpha} dt \\ &\leq 2\mathcal{M}^2\|u_i\|_{\psi_\alpha}^2 \exp\{-\mathcal{M}^\alpha\} + 4\|u_i\|_{\psi_\alpha}^2 \int_{\mathcal{M}}^\infty t \exp\{-t^\alpha\} dt, \end{aligned} \quad (4.12)$$

where the last inequality is due to Markov's inequality that for all $z > 0$

$$\mathbb{P}(|u_i|/\|u_i\|_{\psi_\alpha} \geq z) \leq \exp\{-z^\alpha\} \mathbb{E} \exp\{|u_i|^\alpha/\|u_i\|_{\psi_\alpha}^\alpha\} \leq 2 \exp\{-z^\alpha\}. \quad (4.13)$$

It can be shown from Calculus I that the function $g(t) = t^3 \exp\{-t^\alpha\}$ is decreasing in $[\mathcal{B}, +\infty)$ and is increasing in $[0, \mathcal{B}]$, where \mathcal{B} was earlier defined in (4.10) [Fan et al., 2012]. If $\mathcal{M} \in [\mathcal{B}, \infty)$ we have

$$\begin{aligned} \int_{\mathcal{M}}^\infty t \exp\{-t^\alpha\} dt &= \int_{\mathcal{M}}^\infty t^{-2} t^3 \exp\{-t^\alpha\} dt \\ &\leq \int_{\mathcal{M}}^\infty t^{-2} dt \cdot \mathcal{M}^3 \exp\{-\mathcal{M}^\alpha\} \\ &= \mathcal{M}^{-1} \cdot \mathcal{M}^3 \exp\{-\mathcal{M}^\alpha\} = \mathcal{M}^2 \exp\{-\mathcal{M}^\alpha\}. \end{aligned} \quad (4.14)$$

If $\mathcal{M} \in (0, \mathcal{B})$, we have by setting \mathcal{M} as \mathcal{B} in above

$$\begin{aligned} \int_{\mathcal{M}}^{\infty} t \exp \{-t^{\alpha}\} dt &= \int_{\mathcal{M}}^{\mathcal{B}} t \exp \{-t^{\alpha}\} dt + \int_{\mathcal{B}}^{\infty} t \exp \{-t^{\alpha}\} dt \\ &\leq \int_{\mathcal{M}}^{\mathcal{B}} dt \cdot \mathcal{B} \exp \{-\mathcal{M}^{\alpha}\} + \mathcal{B}^2 \exp \{-\mathcal{B}^{\alpha}\} \\ &\leq (\mathcal{B} - \mathcal{M}) \mathcal{B} \exp \{-\mathcal{M}^{\alpha}\} + \mathcal{B}^2 \exp \{-\mathcal{M}^{\alpha}\} \\ &\leq 2\mathcal{B}^2 \exp \{-\mathcal{M}^{\alpha}\}. \end{aligned} \quad (4.15)$$

Combining (4.12) with the two above displays (4.14) and (4.15) we obtain

$$\begin{aligned} \mathbb{E}(u_i'')^2 &\leq 2\mathcal{M}^2 \|u_i\|_{\psi_{\alpha}}^2 \exp \{-\mathcal{M}^{\alpha}\} + 4\|u_i\|_{\psi_{\alpha}}^2 \int_{\mathcal{M}}^{\infty} t \exp \{-t^{\alpha}\} dt \\ &\leq (6\mathcal{M}^2 + 8\mathcal{B}^2) \|u_i\|_{\psi_{\alpha}}^2 \exp \{-\mathcal{M}^{\alpha}\}, \end{aligned}$$

completing the proof of (4.9) and hence (4.11).

(iv) Putting the pieces together: combining (4.6), (4.8) and (4.11) we obtain for an arbitrary $u \in (0, \infty)$ that

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq n \leq N} T_n \geq 2z \right) &\leq \mathbb{P} \left(\max_{1 \leq n \leq N} T_n' \geq z \right) + \mathbb{P} \left(\max_{1 \leq n \leq N} T_n'' \geq z \right) \\ &\leq \exp \left\{ -\frac{z^2}{8\mathcal{M}^2} \right\} + \frac{6\mathcal{M}^2 + 8\mathcal{B}^2}{z^2} \exp \{-\mathcal{M}^{\alpha}\} \end{aligned} \quad (4.16)$$

We choose \mathcal{M} as, by making the exponents equal in above,

$$\mathcal{M} = \left(\frac{z^2}{8} \right)^{\frac{1}{\alpha+2}} \quad \text{such that} \quad \frac{z^2}{8\mathcal{M}^2} = \mathcal{M}^{\alpha} = \left(\frac{z^2}{8} \right)^{\frac{\alpha}{\alpha+2}}.$$

Plugging this \mathcal{M} back into (4.16) we obtain

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq n \leq N} T_n \geq 2z \right) &\leq \exp \left\{ -\left(\frac{z^2}{8} \right)^{\frac{\alpha}{\alpha+2}} \right\} + \frac{6\mathcal{M}^2 + 8\mathcal{B}^2}{z^2} \exp \left\{ -\left(\frac{z^2}{8} \right)^{\frac{\alpha}{\alpha+2}} \right\} \\ &\leq \left[1 + \left(\frac{1}{8} \right)^{\frac{2}{\alpha+2}} \frac{6}{z^{\frac{2\alpha}{\alpha+2}}} + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{8}{z^2} \right] \exp \left\{ -\left(\frac{z^2}{8} \right)^{\frac{\alpha}{\alpha+2}} \right\} \end{aligned} \quad (4.17)$$

where we plugged in the expression of \mathcal{B} in (4.10). We can further simplify the square-bracket prefactor in the last line of (4.17) which can be tightly bounded by

$$\begin{aligned} 1 + \left(\frac{1}{8} \right)^{\frac{2}{\alpha+2}} \frac{6}{z^{\frac{2\alpha}{\alpha+2}}} + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{8}{z^2} &\leq 1 + \frac{6 \cdot \frac{2}{\alpha+2}}{(8)^{\frac{2}{\alpha+2}}} + \frac{6 \cdot \frac{\alpha}{\alpha+2}}{(8)^{\frac{2}{\alpha+2}} z^2} + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{8}{z^2} \\ &\leq 3 + \left(\frac{0.75 \cdot \frac{\alpha}{\alpha+2}}{(8)^{\frac{2}{\alpha+2}}} + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \right) \frac{8}{z^2} \\ &\leq 3 + \left(0.75 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \right) \frac{8}{z^2} \\ &\leq 3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{16}{z^2}. \end{aligned}$$

where we used an implication of Jensen's inequality: for $\gamma = \alpha/(\alpha+2) \in (0, 1)$ one has $x^{\gamma} \leq 1 - \gamma + \gamma x$ for all $x \geq 0$ (where the equality holds for $x = 1$), as well as

a few elementary algebraic inequalities, including $\gamma 8^{-\gamma} < 0.177$, $(1 - \gamma)8^{-\gamma} < 1$, $(3/\alpha)^{2/\alpha} > 0.78$ for all $\alpha > 0$ and $0 < \gamma = 2/(\alpha + 2) < 1$. Thus, (4.2) is concluded by noticing the relation (4.5) and setting $z/2$ in the place of z , which hence proves Theorem 3.1 via the argument in (i) in our proof.

5 General dimensions result

In many applications we are often more interested in a concentration tail inequality for vector-valued martingales. To proceed, we need a so-called *dimension reduction lemma* for Hilbert space which is inspired from its continuum version proved in [Kallenberg & Sztencel, 1991]. We argue that it is sufficient to prove it for the case $d = 2$. Writing in terms of martingale differences, we have

Lemma 5.1 (Dimension reduction lemma for \mathbb{R}^d or Hilbert space). *Let $\mathbf{u}_i, i = 1, \dots, N$ be a \mathbb{R}^d -valued martingale difference sequence with respect to filtration \mathcal{F}_i , i.e. for each $1 \leq i \leq N$, $\mathbb{E}[\mathbf{u}_i | \mathcal{F}_{i-1}] = \mathbf{0}$. Then there exists a \mathbb{R}^2 -valued martingale difference sequence $\mathbf{u}'_i, i = 1, \dots, N$ with respect to the same filtration so that for each $n = 1, \dots, N$*

$$\left\| \sum_{i=1}^n \mathbf{u}_i \right\| = \left\| \sum_{i=1}^n \mathbf{u}'_i \right\| \quad \text{and} \quad \|\mathbf{u}_n\| = \|\mathbf{u}'_n\|. \quad (5.1)$$

For a proof of Lemma 5.1, see Lemma 2.3 of [Lee et al., 2016], which proves the lemma on a generic Hilbert space.

Theorem 5.2. *Let $\alpha \in (0, \infty)$ be given. Assume that $(\mathbf{u}_i, i = 1, \dots, N)$ is a sequence of \mathbb{R}^d -valued martingale differences with respect to \mathcal{F}_i , i.e. $\mathbb{E}[\mathbf{u}_i | \mathcal{F}_{i-1}] = \mathbf{0}$, and it satisfies $\|\mathbf{u}_i\|_{\psi_\alpha} < \infty$ for each $i = 1, \dots, N$. Then for an arbitrary $N \geq 1$ and $z > 0$,*

$$\mathbb{P} \left(\max_{n \leq N} \left\| \sum_{i=1}^n \mathbf{u}_i \right\| \geq z \right) \leq 4 \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{128 \sum_{i=1}^N \|\mathbf{u}_i\|_{\psi_\alpha}^2}{z^2} \right] \exp \left\{ - \left(\frac{z^2}{64 \sum_{i=1}^N \|\mathbf{u}_i\|_{\psi_\alpha}^2} \right)^{\frac{\alpha}{\alpha+2}} \right\}. \quad (5.2)$$

Theorem 5.2 explicitly argues that the martingale inequality hold with the *dimension-free property*: the bound on the right hand of (5.2) is independent of dimension d and only depends on the martingale differences via $\sum_{i=1}^N \|\mathbf{u}_i\|_{\psi_\alpha}^2$.

Proof of Theorem 5.2. From Lemma 5.1 we have a \mathbb{R}^2 -valued martingale difference sequence $(\mathbf{u}'_i = (u'_{i,1}, u'_{i,2})^\top, i = 1, \dots, N)$ such that for each $i = 1, \dots, N$, (5.1) holds. It is straightforward to justify $\|\mathbf{u}_i\|_{\psi_\alpha} = \|\mathbf{u}'_i\|_{\psi_\alpha}$ for each i . Therefore to prove (5.2), we only need to show

$$\mathbb{P} \left(\max_{n \leq N} \left\| \sum_{i=1}^n \mathbf{u}'_i \right\| \geq z \right) \leq 4 \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{128 \sum_{i=1}^N \|\mathbf{u}'_i\|_{\psi_\alpha}^2}{z^2} \right] \exp \left\{ - \left(\frac{z^2}{64 \sum_{i=1}^N \|\mathbf{u}'_i\|_{\psi_\alpha}^2} \right)^{\frac{\alpha}{\alpha+2}} \right\}. \quad (5.3)$$

Note by definition, for $\ell = 1, 2$ $\|u'_{i,\ell}\|_{\psi_\alpha} \leq \|\mathbf{u}_i\|_{\psi_\alpha}$. Applying Theorem 3.1 to both $(u'_{i,\ell})$ and $(-u'_{i,\ell})$ as supermartingale difference sequences, we have for $\ell = 1, 2$

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq n \leq N} \left| \sum_{i=1}^n u'_{i,\ell} \right| \geq z/\sqrt{2} \right) &\leq \mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n u'_{i,\ell} \geq z/\sqrt{2} \right) + \mathbb{P} \left(\max_{1 \leq n \leq N} \sum_{i=1}^n -u'_{i,\ell} \geq z/\sqrt{2} \right) \\ &\leq 2 \left[3 + \left(\frac{3}{\alpha} \right)^{\frac{2}{\alpha}} \frac{128 \sum_{i=1}^N \|u_{i,\ell}\|_{\psi_\alpha}^2}{z^2} \right] \exp \left\{ - \left(\frac{z^2}{64 \sum_{i=1}^N \|u_{i,\ell}\|_{\psi_\alpha}^2} \right)^{\frac{\alpha}{\alpha+2}} \right\}. \end{aligned}$$

Thus (5.3) follows from union bound. \square

It remains an open question if similar concentration inequalities hold for polynomial-tail martingale differences where \mathbf{u}_i satisfies $\mathbb{P}(\|\mathbf{u}_i\| \geq z) \leq Cz^{-\beta}$ for $\beta \in (2, \infty)$? In the case where \mathbf{u}_i 's are independent, Theorem 6.21 of [Ledoux & Talagrand, 2013] gives a bound on the sum of vectors that can be turned to a tail inequality, but to our best knowledge a general result for martingale differences (even just in one dimension) is *not* available and left for future research.

References

- [Durrett, 2010] Durrett, R. (2010). *Probability: Theory and Examples (4th edition)*. Cambridge University Press.
- [Fan et al., 2012] Fan, X., Grama, I., & Liu, Q. (2012). Large deviation exponential inequalities for supermartingales. *Electronic Communications in Probability*, 17.
- [Fan et al., 2017] Fan, X., Grama, I., & Liu, Q. (2017). Deviation inequalities for martingales with applications. *Journal of Mathematical Analysis and Applications*, 448(1), 538–566.
- [Hayes, 2005] Hayes, T. P. (2005). A large-deviation inequality for vector-valued martingales.
- [Kallenberg & Sztencel, 1991] Kallenberg, O. & Sztencel, R. (1991). Some dimension-free features of vector-valued martingales. *Probability Theory and Related Fields*, 88(2), 215–247.
- [Laib, 1999] Laib, N. (1999). Exponential-type inequalities for martingale difference sequences. application to nonparametric regression estimation. *Communications in Statistics-Theory and Methods*, 28(7), 1565–1576.
- [Ledoux & Talagrand, 2013] Ledoux, M. & Talagrand, M. (2013). *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media.
- [Lee et al., 2016] Lee, J. R., Peres, Y., & Smart, C. K. (2016). A gaussian upper bound for martingale small-ball probabilities. *The Annals of Probability*, 44(6), 4184–4197.
- [Lesigne & Volný, 2001] Lesigne, E. & Volný, D. (2001). Large deviations for martingales. *Stochastic processes and their applications*, 96(1), 143–159.
- [Li et al., 2018] Li, C. J., Wang, M., Liu, H., & Zhang, T. (2018). Near-optimal stochastic approximation for online principal component estimation. *Mathematical Programming*, 167(1), 75–97.
- [Wainwright, 2019] Wainwright, M. J. (2019). *Basic tail and concentration bounds*, (pp. 21–57). In *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press.