LESSON 24 - STUDY GUIDE

Abstract. This will be the first of two lessons focused on the machinery required to prove pointwise almost everywhere convergence of approximate identities or, more generally, sequences of Lebesgue measurable functions. That machinery consists of maximal operators, weak type bounds and the Marcinkiewicz interpolation theorem, and they are already part of the tools of modern harmonic analysis, which we are finally arriving at. For this lesson we will concentrate on the concepts of weak L^p spaces, weak type estimates and the Marcinkiewicz interpolation theorem, which, together with Riesz-Thorin, make up the two fundamental theorems of the theory of interpolation of operators. We will leave the study of maximal operators and, in particular, the Hardy-Littlewood maximal function, to the next lesson.

1. Weak L^p spaces, weak type inequalities and the Marcinkiewicz interpolation theorem.

Study material: The most complete and detailed presentation of this lesson's subject is in Folland's book [1], in section **6.4 Distribution Functions and Weak L^p** and section **6.5 Interpolation of L^p spaces** from chapter **6 L^p Spaces**. Section 6.4 is, in fact, the only section in this chapter that we had not covered yet, because we had skipped it when we did the theory of L^p spaces at the beginning of the course, leaving it for now. It should be pointed out that Folland's complete presentation of the fully general version of the Marcinkiewicz interpolation theorem is one of the few to be found in textbooks at this level. Grafakos [2] also has a nice presentation of these topics in chapter **1 L^p Spaces and Interpolation**, specifically in section **1.1 L^p and Weak L^p** and in section **1.3.1 Real Method: The Marcinkiewicz Interpolation Theorem**, but in spite of the more advanced nature of this book and its exclusive focus in harmonic analysis, nevertheless it only presents the simpler diagonal version of the Marcinkiewicz interpolation theorem, like we do at the end of this lesson, and not the full general version as in Folland. Finally, I also strongly recommend Javier Duoandikoetxea's book [3] where, after an introductory review chapter of basic and classical results of Fourier series and transforms, he starts presenting the modern methods in chapter **2 The Hardy-Littlewood Maximal Function**. Even though the focus of the presentation is on maximal operators and, particularly, the Hardy-Littlewood maximal function, that we will concentrate on in the following lesson, he necessarily starts by introducing the basic tools for its study, namely weak-type estimates and the Marcinkiewicz interpolation, also in the simpler diagonal case.

At the conclusion of the last lesson, we got to the point where we know that partial sums of Fourier series converge in $L^p(\mathbb{T})$ norm if and only if the conjugation operator is well defined (and consequently bounded, by the closed graph theorem) in $L^p(\mathbb{T})$. Recall that we already know that conjugation cannot hold for $p = 1$ or $p = \infty$ for we have concluded that Fourier series do not converge in norm for arbitrary functions in these cases. We were also able to characterize harmonic functions on the unit disk $D \subset \mathbb{C}$ whose $L^p(\mathbb{T})$ norms at fixed radius are uniformly bounded for $0 < r < 1$ - so called Hardy space of functions $h^p(D)$ - as Poisson integrals of $L^p(\mathbb{T})$ functions on the boundary ∂D, for $1 < p \leq \infty$, or measures in the case $p = 1$. So the conjugation problem for $1 < p < \infty$ can be translated to harmonic functions in $h^p(D)$ by asking whether the harmonic conjugate, that vanishes at the origin, of a function $u \in h^p(D)$ is also a function in $h^p(D)$. Observe that, for $p = 1$, even if we started from a particular $f \in L^1(\mathbb{T})$ and then concluded that the harmonic conjugate of its Poisson integral $u = P_r * f \in h^1(D)$

Date: May 26, 2020.

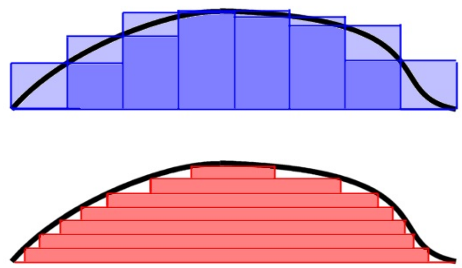

Figure 1. Riemann vs. Lebesgue integration

satisfied $v = Q_r * f \in h^1(D)$, this would still not guarantee that it would lead to a conjugate function $\tilde{f} \in L^1(D)$ at the boundary ∂D because our theory up to this point only implies in this case that $v = P_r * \mu$, $\mu \in \mathcal{M}(\mathbb{T})$, so that the boundary value would generally just be a Borel measure. Unless, of course, we had some type of additional condition, like analyticity, which, from the Riesz brothers' theorem stated at the end of the last lesson, Theorem 1.6, would imply that the measure would be absolutely continuous with respect to the Lebesgue measure and thus representable by a function in $L^1(\mathbb{T})$. Nevertheless, surely that cannot be the general picture because $L^1(\mathbb{T})$ does not admit conjugation.

We will see, however, that in spite of all these drawbacks for $f \in L^1(\mathbb{T})$, it still is true that the harmonic conjugate $v = Q_r * f$ in D always converges pointwise almost everywhere to a function, as $r \to 1$. To show it, we need to finally focus carefully on pointwise convergence of approximate identities, a subject that we have been referring to since Lesson 15, without any proofs, except for Fejér's theorem. For that purpose, we will now introduce several topics that are part of the basic toolkit of modern harmonic analysis, and that we have been avoiding until now in order to not overburden the beginning of the course with technical methods in L^p spaces and approximate identities. The moment has finally arrived for us to concentrate again on fine properties of measurable functions and Lebesgue integration.

We start with the concept of weak L^p spaces and weak-type inequalities. A basic intuitive difference between the Riemann and the Lebesgue integral is that, for integrating, say, a positive function defined on a compact interval of the real line, for the Riemann integral one partitions the domain into subintervals and approximates the area under the graph by a sum of rectangles with the width of the subintervals and the height of the values of the function, whereas for the Lebesgue integral one partitions the range of the function and approximates the area under the graph by a sum of rectangles with height of the partition interval and the width equal to the measure of slicing the graph at that level set.

This is often called the "layer cake representation" and can be rigorously expressed by introducing the concept of distribution function.

Definition 1.1. *Let $f : X \to \mathbb{C}$ be a measurable function on a measure space (X, μ). We define its distribution function $\lambda_f : [0, \infty[\to [0, \infty]$ by*

$$\lambda_f(\alpha) = \mu(\{x : |f(x)| > \alpha\}). \tag{1.1}$$

In other words, the distribution function is the measure of the set referred to above, obtained by slicing the graph of $|f(x)|$, in this general case, at the level of α. It satisfies the following elementary properties.

Proposition 1.2.

(1) *λ_f is decreasing and right continuous.*

(2) *If* $|f(x)| \leq |g(x)|$ *μ-almost everywhere* $x \in X$*, then* $\lambda_f \geq \lambda_g$
(3) *If* $|f_n(x)| \nearrow |f(x)|$ *μ-almost everywhere* $x \in X$*, then* $\lambda_{f_n}(\alpha) \nearrow \lambda_f(\alpha)$.
(4) *If* $f = g + h$ *then* $\lambda_f(\alpha) \leq \lambda_g(\alpha/2) + \lambda_h(\alpha/2)$.
(5) *For all* $c \in \mathbb{C} \setminus \{0\}$ *then* $\lambda_{cf}(\alpha) = \lambda_f(\alpha/|c|)$.

Proof.

(1) That λ_f is decreasing is obvious from the fact that, if $\alpha_1 > \alpha_2 \geq 0$ then $\{x : |f(x)| > \alpha_1\} \subset \{x : |f(x)| > \alpha_2\}$ so that necessarily $\lambda_f(\alpha_1) = \mu(\{x : |f(x)| > \alpha_1\}) \leq \mu(\{x : |f(x)| > \alpha_2\}) = \lambda_f(\alpha_2)$. Also, if $\alpha_n \geq \alpha$ is a decreasing sequence such that $\alpha_n \searrow \alpha$ then $\{x : |f(x)| > \alpha\} = \cup_n\{x : |f(x)| > \alpha_n\}$ so that $\lambda_f(\alpha) = \mu(\{x : |f(x)| > \alpha\}) = \mu(\cup_n\{x : |f(x)| > \alpha_n\}) = \lim_n \mu(\{x : |f(x)| > \alpha_n\}) = \lim_n \lambda_f(\alpha_n)$.
(2) If $|f(x)| \leq |g(x)|$ μ-almost everywhere $x \in X$, then $|f(x)| > \alpha \Rightarrow |g(x)| > \alpha$, except possibly on a set of measure zero, and therefore, for every α, $\lambda_f(\alpha) = \mu(\{x : |f(x)| > \alpha\}) \leq \mu(\{x : |g(x)| > \alpha\}) = \lambda_g(\alpha)$.
(3) If $|f_n(x)| \nearrow |f(x)|$ μ-almost everywhere $x \in X$, then, from the previous property, we have $\lambda_{f_n}(\alpha) \leq \lambda_f(\alpha)$. Also, $\{x : |f(x)| > \alpha\} = \cup_n\{x : |f_n(x)| > \alpha\}$, except possibly on a set of measure zero, and therefore, for every α, $\lambda_f(\alpha) = \mu(\{x : |f(x)| > \alpha\}) = \mu(\cup_n\{x : |f(x)| > \alpha\}) = \lim_n \mu(\{x : |f_n(x)| > \alpha\}) = \lim_n \lambda_{f_n}(\alpha)$.
(4) If $f = g + h$ then, if $|f(x)| > \alpha$, either $|g(x)| > \alpha/2$ or, if not, then $\alpha < |f(x)| \leq |g(x)| + |h(x)| \Rightarrow |h(x)| \geq \alpha - |g(x)| > \alpha/2$. Therefore $\{x : |f(x)| > \alpha\} \subset \{x : |g(x)| > \alpha/2\} \cup \{x : |h(x)| > \alpha/2\}$ which implies $\lambda_f(\alpha) \leq \lambda_g(\alpha/2) + \lambda_h(\alpha/2)$.
(5) For all $c \in \mathbb{C} \setminus \{0\}$ then $|cf(x)| > \alpha$ is equivalent to $|f(x)| > \alpha/|c|$, which means $\{x : |cf(x)| > \alpha\} = \{x : |f(x)| > \alpha/|c|\}$ and this yields $\lambda_{cf}(\alpha) = \lambda_f(\alpha/|c|)$.

□

The distribution function is therefore a decreasing function of its variable α, which should always be thought of as the level at which the graph of $|f|$ is sliced and the measure computed, and it therefore provides information about the size of $|f|$ but none about its pointwise behavior. In particular, if a function is translated, the pointwise values of the function generally all change, but the distribution function remains exactly the same. The way λ_f decreases as α increases to infinity provides information about the largeness of the function and is of local concern, in particular with respect to the rate of blow-up of the function; in the opposite end, the rate at which λ_f increases as α decreases to zero describes how the function behaves "at infinity" and is of global concern. It is not important, though, if a function has compact support, as the measure of that support will be an upper bound for λ_f.

We can now relate Lebesgue integrals of f with its distribution function, rigorously justifying the layer cake representation described above.

Theorem 1.3. *Let* $\Phi : [0, \infty[\to [0, \infty[$ *be a differentiable increasing function with* $\Phi(0) = 0$ *and* f *a measurable function on* (X, μ) *as above, which we now assume to be a σ-finite measure space. Then*

$$\int_X \Phi(|f(x)|) d\mu(x) = \int_0^\infty \Phi'(\alpha)\lambda_f(\alpha) d\alpha.$$

Proof. We can write the left hand side as

$$\int_X \Phi(|f(x)|) d\mu(x) = \int_X \int_0^{|f(x)|} \Phi'(\alpha)\, d\alpha\, d\mu(x) = \int_X \int_0^\infty \chi_{\{\alpha<|f(x)|\}} \Phi'(\alpha)\, d\alpha\, d\mu(x),$$

and we now use Fubini's theorem to exchange the order of integration,

$$\int_0^\infty \int_X \chi_{\{\alpha<|f(x)|\}}\, d\mu(x)\, \Phi'(\alpha)\, d\alpha = \int_0^\infty \mu(\{x : |f(x)| > \alpha\})\, \Phi'(\alpha)\, d\alpha = \int_0^\infty \lambda_f(\alpha)\, \Phi'(\alpha)\, d\alpha,$$

which concludes the proof. ☐

So, in particular, if we make $\Phi(\alpha) = \alpha$ we obtain

$$\int_X |f(x)| d\mu(x) = \int_0^\infty \lambda_f(\alpha)\, d\alpha,$$

which is exactly the formula for the layer cake intuitive argument described above. More generally we have the very useful following corollary.

Corollary 1.4. *Let (X, μ) be a σ-finite measure space and $f : X \to \mathbb{C}$. Then, for $0 < p < \infty$, we have*

$$\|f\|_{L^p(X)}^p = p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha)\, d\alpha. \tag{1.2}$$

while for $p = \infty$,

$$\|f\|_{L^\infty(X)} = \inf \{\alpha \geq 0 : \lambda_f(\alpha) = 0\}.$$

We thus have a totally alternative method for computing L^p norms which is extremely useful, as we will see shortly. Observe in particular, from (1.2), that a function has finite L^p norm, for $p < \infty$, if its distribution function decays just slightly faster than $1/\alpha^p$ as $\alpha \to \infty$, while it explodes to ∞ as $\alpha \to 0$ just slightly slower than $1/\alpha^p$. If a function f has a distribution function of the order exactly $1/\alpha^p$ as $\alpha \to \infty$ or $\alpha \to 0$, then the integral just barely fails to be finite, as it will diverge logarithmically at these extremes. We say such functions are in *weak* $L^p(X)$.

Definition 1.5. *For $0 < p < \infty$ the space of weak L^p functions on X, denoted by $L^p_w(X) = L^p_w(X, \mu)$ or $L^{p,\infty}(X) = L^{p,\infty}(X, \mu)$, is the set of equivalence classes of $\mu-$almost everywhere equal functions for which*[1]

$$\|f\|_{L^{p,\infty}(X)} = \inf \left\{ C \geq 0 : \lambda_f(\alpha) \leq \frac{C^p}{\alpha^p} \quad \text{for all} \quad \alpha > 0 \right\} = \sup_{\alpha > 0} (\alpha^p \lambda_f(\alpha))^{1/p},$$

is finite. The space weak $L^\infty(X) = L^{\infty,\infty}(X)$ is just, by definition, the usual $L^\infty(X)$.

It is easy to see that, for any constant $c \in \mathbb{C}$ we have

$$\|cf\|_{L^{p,\infty}(X)} = \sup_{\alpha>0} (\alpha^p \lambda_{cf}(\alpha))^{1/p} = \sup_{\alpha>0} (\alpha^p \lambda_f(\alpha/|c|))^{1/p} = |c| \|f\|_{L^{p,\infty}(X)},$$

from property (5) in the Proposition 1.2, while

$$\begin{aligned} \|f + g\|_{L^{p,\infty}(X)} &= \sup_{\alpha>0} (\alpha^p \lambda_{f+g}(\alpha))^{1/p} \leq \sup_{\alpha>0} \Big(\alpha^p \big(\lambda_f(\alpha/2) + \lambda_g(\alpha/2) \big) \Big)^{1/p} \\ &\leq 2 \Big(\|f\|_{L^{p,\infty}(X)}^p + \|g\|_{L^{p,\infty}(X)}^p \Big)^{1/p} \leq \max(2, 2^{1/p}) \Big(\|f\|_{L^{p,\infty}(X)} + \|g\|_{L^{p,\infty}(X)} \Big), \end{aligned}$$

from property (4) in the Proposition 1.2. Observe also that $\|f\|_{L^{p,\infty}(X)} = 0$ implies $\lambda_f(\alpha) = 0$ for all α so that $f(x) = 0$ almost everywhere. We conclude therefore that $L^p_w(X)$ is not a normed space, but a quasi-normed space because the triangle inequality fails, just like for $L^p(X)$ with $0 < p < 1$.

So a weak L^p function satisfies the bound

$$\lambda_f(\alpha) \leq \frac{\|f\|_{L^{p,\infty}(X)}^p}{\alpha^p}.$$

[1] The notation $L^{p,\infty}$ comes from identifying the weak L^p_w spaces with the endpoint case of Lorentz spaces usually denoted by $L^{p,q}$ which also include the classical L^p spaces as the case $p = q$. We will not get into Lorentz spaces, but Grafakos [2] has a nice introduction to them in section **1.4 - Lorentz Spaces** in the first chapter of the book **1 - L^p Spaces and Interpolation**.

In $\mathbb{R}^n$, the paradigmatic examples of functions in $L^p_w(\mathbb{R}^n)$ but not in $L^p(\mathbb{R}^n)$ are $\frac{1}{|x|^{n/p}}$, whose p-th power barely fails to be integrable at the origin or at infinity. Their distribution functions satisfy

$$\lambda_{\frac{1}{|x|^{n/p}}}(\alpha) = \left|\left\{x : \frac{1}{|x|^{n/p}} > \alpha\right\}\right| = \left|\left\{x : |x| < \frac{1}{\alpha^{p/n}}\right\}\right| = \left|B_{\frac{1}{\alpha^{p/n}}}(0)\right| = \omega_n \frac{1}{\alpha^p},$$

where ω_n denotes de volume of the unit ball in $\mathbb{R}^n$ and by the absolute value of a set $A \subset \mathbb{R}^n$, $|A|$, we mean its Lebesgue measure. So we have $\frac{1}{|x|^{n/p}} \in L^p_w(\mathbb{R}^n)$ with $\|\frac{1}{|x|^{n/p}}\|_{L^{p,\infty}(\mathbb{R}^n)} = w_n^{1/p}$.

On the other hand, all the functions in the usual $L^p(X)$ spaces - now understandably also called *strong* $L^p(X)$ spaces - are also in $L^p_w(X)$. That's a consequence of the important Chebyshev inequality, for any $0 < p < \infty$ and $\alpha > 0$

$$\alpha^p \lambda_f(\alpha) = \alpha^p \mu(\{x : |f(x)| > \alpha\}) \leq \int_{\{x:|f(x)|>\alpha\}} |f(x)|^p d\mu \leq \int_X |f(x)|^p d\mu = \|f\|^p_{L^p(X)}. \tag{1.3}$$

So we have the following proposition.

Proposition 1.6. *Let $0 < p < \infty$. Then $L^p(X) \subset L^p_w(X) = L^{p,\infty}(X)$ and $\|f\|^{p,\infty}_{L^p(X)} \leq \|f\|^p_{L^p(X)}$ for any $f \in L^p(X)$.*

So, the example seen above, $f(x) = \frac{1}{|x|^{n/p}}$, in $\mathbb{R}^n$, shows that the inclusion $L^p(\mathbb{R}^n) \subset L^p_w(\mathbb{R}^n)$ is strict for $0 < p < \infty$, while for $p = \infty$ they are, by definition, the same spaces $L^\infty(\mathbb{R}^n) = L^\infty_w(\mathbb{R}^n)$.

Weak-type inequalities, or estimates, are bounds that enable control of the L^p_w (quasi)norms. Of particular importance, are those operators from a vector space of measurable functions on a space (X, μ) to measurable functions on another space (Y, ν) for which weak and strong L^q norms on Y can be controlled by strong L^p norms on X. We are interested in a slightly more general class of operators than just the linear ones, so we introduce *sublinear* operators as those that satisfy

$$|T(f + g)| \leq |Tf| + |Tg| \quad \text{and} \quad |T(cf)| = |c||Tf|,$$

for all f, g in its domain and $c \in \mathbb{C}$.

Definition 1.7. *Let $0 < p, q \leq \infty$ and T a sublinear operator from a vector space of measurable functions on (X, μ) that includes $L^p(X, \mu)$ to measurable functions on (Y, ν). We say that*

- *T is weak type (p, q) if $Tf \in L^q_w(Y, \nu)$ for all $f \in L^p(X, \mu)$ and there exists $C \geq 0$ such that*
$$\|Tf\|_{L^q_w(Y,\nu)} \leq C\|f\|_{L^p(X,\mu)}.$$
- *T is strong type (p, q) if $Tf \in L^q(Y, \nu)$ for all $f \in L^p(X, \mu)$ and there exists $C \geq 0$ such that*
$$\|Tf\|_{L^q(Y,\nu)} \leq C\|f\|_{L^p(X,\mu)}.$$

Clearly, from Proposition 1.6 we conclude that strong type (p, q) operators are also weak type (p, q). And, when $q = \infty$, strong type (p, ∞) is the same as weak type (p, ∞).

We have finally built up the definitions and concepts that enable us to conclude this lesson with one of the most important theorems in this course, the second of the two fundamental interpolation theorems: the Marcinkiewicz interpolation. As opposed to the crucial role of the Three-Lines lemma in the Riesz-Thorin interpolation theorem, the Marcinkiewicz interpolation theorem does not use any complex analysis and therefore it is considered the paradigm of the real interpolation method, from which a whole theory of analogous interpolation methods has grown and developed for the last almost 100 years. The general theory of interpolation of operators therefore consists of generalizations of the method used in the Riesz-Thorin interpolation theorem, called complex interpolation, in parallel with generalizations of the method used in the Marcinkiewicz interpolation, called real interpolation.

Theorem 1.8. **(Marcinkiewicz)** *Let (X,μ) and (Y,ν) be σ-finite measure spaces and $0<p_0<p_1\leq\infty$. Let T be a sublinear operator from $L^{p_0}(X,\mu)+L^{p_1}(X,\mu)$ to the space of measurable functions on (Y,ν). Then, if T is weak type (p_0,p_0) and weak type (p_1,p_1), i.e. there exist constants $C_0,C_1\geq 0$ such that*

$$\|Tf\|_{L^{p_0,\infty}(Y,\nu)}\leq C_0\|f\|_{L^{p_0}(X,\mu)},$$

and

$$\|Tf\|_{L^{p_1,\infty}(Y,\nu)}\leq C_0\|f\|_{L^{p_1}(X,\mu)},$$

then T is strong type (p,p) for all $p_0<p<p_1$.

Proof. Let $f\in L^p(X)$. Then, as usual, we split f into a sum $f=f_0^\alpha+f_1^\alpha$, with $f_0^\alpha\in L^{p_0}(X)$ and $f_1^\alpha\in L^{p_1}(X)$, for all $\alpha>0$, by cutting f into high and low parts, respectively

$$f_0^\alpha=f\chi_{\{x:|f(x)|>\alpha\}}\quad\text{and}\quad f_1^\alpha=f\chi_{\{x:|f(x)|\leq\alpha\}}.$$

Due to the sublinearity of T we then have

$$|Tf|\leq|Tf_0^\alpha|+|Tf_1^\alpha|,$$

and just like in property (4) of Proposition 1.2 we have

$$\lambda_{Tf}(\alpha)\leq\lambda_{Tf_0^\alpha}(\alpha/2)+\lambda_{Tf_1^\alpha}(\alpha/2).$$

Now, if $p_1<\infty$, from the weak type estimates we obtain

$$\lambda_{Tf_i^\alpha}(\alpha/2)\leq\left(\frac{2C_i\|f_i^\alpha\|_{L^{p_i}(X)}}{\alpha}\right)^{p_i},\quad i=0,1,$$

and therefore, estimating the $L^p(Y)$ norm of Tf from (1.2) we obtain

$$\begin{aligned}
\|Tf\|_{L^p(Y)}^p &= p\int_0^\infty\alpha^{p-1}\lambda_{Tf}(\alpha)d\alpha\\
&\leq p\int_0^\infty\alpha^{p-1}\big(\lambda_{Tf_0^\alpha}(\alpha/2)+\lambda_{Tf_1^\alpha}(\alpha/2)\big)d\alpha\\
&\leq p\int_0^\infty\alpha^{p-1}\lambda_{Tf_0^\alpha}(\alpha/2)d\alpha+p\int_0^\infty\alpha^{p-1}\lambda_{Tf_1^\alpha}(\alpha/2)d\alpha\\
&\leq (2C_0)^{p_0}p\int_0^\infty\alpha^{p-1-p_0}\|f_0^\alpha\|_{L^{p_0}(X)}^{p_0}d\alpha+(2C_1)^{p_1}p\int_0^\infty\alpha^{p-1-p_1}\|f_1^\alpha\|_{L^{p_1}(X)}^{p_1}d\alpha\\
&= (2C_0)^{p_0}p\int_0^\infty\alpha^{p-1-p_0}\int_X|f_0^\alpha(x)|^{p_0}d\mu\,d\alpha+(2C_1)^{p_1}p\int_0^\infty\alpha^{p-1-p_1}\int_X|f_1^\alpha(x)|^{p_1}d\mu\,d\alpha\\
&= (2C_0)^{p_0}p\int_0^\infty\alpha^{p-1-p_0}\int_{\{x:|f(x)|>\alpha\}}|f(x)|^{p_0}d\mu\,d\alpha+(2C_1)^{p_1}p\int_0^\infty\alpha^{p-1-p_1}\int_{\{x:|f(x)|\leq\alpha\}}|f(x)|^{p_1}d\mu\,d\alpha\\
&= (2C_0)^{p_0}p\int_X\left(\int_0^{|f(x)|}\alpha^{p-1-p_0}d\alpha\right)|f(x)|^{p_0}d\mu+(2C_1)^{p_1}p\int_X\left(\int_{|f(x)|}^\infty\alpha^{p-1-p_1}d\alpha\right)|f(x)|^{p_0}d\mu\\
&= \frac{(2C_0)^{p_0}p}{p-p_0}\int_X|f(x)|^{p-p_0}|f(x)|^{p_0}d\mu+\frac{(2C_1)^{p_1}p}{p_1-p}\int_X|f(x)|^{p-p_1}|f(x)|^{p_1}d\mu\\
&= \left(\frac{(2C_0)^{p_0}p}{p-p_0}+\frac{(2C_1)^{p_1}p}{p_1-p}\right)\int_X|f(x)|^pd\mu.
\end{aligned}$$

If $p_1=\infty$ then we just have a weak type estimate (p_0,p_0)

$$\lambda_{Tf_0^\alpha}(\alpha)\leq\left(\frac{C_0\|f_0^\alpha\|_{L^{p_0}(X)}}{\alpha}\right)^{p_0},$$

and because $\|Tf_1^\alpha\|_{L^\infty(Y)} \leq C_1\|f_1^\alpha\|_{L^\infty(X)}$ we conclude that $|Tf_1^\alpha(y)| \leq C_1\alpha$ for ν-almost everywhere $y \in Y$. Therefore $\lambda_{Tf}(2C_1\alpha) \leq \lambda_{Tf_0^\alpha}(C_1\alpha) + \lambda_{Tf_1^\alpha}(C_1\alpha) = \lambda_{Tf_0^\alpha}(C_1\alpha)$ because the measure of the set where $|Tf_1^\alpha(y)| > C_1\alpha$ is zero. Hence

$$\begin{aligned}
\|Tf\|_{L^p(Y)}^p &= p\int_0^\infty \alpha^{p-1}\lambda_{Tf}(\alpha)d\alpha \\
&= 2C_1p\int_0^\infty (2C_1\alpha)^{p-1}\lambda_{Tf}(2C_1\alpha)d\alpha \\
&\leq (2C_1)^p p\int_0^\infty \alpha^{p-1}\lambda_{Tf_0^\alpha}(C_1\alpha)d\alpha \\
&\leq 2^pC_1^{p-p_0}C_0^{p_0}p\int_0^\infty \alpha^{p-1-p_0}\|f_0^\alpha\|_{L^{p_0}(X)}^{p_0}d\alpha \\
&= 2^pC_1^{p-p_0}C_0^{p_0}p\int_0^\infty \alpha^{p-1-p_0}\int_X |f_0^\alpha(x)|^{p_0}d\mu\, d\alpha \\
&= 2^pC_1^{p-p_0}C_0^{p_0}p\int_0^\infty \alpha^{p-1-p_0}\int_{\{x:|f(x)|>\alpha\}} |f(x)|^{p_0}d\mu\, d\alpha \\
&= 2^pC_1^{p-p_0}C_0^{p_0}p\int_X \left(\int_0^{|f(x)|} \alpha^{p-1-p_0}d\alpha\right)|f(x)|^{p_0}d\mu \\
&= 2^pC_1^{p-p_0}C_0^{p_0}p\int_X |f(x)|^{p-p_0}|f(x)|^{p_0}d\mu \\
&= 2^pC_1^{p-p_0}C_0^{p_0}\frac{p}{p-p_0}\int_X |f(x)|^p d\mu.
\end{aligned}$$

□

To conclude we will just make a couple of observations. The best constant for the strong (p,p) interpolated bound can be obtained by optimizing the point where one splits f into the high and low parts f_0 and f_1 and can be shown to be (see Grafakos [2] or Duoandikoetxea [3])

$$C = 2\left(\frac{p}{p-p_0} + \frac{p}{p_1-p}\right)^{1/p} C_0^{1-\theta}C_1^\theta,$$

where $\theta \in]0,1[$ is the interpolation parameter

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}.$$

Actually, what we proved here is just the diagonal case of the Marcinkiewicz interpolation, with $q_0 = p_0$ and $q_1 = p_1$, which is the most frequently useful. But the theorem is actually more general and, just like the Riesz-Thorin interpolation theorem, can be proved for different p_0, q_0, p_1, q_1 but conditioned, however, to the constraint $q_0 \geq p_0$ and $q_1 \geq p_1$. The full proof of the general case can be found in Folland's book [1].

Finally, it is worth comparing the Riesz-Thorin and the Marcinkiewicz interpolation theorems. To begin with, they yield the same type of conclusion if the operator T is linear and we start with strong type estimates for Marcinkiewicz. But the Riesz-Thorin theorem produces a bound with a much sharper constant. Also, there are ranges of exponents, in particular when $q_0 < p_0$ or $q_1 < p_1$, that can only be interpolated with Riesz-Thorin because not even the general Marcinkiewicz theorem can be used in those cases. On the other hand, the Marcinkiewicz interpolation theorem can be used for sublinear operators and weak type estimates, which the Riesz-Thorin theorem does not cover. They are, therefore, quite independent in their applicability.

For the next lesson, the machinery of weak L^p_w spaces, weak type estimates and the Marcinkiewicz interpolation theorem will be of fundamental importance to obtain pointwise almost everywhere convergence of approximate identities. As we will see, these are typically obtained from weak type bounds for so called maximal operators, which are only sublinear. The Marcinkiewicz interpolation theorem will also be one of the main ingredients in our final proof of the conjugation problem for $L^p(\mathbb{T})$, and thus of the convergence of Fourier series in $L^p(\mathbb{T})$ norm, for $1 < p < \infty$, because it will result from the interpolation of the strong type $(2, 2)$ bound for the conjugation operator, with a crucial weak type $(1, 1)$ estimate that we will still prove.

References

[1] Gerald B. Folland, *Real Analysis, Modern Techniques and Applications,* 2nd Edition, John Wiley & Sons, 1999.
[2] Loukas Grafakos, *Classical Fourier Analysis,* 3rd Edition, Springer, Graduate Texts in Mathematics 249, 2014.
[3] Javier Duoandikoetxea, *Fourier Analysis,* American Mathematical Society, Graduate Studies in Mathematics 29, 2001.