ON THE CONCENTRATION OF MEASURE IN ORLICZ SPACES OF EXPONENTIAL TYPE

VIET NGUYEN

Introduction

The study of Orlicz spaces, first described in 1931 by Wladyslaw Orlicz and Zygmunt Wilhelm Birnbaum [1], became popular in the empirical processes and functional analysis literature, due to the rising interest in chaining arguments to derive probabilistic bounds for stochastic processes, and generalizations of $\mathcal{L}^p$ spaces as well as Sobolev space embeddings, respectively. Orlicz spaces exhibit strong concentration phenomena, inherited from their construction. In particular, they are associated to the sub-Exponential and sub-Gaussian classes of random variables. In this article, we aim to provide a brief introduction to concentration of measure in Orlicz spaces, in particular, Orlicz spaces of exponential type. We begin by the construction of these spaces, and delve into certain concentration guarantees and applications.

1. K_σ-spaces, N-functions, Orlicz spaces, norm

We will closely follow the construction in [4]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be our underlying probability space.

Definition 1.1. (K_σ-space, or K-lineal) A *lattice* $(K(\Omega), \leq) \subset \mathcal{L}^0(\Omega)$ is a K_σ-space with norm $\|\cdot\|$ such that:

a) $\forall \xi, \eta \in K(\Omega), |\xi| \leq |\eta| \implies \|\xi\| \leq \|\eta\|$ (monotonicity of norms)

b) $\{\xi_n\} \subset K(\Omega), \exists \eta \in K(\Omega)$ such that $\sup_{n\geq 1} |\xi_n| < \eta$ a.s. $\implies \sup_{n\geq 1} |\xi_n| \in K(\Omega)$

One can see that a K_σ-space is a Banach lattice with an additional condition b). As an example, the function spaces $(\mathcal{L}^p(\Omega), p \geq 1)$ is a K_σ-space. For our future purposes, we write $K(\Omega)$ not to denote a vector lattice, but to denote a K_σ-space. We state here without proof that in a K_σ-space, convergence in norm implies convergence in probability, namely:

Lemma 1.2. Let $\{\xi_n\} \subset K(\Omega), \xi \in K(\Omega)$. Then,

$$\lim_{n\to\infty} \|\xi_n - \xi\| = 0 \implies \forall \epsilon > 0, \lim_{n\to\infty} \mathbb{P}\left(|\xi_n - \xi| > \epsilon\right) = 0$$

We introduce here the class of N-functions, sometimes called Young's functions and Young-Orlicz modulus beyond Soviet literature, that we will use to construct interesting K_σ-spaces, namely Orlicz spaces.

Definition 1.3. (N-functions) A continuous, even, convex function $\psi = \psi(x), x \in \mathbf{R}$ is an N-function if $\psi(x)$ is monotonically increasing for $x > 0, \psi(0) = 0$, and satisfies:

(n_0) $\lim_{x\to 0} \frac{\psi(x)}{x} = 0$

(n_∞) $\lim_{x\to\infty} \frac{\psi(x)}{x} = \infty$

ψ is an N-function if and only if there exists a càdlàg density $p(t), t \geq 0$ such that $\forall t > 0, p(t) > 0, p(0) = 0, p(t) \to \infty$ as $t \to \infty$:

$$\psi(x) = \int_0^{|x|} p(t)\, dt$$

A really important example of N-functions are functions of the type $\psi(x) = \exp\{\varphi(x)\} - 1$, such that $\varphi(x)$ is an N-function that is *not* x^p for any p, essential in constructing Orlicz spaces of exponential type. The following are important properties of N-functions that we shall use in later proofs.

Lemma 1.4. Let ψ be an N-function. Then, the following hold:

a) $\psi(|x| + |y|) \geq \psi(|x|) + \psi|y|$.

b) If $a > 1$, $\psi(ax) \geq a\psi(x)$.

We are now ready to define Orlicz spaces, whose elements are random variables defined on our underlying probability space:

Definition 1.5. (Orlicz space) Let ψ be an arbitrary N-function. The *Orlicz space of random variables* $\mathcal{L}_\psi(\Omega)$ is the family of equivalence classes of random variables where for each $\xi \in \mathcal{L}_\psi(\Omega), \exists r_\xi > 0$ such that:

$$\mathbb{E}\left[\psi\left(\frac{\xi}{r_\xi}\right)\right] = \int_\Omega \left(\psi \circ \frac{\xi}{r_\xi}\right) d\mathbb{P} < \infty$$

Two random variables are equivalent if they are equal almost surely. The norm

$$\|\xi\|_\psi = \inf\left\{r > 0 : \mathbb{E}\left[\psi\left(\frac{\xi}{r}\right)\right] \leq 1\right\}$$

is the Orlicz norm (sometimes Luxemburg norm) with the understanding that $\inf \emptyset = \infty$. This makes $\mathcal{L}_\psi(\Omega)$ a Banach space. Furthermore, one can show that $\mathcal{L}_\psi(\Omega)$ is a lattice, and hence a K_σ-space. We can characterize $\mathcal{L}_\psi(\Omega) \subset \mathcal{L}^0(\Omega)$ by $\mathcal{L}_\psi(\Omega) = \left\{\xi \in \mathcal{L}^0(\Omega) : \|\xi\|_\psi < \infty\right\}$.

Example 1.6. Suppose $\psi(x) = |x|^p$. Then, $\mathcal{L}_\psi(\Omega) = \mathcal{L}^p(\Omega)$ and $\|\xi\|_\psi = \|\xi\|_p$.

Finiteness of the norm immediately implies a tail bound:

Lemma 1.7. Let $\xi \in \mathcal{L}_\psi(\Omega)$ where ψ is a N-function. Then,

$$\mathbb{P}(|\xi| > x) \leq \psi^{-1}\left(\frac{\xi}{\|\xi\|_\psi}\right)$$

Proof. By Markov's generalized inequality, since $\psi \circ \left(\xi \mapsto \frac{1}{\|\xi\|_\psi}\right)$ is nonnegative and monotonically increasing, we have:

$$\mathbb{P}(|\xi| > x) \leq \frac{\mathbb{E}\left[\psi(\xi / \|\xi\|_\psi)\right]}{\psi(\xi / \|\xi\|_\psi)} \leq \frac{1}{\psi(\xi / \|\xi\|_\psi)}$$

∎

2. Orlicz spaces of exponential type, concentration

Definition 2.1. (Orlicz space of exponential type) Suppose that φ is a N-function that does not necessarily satisfy (n_0) or (n_∞), and let $\psi(x) = \exp\{\varphi(x)\} - 1$. We call the Orlicz space generated by ψ of exponential type, denoting it by $\mathrm{Exp}_\varphi(\Omega)$, and the corresponding norm $\|\cdot\|_{E\varphi}$.

Elements of $\mathrm{Exp}_\varphi(\Omega)$ enjoy an exponential tail bound:

Theorem 2.2. Let $\xi \in \mathrm{Exp}_\varphi(\Omega)$. Then,

$$\mathbb{P}(|\xi| \geq x) \leq 2\exp\left\{-\varphi\left(\frac{x}{\|\xi\|_{E\varphi}}\right)\right\}$$

The converse is also true; if

$$\mathbb{P}(|\xi| \geq x) \leq C\exp\left\{-\varphi\left(\frac{x}{D}\right)\right\}$$

for some $C, D > 0$, then $\xi \in \mathrm{Exp}_\varphi(\Omega)$ with $\|\xi\|_{E\varphi} \leq (1+C)D$.

Proof. Assume $\|\xi\|_{E\varphi} > 0$. Generalized Markov yields:

$$\begin{aligned}
\mathbb{P}(|\xi| \geq x) &\leq \mathbb{E}\left[\exp\left\{\varphi\left(\frac{\xi}{\|\xi\|_{E\varphi}}\right)\right\}\right]\exp\left\{-\varphi\left(\frac{x}{\|\varphi\|_{E\varphi}}\right)\right\} \\
&= \mathbb{E}\left[\psi\left(\frac{\varphi}{\|\varphi\|_{E\varphi}}\right) + 1\right]\exp\left\{-\varphi\left(\frac{x}{\|\varphi\|_{E\varphi}}\right)\right\} \\
&\leq 2\exp\left\{-\varphi\left(\frac{x}{\|\varphi\|_{E\varphi}}\right)\right\}
\end{aligned}$$

For the converse statement, assume $\mathbb{P}(|\xi| \geq x) \leq C\exp\left\{-\varphi\left(\frac{x}{D}\right)\right\}$ holds for $C, D > 0$. Assume $|\xi|$ has a density w.r.t. the Lebesgue measure, let F be the CDF of $|\xi|$ and $G = 1 - F$, and let $a > D$ (in particular, $\varphi(1/a) < \varphi(1/D)$). Remark that by the concentration assumption, we have:

$$\begin{aligned}
\exp\left\{\varphi\left(\frac{x}{a}\right)\right\}G(x) &\leq C\exp\left\{-\left(\varphi\left(\frac{x}{D}\right) - \varphi\left(\frac{x}{a}\right)\right)\right\} \\
&\leq C\exp\left\{-\varphi\left(x\left(\frac{1}{D} - \frac{1}{a}\right)\right)\right\}
\end{aligned}$$

since for N-function φ, $\varphi\left(\frac{x}{D} - \frac{x}{a}\right) \leq \varphi\left(\frac{x}{D}\right) - \varphi\left(\frac{x}{a}\right)$ holds, from Lemma 1.4. As $x \to \infty$, we have that $\exp\left\{\varphi\left(\frac{x}{a}\right)\right\}G(x) \to 0$. Now, consider:

$$\mathbb{E}\left[\exp\left\{\varphi\left(\frac{\xi}{a}\right)\right\}\right] = \int_0^\infty \exp\left\{\varphi\left(\frac{x}{a}\right)\right\}\, dF(x) \tag{1}$$

$$= -\exp\left\{\varphi\left(\frac{x}{a}\right)\right\}G(x)\Big]_0^\infty + \int_0^\infty \exp\left\{\varphi\left(\frac{x}{a}\right)\right\}G(x)\, d\varphi\left(\frac{x}{a}\right) \tag{2}$$

$$\leq -(-1) + C\int_0^\infty \exp\left\{\varphi\left(\frac{x}{a}\right) - \varphi\left(\frac{x}{D}\right)\right\}\, d\varphi\left(\frac{x}{a}\right) \tag{3}$$

$$\leq 1 + C\int_0^\infty \exp\left\{-\frac{a-D}{D}\varphi\left(\frac{x}{a}\right)\right\}\, d\varphi\left(\frac{x}{a}\right) = 1 + \frac{CD}{a-D} \tag{4}$$

where (2) follows from change of variables, (3) by substituting our control on $\exp\left\{\varphi\left(\frac{x}{a}\right)\right\}G(x)$, and (4) follows from the substitution $\varphi\left(\frac{x}{D}\right) = \varphi\left(\frac{x}{a}\frac{a}{D}\right) \geq \frac{a}{D}\varphi\left(\frac{x}{a}\right)$, from Lemma 1.4. We see that

$\mathbb{E}\left[\psi\left(\frac{\xi}{a}\right)\right] = \mathbb{E}\left[\exp\left\{\varphi\left(\frac{\xi}{a}\right)\right\}\right] - 1 \leq \frac{CD}{a-D} < \infty$, and thus $\xi \in \mathrm{Exp}_\varphi(\Omega)$. Choosing $a = (1+C)D$, since for this choice of a, $\mathbb{E}\left[\psi\left(\frac{\xi}{a}\right)\right] < \infty$, and the norm being the infimum over all such a, it follows that $\|\xi\|_{E\varphi} \leq (1+C)D$, giving us the desired conclusion. ∎

From Theorem 3.2., in particular, we can see that choosing $\varphi(x) = |x|^2$, hence $\psi(x) = \exp\left\{|x|^2\right\} - 1$, the finiteness of the Orlicz exponential type norm guarantees the sub-Gaussian property of ξ. Controlling the Orlicz norm directly enables us to control tail bounds in an Orlicz space.

3. Application: norm control for the maxima of finitely many random variables

We can obtain maximal inequalities using Cramér transforms and applying Chernoff bounds to control the maximum of a finite number of random variables. Using Orlicz norm, one is able to achieve the same feat.

Theorem 3.1. Let ψ be a N-function for which there exists $c > 0$ such that $\psi(x)\psi(y) \leq \psi(cxy)$, for all $x, y \geq 1$. Then, for m random variables $\xi_1, \ldots, \xi_m$,

$$\left\|\max_{i\leq m} |\xi_i|\right\|_\psi \leq K\psi^{-1}(m) \max_{i\leq m} \|\xi_i\|_\psi$$

Given some control on the individual Orlicz norms of each random variable, one is able to bound the Orlicz norm of the maximum.

Proof. We have from assumption $\psi\left(\frac{x}{y}\right) \leq \frac{\psi(cx)}{\psi(y)}$. For any $C > 0$,

$$\max_{i\leq m} \psi\left(\frac{|\xi_i|}{Cy}\right) \leq \max_{i\leq m}\left[\frac{\psi(c|\xi_i|/C)}{\psi(y)} \mathbb{1}\left\{\frac{|\xi_i|}{Cy} \geq 1\right\} + \psi\left(\frac{|\xi_i|}{Cy}\right)\mathbb{1}\left\{\frac{|\xi_i|}{Cy} < 1\right\}\right] \tag{1}$$

$$\leq \max_{i\leq m} \frac{\psi(c|\xi_i|/C)}{\psi(y)} + \psi(1) \tag{2}$$

$$\leq \sum_{i=1}^{m} \frac{\psi(c|\xi_i|/C)}{\psi(y)} + \psi(1) \tag{3}$$

where (2) follows from convexity of ψ. We now set $C := \max_{i\leq m} \|\xi_i\|_\psi$, and take expectations:

$$\mathbb{E}\left[\psi\left(\frac{\max_{i\leq m}|\xi_i|}{Cy}\right)\right] \leq \mathbb{E}\left[\max_{i\leq m}\psi\left(\frac{|\xi_i|}{Cy}\right)\right] \tag{4}$$

$$\leq \sum_{i=1}^{m} \mathbb{E}\left[\frac{\psi\left(|\xi_i|/\max_{i\leq m}\|\xi_i\|_\psi\right)}{\psi(y)}\right] + \psi(1) \tag{5}$$

$$\leq \frac{m}{\psi(y)} + \psi(1) \tag{6}$$

where (6) follows from properties of N-functions w.r.t. their corresponding Orlicz norms. We note that when $\psi(1) \leq \frac{1}{2}$, we have that $\frac{m}{\psi(y)} + \psi(1) \leq 1$

$$\begin{aligned}
\frac{m}{\psi(y)} + \psi(1) \leq 1 &\iff \psi(y) \geq \frac{m}{1-\psi(1)} \\
&\iff y \geq \psi^{-1}\left(\frac{m}{1-\psi(1)}\right) \geq \psi^{-1}(2m) \\
&\iff Cy \geq \psi^{-1}(2m)c\max_{i\leq m}\|\xi_i\|_\psi
\end{aligned}$$

Therefore, $\mathbb{E}\left[\psi\left(\frac{\max_{i\le m}|\xi_i|}{Cy}\right)\right] \le 1$ when $Cy \ge \psi^{-1}(2m)c\max_{i\le m}\|\xi_i\|_\psi$. Hence, the norm being the infimum of the previous expectation, we have that:

$$\begin{aligned}\left\|\max_{i\le m}|\xi_i|\right\| &\le \psi^{-1}(2m)c\max_{i\le m}\|\xi_i\|_\psi \\ &\le 2c\psi^{-1}(m)\max_{i\le m}\|\xi_i\|_\psi\end{aligned}$$

which follows again from convexity of ψ. ∎

Recall that the property of being sub-Gaussian is related to how one is able to control the even moments of a random variable. With the Orlicz norm, we are able to make a similar statement.

Remark 3.2. Consider the particular $\psi(x) = \exp\left\{|x|^2\right\} - 1$, and we assume $\|\xi\|_\psi = \sigma$. Then,

$$\mathbb{E}\left[\frac{\xi^{2p}}{\sigma^{2p}}\right] \le p! \cdot \mathbb{E}\left[\frac{\xi^2}{\sigma^2}\right] \le 2p!$$

Thus a bound on $\|\xi\|_\psi$ gives us a generous bound for higher moments of ξ.

4. Bernstein-Orlicz norm and beyond

Motivated by the study of empirical processes, authors of [5] and [7] sought to extend the idea of Orlicz norms beyond exponential types. As guessed, each of the norms mentioned above successfully captures their corresponding concentration inequality. The advantage of control via Orlicz-type norms over direct tail probabilities analyses and/or via conditional expectations (in particular, methods employing Efron-Stein type inequalities) for stochastic processes analyses is that the former tends to be simpler. We mention without proof that the analysis in the previous section can be done using a Bernstein-Orlicz type argument, simplifying several steps in the proof shown.

Definition 4.1. (Bernstein-Orlicz norm) Let $L > 0$ be given. We define:

$$\psi_L(z) := \exp\left\{\frac{\sqrt{1+2Lz}-1}{L}\right\}^2 - 1, \quad z \ge 0$$

It is not hard to see that ψ_L is an N-function. The (L-)Bernstein-Orlicz norm is the ψ-Orlicz norm given by $\psi = \psi_L$.

A particular characteristic of the Bernstein-Orlicz norm is that it interpolates sub-Gaussian and sub-Exponential tail behavior, governed by the constant L:

$$\psi_L(z) \approx \begin{cases}\exp\left\{|z|^2\right\} - 1 & \text{for } Lz \text{ small} \\ \exp\left\{2|z|\ /L\right\} - 1 & \text{for } Lz \text{ large}\end{cases}$$

We can already observe a concentration phenomenon for $\xi \in \mathcal{L}_\psi(\Omega)$:

Lemma 4.2. Let ξ as described, and $\tau := \|\xi\|_{\psi_L}$. We have that for all $t \ge 0$,

$$\mathbb{P}\left(|\xi| > \tau\left(\sqrt{t} + \frac{Lt}{2}\right)\right) \le 2\exp\left\{-t\right\}$$

Conversely, if the above is satisfied for some τ, then $\|\xi\|_{\psi_{\sqrt{3}L}} \le \sqrt{3}\tau$. Hence, concentration implies control on the Bernstein-Orlicz norm and vice versa.

From the above, we have the following extension to Bernstein's inequality, involving the Bernstein-Orlicz norm:

Theorem 4.3. (Bernstein and [5]) Let $\xi_1, \dots, \xi_n$ be independent random variables over $\mathbf{R}$ with zero mean. Suppose that for some σ, K, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|^m\right] \leq \frac{m!}{2}K^{m-2}\sigma^2, \quad m = 1, 2, \dots$$

Then, for all $t > 0$,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\xi_i\right| \geq \sigma\sqrt{2t} + \frac{Kt}{\sqrt{n}}\right) \leq 2\exp\{-t\}$$

Furthermore, for $L := \frac{\sqrt{6}K}{\sigma\sqrt{n}}$, we have

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\right\|_{\psi_L} \leq \sqrt{6}\sigma$$

Now, one can imagine choosing different N-functions could yield different known inequalities. For instance, letting $\psi_L(z) = \exp\left\{\frac{2}{L^2}((1+Lx)\log(1+Lx) - Lx)\right\} - 1$, one can observe that the Orlicz norm of ψ_L captures Bennett's inequality in a very similar fashion; we refer the reader to [7].

5. Final remarks

In fact, in this [7], the author recovers a number of such inequalities, notably Prokhorov's arcsinh exponential bound, and relates them back to the Orlicz-type norms on the underlying Orlicz space of random variables (of exponential type), comparing them to certain classical results from Talagrand and Ledoux. This motivates the continuation of research in 1. developing proof techniques for probabilistically bounding functions of interest in stochastic and empirical processes using Orlicz-type norms and 2. the study of Orlicz spaces and its properties itself, and even more generally, the study of K_σ-spaces. For the latter, authors of [4] have done so extensively, with applications to remarkable subspaces of these K_σ-spaces. Beyond concentration, it is worth exploring the case where Orlicz spaces coincide with Riesz spaces, the latter enjoying nice duality theories when viewed functiorially in the category of complete normed cones as detailed in[2], and study the implications of concentration phenomena while leveraging applied category theory.

Acknowledgments

I would like to send my gratitude to Prof. Lin for masterfully delivering us quality contents in her topics course on concentration phenomena. I would also like to thank Prof. Panangaden for his productive discussions and intuition, and Jordan Paillé for his insightful help throughout the course.

References

[1] Orlicz W. Birnbaum Z. Uber die verallgemeinerung des begriffes der zueinander konjugierten potenzen. *Studia Mathematica*, 3(1):1–67, 1931.

[2] Philippe Chaput, Vincent Danos, Prakash Panangaden, and Gordon Plotkin. Approximating markov processes by averaging. *J. ACM*, 61(1), January 2014.

[3] David Pollard. *MiniEmpirical*. http://www.stat.yale.edu/ pollard/Books/Mini/.

[4] Yu. V. Kozachenko V. V. Buldygin. *Metric Characterization of Random Variables and Random Processes (Translations of Mathematical Monographs)*. American Mathematical Society, 2000.

[5] Sara van de Geer and Johannes Lederer. The bernstein–orlicz norm and deviation inequalities. *Probability theory and related fields*, 157(1-2):225–250, 2013.

[6] Aad W Van Der Vaart and Jon A Wellner. Weak convergence and empirical processes. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.

[7] Jon A Wellner. The bennett-orlicz norm. *Sankhya A*, 79(2):355–383, 2017.