ORLICZ SPACES

CHRISTIAN LÉONARD

(Work in progress)

Contents

Obtaining basic results on Orlicz spaces in the literature is not so easy. Indeed, the old seminal textbook by Krasnosel'skii and Rutickii [1] (1961) contains all the fundamental properties one should know about Orlicz spaces, but the theory is developed in $\mathbb{R}^d$ with the Lebesgue measure. One is left with the question "What can be kept in a more general measure space?". Many of the answers to this question might be found in the more recent textbook by Rao and Ren [3] (1991) where the theory is developed in very general situations including many possible pathologies of the Young functions and the underlying measure. Because of its high level of generality, I find it uneasy to extract from [3] the basic ideas of the proofs of the theorems on Orlicz spaces which are the analogues of the basic results on L_p spaces.

The aim of these notes is to present basic results about Orlicz spaces. I have tried to make the proofs as self-contained and synthetic as possible. I hope for the indulgence of the reader acquainted with Walter Rudin's books, in particular with [4] where wonderful pages are written on the L_p spaces which are renowned Orlicz spaces.

1. Basic definitions and results on Orlicz spaces

The notion of Orlicz space extends the usual notion of L_p space with $p \geq 1$. The function s^p entering the definition of L_p is replaced by a more general convex function $\theta(s)$ which is called a Young function.

Definition 1.1 (Young function). *A function $\theta : \mathbb{R} \to [0, \infty]$ is a Young function if*

(i) θ is a convex lower semicontinuous $[0, \infty]$-valued function on $\mathbb{R}$;
(ii) θ is even and $\theta(0) = 0$;
(iii) θ is non-trivial: it is different from the constant function $0(s) = 0$, $s \in \mathbb{R}$ and its convex conjugate $0^*(s) = \begin{cases} 0, & \textit{if } s = 0 \\ +\infty, & \textit{otherwise.} \end{cases}$

One says that the Young function θ is finite *if its effective domain* $\operatorname{dom}\theta := \{s \in \mathbb{R}; \theta(s) < \infty\}$ *is the whole real line* $\mathbb{R}$.
One says that the Young function θ is strict *if it is finite and* $\lim_{s\to\infty} \theta(s)/s = \infty$.

Date: April 23, 07.

In the literature, Young functions are sometimes not required to be lower semicontinuous or are allowed to be trivial in the sense of condition (iii). These restrictions are introduced in these notes in order to escape from the uninteresting trivial cases and to work with convex conjugacy without trouble. Doing so, one doesn't loose generality.

Examples 1.2. The following functions are Young functions.

(1) $\theta_p(s) := |s|^p/p$ with $p \geq 1$;

(2) $\theta_\infty(s) = \iota_{[-1,+1]}(s) := \begin{cases} 0, & \text{if } -1 \leq s \leq +1 \\ +\infty, & \text{otherwise.} \end{cases}$;

(3) $\theta_{\exp}(s) = e^{|s|} - 1$; $\tilde{\theta}_{\exp}(s) := e^{|s|} - |s| - 1$.

Remarks 1.3.

(1) Properties (i) and (ii) imply that θ achieves its minimum at 0 and is increasing on $[0, \infty)$.
(2) Under (i) and (ii), one has (iii) if and only if there exist $s_0, s_1 > 0$ such that $\theta(s_0) \in [0, \infty)$ and $\theta(s_1) \in (0, \infty]$.
(3) Any Young function satisfies $\lim_{s\to\infty} \theta(s) = +\infty$.
(4) As a Young function is convex on $\mathbb{R}$, it is continuous on the interior of its effective domain. In particular, a finite Young function is continuous.

Convex conjugacy of Young functions will appear to be linked to the duality of Orlicz spaces. Recall that the convex conjugate of θ is defined by

$$\theta^*(t) := \sup_{s\in\mathbb{R}}\{st - \theta(s)\} \in [0, \infty], \quad t \in \mathbb{R}.$$

Proposition 1.4.

(1) θ is a Young function if and only if θ^* is a Young function.
(2) Let θ be a Young function. We have

$$[\theta \textit{ is strict}] \Leftrightarrow [\theta^* \textit{ is strict}] \Leftrightarrow [\theta \textit{ and } \theta^* \textit{ are finite}].$$

Proof. Left in exercise. □

Let the space $\mathcal{X}$ be given with a σ-field and a σ-finite positive measure μ. Recall that μ is said to be σ-finite if there exists a sequence $(\mathcal{X}_k)_{k\geq1}$ of measurable subsets of $\mathcal{X}$ such that

$$\bigcup_{k\geq1} \mathcal{X}_k = \mathcal{X} \quad \text{and} \quad \mu(\mathcal{X}_k) < \infty, \ \forall k \geq 1. \tag{1.5}$$

Without loss of generality, considering $\bigcup_{i\leq k} \mathcal{X}_i$, one can assume that $(\mathcal{X}_k)_{k\geq1}$ is an increasing sequence.

Definition 1.6 (Luxemburg norm)**.** *Let θ be a Young function. For any measurable function f on $\mathcal{X}$,*

$$\|f\|_\theta := \inf\left\{b > 0; \int_\mathcal{X} \theta(f/b)\, d\mu \leq 1\right\} \in [0, \infty]$$

where it is understood that $\inf(\emptyset) = +\infty$.

It is the gauge of the set $U = \{f \text{ measurable}; \int_\mathcal{X} \theta(f)\, d\mu \leq 1\}$.

Examples 1.7 (Fundamental examples). The Young function θ_p at Example 1.2-1 leads to the usual norm $p^{1/p}\|f\|_{\theta_p} = \|f\|_p = [\int_\mathcal{X} |f|^p\, d\mu]^{1/p}$ on $L_p(\mu)$, while θ_∞ at Example 1.2-2 gives the usual norm $\|f\|_{\theta_\infty} = \|f\|_\infty = \mu\text{-ess sup}|f|$ on $L_\infty(\mu)$.

The functions $\theta_{\exp}$ and $\tilde{\theta}_{\exp}$ given at Example 1.2-3 will give rise to function spaces entering the class of Orlicz spaces which are useful to study the relative entropy, as will be seen later.

Lemma 1.8. *For any measurable function f on $\mathcal{X}$, $\|f\|_\theta = 0$ if and only if $f = 0$ μ-almost everywhere.*

Proof. Clearly, $\|f\|_\theta = 0$ if and only if $\int_{\mathcal{X}} \theta(f/\epsilon)\, d\mu \le 1$, $\forall \epsilon > 0$. It follows that the equivalence (a) below is satisfied:

$$\begin{aligned} \|f\|_\theta = 0 \quad &\overset{(a)}{\Leftrightarrow} \quad \int_{\mathcal{X}} \theta(af)\, d\mu = 0, \qquad \forall a > 0 \\ &\overset{(b)}{\Leftrightarrow} \quad \theta(af) = 0, \mu\text{-a.e.} \quad \forall a > 0 \\ &\overset{(c)}{\Leftrightarrow} \quad f = 0, \mu\text{-a.e.} \end{aligned}$$

where (b) holds since $\theta \ge 0$ and (c) is a consequence of the existence of some $s_1 > 0$ such that $\theta(s_1) > 0$, see Remark 1.3-2. □

Identification of almost everywhere equal functions. As with L_p spaces, one identifies the functions which are μ-almost everywhere equal. This means that one works with the equivalence classes of the equivalence relation defined by the μ-almost everywhere equality. From now on, this will be done without further mention.
Consequently, one writes:

$$\|f\|_\theta = 0 \Leftrightarrow f = 0. \tag{1.9}$$

Lemma 1.10. *If $0 < \|f\|_\theta < \infty$, then $\int_{\mathcal{X}} \theta(f/\|f\|_\theta)\, d\mu \le 1$.*
In particular, $\|f\|_\theta \le 1$ is equivalent to $\int_{\mathcal{X}} \theta(f)\, d\mu \le 1$.

Proof. For all $b > \|f\|_\theta$, $\int_{\mathcal{X}} \theta(f/b)\, d\mu \le 1$. Letting b decrease to $\|f\|_\theta$, one obtains the first result by monotone convergence. The second statement follows from this and Lemma 1.8. □

Proposition 1.11. *The gauge $\|\cdot\|_\theta$ is a norm on the vector space of all the measurable functions f such that $\|f\|_\theta < \infty$.*

Proof. It is already seen that (1.9) holds under the identification of a.e. equal functions. It is clear that for all real λ, $\|\lambda f\|_\theta = |\lambda| \|f\|_\theta$.
It remains to prove the triangle inequality. Let f and g be two measurable functions such that $0 < \|f\|_\theta + \|g\|_\theta < \infty$. Then,

$$\begin{aligned} & \int_{\mathcal{X}} \theta\left(\frac{f+g}{\|f\|_\theta + \|g\|_\theta}\right) d\mu \\ = \;& \int_{\mathcal{X}} \theta\left(\frac{\|f\|_\theta}{\|f\|_\theta + \|g\|_\theta}\frac{f}{\|f\|_\theta} + \frac{\|g\|_\theta}{\|f\|_\theta + \|g\|_\theta}\frac{g}{\|g\|_\theta}\right) d\mu \\ \le \;& \frac{\|f\|_\theta}{\|f\|_\theta + \|g\|_\theta}\int_{\mathcal{X}} \theta\left(\frac{f}{\|f\|_\theta}\right) d\mu + \frac{\|g\|_\theta}{\|f\|_\theta + \|g\|_\theta}\int_{\mathcal{X}} \theta\left(\frac{g}{\|g\|_\theta}\right) d\mu \\ \le \;& 1 \end{aligned}$$

where the last but one inequality follows from the convexity of θ and the last inequality from Lemma 1.10. Therefore, $\|f + g\|_\theta \le \|f\|_\theta + \|g\|_\theta$.
As a consequence, the set of all measurable functions f such that $\|f\|_\theta < \infty$ is a vector space. □

Definitions 1.12 (Orlicz spaces). *Let θ be a Young function. One defines the Orlicz spaces*

$$\begin{aligned} L_\theta(\mathcal{X},\mu) &:= \left\{ f \textit{ measurable}; \exists a>0, \int_{\mathcal{X}} \theta(af)\,d\mu < \infty \right\} \\ M_\theta(\mathcal{X},\mu) &:= \left\{ f \textit{ measurable}; \forall a>0, \int_{\mathcal{X}} \theta(af)\,d\mu < \infty \right\} \end{aligned}$$

Let us drop $(\mathcal{X},\mu)$ and write $L_\theta = L_\theta(\mathcal{X},\mu)$ and $M_\theta = M_\theta(\mathcal{X},\mu)$. Clearly

$$M_\theta \subset L_\theta$$

and

$$L_\theta = \{f \text{ measurable}; \|f\|_\theta < \infty\}.$$

Hence, L_θ is a vector space. Of course, M_θ is a vector subspace of L_θ.
One calls respectively L_θ and M_θ the *large* and *small* Orlicz spaces. Very often, the large space L_θ is simply called the Orlicz space.

Examples 1.13. One refers to the Young functions introduced at Examples 1.2.

(1) With θ_p, $M_{\theta_p} = L_{\theta_p} = L_p$;
(2) With θ_∞, $M_{\theta_\infty} = \{0\}$ and $L_{\theta_\infty} = L_\infty$;
(3) With $\theta_{\exp}(s) := e^{|s|} - 1$ or $\tilde\theta_{\exp}(s) := e^{|s|} - |s| - 1$, in general we have the strict inclusions

$$\{0\} \subsetneqq M_{\theta_{\exp}} \subsetneqq L_{\theta_{\exp}}.$$

To see this, consider the example where $\mathcal{X} = \mathbb{R}$, μ is the standard Gaussian probability measure $\mathcal{N}(0,1)$. Then, the function $f(x) = x^2$ is in $L_{\theta_{\exp}}$ but not in $M_{\theta_{\exp}}$.
(4) If μ is a bounded measure, the functions $\theta_{\exp}$ and $\tilde\theta_{\exp}$ define the same Orlicz spaces. On the other hand, if μ is unbounded, this is not true anymore. This will be illustrated later.

It follows from Proposition 1.4 that if L_θ is an Orlicz space, then L_{θ^*} is also an Orlicz space.

Proposition 1.14 (Hölder's inequality). *For all $f \in L_\theta$ and $g \in L_{\theta^*}$,*

$$\int_{\mathcal{X}} |fg|\,d\mu \le 2\|f\|_\theta \|g\|_{\theta^*}.$$

In particular, $fg \in L_1$.

Proof. If $\|f\|_\theta = 0$ or $\|g\|_{\theta^*} = 0$, one concludes with Lemma 1.8. Assume now that $0 < \|f\|_\theta, \|g\|_{\theta^*}$. Because of Young's inequality: $st \le \theta(s) + \theta^*(t)$, $\frac{f}{\|f\|_\theta}\frac{g}{\|g\|_{\theta^*}} \le \theta(f/\|f\|_\theta) + \theta^*(g/\|g\|_{\theta^*})$. Hence,

$$\frac{1}{\|f\|_\theta \|g\|_{\theta^*}} \int_{\mathcal{X}} fg\,d\mu \le \int_{\mathcal{X}} \theta(f/\|f\|_\theta)\,d\mu + \int_{\mathcal{X}} \theta^*(g/\|g\|_{\theta^*})\,d\mu \le 2.$$

□

Proposition 1.15 ($\mu(\mathcal{X}) < \infty$). *Assume that μ is bounded, then*

$$L_\theta \subset L_1$$

where the inclusion is continuous from $(L_\theta, \|\cdot\|_\theta)$ to $(L_1, \|\cdot\|_1)$.

Proof. There exist $u > 0$ and $v \geq 0$ such that $\theta(s) \geq us - v$, $\forall s \geq 0$. Let $f \in L_\theta$. For $a > 0$ small enough, $a\int_{\mathcal{X}} |f|\, d\mu \leq \frac{1}{u}\int_{\mathcal{X}}[\theta(af)+v]\, d\mu = \frac{1}{u}\int_{\mathcal{X}}\theta(af)\, d\mu + v\mu(\mathcal{X})/u < \infty$. Therefore, $L_\theta \subset L_1$ and this inclusion is continuous since with $c > 0$ taken sufficiently small for the inequality $u/c - v\mu(\mathcal{X}) \geq 1$ to hold, we have

$$\int_{\mathcal{X}} \theta\left(\frac{f}{c\|f\|_1}\right) d\mu \geq \int_{\mathcal{X}} \frac{u}{c}\frac{f}{\|f\|_1}\, d\mu - v\mu(\mathcal{X}) = u/c - v\mu(\mathcal{X}) \geq 1.$$

This means that $c\|f\|_1 \leq \|f\|_\theta$ for all $f \in L_\theta$. □

Lemma 1.16. *Let $(f_n)_{n\geq 1}$ be a sequence in L_θ. Then, the following assertions are equivalent:*

(a) $\lim_{n\to\infty} \|f_n\|_\theta = 0$;
(b) For all $a > 0$, $\limsup_{n\to\infty} \int_{\mathcal{X}} \theta(af_n)\, d\mu \leq 1$;
(c) For all $a > 0$, $\lim_{n\to\infty} \int_{\mathcal{X}} \theta(af_n)\, d\mu = 0$.

Proof. The equivalence $(a) \Leftrightarrow (b)$ is a direct consequence of the definition of $\|f\|_\theta$. Of course $(c) \Rightarrow (b)$ is obvious. As θ is convex and $\theta(0) = 0$, for all $s \geq 0$ and $0 < \epsilon \leq 1$, $\theta(s) = \theta((1-\epsilon)0 + \epsilon(s/\epsilon)) \leq (1-\epsilon)\theta(0) + \epsilon\theta(s/\epsilon)$. That is

$$\theta(s) \leq \epsilon\theta(s/\epsilon), \quad s \geq 0, 0 < \epsilon \leq 1 \tag{1.17}$$

from which $(b) \Rightarrow (c)$ follows easily. □

Proposition 1.18. *The normed spaces $(L_\theta, \|\cdot\|_\theta)$ and $(M_\theta, \|\cdot\|_\theta)$ are Banach spaces.*

Proof. Let us begin proving that L_θ is complete. Let $(f_n)_{n\geq 1}$ be a Cauchy sequence in L_θ. As μ is σ-finite there exists a countable partition $(X_k; k \geq 1)$ of measurable subsets of $\mathcal{X}$ such that $\mu(X_k) < \infty$ for all $k \geq 1$. In restriction to each X_k furnished with the measure $\mu_k(\cdot) = \mu(X_k \cap \cdot)$, $(f_n)_{n\geq 1}$ is a Cauchy sequence in $L_\theta(X_k, \mu_k)$. Thanks to Proposition 1.15, it is also Cauchy in $L_1(X_k, \mu_k)$. As L_1 is complete, it is convergent in L_1 and one can extract a subsequence which converges μ_k-a.e. pointwise to f on X_k. Using the diagonal extraction procedure, one can extract a subsequence $(f_{n_k})_{k\geq 1}$ which converges μ-a.e. pointwise to f on the whole space $\mathcal{X}$.

Let $a > 0$. By Lemma 1.16 there exists a large enough integer $N(a)$ such that

$$\int_{\mathcal{X}} \theta(a(f_m - f_n))\, d\mu \leq 1, \quad \forall m, n \geq N(a).$$

With Fatou's lemma this gives

$$\int_{\mathcal{X}} \theta(a(f_m - f))\, d\mu \leq \liminf_{k\to\infty} \int_{\mathcal{X}} \theta(a(f_m - f_{n_k}))\, d\mu \leq 1, \quad \forall m \geq N(a).$$

Therefore, $f_m - f$ belongs to L_θ. But $f_m \in L_\theta$, so that $f \in L_\theta$.
Moreover, as $\limsup_{m\to\infty} \int_{\mathcal{X}} \theta(a(f_m - f))\, d\mu \leq 1$ for all $a > 0$, we have $\lim_{m\to\infty} \|f_m - f\|_\theta = 0$. This proves that L_θ is complete.
It remains to show that M_θ is closed in L_θ. Let $(f_n)_{n\geq 1}$ be a sequence in M_θ which converges to f in L_θ. As θ is convex, for all $a > 0$ and $n \geq 1$,

$$\begin{aligned}\int_{\mathcal{X}} \theta(af/2)\, d\mu &= \int_{\mathcal{X}} \theta\left(\frac{1}{2}af_n + \frac{1}{2}a(f - f_n)\right) d\mu \\ &\leq \frac{1}{2}\underbrace{\int_{\mathcal{X}} \theta(af_n)\, d\mu}_{A} + \frac{1}{2}\underbrace{\int_{\mathcal{X}} \theta(a(f - f_n))\, d\mu}_{B}.\end{aligned}$$

But A is finite for all n and *all* a since $f_n \in M_\theta$ and $B \le 1$ as soon as $n \ge N(a)$ for some large enough $N(a)$. Hence, $\int_{\mathcal{X}} \theta(af/2)\, d\mu < \infty$ for all a, that is $f \in M_\theta$. □

Remark. A direct proof of this result is possible, following the standard proofs of the completeness of L^p. See [4] for instance. Here, we relied on some known properties of L_1 to spare time.

Definition 1.19 (Orlicz norm)**.** *Let us introduce another norm on* L_θ *:*

$$|f|_\theta := \sup\left\{\int_{\mathcal{X}} fg\, d\mu; g \in L_{\theta^*}, \int_{\mathcal{X}} \theta^*(g)\, d\mu \le 1\right\}, \quad f \in L_\theta.$$

Note that by Hölder's inequality, this integral is well-defined. Because of Lemma 1.10, one can also write

$$|f|_\theta := \sup\left\{\int_{\mathcal{X}} fg\, d\mu; g \in L_{\theta^*}, \|g\|_{\theta^*} \le 1\right\}.$$

Proposition 1.20. $|\cdot|_\theta$ *is a norm on* L_θ *which is equivalent to* $\|\cdot\|_\theta$*. More precisely,*

$$\|f\|_\theta \le |f|_\theta \le 2\|f\|_\theta, \quad f \in L_\theta.$$

Proof. • Let us show that $|f|_\theta = 0$ implies that $f = 0$, μ-a.e. Take $f \neq 0$. Since μ is σ-finite, there is some set A such that $0 < \mu(A) < \infty$ and $A \subset \{f \neq 0\}$. Choose $g_o = c\mathbf{1}_A$ with $c > 0$ such that $\int_{\mathcal{X}} \theta^*(g)\, d\mu = \theta^*(c)\mu(A) \le 1$ (this is possible since the Young function θ^* is continuous at zero) and remark that $|f|_\theta \ge \int_{\mathcal{X}} |f| g_o\, d\mu > 0$.
It is clear that $|\lambda f|_\theta = |\lambda||f|_\theta$ and $|f+g|_\theta \le |f|_\theta + |g|_\theta$ for all $f, g \in L_\theta$ and all real λ. We have shown that $|\cdot|_\theta$ is a norm.
• For all $g \in M_{\theta^*}$ such that $\int_{\mathcal{X}} \theta^*(g)\, d\mu \le 1$, we have $\|g\|_{\theta^*} \le 1$ and Hölder's inequality states that $\int_{\mathcal{X}} fg\, d\mu \le 2\|f\|_\theta \|g\|_{\theta^*} \le 2\|f\|_\theta$. That is $|f|_\theta \le 2\|f\|_\theta$.
• To prove that $\|f\|_\theta \le |f|_\theta$, it remains to show that

$$\int_{\mathcal{X}} \theta(f/|f|_\theta)\, d\mu \le 1 \tag{1.21}$$

for any positive f in L_θ. Denote $s_o = \sup \operatorname{dom} \theta \in (0, \infty]$. Let us assume for a while that

$$\text{either } \theta(s_o) = \infty \text{ or } \partial\theta(s_o) \neq \emptyset. \tag{1.22}$$

Under this assumption, for all s in $\operatorname{dom}\theta$, the subdifferential $\partial\theta(s)$ is non-empty and Young's equality is

$$\theta(s) + \theta^*(\theta'(s)) = s\theta'(s), \quad s \in \operatorname{dom}\theta \tag{1.23}$$

where we pick some $\theta'(s) \in \partial\theta(s)$ for all $s \in \operatorname{dom}\theta$. Of course, if θ is differentiable, then $\theta'(s)$ is the usual derivative. As θ is convex, the function θ' is increasing so that it is measurable. Now, considering the measurable function

$$g := \theta'(f/|f|_\theta)\, \mathbf{1}_A,$$

where A is any measurable subset of $\mathcal{X}$, one obtains that

$$\int_A \theta(f/|f|_\theta)\, d\mu + \int_{\mathcal{X}} \theta^*(g)\, d\mu = \int_{\mathcal{X}} \frac{f}{|f|_\theta} g\, d\mu \tag{1.24}$$

where these integrals take their values in $[0, \infty]$ and the convention that $\theta'(0) = 0$ is adopted so that Young's equality is still satisfied outside of A.

If $\int_A \theta^*(g)\, d\mu \le 1$, we get $\int_{\mathcal{X}} fg\, d\mu \le |f|_\theta$, that is $\int_{\mathcal{X}} \frac{f}{|f|_\theta} g\, d\mu \le 1$. As $\int_{\mathcal{X}} \theta^*(g)\, d\mu \ge 0$, one obtains with (1.24) that $\int_A \theta(f/|f|_\theta)\, d\mu \le \int_{\mathcal{X}} \frac{f}{|f|_\theta} g\, d\mu \le 1$.

If $1 < \int_A \theta^*(g)\,d\mu < \infty$, let us consider the function $g/\int_{\mathcal{X}} \theta^*(g)\,d\mu$. Since θ^* is a Young function, (1.17) tells us that for all $0 < \alpha \le 1$, $\theta^*(\alpha t) \le \alpha\theta^*(t)$, $t \ge 0$. In particular, $\theta^*(g/\int_{\mathcal{X}} \theta^*(g)\,d\mu) \le \theta^*(g)/\int_{\mathcal{X}} \theta^*(g)\,d\mu$ and $\int_{\mathcal{X}} \theta^*(g/\int_{\mathcal{X}} \theta^*(g)\,d\mu)\,d\mu \le 1$. Hence, $\int_{\mathcal{X}} fg/(\int_{\mathcal{X}} \theta^*(g)\,d\mu)\,d\mu \le |f|_\theta$. In other words, $\int_{\mathcal{X}} \frac{f}{|f|_\theta} g\,d\mu \le \int_{\mathcal{X}} \theta^*(g)\,d\mu$. Together with (1.24), this implies that $\int_A \theta\left(\frac{f}{|f|_\theta}\right)\,d\mu = 0$.

At this point of the proof, we have shown that $\int_A \theta\left(\frac{f}{|f|_\theta}\right)\,d\mu \le 1$ provided that $\int_A \theta^*(\theta'(f/|f|_\theta))\,d\mu$ is finite.

• Now, let's have a look at the situation where $\int_{\mathcal{X}} \theta^*(\theta'(f/|f|_\theta))\,d\mu = \infty$. If (1.22) is satisfied, the function $\theta^*(\theta'(s))$ is finite everywhere by Proposition 1.37-c of [2]. This implies that $(B_k)_{k\ge1}$ with $B_k = \{\theta^*[\theta'(f/|f|_\theta)] \le k\}$ is an increasing sequence such that $\cup_{k\ge1} B_k = \mathcal{X}$. As μ is σ-finite there exists another *increasing* sequence $(\mathcal{X}_k)_{k\ge1}$ of measurable subsets of $\mathcal{X}$ satisfying (1.5). Taking $A_k = B_k \cap \mathcal{X}_k$, it follows that $(A_k)_{k\ge1}$ is an increasing sequence such that $\lim_{k\to\infty} A_k = \mathcal{X}$ and $\int_{A_k} \theta^*[\theta'(f/|f|_\theta)]\,d\mu < \infty$ for each k. With our previous result, this implies that $\int_{A_k} \theta\left(\frac{f}{|f|_\theta}\right)\,d\mu \le 1$ for all k. One concludes by monotone convergence that $\int_{\mathcal{X}} \theta\left(\frac{f}{|f|_\theta}\right)\,d\mu \le 1$ which is the desired result: $\|f\|_\theta \le |f|_\theta$.

• It remains to get rid of the technical requirement (1.22). The remaining case is: $\theta(s_o) < \infty$ with $\theta'(s_o) = \infty$. Instead of g as above, consider the sequence $(g_n)_{n\ge1}$ defined by

$$g_n := \theta'(f/|f|_\theta)\,\mathbf{1}_A + n\mathbf{1}_B,$$

where A and B are any measurable subset of $\mathcal{X}$ such that $A \subset \{f/|f|_\theta \ne s_o\}$ and $B \subset \{f/|f|_\theta = s_o\}$ with $\mu(B) < \infty$.

Now, (1.24) with g_n instead of g is not an equality anymore, the difference providing from the contribution of the set B where Young's equality (1.23) is replaced by the asymptotic equality

$$\theta(s_o) + \theta^*(n) = s_o n + \epsilon_n$$

with $\lim_{n\to\infty} \epsilon_n = 0$. With g_n, (1.24) is replaced by

$$\begin{aligned}
&\int_{A\cup B} \theta(f/|f|_\theta)\,d\mu \\
=\ &\int_A \theta(f/|f|_\theta)\,d\mu + \mu(B)\theta(s_o) \\
=\ &\int_A \frac{f}{|f|_\theta} g_n\,d\mu - \int_A \theta^*(g_n)\,d\mu + \mu(B)(s_o n - \theta^*(n) + \epsilon_n) \\
=\ &\int_{\mathcal{X}} \frac{f}{|f|_\theta} g_n\,d\mu - \int_{\mathcal{X}} \theta^*(g_n)\,d\mu + \mu(B)\epsilon_n
\end{aligned}$$

Reasoning as above, one obtains that $\int_{A\cup B} \theta\left(\frac{f}{|f|_\theta}\right)\,d\mu \le 1 + \mu(B)\epsilon_n$ provided that $\int_A \theta^*(\theta'(f/|f|_\theta))\,d\mu$ and $\mu(B)$ are finite. As this is valid for all $n \ge 1$, letting n tend to infinity leads us to $\int_{A\cup B} \theta\left(\frac{f}{|f|_\theta}\right)\,d\mu \le 1$. But now, the function $\theta^*(\theta'(f/|f|_\theta))$ is finite on any $A \subset \{f/|f|_\theta \ne s_o\}$ and one concludes as before, taking advantage of the σ-finiteness of μ, that (1.21) holds true. □

Remark 1.25. One may think of

$$N_\theta(f) := \sup\left\{\int_{\mathcal{X}} fg\,d\mu; g \in M_{\theta^*}, \|g\|_{\theta^*} \le 1\right\}$$

and

$$N_\theta^B(f) := \sup\left\{\int_{\mathcal{X}} fg\,d\mu; g \in B_{\mathcal{X}}, \|g\|_{\theta^*} \le 1\right\}$$

which are similar to $|f|_\theta$, but where the suprema are taken over $g \in M_{\theta^*}$ and $g \in B_{\mathcal{X}}$ rather than L_{θ^*}. Following the previous proof, one sees that N_θ^B is a norm.
But, one must be careful with N_θ since $[N_\theta(f) = 0 \Rightarrow f = 0]$ fails unless it is assumed that $\operatorname{dom}\theta^* = \mathbb{R}$. Indeed, if $\operatorname{dom}\theta^* \subsetneq \mathbb{R}$, $M_{\theta^*} = \{0\}$ and $N_\theta(f) = 0$ for all f. On the other hand, $\operatorname{dom}\theta^* = \mathbb{R}$ insures that $g_o = c\mathbf{1}_A$ at the first step of the proof is in $B_{\mathcal{X}}$, so that $[N_\theta(f) = 0 \Rightarrow f = 0]$ holds.
The rest of the proof still works with the norms $N_\theta(f)$ and $N_\theta^B(f)$. This leads us to

$$\|f\|_\theta \le N_\theta(f) \le 2\|f\|_\theta$$

and

$$\|f\|_\theta \le N_\theta^B(f) \le 2\|f\|_\theta$$

for all $f \in L_\theta$.

2. A first step towards duality

Because of Hölder's inequality, any $h \in L_{\theta^*}$ defines a continuous linear form via

$$\ell_h(f) := \int_{\mathcal{X}} hf\,d\mu, \quad f \in L_\theta \tag{2.1}$$

with an operator norm $\|\ell_h\| \le 2\|h\|_{\theta^*}$. The main result of this section is the following

Theorem 2.2 ($\operatorname{dom}\theta = \mathbb{R}$)**.** *Assume that θ is finite. Then, the topological dual space M_θ' of $(M_\theta, \|\cdot\|_\theta)$ is isomorphic to L_{θ^*} :*

$$M_\theta' \cong L_{\theta^*}.$$

This means that one can associate to any $\ell \in M_\theta'$, a unique $h \in L_{\theta^*}$ such that $\ell = \ell_h$ where ℓ_h is defined at (2.1).

Remarks 2.3.

1. When θ takes infinite values, M_θ is the null space and Theorem 2.2 obviously fails.
2. Defining for any $1 \le p < \infty$, $\theta_p(s) = |s|^p/p, s \in \mathbb{R}$, we have for all $1 < p < \infty$, $M_{\theta_p} = L_{\theta_p} = L_p$ and $\theta_p^* = \theta_{p^*}$ where

   $$1/p + 1/p^* = 1.$$

   Theorem 2.2 appears to be an extension of the usual dual representation theorem of L_p spaces which states that

   $$L_p' \cong L_{p^*}$$

   for all $1 < p < \infty$.

The proof of Theorem 2.2 will be done after some preliminary lemmas are established.

Lemma 2.4. *Let h be a measurable function.*

(a) If $hf \in L_1$ for all $f \in L_\theta$, then $h \in L_{\theta^*}$.
(b) Suppose that θ is finite. If $hf \in L_1$ for all $f \in M_\theta$, then $h \in L_{\theta^*}$.

Proof. •*Proof of (a).* For all h measurable, $n,k\geq 1$, denote $h_n := |h|\wedge n$ and define the linear form

$$\ell_{n,k}(f) := \int_{\mathcal{X}} \mathbf{1}_{\mathcal{X}_k} h_n f\, d\mu, \quad f\in L_\theta$$

where $(\mathcal{X}_k)_{k\geq 1}$ is the increasing sequence already encountered at (1.5). As by Hölder's inequality $|\ell_{n,k}(f)| \leq 2\|\mathbf{1}_{\mathcal{X}_k} h_n\|_{\theta^*}\|f\|_\theta$, $\ell_{n,k}$ is a *continuous* linear form on the Banach space L_θ and for all $n,k\geq 1$

$$|\ell_{n,k}(f)| \leq \int_{\mathcal{X}} \mathbf{1}_{\mathcal{X}_k} h_n |f|\, d\mu \leq \int_{\mathcal{X}} |h||f|\, d\mu := A(f) < \infty.$$

Therefore, one can apply the uniform boundedness principle (Banach-Steinhaus theorem) to assert that

$$B := \sup_{n,k\geq 1} \|\ell_{n,k}\| < \infty$$

where $\|\ell_{n,k}\| = \sup\{\ell_{n,k}(f); f\in L_\theta, \|f\|_\theta \leq 1\}$. It follows that

$$\begin{aligned}
|h|_{\theta^*} &= \sup\left\{\int_{\mathcal{X}} |h| f\, d\mu; f\in L_\theta, \|f\|_\theta\leq 1\right\} \\
&\overset{(a)}{=} \sup\left\{\lim_{n,k\to\infty} |\ell_{n,k}(f)|; f\in L_\theta, \|f\|_\theta\leq 1\right\} \\
&\leq \sup\{B\|f\|_\theta; f\in L_\theta, \|f\|_\theta \leq 1\} \\
&= B < \infty.
\end{aligned}$$

By Proposition 1.20, $\|h\|_{\theta^*} \leq |h|_{\theta^*} < \infty$. This shows that h belongs to L_{θ^*}.
It remains to justify equality (a) in the previous sequence of inequalities. Clearly, one can restrict the proof to the case where f is nonnegative and the result follows by monotone convergence.
•*Proof of (b).* The same proof works, replacing $|h|_{\theta^*}$ by $N_{\theta^*}(h)$ and taking Remark 1.25 into account. □

Lemma 2.5. *The space $B_{\mathcal{X}}\cap M_\theta$ of all bounded functions in M_θ is everywhere dense in M_θ.*

Proof. First consider the case where θ is not finite. Then, M_θ reduces to $\{0\}$ and there is nothing to prove.
Now, let us assume that θ is finite and let f be in M_θ. For all $n\geq 1$, denote $f_n := (-n)\vee f\wedge n \in B_{\mathcal{X}}\cap M_\theta$. Clearly, $\lim_{n\to\infty} f_n = f$ pointwise and $\|f-f_n\|_\theta = \|f\mathbf{1}_{\{|f|>n\}}\|_\theta$. But, for *all* $a>0$, $\lim_{n\to\infty}\int_{\mathcal{X}} \theta(af\mathbf{1}_{\{|f|>n\}})\, d\mu = 0$ by dominated convergence. Hence, $\lim_{n\to\infty}\|f-f_n\|_\theta = 0$, by Lemma 1.16. □

Remark 2.6. One already knows that M_θ is closed in L_θ (see Proposition 1.18) and that it happens that $B_{\mathcal{X}}\subset M_\theta \subsetneq L_\theta$ (see Example 1.13-3 where μ is bounded), therefore it happens that $B_{\mathcal{X}}\cap L_\theta$ isn't everywhere dense in L_θ.

We are now ready to write the proof of Theorem 2.2.

Proof of Theorem 2.2. • The easy inclusion is $L_{\theta^*}\subset M'_\theta$. Indeed, Hölder's inequality implies that for all $h\in L_{\theta^*}$, ℓ_h defined at (2.1) is continuous on M_θ.
• Let us prove the converse inclusion: $M'_\theta\subset L_{\theta^*}$. Let ℓ belong to M'_θ. One first shows that ℓ is a measure which is absolutely continuous with respect to μ :

$$\ell \ll \mu.$$

Clearly, ℓ is an additive functional on M_θ. To be a measure, it still needs to be σ-additive. To see this, we are going to prove that

$$\lim_{n\to\infty} \ell(g_n f) = 0 \tag{2.7}$$

for any $f \in M_\theta$ and any decreasing sequence $(g_n)_{n\geq 1}$ in $B_{\mathcal{X}}$ such that $0 \leq g_n \leq 1$ and $\lim_{n\to\infty} g_n(x) = 0$, $\forall x \in \mathcal{X}$. As $\operatorname{dom}\theta = \mathbb{R}$, by dominated convergence one sees that for *all* $a \geq 0$, $\lim_{n\to\infty} \int_{\mathcal{X}} \theta(ag_n f)\,d\mu = 0$ since for all $n \geq 1$, $0 \leq \theta(ag_n f) \leq \theta(af)$ and $\int_{\mathcal{X}} \theta(af)\,d\mu < \infty$. Hence, by Lemma 1.16, $\lim_{n\to\infty} \|g_n f\|_\theta = 0$. But ℓ is continuous on M_θ, so that (2.7) is satisfied. This proves that for all $f \in M_\theta$, there exists a signed measure λ_f such that $\ell(gf) = \int_{\mathcal{X}} g\,d\lambda_f$, $\forall g \in B_{\mathcal{X}}$.
On the other hand, for any $g \in B_{\mathcal{X}}$ such that $g = 0$ μ-a.e., we have $\ell(gf) = \ell(0) = 0$. This implies that λ_f is absolutely continuous with respect to μ and by Radon-Nykodym's theorem there exists some function h_f in $L_1(\mathcal{X}, \lambda_f)$ such that

$$\ell(gf) = \int_{\mathcal{X}} g h_f\,d\mu, \quad \forall g \in B_{\mathcal{X}}.$$

With $f \in B_{\mathcal{X}} \cap M_\theta$, as $fg \in B_{\mathcal{X}}$ for all $g \in B_{\mathcal{X}}$, writing $f = 1 \times f$, one obtains that $h_f = fh_1$, μ-a.e. Hence,

$$\ell(f) = \int_{\mathcal{X}} fh\,d\mu, \quad \forall f \in B_{\mathcal{X}} \cap M_\theta \tag{2.8}$$

where we have taken $h = h_1$.
Denote $f_+ = 0 \vee f$, $f_- = 0 \vee (-f)$, $\eta_+ = \mathbf{1}_{\{h\geq 0\}}$ and $\eta_- = \mathbf{1}_{\{h<0\}}$. Of course, for all $f \in M_\theta$, $\ell(f) = \ell(\eta_+ f_+) + \ell(\eta_- f_+) - \ell(\eta_+ f_-) - \ell(\eta_- f_-)$ where the functions $\eta_+ f_+, \eta_- f_+, \eta_+ f_-$ and $\eta_- f_-$ are nonnegative. As their products with h have constant sign, one can restrict our attention to the case where $f \geq 0$ and $h \geq 0$.
For all $f \geq 0$ in M_θ, as a consequence of the dominated convergence theorem (see Lemma 2.5), we have $\lim_{k\to\infty} f \wedge k = f$ in M_θ. Since ℓ is continuous, we obtain $\lim_{k\to\infty} \ell(f \wedge k) = \ell(f)$. By monotone convergence, $\lim_{k\to\infty} \int_{\mathcal{X}} (f \wedge k)h\,d\mu = \int_{\mathcal{X}} fh\,d\mu \in [0, \infty]$. These two limits, (2.8) and the decomposition into functions with constant sign lead us to

$$\ell(f) = \int_{\mathcal{X}} fh\,d\mu, \quad \forall f \in M_\theta.$$

One concludes with Lemma 2.4-b that $h \in L_{\theta^*}$. This completes the proof of the theorem. $\square$

References

[1] M.A. Krasnosel'skii and Ya.B. Rutickii. *Convex functions and Orlicz spaces.* P. Noordfoff Ltd., Groningen, 1961.
[2] C. Léonard. A set of lecture notes on convex optimization with some applications to probability theory. Incomplete draft. Available online via `http://www.cmap.polytechnique.fr/~leonard/`.
[3] M.M. Rao and Z.D. Ren. *Theory of Orlicz spaces*, volume 146 of *Pure and Applied Mahtematics*. Marcel Dekker, Inc., 1991.
[4] W. Rudin. *Real and complex analysis.* McGraw-Hill, 1987.

Modal-X, Université Paris 10. Bât. G, 200 av. de la République. 92001 Nanterre Cedex, France

CMAP, École Polytechnique. 91128 Palaiseau Cedex, France
E-mail address: `christian.leonard@polytechnique.fr`