



HAL
open science

Orlicz norms and concentration inequalities for β -heavy tailed random variables

Linda Chamakh, Emmanuel Gobet, Wenjun Liu

► **To cite this version:**

Linda Chamakh, Emmanuel Gobet, Wenjun Liu. Orlicz norms and concentration inequalities for β -heavy tailed random variables. 2021. hal-03175697v1

HAL Id: hal-03175697

<https://hal.science/hal-03175697v1>

Preprint submitted on 20 Mar 2021 (v1), last revised 30 Jun 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Orlicz norms and concentration inequalities for β -heavy tailed random variables*

Linda Chamakh[†] Emmanuel Gobet[‡] Wenjun Liu[§]

March 20, 2021

Abstract

We establish a new concentration-of-measure inequality for the sum of independent random variables with β -heavy tail. This includes exponential of Gaussian distributions (a.k.a. log-normal distributions), or exponential of Weibull distributions, among others. These distributions have finite polynomial moments at any order but many not have finite α -exponential moments. We exhibit a new Orlicz norm adapted to this setting of β -heavy tails, we prove a new Talagrand inequality for the sum and a new maximal inequality. As consequence, a deviation probability of the sum from its mean is obtained.

Keywords: heavy tails; deviation inequality; Orlicz norm; Talagrand inequality; maximal inequality; empirical process.

Mathematics Subject Classification (2020): 60E15, 60F10

1 Introduction

Concentration inequalities. Understanding how sample statistical fluctuations impact prediction errors is crucial in learning algorithms. Typically, we are interested in bounding the probability that a sum of random variables exceeds a certain threshold, essentially in quantifying the deviation of the sum from its expectation. In other words, we aim at analyzing how fast the sum concentrates around its expectation. Take the notation $[M]$ for all integers from 1 to M included. For independent and centered random variables $(Y_m)_{m \in [M]}$ taking value in a Banach space $(B, \|\cdot\|_B)$, the quantity of interest takes the form $\mathbb{P}\left(\left\|\sum_{m \in [M]} Y_m\right\|_B > \epsilon\right) \leq f(\epsilon, M)$ for the most explicit and tightest possible function f . The bounded, sub-Gaussian or the sub-exponential random variables have been largely covered by the literature (for example, via Bennett inequality and via Bernstein inequality - see [BLM13] for an extensive review of main concentration inequalities techniques), as well as the case of alpha-exponential tails [CGS20] (random variables Y s.t. there exists $\alpha > 0, c > 0$, such that $\mathbb{E}\left[\exp\left(\left(\frac{\|Y\|_B}{c}\right)^\alpha\right)\right] < \infty$). The fat-tailed case, for which the moment generating function does not exist but some polynomial moments exist, can be tackled for example via Burkholder or Fuk-Nagaev type of inequalities [Rio17, Mar17]. These inequalities are based on the existence and on the bounding of polynomial moments of the random variables. In this chapter, we focus on the heavy-tailed random variables case, in the limit case when no α -exponential moment is defined but every polynomial moment exist.

*This research is supported by the *Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique* and by the *Association Nationale de la Recherche Technique*. This work is part of the first author's doctoral thesis (funded by BNP Paribas), under the supervision of the second author.

[†]Global Markets Quantitative Research – BNP Paribas and CMAP, CNRS, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France. Email: linda.chamakh@polytechnique.edu

[‡]CMAP, CNRS, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France. Email: emmanuel.gobet@polytechnique.edu

[§]École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France. Email: wenjun.liu@polytechnique.edu

Orlicz norm. Orlicz norm [KR61] provides a nice tool to study the statistical fluctuations of an estimator for a given family of distributions. Consider an Orlicz function $\Psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, that is a continuous non-decreasing function, vanishing in zero and with $\lim_{x \rightarrow +\infty} \Psi(x) = +\infty$, and define the Ψ -Orlicz norm of the B -valued random variable Y by

$$\|Y\|_\Psi := \inf \left\{ c > 0 : \mathbb{E} \left[\Psi \left(\frac{\|Y\|_B}{c} \right) \right] \leq 1 \right\}. \quad (1.1)$$

With the additional property that Ψ is convex, Orlicz functions are commonly referred to as "Young functions" (or "N-functions" as in [KR61]). Van de Geer and Lederer [vdGL13] exhibit in their work a "Bernstein-Orlicz" norm (the "(L)-Bernstein-Orlicz" norm) adapted to sub-Gaussian and sub-exponential tails and provide deviation inequalities for suprema of functions of random variables. The (L)-Bernstein-Orlicz norm is the Ψ_L -Orlicz norm with

$$\Psi_L(z) = \exp \left[\frac{\sqrt{1 + 2Lz} - 1}{L} \right]^2 - 1.$$

We see that $\|Y\|_{\Psi_L} < \infty$ implies the existence of exponential moment. One of their main results [vdGL13, Theorem 8] is a uniform concentration result. As shown in Wellner [Wel17], it is possible to generalize these results to any Orlicz function $\Psi(x) = e^{h(x)} - 1$ with h convex. It requires again the existence of exponential moment which is not our purpose. We would like to go beyond and do not assume any α -exponential moment. As a new Orlicz function able to handle heavy-tail situations, we will consider:

$$\Psi_\beta^{\text{HT}}(x) := \exp((\ln(x+1))^\beta) - 1, \quad x \geq 0, \quad (1.2)$$

for a parameter $\beta > 1$. We say that Y is β -heavy tailed if there exists a $c > 0$ s.t.

$$\mathbb{E} \left[\Psi_\beta^{\text{HT}} \left(\frac{\|Y\|_B}{c} \right) \right] < \infty.$$

Typically, we aim at encompassing situations like $Y = \exp(|G|^{\frac{2}{\beta}})$ where G is a Gaussian random variable; the case $\beta = 2$ corresponds to log-normal tails. See Section 2.2 for various examples.

Observe that when (1.1) is finite with $\Psi = \Psi_\beta^{\text{HT}}$, Y has finite polynomial moment of order p for any $p > 0$, but may not have α -exponential moments. Besides, our β -heavy tailed setting is closely related to *long-tail* modelling¹, which is used for instance in queuing applications [Asm03, Chapter 10].

Deviation inequalities for sum of processes via Talagrand and Markov inequalities. What we call Talagrand inequality is an inequality of type:

$$\left\| \sum_{m \in [M]} Y_m \right\|_\Psi \leq C_\Psi \left(\left\| \sum_{m \in [M]} Y_m \right\|_{L_1(B)} + \left\| \max_{m \in [M]} \|Y_m\|_B \right\|_\Psi \right). \quad (1.3)$$

Talagrand [Tal89, Theorem 3] showed that this inequality is satisfied with $\Psi_\alpha(x) := e^{x^\alpha} - 1$.

For the sake of presentation, let us consider i.i.d. $(Y_m)_{m \in [M]}$. The first term is then $\left\| \sum_{m \in [M]} Y_m \right\|_{L_1(B)} \leq \mathcal{O}(\sqrt{M})$ by Bukholder inequality when $B \subseteq \mathbb{R}$ or the more general inequality of [Pis16, Proposition 4.35] when B is a Hilbert space or a Banach space of type 2.

When the maximal inequality is satisfied, that is under the form of [vdVW96, Lemma 2.2.2.], the second term is bounded by

$$\left\| \max_{m \in [M]} \|Y_m\|_B \right\|_\Psi \leq K_\Psi \Psi^{-1}(M) \max_{m \in [M]} \|Y_m\|_\Psi.$$

¹typically $S(x) := \mathbb{P}(\|Y\|_B > x) = \exp(-(\ln(1+x))^\beta)$ for which $\lim_{x \rightarrow +\infty} S(x+t)/S(x) = 1$ for any $t > 0$.

Hence, for any $\epsilon > 0$, denoting $X := \frac{1}{M} \left\| \sum_{m \in [M]} Y_m \right\|_B$, thanks to the Markov inequality, the Talagrand inequality and the two previous norm controls, we get

$$\begin{aligned} \mathbb{P}(X \geq \epsilon) &\stackrel{\text{Sect. 2.1-(iii)}}{\leq} \frac{2}{1 + \Psi(\epsilon/\|X\|_\Psi)} = 2 \left(1 + \Psi \left(\frac{\epsilon M}{\left\| \sum_{m \in [M]} Y_m \right\|_\Psi} \right) \right)^{-1} \\ &\leq 2 \left(1 + \Psi \left(\frac{\epsilon M}{C'_\Psi(\Psi^{-1}(M) + \sqrt{M})} \right) \right)^{-1}. \end{aligned} \quad (1.4)$$

In particular, for $\Psi = \Psi_\alpha$, the above inequality simplifies to:

$$\mathbb{P}(X \geq \epsilon) \leq 2 \exp \left(-C'_\alpha \left(\epsilon \sqrt{M} \right)^\alpha \right).$$

By leveraging union bound techniques and Hoffman-Jorgensen inequality, we can extend this inequality to suprema of function as done in [CGS20], in the spirit of [Ada08]. In any case, a key element to derive these concentration inequalities is the Talagrand inequality (1.3).

Our contribution. The purpose of this work is to establish the Talagrand inequality for $\Psi = \Psi_\beta^{\text{HT}}$, to tackle β -heavy tailed random variables as a difference with previous contributions available in the literature. Note that this particular Orlicz function (1.2) is not at all part of the general result established by Talagrand [Tal89, Proposition 12], which states that the inequality (1.3) holds for Orlicz function of the form $\Psi(x) := e^{x\zeta(x)}$ with ζ *non-decreasing* for x large enough and satisfying $\limsup_{u \rightarrow +\infty} \frac{\zeta(e^u)}{\zeta(u)} < +\infty$; indeed, in our setting, one easily checks that $\zeta(x) = x^{-1} \ln(\Psi_\beta^{\text{HT}}(x)) = x^{-1} (\ln(\exp((\ln(x+1))^\beta) - 1))$ is *decreasing* for x large.

Outline. In the first Section, we recall the motivating example and the adapted Orlicz function for which our main result is stated. In Section 3, the proof of the main result is given. The proof is based on a key inequality and on the application of subsidiary results from [Tal89].

2 Motivating examples and main results

2.1 Orlicz norm properties

Although $\|\cdot\|_\Psi$ defined in (1.1) may not satisfy in general the triangle inequality, we keep calling it Orlicz norm for the sake of simplicity. For a given Banach space $(B, \|\cdot\|_B)$ over the field \mathbb{R} , we denote $L_\Psi(B) := \{Y : \Omega \rightarrow B \text{ s.t. } \|Y\|_\Psi < +\infty\}$ the set of B -valued random variables with finite Ψ -Orlicz norm. For self-containedness we summarize a few well-known properties of the $\|\cdot\|_\Psi$ norm, for a given Orlicz function Ψ , which hold independently of the convexity of Ψ (unless explicitly required). See [KR61], or more recently [CGS20, Section 4].

- (i) Normalization: If $Y \in L_\Psi(B)$ then $\mathbb{E} \left[\Psi \left(\frac{\|Y\|_B}{\|Y\|_\Psi} \right) \right] \leq 1$.
- (ii) Homogeneity: If $Y \in L_\Psi(B)$ and $c \in \mathbb{R}$ then $cY \in L_\Psi(B)$ and $\|cY\|_\Psi = |c| \|Y\|_\Psi$.
- (iii) Deviation inequality: If $Y \in L_\Psi(B)$ then $\mathbb{P}(\|Y\|_B \geq c) \leq \frac{2}{\Psi(c/\|Y\|_\Psi) + 1}$ for any $c \geq 0$.
- (iv) If Ψ is convex, $\|\cdot\|_\Psi$ satisfies to the triangle inequality.

2.2 Motivating examples of heavy-tailed distributions and adapted Orlicz norm

Log-normal distribution. Let Y be a scalar random variable with log-normal distribution, i.e.

$$\ln(Y) \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2),$$

with $\sigma > 0$. The distribution of Y admits the density

$$f_Y(y; \mu, \sigma) := \frac{1}{\sigma\sqrt{2\pi}y} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \mathbf{1}_{y>0}.$$

We now investigate what kind of Orlicz function Ψ can be used to have $\|Y\|_\Psi < \infty$. In particular, we search for $\Psi(x) = \exp(\xi(x)) - 1$ such that ξ is non-decreasing, $\xi(0) = 0$ and $\lim_{x \rightarrow +\infty} \xi(x) = +\infty$ in order to ensure that $\Psi(0) = 0$ and $\lim_{x \rightarrow +\infty} \Psi(x) = +\infty$. Let $c > 0$, observe that

$$\mathbb{E} \left[\exp \left(\left| \xi \left(\frac{|Y|}{c} \right) \right| \right) \right] < \infty \implies \liminf_{x \rightarrow \infty} \xi \left(\frac{x}{c} \right) - \frac{(\ln x)^2}{2\sigma^2} = -\infty. \quad (2.1)$$

Consider the following functions for $\beta > 0$:

1. $\xi_\beta(x) = (\ln(x+1))^\beta$, $x \geq 0$. Note that the case $\beta \leq 1$ is not much interesting in our setting since it quantifies tails with finite expectation at most (fat tail cases).
2. $\xi_\beta(x) = (\ln(x+1))^\beta (\ln(\ln(x+1) + 1))^\alpha$, $x \geq 0$, $\alpha \in \mathbb{R}$. This second case is a scale refinement of the first case. It is not studied here.

These functions satisfy the necessary condition (2.1) if $\beta < 2$ and for a large c^2 . Furthermore, since for any $c > 0$, $\xi_\beta \left(\frac{x}{c} \right) < \epsilon (\ln x - \mu)^2$ for any $\epsilon > 0$ for x large enough, $\mathbb{E} \left[\exp \left(\left| \xi_\beta \left(\frac{|Y|}{c} \right) \right| \right) \right] < +\infty$.

Other distributions satisfying $\mathbb{E} \left[\exp \left(\xi_\beta \left(\frac{\|Y\|_B}{c} \right) \right) \right] < +\infty$. The associated Orlicz function $\Psi_\beta^{\text{HT}}(x) := \exp(\xi_\beta(x)) - 1$ is adapted to other distributions than just the log-normal distribution. For any random variable X admitting finite α -exponential moment with $\alpha > 1$, then Y defined by $\ln(Y) = X$ will admit β -heavy tailed for any $1 < \beta < \alpha$. We refer the reader to [CGS20, Table 2] for an exhaustive list of distributions admitting α -exponential moments. Here are a few examples:

- The Generalized normal distribution with parameters $c \in \mathbb{R}$, $b > 0$, $\alpha > 0$ has a density $f(x) = c_f e^{-\frac{1}{2} \left(\frac{|x-c|}{b} \right)^\alpha}$ up to a positive normalization constant c_f : it clearly admits a finite α -exponential moment. Hence $Y = \exp(X)$ where X has density f hence admits β -heavy tails for $\beta < \alpha$.
- The Skew normal distribution with parameters $b \in \mathbb{R}$, $c \in \mathbb{R}$, $v \in (0, +\infty)$ has a density $f(x) = c_f e^{-\frac{(x-c)^2}{2v}} \Phi \left(\frac{b(x-c)}{\sqrt{v}} \right)$, where Φ denotes the standard Gaussian cumulative distribution function and c_f is a positive normalization constant: it admits 2-exponential moment. If X has this density, then $Y = \exp(X)$ has β -heavy tails for $\beta < 2$.
- The Weibull distribution with parameters $\lambda > 0$, $k > 0$ has a density $f(x) = c_f x^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \mathbf{1}_{x \geq 0}$ up to a positive normalization constant c_f : it has finite k -exponential moment. Consequently, $Y = \exp(X)$ where X has the density as above, admits β -heavy tails for $\beta < k$. Such distributions are used, for instance, for earthquake magnitude modelling [HR99].

2.3 Ψ_β^{HT} -Orlicz norm: properties and inequalities

We state different properties of the Orlicz function to be used for β -heavy tailed distribution.

Proposition 2.1. For $\beta > 0$ define $\Psi_\beta^{\text{HT}} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ by

$$\Psi_\beta^{\text{HT}}(x) := \exp(\xi_\beta(x)) - 1 \quad \text{with} \quad \xi_\beta(x) := (\ln(1+x))^\beta, \quad x \geq 0. \quad (2.2)$$

The following properties hold:

1. The application $\beta \mapsto \Psi_\beta^{\text{HT}}$ defines a group isomorphism between $((0, +\infty), \times)$ and $((\Psi_\beta^{\text{HT}} : \beta > 0), \circ)$, and in particular, $(\Psi_\beta^{\text{HT}})^{-1} = \Psi_{1/\beta}^{\text{HT}}$.

² $\beta = 2$ is possible under restriction on σ : if $\sigma < \frac{1}{\sqrt{2}}$, then $\liminf_{x \rightarrow \infty} \xi_2(x) - \frac{(\ln x)^2}{2\sigma^2} = -\infty$.

2. For $\beta > 0$, Ψ_β^{HT} is an Orlicz function.
3. For $\beta > 1$, Ψ_β^{HT} is convex.

As a consequence, the associated Ψ_β^{HT} -Orlicz norm satisfies to the triangle inequality for $\beta > 1$.

Proof. Item 1. Observe that $\Psi_1^{\text{HT}}(x) = x$ and $\Psi_{\beta_1}^{\text{HT}}(\Psi_{\beta_2}^{\text{HT}}(x)) = \Psi_{\beta_1\beta_2}^{\text{HT}}(x)$ for any $x \geq 0$; the property of group isomorphism readily follows.

Item 2. This is straightforward to verify.

Item 3. Ψ_β^{HT} is a C^∞ -function on $(0, \infty)$, with a second derivative equal to

$$\Psi_\beta^{\text{HT}''}(x) = \frac{\exp((\ln(1+x))^\beta) (\ln(1+x))^{\beta-2}}{(1+x)^2} \times \beta \times \underbrace{\left(\beta (\ln(1+x))^\beta + (\beta-1) - \ln(1+x) \right)}_{=:g(\ln(1+x))}.$$

The function g is continuously differentiable on \mathbb{R}^+ , strictly positive at 0 ($g(0) = \beta - 1 > 0$) and goes to infinity at infinity (since $\beta > 1$); the critical points of g' are solutions to $\beta^2 y^{\beta-1} - 1 = 0$, therefore it is unique (equal to $y_\beta := \beta^{-\frac{2}{\beta-1}}$) and corresponds to the minimum of g . Let us evaluate the sign of g at the minimum:

$$g(y_\beta) = \beta y_\beta^\beta + (\beta - 1) - y_\beta = \frac{y_\beta}{\beta} + (\beta - 1) - y_\beta = (\beta - 1) \left(1 - \frac{y_\beta}{\beta} \right) = (\beta - 1) \left(1 - \frac{1}{\beta^{\frac{\beta+1}{\beta-1}}} \right) > 0.$$

All in all, we have proved that $\Psi_\beta^{\text{HT}''}(x) > 0$ for any $x > 0$. \square

We now restrict ourselves to the more interesting case $\beta > 1$. Let us state our main result: the Talagrand inequality (1.3) for the Ψ_β^{HT} -Orlicz norm.

Theorem 2.1 (Talagrand type inequality). *Let $\beta \in (1, +\infty)$. Then there is a universal constant $\mathcal{K}_{\beta,(2.3)}$ s.t. for all independent, mean zero, random variables sequence $(Y_m)_{m \in [M]}$ with $Y_m \in L_{\Psi_\beta^{\text{HT}}}(B)$ for all $m \in [M]$, we have*

$$\left\| \sum_{m \in [M]} Y_m \right\|_{\Psi^{\text{HT}}} \leq \mathcal{K}_{\beta,(2.3)} \left(\left\| \sum_{m \in [M]} Y_m \right\|_{L_1(B)} + \left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi^{\text{HT}}} \right). \quad (2.3)$$

- One of the consequence of this inequality is that if $\left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi^{\text{HT}}}$ and $\left\| \sum_{m \in [M]} Y_m \right\|_{L_1(B)}$ are finite, then $\sum_{m \in [M]} Y_m$ is β -heavy tailed, which some tractable controls.
- Recall that this inequality is a key step to derive concentration inequality of type (1.4).

We also establish that the general maximal inequality [vdVW96, Lemma 2.2.2.] (recalled in Lemma 3.6) holds for the Ψ_β^{HT} function:

Theorem 2.2 (A Ψ_β^{HT} maximal inequality). *Let $\beta \in (1, +\infty)$. Then there exists a universal constant $\mathcal{C}_{\beta,(2.4)}$ s.t. for any random variables Y_1, \dots, Y_M in $L_{\Psi_\beta^{\text{HT}}}(B)$,*

$$\left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi^{\text{HT}}} \leq \mathcal{C}_{\beta,(2.4)} (\Psi_\beta^{\text{HT}})^{-1}(M) \max_{m \in [M]} \|Y_m\|_{\Psi^{\text{HT}}}. \quad (2.4)$$

Recall that $(\Psi_\beta^{\text{HT}})^{-1}(M) = \Psi_{1/\beta}^{\text{HT}}(M)$. As a consequence of the Talagrand inequality (2.3) and the maximal inequality (2.4), by following the same steps as described in (1.4), we can derive the following concentration inequality:

Corollary 2.3 (A concentration inequality for sum of independent β -heavy tailed random variables). *Let $\beta \in (1, +\infty)$. Assume that B is an Hilbert space or a Banach space of type 2. Then for any Y_1, \dots, Y_M independent and centered random variables in $L_{\Psi_\beta^{\text{HT}}}(B)$, for any $\epsilon > 0$,*

$$\mathbb{P} \left(\frac{1}{M} \left\| \sum_{m \in [M]} Y_m \right\|_B \geq \epsilon \right) \leq 2 \exp \left(- \ln \left(1 + \frac{\epsilon M}{\mathcal{K}_{\beta, (2.3)} \left(C(2)^{1/2} \mu_2 \sqrt{M} + \mathcal{C}_{\beta, (2.4)} \mu_{\Psi_\beta^{\text{HT}}} \Psi_{1/\beta}^{\text{HT}}(M) \right)} \right)^\beta \right),$$

where $\mu_{\Psi_\beta^{\text{HT}}} := \max_{m \in [M]} \|Y_m\|_{\Psi_\beta^{\text{HT}}}$ and $\mu_2 := \max_{m \in [M]} \|Y_m\|_{L_2(B)}$, $C(2)$ denotes the universal constant in the Pisier inequality [Pis16, Proposition 4.35], $\mathcal{K}_{\beta, (2.3)}$ the Talagrand constant in (2.3) and $\mathcal{C}_{\beta, (2.4)}$ the maximal inequality constant in (2.4).

If Y_1, \dots, Y_M are i.i.d., $\mu_{\Psi_\beta^{\text{HT}}}$ and μ_2 do not depend on M .

3 Proofs

3.1 Preliminary results

Here, we recall Lemmas 8 and 9 of [Tal89], as well as the "Basic Estimate", which will enable us to prove our Theorem 2.1. In addition to the independent B -valued random variables $(Y_m)_{m \in [M]}$, we will need to consider extra independent Rademacher random variables. Everything is defined as follows. Let $(\Omega^M \times \Omega', \sum^M \otimes \sum', \mathbb{P})$ the basic probability space, where $\mathbb{P} = \mathbb{P} \otimes \mathbb{P}'$ such that the variables Y_m are defined on Ω^M and for $\omega = (\omega_m)_{m \in [M]}$, $Y_m(\omega)$ depends only on ω_m . Let $(\varepsilon_m)_{m \in [M]}$ be a set of random variables defined on Ω' with a Rademacher distribution independent of $(Y_m)_{m \in [M]}$. The following inequalities can be proven independently apart from the context of Orlicz norms.

Lemma 3.1 ([Tal89, Lemma 8]). *If $\mathbb{P}(\max_{m \in [M]} \|Y_m\|_B \geq t) \leq \frac{1}{2}$, then*

$$\sum_{m \in [M]} \mathbb{P}(\|Y_m\|_B \geq t) \leq 2\mathbb{P}\left(\max_{m \in [M]} \|Y_m\|_B \geq t\right).$$

Lemma 3.2 ([Tal89, Lemma 9]). *Set $X^{(r)}$ the r -th largest term of $(\|Y_m\|_B)_{m \in [M]}$. Then*

$$\mathbb{P}\left(X^{(r)} \geq t\right) \leq \frac{1}{r!} \left(\sum_{m \in [M]} \mathbb{P}(\|Y_m\|_B \geq t) \right)^r.$$

Set

$$\mu := \mathbb{E} \left[\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B \right],$$

$\mu > 0$ because the Y_m 's are not all zero random variables (to avoid trivial situations). We now recall a key inequality, which combined with the previous lemma will enable us to prove our main result.

Theorem 3.1 ([Tal89, Equation (2.5)]). *For k, q positive integers s.t. $k \geq q$, $u > 0$ and $u' > 0$, we have*

$$\mathbb{P} \left(\left\| \sum_{i \in [M]} \varepsilon_i Y_i \right\|_B \geq 4q\mu + u + u' \right) \leq 4 \exp \left(- \frac{u^2}{64q\mu^2} \right) + \left(\frac{K_0}{q} \right)^k + \mathbb{P} \left(\sum_{r \leq k} X^{(r)} > u' \right)$$

where the constant K_0 is a universal constant.

3.2 Symmetrisation argument for Ψ convex

In the next Subsection, because we rely on Theorem 3.1, we are going to prove the inequality (2.3) on symmetric random variables first (e.g. variables Y_m s.t. $\varepsilon_m Y_m \stackrel{d}{=} Y_m$). The extension to non-symmetric variables will be direct thanks to Lemma 3.4 which establishes an "equivalence in norms" relationship between the Orlicz norm of the sum of random variables and its associated Rademacher average.

Lemma 3.3. *Let Ψ be convex Orlicz function and $\|\cdot\|_\Psi$ the associate Orlicz norm. For any mean zero random variable $Z \in L_\Psi(B)$, we have $\|Z\|_\Psi \leq \|Z - Z'\|_\Psi$, with Z' any B -valued random variable such that $\mathbb{E}[Z'|Z] = 0$.*

Proof. Let $c > 0$,

$$\begin{aligned} \mathbb{E}[\Psi(\|Z\|_B/c)] &\stackrel{(a)}{=} \mathbb{E}\left[\Psi\left(\frac{\|\mathbb{E}[Z - Z' | Z]\|_B}{c}\right)\right] \stackrel{(b)}{\leq} \mathbb{E}\left[\Psi\left(\frac{\mathbb{E}[\|Z - Z'\|_B | Z]}{c}\right)\right] \\ &\stackrel{(c)}{\leq} \mathbb{E}\left[\mathbb{E}\left[\Psi\left(\frac{\|Z - Z'\|_B}{c}\right) \mid Z\right]\right] = \mathbb{E}\left[\Psi\left(\frac{\|Z - Z'\|_B}{c}\right)\right] \end{aligned}$$

where in (a) we use Z' has a zero conditional mean, in (b) we use that Ψ is non decreasing and the triangular inequality holds for the $\|\cdot\|_B$, in (c) we apply the Jensen inequality. Hence by taking $c = \|Z - Z'\|_\Psi > 0$, the right hand side is smaller than 1 (using Property (i) in Section 2.1), and therefore $\|Z\|_\Psi \leq c = \|Z - Z'\|_\Psi$. \square

Lemma 3.4. *Let Ψ be as in Lemma 3.3. Let $(Y_m)_{m \in [M]}$ be a sequence of independent mean-zero random variables in $L_\Psi(B)$. Let $(\varepsilon_m)_{m \in [M]}$ be independent Rademacher random variables, and let $(Y'_m)_{m \in [M]}$ be an independent copy of the sequence $(Y_m)_{m \in [M]}$. Then*

$$\left\| \sum_{m \in [M]} Y_m \right\|_\Psi \leq \left\| \sum_{m \in [M]} Y_m - \sum_{m \in [M]} Y'_m \right\|_\Psi = \left\| \sum_{m \in [M]} \varepsilon_m (Y_m - Y'_m) \right\|_\Psi \leq 2 \left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_\Psi \leq 4 \left\| \sum_{m \in [M]} Y_m \right\|_\Psi.$$

Later on, we will apply these inequalities with $\Psi = \Psi_\beta^{\text{HT}}$ and $\Psi(x) = x$ (the associated Orlicz norm corresponds then to the L_1 norm).

Proof. The first inequality comes from the application of Lemma 3.3 with $Z = \sum_{m \in [M]} Y_m$ and $Z' = \sum_{m \in [M]} Y'_m$. Since ε_m takes values ± 1 independently of Z, Z' , we have $Y_m - Y'_m \stackrel{d}{=} Y'_m - Y_m \stackrel{d}{=} \varepsilon_m (Y_m - Y'_m)$. Since the sequences are independent in m , we obtain the equality of Lemma 3.4. The second inequality is a consequence of the triangular inequality (iv) and the previous identities in distribution.

The last inequality is a consequence of the application of Lemma 3.3 with $Z = \sum_{m \in [M]} \varepsilon_m Y_m$ and $Z' = \sum_{m \in [M]} \varepsilon_m Y'_m$ satisfying $\mathbb{E}[Z'|Z] = \mathbb{E}\left[\mathbb{E}\left[\sum_{m \in [M]} \varepsilon'_m Y'_m \mid \varepsilon'_m, Z'_m, m \in [M]\right] \mid Z\right] = 0$ and of the triangular inequality: $\|Z\|_\Psi \leq \|Z - Z'\|_\Psi = \left\| \sum_{m \in [M]} (Y_m - Y'_m) \right\|_\Psi \leq 2 \left\| \sum_{m \in [M]} Y_m \right\|_\Psi$. \square

3.3 Proof of the main result, Theorem 2.1

We will denote $K(\beta)$ a constant depending only on β , that may vary from line to line. We assume that at least one of the Y_m 's is not zero *a.s.*, otherwise the announced inequality (2.3) is obvious.

In view of the inequalities of Lemma 3.4, it is enough to do the reasoning and show the inequality (2.3) with the variables $(\varepsilon_m Y_m, m \in [M])$ instead of $(Y_m, m \in [M])$.

\triangleright **Rescaling.** Note that (2.3) is invariant by homogeneous rescaling (see Property (ii) of Section 2.1), i.e. the inequality remains the same for the random variables $\tilde{Y}_m := \frac{\varepsilon_m Y_m}{C}$ for any $C > 0$. For the choice

$$C := \left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_{L_1(B)} + \left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi^{\text{HT}}} > 0,$$

observe that

$$\left\| \sum_{m \in [M]} \tilde{Y}_m \right\|_{L_1(B)} \leq 1 \quad \text{and} \quad \left\| \max_{m \in [M]} \|\tilde{Y}_m\|_B \right\|_{\Psi_{\text{HT}}} \leq 1, \quad (3.1)$$

therefore the inequality (2.3) writes

$$\left\| \sum_{m \in [M]} \tilde{Y}_m \right\|_{\Psi_{\text{HT}}} \leq 2K(\beta).$$

Conversely, if the above holds for some $K(\beta)$ (independent from the \tilde{Y}_m 's verifying (3.1)), then (2.3) holds for the Y_m 's. All in all, it means that without loss of generality, we can assume

$$\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_{L_1(B)} \leq 1 \quad \text{and} \quad \left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi_{\text{HT}}} \leq 1,$$

and then show, under these assumptions, the existence of $K(\beta) \in \mathbb{R}$ (independent on Y_m 's) such that

$$\mathbb{E} \left[\exp \left(\xi_\beta \left(\frac{\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B}{K(\beta)} \right) \right) \right] \leq 2.$$

▷ **Deviation bounds.** By Property (iii) of Section 2.1 and since we assumed $\left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi_{\text{HT}}} \leq 1$,

$$\mathbb{P} \left(\max_{m \in [M]} \|Y_m\|_B \geq t \right) \leq 2 \exp(-\xi_\beta(t)), \quad t \geq 0.$$

The function $\xi_\beta(\cdot) = (\ln(1 + \cdot))^\beta$ being continuously increasing from 0 to $+\infty$, there exists t_0 s.t. $\xi_\beta(t_0) = 2 \ln 2$ and $\forall t \geq t_0$, $2 \exp(-\xi_\beta(t)) \leq 1/2$; for further use, notice the value $t_0 = e^{(2 \ln 2)^{\frac{1}{\beta}}} - 1$. Then applying Lemma 3.1, for $t \geq t_0$, we have

$$\sum_{m \in [M]} \mathbb{P}(\|Y_m\|_B \geq t) \leq 2 \mathbb{P} \left(\max_{m \in [M]} \|Y_m\|_B \geq t \right) \leq 4 \exp(-\xi_\beta(t)).$$

Hence Lemma 3.2 yields for $r \in \mathbb{N}^*$, $t \geq t_0$

$$\mathbb{P} \left(X^{(r)} \geq t \right) \leq \frac{4^r \exp(-r \xi_\beta(t))}{r!}. \quad (3.2)$$

Denote $\tilde{\beta} = \lfloor \beta \rfloor + 1 \geq 2$. Equation (3.2) yields for $t \geq (e^{\tilde{\beta}} - 1) r^{\frac{\tilde{\beta}}{\beta}}$ (notice that $t \geq e^2 - 1 \geq t_0$ as requested)

$$\mathbb{P} \left(X^{(r)} \geq t r^{-\frac{\tilde{\beta}}{\beta}} \right) \leq \frac{4^r \exp \left(-r [\ln(1 + t r^{-\frac{\tilde{\beta}}{\beta}})]^\beta \right)}{r!} =: f(r, t).$$

Since $\tilde{\beta}/\beta > 1$, the sequence $(r^{-\frac{\tilde{\beta}}{\beta}})_{r \geq 1}$ is summable. Set $S_\beta := \sum_{r \geq 1} r^{-\frac{\tilde{\beta}}{\beta}} < +\infty$ and $g(t) := (t/(e^{\tilde{\beta}} - 1))^{\beta/\tilde{\beta}}$.

From the inclusion $\{\sum_{r \leq g(t)} X^{(r)} \geq t S_\beta\} \subset \bigcup_{r \leq g(t)} \{X^{(r)} \geq t r^{-\frac{\tilde{\beta}}{\beta}}\}$ and writing a union bound, we get

$$\mathbb{P} \left(\sum_{r \leq g(t)} X^{(r)} \geq t S_\beta \right) \leq \sum_{r \leq g(t)} f(r, t). \quad (3.3)$$

We claim that for all $1 \leq r \leq g(t)$

$$r^{1/\beta} \ln(1 + t r^{-\frac{\tilde{\beta}}{\beta}}) \geq \ln(1 + t). \quad (3.4)$$

This is a consequence of the above lemma applied with $\rho = r^{\frac{1}{\beta}} \geq 1$ and $\tau = t r^{-\frac{\tilde{\beta}}{\beta}} \geq e^{\tilde{\beta}} - 1$.

Lemma 3.5. For all $\rho \geq 1$ and $\tau \geq e^{\tilde{\beta}} - 1$, we have $\rho \ln(1 + \tau) \geq \ln(1 + \tau \rho^{\tilde{\beta}})$.

Proof. The function $f(\rho) := \rho \ln(1 + \tau) - \ln(1 + \tau \rho^{\tilde{\beta}})$ vanishes at $\rho = 1$, let us prove that it is non-decreasing in ρ provided that $\tau \geq e^{\tilde{\beta}} - 1$. Indeed,

$$f'(\rho) = \ln(1 + \tau) - \frac{\tilde{\beta} \rho^{\tilde{\beta}-1} \tau}{1 + \rho^{\tilde{\beta}} \tau}.$$

Since $\rho^{\tilde{\beta}} \tau \geq 0$ and $\rho \geq 1$, we have $\frac{\tilde{\beta} \rho^{\tilde{\beta}-1} \tau}{1 + \rho^{\tilde{\beta}} \tau} \leq \frac{\tilde{\beta}}{\rho} \leq \tilde{\beta}$. Hence, $f'(\rho) \geq \ln(1 + \tau) - \tilde{\beta} \geq 0$. We are done. \square

Plugging (3.4) into (3.3) yields

$$\mathbb{P} \left(\sum_{r \leq g(t)} X^{(r)} \geq tS_\beta \right) \leq \sum_{r \leq g(t)} \frac{4^r}{r!} \exp(-[\ln(1+t)]^\beta) \leq \exp(4) \exp(-\xi_\beta(t)).$$

Let us recall that $\mu = \left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_{L_1(B)} \leq 1$. We are now at the point to apply Theorem 3.1 with $q = \lceil eK_0 \rceil$, $u = t$, $u' = tS_\beta$, $2q\mu \leq t$ and $k = \lfloor g(t) \rfloor$:

$$\begin{aligned} \mathbb{P} \left(\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B \geq t(S_\beta + 3) \right) &\leq \mathbb{P} \left(\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B \geq 4q\mu + u + u' \right) \\ &\leq 4 \exp\left(-\frac{t^2}{64q\mu^2}\right) + \exp(-\lfloor g(t) \rfloor) + \mathbb{P} \left(\sum_{r \leq g(t)} X^{(r)} \geq tS_\beta \right) \\ &\leq 4 \exp\left(-\frac{t^2}{64q\mu^2}\right) + \exp(-\lfloor g(t) \rfloor) + \exp(4 - \xi_\beta(t)). \end{aligned}$$

The above inequality is valid for any $t \geq t_0 \vee (2\lceil eK_0 \rceil \mu)$. Besides, in the above upper bound, the last third is asymptotically the largest one, therefore there exists $K(\beta) > 0$ such that

$$\mathbb{P} \left(\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B \geq K(\beta)t \right) \leq K(\beta) \exp(-\xi_\beta(t)), \quad t \geq 0.$$

▷ **Orlicz norm bounds.** The estimate implies for all $c > 0$:

$$\begin{aligned} \mathbb{E} \left[\exp \left(\xi_\beta \left(\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B / (cK(\beta)) \right) \right) \right] - 1 &= \int_0^\infty \exp(\xi_\beta(t)) \xi'_\beta(t) \mathbb{P} \left(\frac{\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B}{cK(\beta)} \geq t \right) dt \\ &\leq K(\beta) \int_0^\infty \xi'_\beta(t) \exp(\xi_\beta(t) - \xi_\beta(ct)) dt. \end{aligned} \quad (3.5)$$

Let us check that the above integral is finite for $c > 1$. Only the integrability at $t \rightarrow +\infty$ is questionable. Write

$$\xi_\beta(t) - \xi_\beta(ct) = (\ln(1+t))^\beta \left(1 - \left[1 + \frac{\ln\left(\frac{1+ct}{1+t}\right)}{\ln(1+t)} \right]^\beta \right) \approx_{t \rightarrow +\infty} -\beta (\ln(1+t))^{\beta-1} \ln(c).$$

Therefore, the function to integrate is bounded for t large by (up to constant)

$$g(t) := \frac{(\ln(1+t))^{\beta-1}}{(1+t)} e^{-\frac{1}{2}\beta(\ln(1+t))^{\beta-1} \ln(c)}.$$

We easily check that $\int_0^{+\infty} g(t)dt = \int_0^{+\infty} y^{\beta-1} e^{-\frac{1}{2}\beta y^{\beta-1} \ln(c)} dy < +\infty$ since $\beta > 1$ and $c > 1$. Furthermore, by monotone convergence theorem, the bound (3.5) converges to 0 as $c \rightarrow +\infty$, consequently

$$\mathbb{E} \left[\exp \left(\xi_\beta \left(\frac{\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_B}{cK(\beta)} \right) \right) \right] \leq 2$$

for a $c = c_\beta$ large enough. We have proved that $\left\| \sum_{m \in [M]} \varepsilon_m Y_m \right\|_{\Psi^{\text{HT}}} \leq c_\beta K(\beta)$. \square

3.4 Proof of Theorem 2.2: Ψ_β^{HT} -maximal inequality

We start by recalling the general maximal inequality on which our proof is based.

Lemma 3.6 ([vdVW96, Lemma 2.2.2]). *Let Ψ be a convex Orlicz function satisfying*

$$\limsup_{x, y \rightarrow +\infty} \Psi(x)\Psi(y)/\Psi(c_\Psi xy) < +\infty \quad (3.6)$$

for some constant $c_\Psi > 0$. Then, there is a constant $K > 0$ such that for any B -valued random variables Y_1, \dots, Y_M ,

$$\left\| \max_{m \in [M]} \|Y_m\|_B \right\|_{\Psi} \leq K\Psi^{-1}(M) \max_{m \in [M]} \|Y_m\|_{\Psi}.$$

For $\beta > 1$, Ψ_β^{HT} is a convex Orlicz function, thus it remains to establish (3.6) to get Theorem 2.2. We prove that one can take $c_\Psi = 1$. Let $c \geq 1^3$ s.t. $\Psi_\beta^{\text{HT}}(c^2) \geq 1$. Let x, y st. $x \geq c$ and $y \geq c$: then

$$\begin{aligned} \Psi_\beta^{\text{HT}}(x)\Psi_\beta^{\text{HT}}(y) &\leq e^{(\ln(1+x))^\beta} e^{(\ln(1+y))^\beta} \\ &\leq e^{(\ln(x))^\beta + (\ln(y))^\beta - (\ln(xy))^\beta} e^{(\ln(1+xy))^\beta} \underbrace{e^{2 \sup_{z \geq c} (\ln(1+z))^\beta - (\ln z)^\beta}}_{:=C(c)}. \end{aligned}$$

- $C(c)$ is finite: indeed, by standard equivalents, we have that

$$(\ln(1+z))^\beta - (\ln z)^\beta = (\ln z)^\beta \left(\left[1 + \frac{\ln(1+z^{-1})}{\ln(z)} \right]^\beta - 1 \right) \sim_{z \rightarrow \infty} \beta \frac{(\ln z)^{\beta-1}}{z}$$

which converges to 0 at infinity.

- Notice that $(\ln(xy))^\beta - (\ln(x))^\beta \geq (\ln(y))^\beta$ for any $x, y \geq 1$. Indeed, setting $u = \ln x \geq 0$, $v = \ln y \geq 0$,

$$(u+v)^\beta - u^\beta = \int_u^{u+v} \beta z^{\beta-1} dz \geq \int_0^v \beta z^{\beta-1} dz = v^\beta$$

(because $z \mapsto \beta z^{\beta-1}$ is increasing since $\beta > 1$).

- Last, since $e^{(\ln(1+xy))^\beta} = \Psi_\beta^{\text{HT}}(xy) + 1 \geq \Psi_\beta^{\text{HT}}(c^2) + 1 \geq 2$, one has

$$e^{(\ln(1+xy))^\beta} = \frac{e^{(\ln(1+xy))^\beta}}{e^{(\ln(1+xy))^\beta} - 1} \Psi_\beta^{\text{HT}}(xy) \leq 2\Psi_\beta^{\text{HT}}(xy).$$

All in all, we conclude that $\Psi_\beta^{\text{HT}}(x)\Psi_\beta^{\text{HT}}(y) \leq 2C(c)\Psi_\beta^{\text{HT}}(xy)$, for any $x, y \geq c$. We are done. \square

³one can take $c = \sqrt{e^{(\ln 2)^{1/\beta}} - 1} \geq 1$ for which $\Psi_\beta^{\text{HT}}(c^2) = 1$.

4 Conclusion and perspectives

To conclude, we have extended the Talagrand inequality for an Orlicz norm adapted to variables with β -heavy tails (Theorem 2.1 and Equation (2.2)). We have also shown that a maximal inequality was satisfied (Theorem 2.2), which enabled us to establish a concentration inequality for sum of independent and centered random variables with β -heavy tails (Corollary 2.3). A next step would be to extend this inequality to supremum of functions of random variables with β -heavy tails, which follows from a standard routine using a truncation argument as in [Ada08, CGS20] and union bounding technique as in [SS15] for instance.

References

- [Ada08] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.
- [Asm03] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [CGS20] L. Chamakh, E. Gobet, and Z. Szabó. Orlicz random Fourier feature. *Journal of Machine Learning Research (JMLR)*, 21:1–37, 2020.
- [HR99] T. Huillet and H.-F. Raynaud. Rare events in a log-Weibull scenario-Application to earthquake magnitude data. *The European Physical Journal B - Condensed Matter and Complex Systems*, 12(3):457–469, 1999.
- [KR61] M.A. Krasnoselskii and Ya.B. Rutickii. *Convex functions and Orlicz spaces*. Translated from the first Russian edition by Leo F. Boron. P. Noordhoff Ltd., Groningen, 1961.
- [Mar17] A. Marchina. Concentration inequalities for suprema of unbounded empirical processes. To appear in *Annales Henri Lebesgue*, hal-01545101, 2021.
- [Pis16] G. Pisier. *Martingales in Banach spaces*, volume 155 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2016.
- [Rio17] E. Rio. About the constants in the Fuk-Nagaev inequalities. *Electronic Communications in Probability*, 22, 2017.
- [SS15] B.K. Sriperumbudur and Z. Szabó. Optimal rates for random Fourier features. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1144–1152, 2015.
- [Tal89] M. Talagrand. Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *The Annals of Probability*, 17(4):1546–1570, 1989.
- [vdGL13] S. van de Geer and J. Lederer. The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2):225–250, 2013.
- [vdVW96] A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.
- [Wel17] J.A. Wellner. The Bennett-Orlicz norm. *Sankhya A. The Indian Journal of Statistics*, 79(2):355–383, 2017.