Printed: 6 July 2025 at 21:12

Chapter 5

Orlicz spaces

Orlicz::Orlicz

SECTION 5.1 *introduces* $\mathcal{L}^\Psi$ *spaces, generalizations of the usual* $\mathcal{L}^p$ *spaces, which have proved useful in the study of sample paths of stochastic processes and in the theory of empirical processes.*

SECTION 5.2 *describes some useful properties of* ORLICZ *functions* Ψ *on* $\mathbb{R}^+$*, the convex functions that are used to define the* $\mathcal{L}^\Psi$ *spaces. They are shown to have left- and right-derivatives everywhere, which leads to useful integral representation.*

SECTION 5.3 *describes a set of* ORLICZ *functions that grow much faster than any pth power, a property that leads to several useful inequalities.*

SECTION 5.4 *presents a few simple methods for proving maximal inequalities using* ORLICZ *norms. These inequalities are the basis for some chaining bounds that were popular in empirical process theory.*

SECTION 5.5 *shows how the conjugate of* Φ *of an* ORLICZ *function* Ψ *is related to the* YOUNG *inequality, an analog of the* HÖLDER *inequality. The* ORLICZ *space* $\mathcal{L}^\Phi$ *is then identified with a subset of the space of continuous linear functionals on* $\mathcal{L}^\Psi$*.*

*SECTION 5.6 *describes some of the differences and similarities between* $\mathcal{L}^p$ *and* $\mathcal{L}^\Psi$ *if* Ψ *grows rapidly enough.*

5.1 What is an orlicz space?

Orlicz::S:definition

During the 1980's, bounds involving norms on ORLICZ spaces of functions became very popular in the empirical process literature, in part I suspect because chaining arguments (see Chapter 10) are cleaner with norms than with tail probabilities. More recently, the literature seems to have turned back to working with tail probabilities.

An ORLICZ space is a most useful generalization of the concept of an $\mathcal{L}^p$ space. It replaces the pth power function by a general ORLICZ function, that is, a convex, increasing function $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\Psi(0) = 0$ and $\Psi(x) \to \infty$ as $x \to \infty$. The concept applies with any measure but I have used it mostly for probability measures. The special case where $\Psi_2(x) = \exp(x^2) - 1$ is particularly useful in the study of subgaussianity.

Remark. Here and subsequently in this Chapter "increasing" is treated as a synonym for the clumsier "nondecreasing". It allows an ORLICZ function Ψ to be zero on an interval $[0, x_0]$ but then convexity forces Ψ to be strictly increasing on $[x_0, \infty)$.

Henceforth $\mathcal{Y}$ will denote the set of all ORLICZ functions on $\mathbb{R}^+$, with $\mathcal{Y}_\infty$ the subset where $\Psi(x)/x \to \infty$ as $x \to \infty$.

Orlicz::Luxemburg.norm <1> **Definition.** *For a measure space $(\mathcal{X}, \mathcal{A}, \mu)$ and an ORLICZ function Ψ define $\mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ to be the set of all $\mathcal{A}$-measurable real-valued functions f on $\mathcal{X}$ for which $\mu\Psi(|f|/c_0) < \infty$ for at least one positive constant c_0. For each such f the ORLICZ **norm** $\|f\|_\Psi$ (or Ψ-norm) is defined as*

$$\|f\|_\Psi := \inf\{c > 0 : \mu\Psi(|f(x)|/c) \le 1\}.$$

□

Remark. Some authors define $\|f\|_\Psi$ for every measurable f, interpreting the infimum as $+\infty$ when $\mu\Psi(|f(x)|/c) = \infty$ for each $c > 0$. The space $\mathcal{L}^\Psi$ then consists of all f for which $\|f\|_\Psi < \infty$. Actually, it was Luxemburg (1955) who first introduced the norm $\|f\|_\Psi$; Orlicz (1932, 1936) had worked with another norm, which I will temporarily denote by $\|\cdot\|_\bullet$ to avoid confusion. However the name Orlicz seems to be firmly attached to $\|\cdot\|_\Psi$ in a lot of the literature. As shown in Problem [17], the two norms are equivalent, in the sense that $\|f\|_\Psi \le \|f\|_\bullet \le 2\|f\|_\Psi$ for all f in $\mathcal{L}^\Psi$. I like to use the letter $\mathcal{Y}$ as an indirect nod to the work of William Henry Young. See Section 5.5 for more about that remarkable mathematician.

If $\mu\Psi(|f|/c_0) < \infty$, a dominated convergence argument shows that $\mu\Psi(|f|/(c_0 + n)) \to 0$ as $n \to \infty$. The set $\{c > 0 : \mu(|f|/c) \le 1\}$ is either $(0, \infty)$, in which case $\|f\|_\Psi = 0$ and $f = 0$ ae$[\mu]$, or an interval of the form $[c_1, \infty)$ in which case $\|f\|_\Psi = c_1$ (Problem [1]).

For the special case where $\Psi(x) = x^p$, the inequality $\mu\Psi(|f|/c)) \le 1$ is equivalent to $\mu|f|^p \le c^p$, which ensures that the ORLICZ norm agrees with the usual $\|f\|_p$ norm (for $p \ge 1$) and the ORLICZ space is the same as the usual $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu)$ space.

For the general case it is important to keep the constant c inside that Ψ function. For example, here is how to prove that $\|f + g\|_\Psi \le \|f\|_\psi + \|g\|_\Psi$ for $f, g \in \mathcal{L}^\Psi$: for each $c > \|f\|_\Psi$ and $d > \|g\|_\Psi$, convexity and monotonicity of Ψ give

$$\Psi\left(\frac{|f+g|}{c+d}\right) \le \Psi\left(\frac{c|f|/c}{c+d} + \frac{d|f|/d}{c+d}\right) \le \frac{c}{c+d}\Psi(|f|/c) + \frac{d}{c+d}\Psi(|g|/d).$$

Integrate both sides with respect to μ, using the fact that $\mu\Psi(|f|/c) \le 1$ and $\mu\Psi(|g|/d) \le 1$ to deduce that

$$\mu\Psi\left(\frac{|f+g|}{c+d}\right) \le 1 \qquad \text{if } c > \|f\|_\Psi \text{ and } d > \|g\|_\Psi.$$

It follows that $\|f + g\|_\Psi \le c+d$. Complete the argument by letting c decrease to $\|f\|_\Psi$ and d decrease to $\|g\|_\Psi$. (If you believe the result from Problem [1] you could also take $c = \|f\|_\Psi$ and $d = \|g\|_\Psi$, except in some trivial cases.)

Remark. As usual $\|\cdot\|_\Psi$ fails to be a norm on $\mathcal{L}^\Psi$ only because $\|f\|_\Psi = 0$ if and only if $\mu\{x : f(x) \neq 0\} = 0$. The usual trick of working with the space $L^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ of μ-equivalence classes makes $\|\cdot\|_\Psi$ a true norm and L^Ψ a BANACH space.

The most important ORLICZ functions for my purposes are the power functions, $\Psi(x) = x^p$ for a fixed $p \geq 1$, and the exponential power functions $\Psi_\alpha(x) := \exp(x^\alpha) - 1$ for $\alpha \geq 1$. The function Ψ_2 corresponds to subgaussian, and Ψ_1 to subexponential, tail behaviour.

Remark. It is also possible (Problem [9]) to define an ORLICZ function Ψ_p, for $0 < p < 1$, that behaves like $\exp(x^p)$ far enough from the origin.

When seeking to bound $\|f\|_\Psi$ for some f in $\mathcal{L}^\Psi$ I often find it easier to first obtain a constant c_0 for which $\mu\Psi(|f|/c_0) \leq K_0$, where K_0 is a constant larger than 1. As the next Lemma shows, a simple convexity argument turns such an inequality into a bound on $\|f\|_\Psi$.

Orlicz::Psi.bound <2> **Lemma.** *Suppose Ψ is an ORLICZ function. If $\mu\Psi(|f|/c_0) \leq K_0$ for some finite constants c_0 and $K_0 > 1$ then $\|f\|_\Psi \leq c_0 K_0$.*

Proof. For each θ in $[0, 1]$ convexity of Ψ gives

$$\mu\Psi\left(\frac{\theta|f|}{c_0}\right) \leq \theta\mu\Psi\left(\frac{|f|}{c_0}\right) + (1-\theta)\Psi(0) \leq \theta K_0.$$

□ The choice $\theta = 1/K_0$ makes the last bound equal to 1.

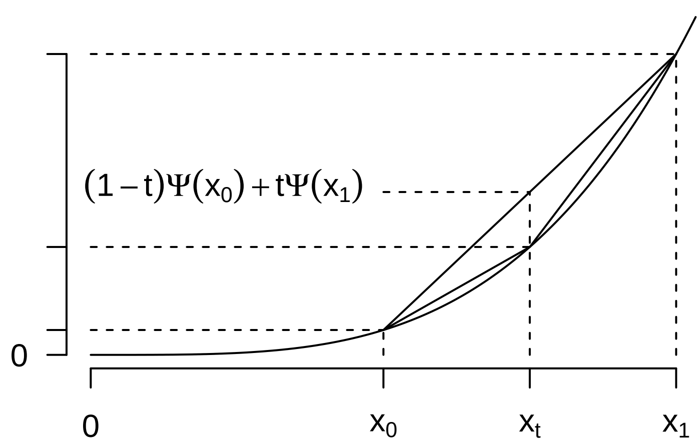
5.2 Facts about orlicz functions

Orlicz::S:convex.facts

Suppose $\Psi \in \mathcal{Y}$. Define the ***slope function***

$$S(x, y) := \frac{\Psi(y) - \Psi(x)}{y - x} \qquad \text{for } 0 \leq x < y.$$

The slope is always nonnegative because Ψ is increasing.



Most of the useful properties of ORLICZ functions follow from a pair of inequalities for the slope function. For given values $0 \le x_0 < x_1 = x_0 + D$ and $0 < t < 1$ define $x_t := (1-t)x_0 + tx_1$. Note that $x_t - x_0 = tD$ and $x_1 - x_t = (1-t)D$. Convexity of Ψ gives

$$\Psi(x_t) \le (1-t)\Psi(x_0) + t\Psi(x_1),$$

which implies

$$\begin{aligned} S(x_0, x_t) &= \frac{\Psi(x_t) - \Psi(x_0)}{x_t - x_0} \le \frac{t\Psi(x_1) - t\Psi(x_0)}{tD} = S(x_0, x_1), \\ S(x_t, x_1) &\ge \frac{(1-t)\Psi(x_1) - (1-t)\Psi(x_0)}{(1-t)D} = S(x_0, x_1). \end{aligned}$$

Re-expressed more suggestively, (first with $x = x_0$ and $y = x_t$ and $y' = x_1$ then with $x = x_0$ and $x' = x_t$ and $y = x_1$) the inequalities become

$$\begin{aligned} &S(x,y) \le S(x,y') &&\text{if } 0 \le x < y < y' \\ &S(x,y) \le S(x',y) &&\text{if } 0 \le x < x' < y. \end{aligned}$$

That is,

\E@ slope.inc <3>

the function $y \mapsto S(x,y)$ is increasing for each fixed x;
the function $x \mapsto S(x,y)$ is increasing for each fixed y.

Remark. The preceding argument used only the convexity of Ψ. Properties <3> are general facts about convex functions on $\mathbb{R}$ or a subinterval thereof. If Ψ is also increasing then $S \ge 0$.

Orlicz::Young.derivs <4>

Theorem. *If $\Psi \in \mathfrak{Y}$ then*

(i) At each $x \ge 0$ the right-derivative

$$\dot{\Psi}_R(x) = \lim_{t \searrow 0} \left(\Psi(x+t) - \Psi(x)\right)/t$$

exists. The function $\dot{\Psi}_R$ is nonnegative, increasing, and continuous from the right.

(ii) At each $x > 0$ the left-derivative

$$\dot{\Psi}_L(x) = \lim_{w \nearrow 0} \left(\Psi(x) - \Psi(w)\right)/(x-w)$$

exists. The function $\dot{\Psi}_L$ is nonnegative, increasing, and continuous from the left.

(iii) The function Ψ is continuous on $\mathbb{R}^+$.

(iv) If $0 \le x < y$ then $\dot{\Psi}_R(x) \le \dot{\Psi}_L(y) \le \dot{\Psi}_R(y)$ and

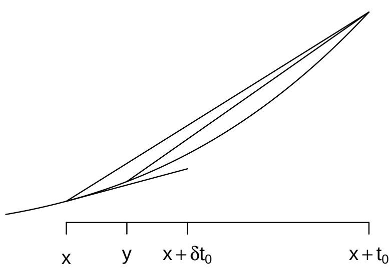
$$(y-x)\dot{\Psi}_R(x) \le \Psi(y) - \Psi(x) \le (y-x)\dot{\Psi}_L(y).$$

(v) $\Psi(x) = \int_0^x \dot\Psi_R(t)\,dt = \int_0^x \dot\Psi_L(t)\,dt$ *for each* $x > 0$. *Consequently, the set* $\{t > 0 : \dot\Psi_R(t) > \dot\Psi_L(t)\}$ *has zero* LEBESGUE *measure.* □

Proof.
For (i): The function $t \mapsto S(x, x+t)$ is increasing and nonnegative. Thus $S(x, x+t)$ decreases to a finite limit $\dot\Psi_R(x) \geq 0$ as $t \searrow 0$.

If $x < x'$ and $t > 0$ then $S(x, x+t) \leq S(x, x'+t) \leq S(x', x'+t)$ Let t decrease to 0 to deduce that $\dot\Psi_R(x) \leq \dot\Psi_R(x')$.

For right-continuity at a given x, we need to show that $\dot\Psi_R(y) \approx \dot\Psi_R(x)$ if y lies in a small enough interval to the right of x. Given a small $\epsilon > 0$ first choose a $t_0 > 0$ for which $\dot\Psi_R(x) \leq S(x, x+t_0) \leq \dot\Psi_R(x) + \epsilon$ then consider $\delta > 0$ and y for which $x < y < x + \delta t_0 < x + t_0$, so that



$$\begin{aligned}
\dot\Psi_R(x) \leq \dot\Psi_R(y) &= \lim_{t\searrow 0} S(y, y+t) \qquad \text{decreasing limit} \\
&\leq S(y, x+t_0) = \frac{\Psi(x+t_0) - \Psi(y)}{x + t_0 - y} \\
&\leq \frac{\Psi(x+t_0) - \Psi(x)}{x + t_0 - (x + \delta t_0)} = \frac{S(x, x+t_0)}{1-\delta} \\
&\leq \frac{(1-\delta)\dot\Psi_R(x) + \delta\dot\Psi_R(x) + \epsilon}{1-\delta} \\
&\leq \dot\Psi_R(x) + 2\epsilon \qquad \text{if } \delta \text{ is small enough.}
\end{aligned}$$

For (ii): For $0 \leq w < x$ the function $\mapsto S(w, x)$ is increasing and nonnegative. Thus $S(w, x)$ increases to a limit $\dot\Psi_L(x) \geq 0$ as $w \nearrow 0$. The limit is finite because

$$\begin{aligned}
S(w, x) &\leq S(w, x+t) \qquad \text{for each } t > 0 \\
&\leq S(x, x+t) \searrow \dot\Psi_R(x) \qquad \text{as } t \searrow 0,
\end{aligned}$$

showing that $\dot\Psi_L(x) \leq \dot\Psi_R(x) < \infty$. The rest of the argument is analogous to the argument for (i), except that approximations are made from below instead of from above.
For (iii): Existence of left and right derivatives implies continuity.
For (iv): The inequality $\dot\Psi_L(y) \leq \dot\Psi_R(y)$ was already established in the proof of (ii). If $x < y$ then

$$\dot\Psi_R(x) = \lim_{w\searrow x} S(x, w) \leq S(x, y) \leq \lim_{w\nearrow y} S(w, y) = \dot\Psi_L(y).$$

Multiply through by $y - x$ to get the other inequality.
For (v): For a positive integer n define $\delta = x/n$ and $x_i = i\delta$ for $i = 0, 1, 2, \ldots$.

Then we have

$$\begin{aligned}\Psi(x) &= \sum\nolimits_{i=1}^{n}\Big(\Psi(x_i)-\Psi(x_{i-1})\Big)\\ &\ge \sum\nolimits_{i=1}^{n}\delta\dot\Psi_R(x_{i-1}) \qquad \text{by (iv)}\\ &\ge \delta\dot\Psi_R(0)+\sum\nolimits_{i=2}^{n}\int_{x_{i-2}}^{x_{i-1}}\dot\Psi_R(t)\,dt = \delta\dot\Psi_R(0)+\int_0^{x_{n-1}}\dot\Psi_R(t)\,dt\\ &\to \int_0^x \dot\Psi_R(t)\,dt \qquad \text{as } n\to\infty, \text{ by dominated donvergence.}\end{aligned}$$

Similarly,

$$\Psi(x) \le \sum\nolimits_{i=1}^{n}\delta\dot\Psi_R(x_i) \le \sum\nolimits_{i=1}^{n}\int_{x_i}^{x_{i+1}}\dot\Psi_R(t)\,dt = \int_{x_2}^{x_{n+1}}\dot\Psi_R(t)\,dt,$$

which also converges to $\int_0^x \dot\Psi_R(t)\,dt$ as $n\to\infty$. Together the upper and lower bounds establish the first asserted integral representation. The argument for $\dot\Psi_L$ is similar.

By subtraction, $0=\int_0^x\Big(\dot\Psi_R(t)-\dot\Psi_L(t)\Big)\,dt=0$. By (iv), the integrand is nonegative. Hence the integrand must equal 0 except on a set of zero LEBESGUE measure. □

Remarks.

(i) Problem [4] strengthens part (v) by showing that the both the nondecreasing functions $\dot\Psi_L$ and $\dot\Psi_R$ can have, at worst, only countably many jumps. Thus the set $\{t>0:\dot\Psi_R(t)>\dot\Psi_L(t)\}$ is not just LEBESGUE negligible but is also, at worst, countably infinite.

(ii) Problem [5] show that every Ψ on $\mathbb{R}^+$ defined by $\Psi(x)=\int_0^x\psi(t)\,dt$, with ψ nonnegative and nondecreasing, is an ORLICZ function. Moreover, such a ψ must satisfy the inequalities

$$\dot\Psi_L(t)\le\psi(t)\le\dot\Psi_R(t) \qquad \text{for all } t>0.$$

Thus ψ can differ from $\dot\Psi_L$ or $\dot\Psi_R$ only at the (at worst countable) set of t's at which $\dot\Psi_R(t)\ne\dot\Psi_L(t)$. In this sense, the function Ψ essentially determines ψ uniquely.

(iii) The condition $\Psi(x)/x\to\infty$ as $x\to\infty$ that is also required for ORLICZ functions in $\mathcal{Y}_\infty$ can also be written as $\psi(t)\to\infty$ as $t\to\infty$ because

$$\psi(x)\ge x^{-1}\Psi(x)\ge x^{-1}\int_{x_0}^{x}\psi(t)\,dt\ge x^{-1}(x-x_0)\psi(x_0)$$

for $x\ge x_0$, for arbitrarily large x_0.

5.3 Orlicz functions that grow exponentially fast

Orlicz::S:expOrlicz

To avoid a lot of nitpicking, each ORLICZ function Ψ in this Section should be assumed to have $\{\Psi = 0\} = \{0\}$.

For some purposes, it helps to have $\Psi(x)/x^p \to \infty$ as $x \to \infty$, for each finite p. For example, Ledoux and Talagrand (1991, page 310) required

\E@ LTinv <5>
$$\Psi^{-1}(ab) \le C_0 \left(\Psi^{-1}(a) + \Psi^{-1}(b)\right) \qquad \text{if } \min(a,b) \ge a_0,$$

for constants C_0 and $a_0 > 0$. Equivalently,

\E@ LTgrowth <6>
$$\Psi(x)\Psi(y) \le \Psi\left(C_0(x+y)\right) \qquad \text{if } \min(x,y) \ge x_0.$$

Write $\mathcal{Y}_{\text{LT}}$ for the set of all ORLICZ functions that satisfy this condition for each strictly positive x_0, with C_0 allowed to depend on x_0. (Often such a bound also holds for x_0 equal to 0.)

The main result in this Section provides a simple sufficient condition for <6>.

Van der Vaart and Wellner (1996, Lemma 2.2.2) imposed a similar-looking condition to control the Ψ-norm of a maximum:

\E@ VWgrowth <7>
$$\Psi(x)\Psi(y) \le \Psi\left(C_1 xy\right) \qquad \text{if } \min(x,y) \ge x_1.$$

for constants C_1 and $x_1 > 0$. Write $\mathcal{Y}_{\text{VW}}$ for the set of all ORLICZ functions that satisfy this condition for each strictly positive x_1, with C_1 depending on x_1.

It is easy to show for the ORLICZ function $\Psi(x) = x^p$, with $p \ge 1$, that $\Psi \in \mathcal{Y}_{\text{VW}}$ but $\Psi \notin \mathcal{Y}_{\text{LT}}$: consider the case where $x = y$. Thus <7> does not require exponential growth, although its most interesting uses do involve rapidly growing ORLICZ functions—see Example <13>. However, the $\mathcal{Y}_{\text{LT}}$ condition implies $\Psi(x)^2 \le \Psi(2C_0 x)$ for $x \ge x_0 > 0$, an exponential growth rate. (It also forces $2C_0 \ge 1$ because $\Psi(x) \le \Psi(x)^2$ when $\Psi(x) \ge 1$.)

For the purposes of deriving sufficent conditions for rapid growth it is cleaner to work with $g(x) := \log(1 + \Psi(x))$, a nonnegative, increasing function with $g(0) = 0$. The g function inherits continuity and existence of left and right derivatives from Ψ. Conversely, a quick way to generate ORLICZ functions is to start from a twice-differentiable, increasing function g with $g(0) = 0$. If we define $\Psi(x) = e^{g(x)} - 1$ then $\ddot{\Psi}(x) = \left(\dot{g}(x)^2 + \ddot{g}(x)\right)\Psi(x)$, which shows that $\Psi \in \mathcal{Y}$ if $\dot{g}(x)^2 + \ddot{g}(x) \ge 0$ for all $x > 0$.

yyexp <8> **Definition.** *Define $\mathcal{Y}_{\exp}$ to be the set of ORLICZ functions $\Psi = e^g - 1$ for which there exists a constant C_g such that $g(x) + g(y) \le g\left(C_g \max(x,y)\right)$ for all x, y in $\mathbb{R}^+$.*

As you will see in later Chapters, the $\mathcal{Y}_{\exp}$ property turns out to be a very natural assumption for a lot of theory developed in the past few decades.

Orlicz::g.growth <9> **Lemma.** *Suppose $\Psi = e^g - 1$ is an ORLICZ function with $\{\Psi = 0\} = \{0\}$.*

(i) *If there exist constants K_0 and $y_0 > 0$ such that $2g(x) \le g(K_0 x)$ whenever $x \ge y_0$ then Ψ is a member of $\mathcal{Y}_{\text{exp}}$.*

(ii) *If $\Psi \in \mathcal{Y}_{\text{exp}}$ then $\Psi(x)\Psi(y) \le \Psi(C_g \max(x,y))$ for all x, y in $\mathbb{R}^+$ and (equivalently) $\Psi^{-1}(ab) \le C_g \max\left(\Psi^{-1}(a), \Psi^{-1}(b)\right)$ for all a, b in $\mathbb{R}^+$. Consequently, $\Psi \in \mathcal{Y}_{\text{LT}}$.*

(iii) $\mathcal{Y}_{\text{exp}} \subset \mathcal{Y}_{\text{VW}}$

(iv) *If $\Psi \in \mathcal{Y}_{\text{exp}}$ then there is a positive x_0 such that $g(x) \ge (x/C_g x_0)^\alpha g(x_0)$ for $x \ge x_0$, where $\alpha := 1/\log_2 C_g$.*

Proof. For (i), first extend the upper bound on $2g(x)$ to cover a neighborhood of the origin. Define $x_1 := \Psi^{-1}(1)/4$. Then, from the facts that $y \mapsto \Psi(y)/y$ is an increasing function on $(0,\infty)$ and $y/(1+y) \le \log(1+y) \le y$ for $y > 0$, we have

$$g(4x) \ge \frac{\Psi(4x)}{1+\Psi(4x)} \ge \frac{4\Psi(x)}{1+\Psi(4x_1)} = 2\Psi(x) \ge 2g(x) \qquad \text{for } 0 \le x \le x_1.$$

Extend to cover the range $[x_1, y_0]$. Define K_1 by $2g(y_0) = g(K_1 x_1)$, so that

$$\max_{x_1 \le x \le y_0} 2g(x)/g(K_1 x) \le 2g(y_0)/g(K_1 x_1) = 1.$$

With $C_g := \max(K_0, 4, K_1)$ we then have $2g(x) \le g(C_g x)$ for all $x$in $\mathbb{R}^+$, implying

\E@ g.claim <10> $$g(x) + g(y) \le 2g(x) \le g\left(C_g \max(x,y)\right) \qquad \text{if } 0 \le y \le x.$$

For (ii) and (iii), first note that

$$\begin{aligned}\Psi(x)\Psi(y) &= e^{g(x)+g(y)} - e^{g(x)} - e^{g(y)} + 1 \\ &\le e^{g(x)+g(y)} - 1 \le e^{g(C_g \max(x,y))} - 1 = \Psi\left(C_g \max(x,y)\right).\end{aligned}$$

For Ψ in $\mathcal{Y}_{\text{LT}}$ use $\max(x,y) \le x+y$ and for Ψ in $\mathcal{Y}_{\text{VW}}$ use $\max(x,y) \le xy/x_1$ when $\min(x,y) \ge x_1$.

For (iv), suppose $C_g^{k-1}x_0 \le x < C_\Psi^k x_0$ with $k \in \mathbb{N}$. By induction, $2^p g(x_0) \le g(C_g^p x_0)$ for each p in $\mathbb{N}_0$. Thus

$$g(x) \ge g(C_g^{k-1}x_0) \ge 2^{k-1}g(x_0) = \left(C_g^\alpha\right)^{(k-1)} g(x_0).$$

□ By definition of k we have $C_g^{k-1} > x/(C_\Psi x_0)$.

Orlicz::e^g <11> **Example.** The constant C_g calculated for the previous Lemma might be unnecessarily large. Consider the case where $g_\alpha(x) := x^\alpha$ for some $\alpha \ge 1$. By direct calculation, $\dot{g}(x)^2 + \ddot{g}(x) = \alpha x^{2\alpha-1}\left(\alpha x + (\alpha - 1)\right) \ge 0$ and $2x^\alpha \le (Kx)^\alpha$ if $K = 2^{1/\alpha}$. Thus $\Psi_\alpha(x) := \exp(x^\alpha) - 1$ is an ORLICZ function in $\mathcal{Y}_{\text{exp}}$ if $1 \le \alpha < \infty$.

The simple form of the g_α function allows us to derive a result slightly sharper than Lemma <9>. As before, for all x, y in $\mathbb{R}^+$,

$$\Psi_\alpha(x)\Psi_\alpha(y) \le \exp\left(g_\alpha(x) + g_\alpha(y)\right) - 1.$$

If $x + y > 0$ and $x/(x+y) = r$ then

$$\frac{g_\alpha(x) + g_\alpha(y)}{g_\alpha(x+y)} = R_\alpha(r) := r^\alpha + (1-r)^\alpha \qquad \text{for } 0 \le r \le 1.$$

The function R_α is convex on $[0, 1]$, with maximum value 1 achieved at the endpoints $\{0, 1\}$ and minimum value $(1/2)^{\alpha-1}$ at $r = 1/2$, which implies

$$g_\alpha(x) + g_\alpha(y) \le g_a(x+y) \qquad \text{for all } x, y \text{ in } \mathbb{R}^+.$$

Consequently $\Psi_\alpha \in \mathfrak{Y}_{\text{LT}}$. We also have

$$\Psi_\alpha(x)\Psi_\alpha(y) \le e^{g_\alpha(x+y)} - 1 = \Psi_\alpha(x+y) \qquad \text{for all } x, y \in \mathbb{R}^+$$

and

$$\Psi_\alpha^{-1}(ab) \le \Psi_\alpha^{-1}(a) + \Psi_\alpha^{-1}(b) \qquad \text{for all } a, b \in \mathbb{R}^+,$$

□ which is much cleaner than <5>.

Problem [9] extends the Example to values of α in $(0, 1)$: there is a Ψ_α in $\mathfrak{Y}_{\exp}$ for which $\Psi_\alpha(x) = \exp(x^\alpha) - 1$ when x is large enough.

5.4 Norms of maxima of finitely many variables

Orlicz::S:OrliczMax

As noted near the start of the Chapter, my interest in ORLICZ norms was first aroused by their use with the chaining method for constructing simple uniform approximations to stochastic processes $\{X_t : t \in T\}$. That method requires some sort of probabilistic control over the increments $X_t - X_s$, expressible in terms of some distance function on T. When that distance takes the form of an ORLICZ norm $\|X_t - X_s\|_\Psi$ and when the probabilistic approximations are expressed using some other ORLICZ norm, the chaining method takes a particularly simple form. The key idea is to control the maximum of finite sets of random variables. The following Theorem collects together a few techniques that I have extracted from various chaining proofs in the literature. The results are stated for probability measures.

Orlicz::Orlicz.maximal *<12>* **Theorem.** *Let $X_1, \dots, X_N$ be nonnegative random variables, **not necessarily independent**, but all defined on the same probability space. Define $M := \max_{i \le N} X_i$. Suppose $\max_{i \le N} \|X_i\|_\Psi \le 1$ for some ORLICZ function Ψ.*

(i) $\mathbb{P}M \le \Psi^{-1}(N)$.

(ii) $\mathbb{P}_B M \le \Psi^{-1}(N/\mathbb{P}B)$ where $\mathbb{P}_B$ denotes conditional expectation given an event B with $\mathbb{P}B > 0$.

(iii) If $\Psi \in \mathfrak{Y}_{\text{VW}}$ then $\|M\|_\Psi \le 2C\Psi^{-1}(N)$ for some constant C.

Proof. The assumption that $\max_{i \le N} \|X_i\|_\Psi \le 1$ implies $\mathbb{P}\Psi(X_i) \le 1$ for each i.

For (i):
Even though the assertion is a special case of assertion (ii), I think it helps build intuition to see this proof first. Jensen's inequality then non-negativity and monotonicity of Ψ imply

$$\Psi\left(\mathbb{P}M\right) \le \mathbb{P}\max_{i\le N}\Psi(X_i) \le \sum\nolimits_{i\le N}\mathbb{P}\Psi(X_i) \le N.$$

For (ii):
Partition B into disjoint subsets B_i such that $M = X_i$ on B_i. Without loss of generality assume $\mathbb{P}B_i > 0$ for each i. (Alternatively, just discard those B_i with zero probability.) By definition,

$$\mathbb{P}_B M = \mathbb{P}_B\left(\sum\nolimits_i B_i X_i\right) = \sum\nolimits_i \frac{\mathbb{P}B_i}{\mathbb{P}B}\mathbb{P}_{B_i}X_i.$$

By Jensen's inequality,

$$\Psi(\mathbb{P}_{B_i}X_i) \le \mathbb{P}_{B_i}\Psi(X_i) = \frac{\mathbb{P}B_i\Psi(X_i)}{\mathbb{P}B_i} \le \frac{\mathbb{P}\Psi(X_i)}{\mathbb{P}B_i} \le \frac{1}{\mathbb{P}B_i}.$$

Thus

$$\begin{aligned}\Psi\left(\mathbb{P}_B M\right) &= \Psi\left(\sum\nolimits_i \frac{\mathbb{P}B_i}{\mathbb{P}B}\mathbb{P}_{B_i}X_i\right)\\ &\le \sum\nolimits_i \frac{\mathbb{P}B_i}{\mathbb{P}B}\Psi\left(\mathbb{P}_{B_i}X_i\right) \qquad \text{by convexity of } \Psi\\ &\le \sum\nolimits_i \frac{\mathbb{P}B_i}{\mathbb{P}B}\frac{1}{\mathbb{P}B_i} = \frac{N}{\mathbb{P}B}.\end{aligned}$$

Perhaps it would be better to write the last equality as an inequality, to cover the case where some $\mathbb{P}B_i$ are zero.

For (iii):
Choose $x_1 = \min(1, \Psi^{-1}(1))$, so that $\beta := \Psi^{-1}(N) \ge x_1$. On the set $\{M \ge \beta C_1\}$ we then have

$$N\Psi(M/\beta C_1) = \Psi(\beta)\Psi(M/\beta C_1) \le \Psi(M) \le \sum\nolimits_i \Psi(X_i).$$

Thus $\Psi(M/\beta C_1) \le \Psi(1) + N^{-1}\sum_i \Psi(X_i)$ for all values of M. Take expected values to deduce that $\mathbb{P}\Psi(M/\beta C_1) \le \Psi(1) + 1$, which, by Lemma <2>, implies $\|M\|_\Psi \le C_1\left(2\Psi(1) + 1\right)\beta$. □

See Section 10.5 for an example that demonstrates the surprising power of inequality (ii) from the Theorem.

Orlicz::pnorm.max <13> **Example.** Suppose $\{Z(t) : t \in T\}$ is a stochastic process indexed by a set equipped with a metric for which $\|Z(s) - Z(t)\|_{\Psi_2} \le d(s,t)$ for $s, t \in T$.

Lemma <9> shows that each of the ORLICZ functions $\Psi_p(x) = \exp(x^p) - 1$, with $p \ge 1$ satisfy the assumption of inequality (iii). In particular, it holds for Ψ_2. If $p \ge 1$ and $k = \lceil p/2 \rceil$ then, for a constant C_p depending on p,

$$\|X\|_p \le \|X\|_{2k} \le C_p \|X\|_{\Psi_2} \qquad \text{for each random variable } X$$

by virtue of the fact that $\mathbb{P}|X/c|^{2k} \le k!\mathbb{P}\Psi_2(|X|/c)$. The constant $(k!)^{1/2k}$ is of order $O(\sqrt{p})$.

In consequence, if s_i, t_i are pairs of points with $d(s_i, t_i) > 0$ then

$$\left\| \max_{i \le N} \frac{|X(s_i) - X(t_i)|}{d(s_i, t_i)} \right\|_p \le K_p \sqrt{1 + \log N}.$$

This bound can be used to derive the inequalities from Pollard (1989, Section 3) that became the main technical tool for the cube-root asymptotic theory developed by Kim and Pollard (1990). □

5.5 Conjugacy and the YOUNG inequality

Orlicz::S:conjugate

This Section explains a connection between ORLICZ functions and the conjugates from Chapter 2. Before writing this Chapter, I regarded conjugacy as a slightly exotic topic that would not not really be needed in the rest of the book. Subsequently I came to think that this convexity idea is actually closely related to several constructions that appear in later Chapters. For example, the splitting idea of Fernique for deriving maximal inequalities, which is described and extended in Section 11.5, seems to be closely related to the YOUNG inequality.

As noted in Section 2.4, it is convenient to define the conjugate of an ORLICZ function Ψ by first extending its domain of definition to the whole real line: define $\Psi(x) = +\infty$ for $x < 0$. The conjugate Ψ^* is defined by

$$\Psi^*(y) := \sup\nolimits_{x \in \mathbb{R}} F(x, y) \qquad \text{where } F(x, y) := yx - \Psi(x) \text{ for } x, y \in \mathbb{R}.$$

Write Λ for the restriction of Ψ^* to $\mathbb{R}^+$. The main task in this Section is to find conditions to ensure that Λ is also an ORLICZ function and then to show how Ψ and Λ fit together to produce a result due to Young (1912).

The choice $x = 0$ gives the value $F(0, y) = 0$, and the choice $x < 0$ gives the value $F(x, y) = -\infty$. It follows that

\E@ Psi* <14>

$$\Lambda(y) = \sup\nolimits_{x \ge 0} \Big(yx - \Psi(x) \Big) \qquad \text{for } y \ge 0.$$

Temporarily write $\psi(t)$ for $\dot{\Psi}_R(t)$. By Theorem <4> we have $\Psi(x) = \int_0^x \psi(t)\,dt$ for all $x \ge 0$. If $c := \sup_t \psi(t)$ were finite then we would have $\Psi(x) \le cx$ for all x, implying $\Lambda(y) = +\infty$ for $y > c$. As we want Λ to be an ORLICZ function we need $\psi(t) \to \infty$ as $t \to \infty$. Equivalently, we need $\Psi(x)/x \to \infty$ as $x \to \infty$, that is, $\Psi \in \mathfrak{Y}_\infty$.

Orlicz::Lam.properties <15> **Theorem.** *If* $\Psi \in \mathcal{Y}_\infty$ *then* $\Lambda \in \mathcal{Y}_\infty$ *with* $\Lambda(y) = \int_0^\infty (y - \dot{\Psi}_R(t))^+ \,dt$.

Proof. As above, write $\psi(t)$ for $\dot{\Psi}_R(t)$ and define $q(y) := \inf\{t \geq 0 : \psi(t) \geq y\}$, with $q(y) = 0$ if $\psi(0) > y$. The assumption that Ψ belongs to $\mathcal{Y}_\infty$ ensures that $q(y) \to \infty$ as $y \to \infty$.

For $x \geq 0$ and $y \geq 0$ we have

$$F(x, y) = xy - \Psi(x) = \int_0^x y - \psi(t)\,dt.$$

By definition, $y - \psi(t) > 0$ if $t < q(y)$ and $y - \psi(t) \leq 0$ if $t > q(y)$. Thus $x \mapsto F(x, y)$ is increasing while $0 \leq x < q(y)$ and nonincreasing for $x > q(y)$. (There is actually be a flat spot if $\{t : \psi(t) = y\}$ is a nondegenerate interval.) It follows that

$$\Lambda(y) = \sup_x F(x, y) = \int_0^{q(y)} y - \psi(t)\,dt = \int_0^\infty (y - \psi(t))^+ \,dt.$$

Also, for $y > 0$ we have

$$\Lambda(y)/y = \sup_x F(x, y) = \int_0^\infty (1 - \psi(t)/y)^+ \,dt.$$

At each fixed t the integrand increases monotonely to 1 as $y \uparrow \infty$. By monotone convergence it follows that $\Lambda(y)/y \uparrow \infty$ as $y \uparrow \infty$. □

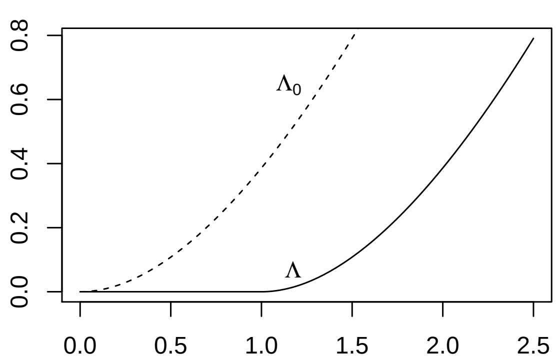
Orlicz::dual <16> **Example.** Everything is much simpler when the ORLICZ function Ψ is everywhere differentiable. For example, if $\Psi(x) = x^p/p$ with $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$ then the maximizing x in the definition of $\Lambda(y)$ equals $y^{q/p}$, so that $\Lambda(y) = y^q/q$. Without the $1/p$ in the definition of Ψ the function would be a messier multiple of y^q.

A tad more interesting is the case where $\Psi(x) = e^x - 1$. For $y \leq 1 = \dot{\Psi}(0)$ the maximum is achieved at $x = 0$, so that $\Lambda(y) = 0$. If $y > 1$ the maximum is now achieved at $x = \log y$ so that $\Lambda(y) = 1 + y \log y - y$. That is

$$\Lambda(y) = 0\{0 \leq y \leq 1\} + (1 + y \log y - y)\{y > 1\}$$

The flat spot on $[0, 1]$ comes from the fact that $\Psi'(0) > 0$.

The flat spot can be removed by subtracting off a linear function to get a zero right-hand derivative at the origin: if $\Psi_0(x) := e^x - 1 - x$ then $\dot{\Psi}_0(x) = e^x - 1$. Now the maximum is achieved at $x = \log(1 + y)$ so that $\Lambda_0(y) = \Lambda(1 + y) = (1 + y)\log(1 + y) - y$.



Problem [11] shows that $\mathcal{L}^{\Psi_0}(\mathcal{X},\mathcal{A},\mathbb{P}) = \mathcal{L}^{\Psi}(\mathcal{X},\mathcal{A},\mathbb{P})$, for probability measures $\mathbb{P}$, with

$$K_\Psi \|X\|_{\Psi_0} \geq \|X\|_\Psi \geq \|X\|_{\Psi_0} \qquad \text{for all } X \in \mathcal{L}^\Psi$$

and $\mathcal{L}^\Lambda = \mathcal{L}^{\Lambda_0}$ with

$$\|Y\|_\Lambda \leq \|Y\|_{\Lambda_0} \leq K_\Lambda \|Y\|_\Lambda \qquad \text{for all } Y \in \mathcal{L}^\Lambda.$$

□ where $K_\Psi = 1 + \Psi^{-1}(1) \approx 1.7$ and $K_\Lambda = 1 + \Lambda(2) \approx 2.3$.

Theorem <15> has another, more impressive, consequence. The conclusion from the Theorem,

$$\Lambda(y) = \sup\nolimits_{x \geq 0} xy - \Psi(x) \qquad \text{with equality when } x = q(y) := \inf\{t \geq 0 : \dot{\Psi}_R(t) \geq y\},$$

can be rewritten as

\E@ Young <17>
$$\Lambda(y) + \Psi(x) \geq xy \qquad \text{for all } x, y \geq 0,$$

with equality if $x = q(y)$. This result is often called "Young's inequality". A very similar result was derived by William Henry Young, whom I cannot resist quoting:

> The method [...] has the great advantage as regards the details of the work that it is based on an extremely simple inequality. From this point of view it is far superior to the other methods of proof, almost intuitive as they are, when the necessary preliminary theorems have been proved, that I have indicated elsewhere. Instead of requiring the generalisation of Schwarz's well-known inequality, it suffices to be acquainted with the generalisation of the relation
>
> $2ab \leq a^2 + b^2$, namely, that $(p+1)ab \leq a^{p+1} + pb^{1+1/p}$.
>
> In the prosecution of research one is often hampered by the difficulty of generalising a known formula, owing to the very simplicity and obviousness of its statement. The inequality just written down possesses, however, the great advantage over Schwarz's inequality that its generalisation is almost immediate.
>
> [Young, 1912, pp 225-226]

Why don't mathematicians write like that any more?

Remark. The equality $\Psi^{**} = \Psi$ ensures that there is a symmetry in the roles of Λ and Ψ. That is, $\Lambda^* = \Psi$. There is also equality in <17> if $y = \Psi'_R(x)$.

The YOUNG inequality tells us something about linear functionals on ORLICZ spaces. Remember that a linear map $T : \mathcal{Z} \to \mathbb{R}$ from a vector space $\mathcal{Z}$ equipped with a norm (or semi-norm) $\|\cdot\|$ is continuous if and only if

$$\|T\| := \sup\{|Tz| : \|z\| \leq 1\} \qquad \text{is finite.}$$

For ancient reasons, such a T is also called a continuous linear ***functional***. (Once upon a time, functions whose domains were sets of functions were considered scary objects.)

For example, if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$ then each g in $\mathcal{L}^q$ defines a continuous linear functional on $\mathcal{L}^p$ by

$$T_g(f) := \mu(fg) \qquad \text{for each } f \in \mathcal{L}^p.$$

The continuity comes directly from the HÖLDER inequality:

$$|T_g(f)| = |\mu(fg)| \le \|f\|_p \|g\|_q,$$

which implies $\|T_g\| \le \|g\|_q$. In fact more is true: $\|T_g\| = \|g\|_q$ and every continuous linear functional T on $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu)$ can be represented (μ-almost uniquely) by means of a function $g \in \mathcal{L}^q(\mathcal{X}, \mathcal{A}, \mu)$ (Folland, 1999, Theorem 6.15).

The analogous results for general ORLICZ spaces take a slightly less pleasing form. The role played by the HÖLDER inequality is taken over by the YOUNG inequality

Orlicz::norm.via.Young <18> **Example.** Suppose Ψ and Λ are conjugate ORLICZ functions, both in $\mathcal{Y}_\infty$. Suppose also that $f \in \mathcal{L}^\Psi$ and $g \in \mathcal{L}^\Lambda$. For finite constants $c > \|f\|_\Psi$ and $d > \|g\|_\Lambda$, the YOUNG inequality gives

$$|fg|/(cd) \le \Psi(|f|/c) + \Lambda(|g|/d),$$

which integrates to give

$$\mu|fg|/(cd) \le \mu\Psi(|f|/c) + \mu\Lambda(|g|/d) \le 2.$$

Rearrange then take infima over c and d to deduce that

$$\mu|fg| \le 2\|f\|_\Psi \|g\|_\Lambda.$$

This inequality tells us that $T_g(f) := \mu(fg)$ is a well defined linear functional on $\mathcal{L}^\Psi$ for which satsifies

$$\|T_g\| := \sup\{|T_g(f)| : \|f\|_\Psi \le 1\} \le 2\|g\|_\Lambda.$$

That is, g defines a continuous linear functional on $\mathcal{L}^\Psi$.

Note the extra factor of 2. Problem [16] provides the corresponding lower bound, $\|g\|_\Lambda \le \|T_g\|$.

The gap between the upper and lower bounds for $\|T_g\|$ might seem a small price to pay for the generalization from $\mathcal{L}^p$ to $\mathcal{L}^\Psi$. Unfortunately there is worse news. These bounds leave open the possibility that there might be other continuous linear functionals on $\mathcal{L}^\Psi$ that cannot be represented as $f \mapsto \mu(fg)$ for some function g. Indeed, as shown in the next Section, if Ψ increases rapidly enough then such functionals do exist. □

*5.6 Comparison of $\mathcal{L}^p$ and $\mathcal{L}^\Psi$

Orlicz::S:LpLPsi

As you saw from the Young quotation, the original motivation for the YOUNG inequality was generalization of the theory of $\mathcal{L}^p$ spaces. In some ways $\mathcal{L}^\Psi$ is better behaved than $\mathcal{L}^p$; in other ways worse. To keep the story as simple as possible, in this Section I consider only the case where the underlying measure is a probability measure, even though much of the discussion could be extended to other measures.

The use of the norms to bound tail probabilities is an example where $\mathcal{L}^\Psi$ can be better behaved than $\mathcal{L}^p$.

Orlicz::llp.tails <19> **Example.** If $f \in \mathcal{L}^p$ with $0 < \sigma = \|f\|_p$ then

$$\mathbb{P}\{|f| \geq t\} \leq \mathbb{P}|f/t|^p = \sigma^p t^{-p} \qquad \text{for } t > 0.$$

The implication does not go the other way. If $\mathbb{P}\{|f| \geq t\} \leq Ct^{-p}$ for all $t > 0$ then, by the TONELLI inequality,

$$\mathbb{P}|f|^p = \int_0^\infty pt^{p-1}\mathbb{P}\{|f| > t\}\,dt \leq Cp\int_0^\infty t^{-1}dt = \infty.$$

The problem occurs as t gets large; for t near zero the tail bound can be improved to $\min(1, Ct^{-p})$. Of course an infinite upper bound does not make $\mathbb{P}|f|^p$ infinite. The example where $U \sim \text{UNIF}(0,1)$ and $X = (1/U)^{1/p}$ provides a specific example where $\mathbb{P}\{X \geq t\} = \min(1, t^{-p})$ but $\mathbb{P}X^p = \infty$. □

Remark. If $p \geq 1 + \delta$ with $\delta > 0$ and $\mathbb{P}\{|f| \geq t\} \leq Ct^{-p}$ for $t > 0$ then $\mathbb{P}|f|^{p-\delta} < \infty$.

The story for $\mathcal{L}^\Psi$ is slightly more encouraging. Suppose $\Psi(x) = \int_0^x \psi(t)\,dt$ for each $x \geq 0$ (as in Section 5.2) and Ψ increases rapidly enough to ensure that

\E@ rapid.increase <20>

$$\int_0^\infty \frac{\psi(t)}{\max(1, \Psi(At))}\,dt < \infty \qquad \text{for some constant } A > 0.$$

Then the next Example shows that there is an equivalence between the tail bound and finiteness of the $\mathcal{L}^\Psi$ norm. It is not hard to show that each Ψ_p, for $p \geq 1$, has property <20>.

Orlicz::llPsi.tails <21> **Example.** If $f \in \mathcal{L}^\Psi$ with $0 < \sigma = \|f\|_\Psi$ then

$$\mathbb{P}\{|f| \geq t\} \leq \min\left(1, \mathbb{P}\frac{\Psi(|f|/\sigma)}{\Psi(t/\sigma)}\right) \leq \min\left(1, \frac{1}{\Psi(t/\sigma)}\right) \qquad \text{for } t > 0.$$

Conversely, suppose Ψ has property <20> and there is a positive constant a for which

$$\mathbb{P}\{|f| \geq t\} \leq \min(1, 1/\Psi(at)) = 1/\max(1, \Psi(at)) \qquad \text{for each } t > 0.$$

Then the TONELLI inequality gives

$$\begin{aligned}\mathbb{P}\Psi(|f|/c) &= \mathbb{P}\int_0^\infty \psi(t)\{|f|/c > t\}\,dt \\ &= \int_0^\infty \psi(t)\mathbb{P}\{|f|/c > t\}\,dt \\ &\le \int_0^\infty \frac{\psi(t)}{\max(1, \Psi(act))}\,dt \quad .\end{aligned}$$

□ The final integral is finite if $ac \ge A$.

As Problem [14] shows, the characterization of $\mathcal{L}^\Psi$ in Example <21> via tail probabilities comes at a cost: under the rapid-growth assumption <20> there always exist continuous linear functionals that are not of the form $f \mapsto \mathbb{P}(fg)$ for some g.

In trying to understand what goes wrong, I found it instructive to try to adapt the $\mathcal{L}^p$ proof for $1 < p < \infty$ (**?**, Section 6.2) to a general ORLICZ function Ψ and a probability space $(\mathcal{X}, \mathcal{A}, \mathbb{P})$. We start with a continuous linear map T from $\mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mathbb{P})$ into the real line. There is no loss of generality (Bourbaki, 2004, Section II.2.2) in assuming that $T(f) \ge 0$ whenever $f \ge 0$.

Here is an outline the steps of that proof. Notice the switch from $\mathcal{L}^\Psi$ to $\mathcal{L}^p$ at step (iii).

(i) The indicator function $\mathbb{1}_A$ of a set $A \in \mathcal{A}$ has $\mathbb{P}\Psi(\mathbb{1}_A/c) = \mathbb{P}A\Psi(1/c)$. It follows that $\|\mathbb{1}_A\|_\Psi = 0$ if $\mathbb{P}A = 0$ and $1/\Psi^{-1}(1/\mathbb{P}A)$ otherwise. The set function $\nu A := T(\mathbb{1}_A)$ inherits finite additivity from the linearity of T and countable additivity from the continuity of T: If $A_n \downarrow \emptyset$ then

$$|\nu(A_n)| = |T(\mathbb{1}_{A_n})| \le \|T\|\,\|\mathbb{1}_{A_n}\|_\Psi \to 0 \qquad \text{because } \mathbb{P}A_n \to 0.$$

Thus ν is a countably additive measure with $\nu\mathcal{X} = T(\mathbb{1}) < \infty$ and for which $\nu A = 0$ whenever $\mathbb{P}A = 0$. By the RADON-NIKODYM theorem (**?**, Section 3.2), there exists a nonnegative, $\mathbb{P}$-integrable function Δ for which $\nu A = \mathbb{P}(A\Delta)$ for each $A \in \mathcal{A}$.

(ii) If $h(x) = \sum_i \alpha_i\{x \in A_i\}$ is a simple function then, by linearity of T and of the integrals with respect to ν and $\mathbb{P}$,

$$Th = \sum\nolimits_i \alpha_i T(\mathbb{1}_{A_i}) = \sum\nolimits_i \alpha_i \nu(A_i) = \nu f = \mathbb{P}(f\Delta).$$

(iii) The set of all simple functions is dense in $\mathcal{L}^p$. That is, for each f in $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mathbb{P})$ and each $\epsilon > 0$ there exists a simple function h_ϵ for which $\|f - h_\epsilon\|_p < \epsilon$. Continuity of T extends the equalities from step (ii) from simple functions to all of $\mathcal{L}^p$.

(iv) We then play around with various choices for f in $\mathcal{L}^p$ to show that $\mathbb{P}\Delta^q < \infty$, where $p^{-1} + q^{-1} = 1$ and $\|\Delta\|_q = \|T\|$.

It is step (iii) that fails for an ORLICZ function Ψ satisfying the rapid-growth condition. As shown in Problem [14], under that condition there are always simple probability spaces $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ with some function in $f_0 \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mathbb{P})$ that cannot be closely approximated in $\mathcal{L}^\Psi$ norm by any bounded measurable function, let alone by a simple functions.

The main difficulty is well illustrated by the following Example for the ORLICZ function $\Psi_2(x) = \exp(x^2) - 1$.

Orlicz::simple.dense <22> **Example.** Let $\mathbb{P}$ be the $N(0,1)$ distribution on $\mathcal{B}(\mathbb{R})$. Consider $f_0(x) = x$, which belongs to $\mathcal{L}^{\Psi_2}$ because

$$1 + \mathbb{P}\Psi_2(f_0^2/c^2) = (2\pi)^{-1/2}\int_{\mathbb{R}} \exp(-x^2/2 + x^2/c^2)\,dx = \begin{cases} c/\sqrt{c^2-2} & \text{if } c^2 > 2 \\ +\infty & \text{if } 0 < c^2 \le 2 \end{cases}.$$

Consequently, $\|f_0\|_{\Psi_2} = \sqrt{8/3}$ and $f_0 \in \mathcal{L}^{\Psi_2}$.

Surprisingly, no bounded function (let alone a simple function) can approximate f_0 very closely. Suppose $|h(x)| \le K < \infty$. The fact that $x - K \ge x/2$ when $x \ge 2K$ implies

$$1 + \mu\Psi_2(|f_0 - h|/c) \ge (2\pi)^{-1/2}\int_{2K}^{\infty} \exp(-x^2/2 + x^2/4c^2)\,dx,$$

which is infinite if $0 < c \le 1/\sqrt{2}$. Thus $\|f_0 - h\|_\Psi > 1/\sqrt{2}$.

It follows that f_0 does not belong to the smallest closed subspace $\mathcal{H}$ of $\mathcal{L}^{\Psi_2}$ that contains all the bounded, measurable functions. By the HAHN-BANACH theorem **?**, Theorem 5.8 there exists a continuous linear functional T on $\mathcal{L}^{\Psi_2}$ with $T(h) = 0$ for all h in $\mathcal{H}$ and $T(f_0) \ne 0$. There can be no measurable function Δ for which $f\Delta \in \mathcal{L}^1$ and $T(f) = \mathbb{P}(f\Delta)$ for all $f \in \mathcal{L}^{\Psi_2}$: the choice $h = \mathrm{sgn}(\Delta)$ would force $\mathbb{P}|\Delta| = 0$ and consequently $\mathbb{P}(f_0\Delta) = 0 \ne T(f_0)$.

□

5.7 Problems

Orlicz::S:Problems

Orlicz::P:achieved [1] Suppose $f \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ and $c_1 := \|f\|_\Psi > 0$. Show that $\mu\Psi(|f|/c_1) \le 1$. Hint: Use dominated convergence and $\mu\Psi\left(|f|/(c_1 + n^{-1})\right) \le 1$ for $n \in \mathbb{N}$.

Orlicz::P:increasing [2] Suppose $\{f_n : n \in \mathbb{N}\} \subset \mathcal{L}^\Psi = \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ and $0 \le f_n \uparrow f$ pointwise. If $c_0 := \sup_n \|f\|_\Psi < \infty$ show that $f \in \mathcal{L}^\Psi$ and $\|f\|_\Psi = c_0$. Hint: Use monotone convergence for $\mu\Psi(f_n/c)$ with $c > c_0$.

Orlicz::P:e.to.g [3] Suppose $h(x) = e^{g(x)}$, where g is a twice differentiable real-valued function on $\mathbb{R}^+$. Show that h is convex on any interval J for which $(g'(x))^2 + g''(x) \ge 0$ for $x \in \mathrm{int}(J)$.

Orlicz::P:convex.derivs

[4] Let $\dot{\Psi}_L$ and $\dot{\Psi}_R$ be the left- and right-derivative functions for an ORLICZ function Ψ. For $x > 0$ let $\dot{\Psi}(x-)$ denote the limit of $\dot{\Psi}_R(y)$ as $y \nearrow x$, which exists because $\dot{\Psi}_R$ is an increasing function.

(i) For an $x > 0$, let $\{y_i\}$ be a strictly increasing sequence in $\mathbb{R}^+$ for which $0 \le y_1 < y_2 < \cdots \nearrow x$. Explain why $\dot{\Psi}_R(y_i) \le \dot{\Psi}_L(y_{i+1})$ for each i. Deduce that $\dot{\Psi}_R(x-) = \dot{\Psi}_L(x)$, so that Ψ is differentiable at x if and only if the size $\dot{\Psi}(x) - \dot{\Psi}_L(x)$ of the jump at x is zero.

(ii) For each $\epsilon > 0$ and $A > 0$ show that there are at most finitely many x in $(0, A]$ for which $\dot{\Psi}_R(x) - \dot{\Psi}_L(x) > \epsilon$. Deduce that $\dot{\Psi}_R$ has at most countably many jumps in $\mathbb{R}^+$ and hence $\dot{\Psi}_R(x) \neq \dot{\Psi}_L(x)$ for, at worst, a countable set of x in $\mathbb{R}^+$.

Orlicz::P:convex.integral

[5] Suppose $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ is nondecreasing. For x in $\mathbb{R}^+$ define $\Psi(x) = \int_0^x \psi(r)\,dr$. Show that $\Psi \in \mathcal{Y}$. Hint: Suppose $x_\theta = \bar{\theta}x_0 + \theta x_1$, where $\bar{\theta} = 1 - \theta$ with $x_0 < x_1$ and $0 < \theta < 1$. Show that

$$\bar{\theta}\Psi(x_0) + \theta\Psi(x_1) - \Psi(x_\theta) = -\bar{\theta}\int_{x_0}^{x_\theta} \psi(r)\,dr + \theta\int_{x_\theta}^{x_1} \psi(r)\,dr.$$

Use the bounds $\psi(r_0) \le \psi(x_\theta) \le \psi(r_1)$ for $r_0 \le x_\theta \le r_1$.

Orlicz::P:Psi.rep

[6] Suppose $\Psi \in \mathcal{Y}$ has a representation $\Psi(x) = \int_0^x \psi(t)\,dt$ with ψ an increasing function on $\mathbb{R}^+$. Show that $\dot{\Psi}_R(t) \ge \psi(t)$ for all $t \ge 0$ and $\dot{\Psi}_L(t) \le \psi(t)$ for all $t > 0$. Hint: For $\delta > 0$ we have

$$\delta\dot{\Psi}_R(t_0 + \delta) \ge \int_{t_0}^{t_0+\delta} \dot{\Psi}_R(t)\,dt = \Psi(t_0 + \delta) - \Psi(t_0) = \int_{t_0}^{t_0+\delta} \psi(t)\,dt.$$

Orlicz::P:xlogx

[7] Suppose $\Psi(x) = e^{g(x)} - 1$, where $g(x) = (x \log x)^+$.

(i) Show that g is convex and $g(x) + g(y) \le g(x + y)$ for all $x, y \ge 0$. Deduce that $\Psi \in \mathcal{Y}_{\exp}$.

(ii) Show that $\|X\|_\Psi < \infty$ if X has a POISSON(θ) distribution.

Orlicz::P:extend.to.Young

[8] Suppose Ψ_0 is a convex, nonnegative, increasing function defined on an interval $[a, \infty)$, with $a > 0$. Suppose also that there exists an $x_0 \in [a, \infty)$ for which $\Psi_0(x_0)/x_0 = \gamma := \inf_{x \ge a} \Psi_0(x)/x$. Show that the extension $\Psi(x) := \gamma x\{0 \le x < x_0\} + \Psi_0(x)\{x \ge x_0\}$ is an ORLICZ function.

Orlicz::P:alpha.lt.1

[9] For a fixed α in $(0, 1)$ define $f_\alpha(x) = e^{x^\alpha}$ for $x \ge 0$.

(i) Show that $x \mapsto f_\alpha(x)$ is convex for $x \ge z_\alpha := (\alpha^{-1} - 1)^{1/\alpha}$ and concave for $0 \le x \le z_\alpha$.

(ii) Define $x_\alpha = (1/\alpha)^{1/\alpha}$ and $\tau_\alpha = f_\alpha(x_\alpha)/x_\alpha = (\alpha e)^{1/\alpha}$. Show that

$$\Psi_\alpha(x) := \tau x\{x < x_\alpha\} + f_\alpha(x)\{x \ge x_\alpha\}$$

is an ORLICZ function.

(iii) Show that there exists a constant K_α for which

$$\Psi_\alpha(x) \le \exp(x^\alpha) \le K_\alpha + \Psi_\alpha(x) \qquad \text{for all } x \ge 0.$$

(iv) Show that $\Psi_\alpha(x)\Psi_\alpha(y) \le \Psi_\alpha(2^{1/\alpha}xy)$ for $x \wedge y \ge x_\alpha$.

Orlicz::P:ll1.llPsi [10] Suppose $\Psi \in \mathcal{Y}$ and $X \in \mathcal{L}^\Psi(\mathfrak{X}, \mathcal{A}, \mathbb{P})$, for a probability measure $\mathbb{P}$. Show that $\mathbb{P}|X| \le \Psi^{-1}(1)\, \|X\|_\Psi$. Hint: the JENSEN inequality.

Orlicz::P:slope.zero [11] Suppose $\Psi \in \mathcal{Y}_\infty$ with $\dot{\Psi}_R(0) = \tau > 0$. Define $\Psi_0(x) = \Psi(x) - \tau x$ for $x \ge 0$.

(i) Show that $\Psi_0 \in \mathcal{Y}$. Suppose $X \in \mathcal{L}^{\Psi_0}(\mathfrak{X}, \mathcal{A}, \mathbb{P})$, for a probability measure $\mathbb{P}$. For $K_\Psi := 1 + \tau\Psi_0^{-1}(1)$ show that $K_\Psi \|X\|_{\Psi_0} \ge \|X\|_\Psi \ge \|X\|_{\Psi_0}$. Deduce that $\mathcal{L}^\Psi = \mathcal{L}^{\Psi_0}$. Hint: previous Problem.

(ii) Suppose Ψ has conjugate Λ and that Ψ_0 has conjugate Λ_0. Show that $\Lambda_0(y) = \Lambda(y + \tau)$ for $y \ge 0$.

(iii) Define $K_\Lambda := 1 + \Lambda(2\tau)$. For each $Y \in \mathcal{L}^\Lambda$ show that

$$\|Y\|_\Lambda \le \|Y\|_{\Lambda_0} \le K_\Lambda \|Y\|_\Lambda$$

Deduce that $\mathcal{L}^{\Lambda_0} = \mathcal{L}^\Lambda$. Hint: The first inequality follows from $\Lambda_0 \ge \Lambda$. For the second inequality first show that $\Lambda_0(|Y|/c) \le \Lambda(2\tau) + \Lambda(2|Y|/c)$ then take expected values with $c > \|Y\|_\Lambda$.

Orlicz::P:interp [12] Suppose $n_0 = 1 \le n_1 < n_2 < \dots$ is an increasing sequence of positive integers with $\sup_{k\in\mathbb{N}} n_k/n_{k-1} = B < \infty$ and $f : [1,\infty) \to \mathbb{R}^+$ is an increasing function with $\sup_{t\ge1} f(Bt)/f(t) = A < \infty$. For each random variable X show that $\sup_{r\ge1} \|X\|_r / f(r) \le A \sup_{k\in\mathbb{N}} \|X\|_{n_k} / f(n_k)$.

Orlicz::P:moment.Psia [13] For $p \ge 1$ prove the existence of positive constants c_p an C_p for which

$$c_p \|X\|_{\Psi_p} \le \sup_{r\ge1} \frac{\|X\|_r}{r^{1/p}} \le C_p \|X\|_{\Psi_p}$$

for every random variable X. Hint: previous Problem.

Orlicz::P:non.dense [14] Suppose $\mathbb{P}$ is LEBESGUE measure on the $\mathcal{B}(0,1)$ and $\Psi \in \mathcal{Y}_\infty$ satisfies the rapid-growth condition <20>:

$$\int_0^\infty \frac{\psi(t)}{\max(1, \Psi(At))}\, dt < \infty \qquad \text{for some constant } A > 0.$$

Define $f_0(x) = \Psi^{-1}(1/x)$.

(i) Show that $\mathbb{P}\{x : f_0(x) \ge t\} = \mathbb{P}\{x \le 1/\Psi(t)\} = \min(1, 1/\Psi(t))$. Deduce from Example <21> that $f \in \mathcal{L}^\Psi$.

(ii) Suppose h is a $\mathcal{B}(0,1)$-measurable function that is bounded in absolute value by a constant K. Define $A_k = \{x \le 1/\Psi(2K)\}$. Argue that $f_0(x) \ge 2K$ on A_k, so that $f_0(x)/2 \ge |h(x)|$

$$\mathbb{P}\Psi(2|f_0 - h|) \ge \mathbb{P}\Psi(f_0)\{x \in A_k\} = \mathbb{P}\{x \in A_k\}x^{-1} = \infty.$$

Deduce that $\|f_0 - h\|_\Psi \ge 1/2$.

(iii) Let $\mathcal{H}$ denote the closure (in $\mathcal{L}^\Psi$) of the set of all bounded functions in $\mathcal{L}^\Psi$. Explain why $f_0 \notin \mathcal{H}$. Use the HAHN-BANACH theorem (**?**, Theorem 5.8) to prove the existence of a continuous linear functional T on $\mathcal{L}^\Psi$ for which $T(h) = 0$ for $h \in \mathcal{H}$ and $T(f) \ne 0$. Explain why there cannot exist a function g in $\mathcal{L}^\Lambda$ for which $T(f) = \mathbb{P}(fg)$ for each f in $\mathcal{L}^\Psi$. Hint: Consider $h = \text{sgn}(g)$.

Orlicz::P:Psinorm.bnd

[15] For $f \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ show that $\|f\|_\Psi \le 1 + \mu\Psi(|f|)$. Hint: If $c = \|f\|_\Psi > 1$ then convexity gives $\Psi(t/c) \le \Psi(t)/c$ for $t > 0$.

Orlicz::P:Tg

[16] (cf. Garling, 2007, Theorem 6.3.2) Suppose $\Psi \in \mathcal{Y}_\infty$ and $(\mathcal{X}, \mathcal{A}, \mu)$ is a σ-finite measure space. Suppose a function $g : \mathcal{X} \to \mathbb{R}$ is $\mathcal{A}$-measurable and $fg \in \mathcal{L}^1(\mathcal{X}, \mathcal{A}, \mu)$ whenever $f \in \mathcal{L}^\Psi$. Suppose also that the linear functional $T(f) := \mu(fg)$ on $\mathcal{L}^\Psi$ is continuous. Let Λ denote the conjugate of Ψ, as in Section 5.5. Show that $g \in \mathcal{L}^\Lambda$ and

$$\|g\|_\Lambda \le \|T\| := \sup\{|T(f)| : \|f\|_\Psi \le 1\}.$$

Follow these steps.

(i) Replace $f \in \mathcal{L}^\Psi$ by $|f|\text{sgn}(g)$ to show that there is no loss of generality in assuming that $g \ge 0$ and $\|T\| = \sup\{\mu(fg) : 0 \le f \text{ and } \|f\|_\Psi \le 1\}$.

(ii) By σ-finiteness there exist sets B_n in $\mathcal{A}$ for which $B_n \uparrow \mathcal{X}$ and $\mu B_n < \infty$ for each n. Define $g_n(x) = \{x \in B_n\}(n \wedge g(x))$, so that $d_n := \|g_n\|_\Lambda \le \Lambda(n)\mu B_n < \infty$. Also define $f_n(x) = \Lambda'_R(g_n(x)/d_n)$, which is also bounded by a constant multiple of $\{x \in B_n\}$, so that $c_n := \|f_n\|_\Psi < \infty$. (Here I am tacitly assuming that $d_n \ne 0$ and, later, $c_n > 0$. I leave it to you to handle the trivial cases where either $d_n = 0$ or $c_n = 0$.) Use the YOUNG inequality (for the case where there is equality) and Problem [15] to show that

$$\|T\| / d_n \ge \mu\left(f_n g_n/(c_n d_n)\right) = \Big[\mu\Psi(f_n) + \mu\Lambda(g_n/d_n)\Big]/c_n \ge 1.$$

(iii) Invoke Problem [2] to deduce that $\|T\| \ge \lim_{n\to\infty} c_n = \|g\|_\Lambda$.

Orlicz::P:Onorm

[17] Let Ψ and Λ be conjugate members of $\mathcal{Y}_\infty$. For $f \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ define

$$\|f\|_\bullet = \sup\{\|fg\|_1 : \|g\|_\Lambda \le 1\}.$$

If μ is σ-finite, appeal to Example <18> and Problem [16], with the roles of Ψ and Λ interchanged, to deduce $\|f\|_\Psi \le \|f\|_\bullet \le 2\|f\|_\Psi$ for all f in $\mathcal{L}^\Psi$.

5.8 Notes

Orlicz::S:Notes

If I remember rightly, my introduction to the virtues of using expected values instead of tail probabilities in empirical process theory came from Giné and Zinn (1984, Lemmas 2.8 and 2.9). That paper (in its preprint form) and advice from its authors led me to the work of Pisier (1980, 1983), who used ORLICZ norms to establish sufficient conditions for existence of versions of stochastic processes with continuous sample paths. It was then just a short step to the amazing work of Ledoux and Talagrand (1991, Chapter 11).

My subsequent ORLICZ education came from Dudley (1999, Appendix H) and Garling (2007, Chapter 6), whose brief accounts led me to the beautiful PhD thesis of Luxemburg (1955), where the $\|\cdot\|_\Psi$ is defined (page 43). My remarks about the difference between the norms defined by Orlicz and by Luxemburg are based on a rather quick reading of Orlicz (1932, 1936). There appears to be some controversy about the assignment of priority in this area. I do not claim to have made any careful study of the early contributions; my historical remarks should be treated with caution.

There also seems to be some disagreement in the literature about the precise definition of an ORLICZ function. Garling (2007, page 73) used the name *Young function* for the members of my $\mathcal{Y}_\infty$. Dudley (1999, Appendix H) used the name *Young-Orlicz modulus* for a my $\mathcal{Y}$ and *Orlicz modulus* for those Ψ in my $\mathcal{Y}_\infty$ for which $\Psi'_R(0) = 0$. The latter also correspond to the N-functions of Krasnosel'skiĭ and Rutickiĭ (1961).

The special ORLICZ functions $\Psi(x) = x^p$ for $p = 1, 2$ and $\Psi(x) = \Psi_2(x) = e^{x^2} - 1$ were behind the inequalities from Pollard (1989, Section 3). I was (perhaps overly) pleased with myself for being able to replace the $\mathcal{L}^p$ norms by the $\mathcal{L}^{\Psi_2}$ norm in Pollard (1990, Section 3). Subsequently van der Vaart and Wellner (1996, Lemma 2.2.2) extended the inequality to a large collection of ORLICZ functions. It now seems that the inequality in part (ii) of Theorem <12>, which I learned from Ledoux and Talagrand (1991, Chapter 11), is even better at delivering maximal inequalities. In the notes to that chapter, L&T gave credit to Fernique (1983), by noting that their "Theorem 11.2 is equivalent to the (perhaps somewhat unorthodox) formulation of [Fernique]". I think they were referring to Fernique's Lemme 2.2.1 and Corollaire 3.1.

More recently, van de Geer and Lederer (2013) used the ORLICZ function

$$\Psi(x) = \exp\left((\sqrt{1 + 2Lx} - 1)^2/L\right),$$

for positive constants L, and Wellner (2017) used the ORLICZ function

$$\Psi(x) = \exp\left(2L^{-2}\mathbb{h}(Lx)\right) - 1 \qquad \text{where } \mathbb{h}(x) := (1 + x)\log(1 + x) - x,$$

to derive empirical process maximal inequalities of the BERNSTEIN\BENNETT variety (see Chapter 8).

My enthusiasm for things ORLICZ grew after I finally understood the role played by the YOUNG inequality as an extension of the HÖLDER inequality, a powerful tool for generalizing from $\mathcal{L}^p$ to $\mathcal{L}^\Psi$. It was with some disappointment that I eventually learned that the dual space of $\mathcal{L}^{\Psi}$ could not always be identified with $\mathcal{L}^{\Lambda}$, for the ORLICZ function Λ that is conjugate to Ψ. Whence Section 5.6. I have no idea whether the counterexamples in that Section are well known. It would be better to call the YOUNG inequality "one of the inequalities to which the name of W. H. Young has been attached". FOURIER theorists think of the YOUNG inequality as another result involving convolutions (Folland, 1999, p 247).

By the way, Dudley (2003, Notes to Section 5.1) made a good case for adding the name Leonard James Rogers to that of Hölder for the HÖLDER inequality.

References

Bourbaki_integration_I Bourbaki, N. (2004). *Integration I: Chapters 1–6*. Elements of Mathematics. Springer-Verlag. English translation by S. K. Berberian of *Intégration* 1965, 1967, 1959.

Dudley1999UCLT Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

Dudley2003RAP Dudley, R. M. (2003). *Real Analysis and Probability* (2nd ed.), Volume 74 of *Cambridge studies in advanced mathematics*. Cambridge University Press.

Fernique83StFlour Fernique, X. (1983). Regularité de fonctions aléatoires non gaussiennes. *Springer Lecture Notes in Mathematics 976*, 1–74. École d'Été de Probabilités de St-Flour XI, 1981.

Folland1999real Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*. Wiley.

Garling2007CUP Garling, D. J. H. (2007). *Inequalities: A Journey into Linear Analysis*. Cambridge University Press.

GineZinn1984AnnProb Giné, E. and J. Zinn (1984). Some limit theorems for empirical processes. *Annals of Probability 12*, 929–989. (plus 9 pages of contributed discussion).

KimPollard90cuberoot Kim, J. and D. Pollard (1990). Cube root asymptotics. *Annals of Statistics 18*, 191–219.

KrasnoselskiiRutickii1961 Krasnosel'skiĭ, M. A. and Y. B. Rutickiĭ (1961). *Convex Functions and Orlicz Spaces*. Noordhoff. Translated from the first Russian edition by Leo F. Boron.

LedouxTalagrand91book Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. New York: Springer.

Luxemburg1955thesis Luxemburg, W. A. J. (1955). *Banach function spaces*. Ph. D. thesis, TU Delft. http://resolver.tudelft.nl/uuid:252868f8-d63f-42e4-934c-20956b86783f.

Orlicz1932BIAP Orlicz, W. (1932). Über eine gewisse Klasse von Räumen vom Typus B. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres, Classe de Mathématique et Naturelles, série A 8*(9), 207–220.

Orlicz1936BIAP Orlicz, W. (1936). Über Räume (L^M). *Bulletin International de l'Académie Polonaise des Sciences et des Lettres, Classe de Mathématique et Naturelles, série A, 93*, 107.

Pisier7980 Pisier, G. (1980). Conditions d'entropie assurant la continuité de certains processus et applications à l'analyse harmonique. In *Séminaire d'analyse fonctionnelle, 1979-80*, pp. 1–41. École Polytechnique Palaiseau. Available from http://archive.numdam.org/.

Pisier83metricEntropy Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Springer Lecture Notes in Mathematics 995*, 123–154. (A collection of lecture notes from the 1980-81 Special Year in Analysis at the University of Connecticut Department of Mathematics).

Pollard89StatSci Pollard, D. (1989). Asymptotics via empirical processes (with discussion). *Statistical Science 4*, 341–366.

Pollard90Iowa Pollard, D. (1990). *Empirical Processes: Theory and Applications*, Volume 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Hayward, CA: Institute of Mathematical Statistics.

vanderGeerLederer2013PTRF van de Geer, S. and J. Lederer (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probability theory and related fields 157*(1-2), 225–250.

vaartwellner96book van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Process: With Applications to Statistics*. Springer-Verlag.

Wellner2017Sankhya Wellner, J. A. (2017). The Bennett-Orlicz norm. *Sankhya A 79*(2), 355–383.

Young1912 Young, W. H. (1912). On classes of summable functions and their Fourier series. *Proc. R. Soc. Lond. A 87*(594), 225–229.

