LECTURE NOTES 1 FOR 247A

TERENCE TAO

1. Introduction

The aim of this course is to introduce the basic tools and theory of *real-variable harmonic analysis* - very roughly speaking, the art of estimating the size of an output function in terms of the size of an input function, when a known transformation (linear, multilinear, or nonlinear) is applied. In particular we shall focus on the classical Calderón-Zygmund-Stein theory, in which we study such operations as singular integrals, maximal functions, fractional integrals, pseudodifferential operators, and so forth. This subject is intimately tied together with Fourier analysis, and to a lesser extent real, functional, and complex analysis; see for instance the printed supplement to these notes for some discussion. There are many applications of harmonic analysis, for instance to ergodic theory, analytic number theory, PDE, complex analysis, and geometric measure theory, although we shall only give some very few selected applications in this course. In the sequel 247B to this course we shall focus more on the Fourier-analytic side of things, for instance the connection with representation theory.

2. What is harmonic analysis?

Harmonic analysis is, roughly speaking, the quantitative study of functions on domains (e.g. a function $f : \mathbf{R}^d \to \mathbf{R}$) and similar objects (e.g. measures, distributions, subsets of domains, or maps from one domain to another). For sake of discussion let us restrict attention to functions. One either studies a function in isolation (for instance, asking what is the most efficient way to decompose it in a certain manner, or how the size of a function in one norm is related to the size in another), or else one considers *operators* or *transforms* that take one or more functions as input and returns another as output, and one tries to understand how the size of the output (as measured in various norms) relates to the size of the input. Note that in many applications, the input is not given in a usefully explicit fashion (e.g. it might be the solution to a very nonlinear PDE, or perhaps it is the accumulated "noise" in some real-life system); the only information we have on the input is some bounds on its size, as measured in various norms. So it is generally hopeless to try to compute things exactly; the best we can hope for are estimates on the size of things.

A very typical problem is the following. Suppose we are given an explicit linear operator T from one Banach space V of functions to another W; this linear operator might initially not be defined for *all* functions in V, but only in some dense subclass, such as test functions. A typical example would be the *Hilbert transform* H, defined on test functions on $\mathbf{R}$ by the formula

$$Hf(x) := p.v.\frac{1}{\pi}\int_{\mathbf{R}} \frac{f(y)}{x-y}\ dy = \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|x-y|>\varepsilon}\frac{f(y)}{x-y}\ dy.$$

This transform comes up in several places, most notably in complex analysis and the theory of Fourier summation, but let us not discuss these matters in detail here. The $\frac{1}{\pi}$ normalisation factor is natural (for instance, it implies that $H^2 = -1$) but can be easily omitted for the sake of this present discussion. It is not hard to show that the limit here is well-defined when f is very nice (for instance, if it lies in $C^0_c(\mathbf{R})$, the space of continuous and compactly supported functions). But on a larger space such as $L^2(\mathbf{R})$, it is not obvious initially that H is well-defined.

Given such a densely defined operator T from V to W, a natural question is whether this operator can be continuously extended to the entire domain V; the density of the initial domain implies that such a continuous extension, if it exists, is unique, and so would give a *canonical* extension of T to this larger domain. If T is linear, it turns out that this is true if and only if T is *bounded* on its dense class, i.e. there exists a constant $C > 0$ for which we have the estimate

$$\|Tf\|_W \leq C\|f\|_V$$

for all f in the dense class. For instance, we shall eventually show that the Hilbert transform is bounded on $L^p(\mathbf{R})$ for any $1 < p < \infty$, thus there exists C_p such that

$$\|Hf\|_{L^p(\mathbf{R})} \leq C_p\|f\|_{L^p(\mathbf{R})}$$

for all $f \in C^0_c(\mathbf{R})$. Thus there is a canonical extension of the Hilbert transform which applies to any L^p function. On the other hand, we shall show that this estimate fails at $p = 1$ or $p = \infty$, and so the Hilbert transform cannot be meaningfully extended as a map from $L^1(\mathbf{R})$ to itself, or from $L^\infty(\mathbf{R})$ to itself. (We will however show that it maps $L^1(\mathbf{R})$ to another space, namely $L^{1,\infty}(\mathbf{R})$ (weak L^1), and that it maps $L^\infty(\mathbf{R})$ to another space, $BMO(\mathbf{R})$.)

Thus we see that a *qualitative* question (existence of a continuous extension) can be equivalent to a *quantitative* question (existence of a concrete estimate). Indeed, a major purpose of harmonic analysis is to provide the quantitative estimates needed to obtain qualitative properties of functions and operators (e.g. continuity, integrability, convergence, etc.), which in turn are often needed to justify various formal manipulations for many applications (most notably in PDE). However, quantitative estimates are of interest in their own right. For instance, they often demonstrate the *robustness* of various operations to the presence of unpredictable noise. For instance, once we know that the Hilbert transform H is bounded on L^p, this assures us that small perturbations of the input f (as measured in L^p norm) are guaranteed to only cause small perturbations of the output Hf (again measured in L^p norm). This is important for rigorously justifying numerical simulations of operators such as the Hilbert transform, as one always expects to have small errors (arising from measurement error, roundoff error, or other sources). More intangibly, estimates help convey intuition on what transforms such as the Hilbert transform actually *do*:

for instance, can they transform shallow broad functions into spiky narrow functions or vice versa? These are questions which are often difficult to read off from the explicit formula defining these operators, but can instead be seen via the estimates. (For instance, the boundedness of H for high values of p basically prevents shallow broad functions from being transformed into spiky narrow functions, whereas the boundedness for small values of p basically prevents the reverse phenomenon.)

Despite the close partnership between the qualitative and quantitative aspects of analysis, they differ in some key respects. Qualitative analysis is often concerned with very rough functions (or even objects that are not functions at all, such as measures or distributions), and a large part of the difficulty lies in actually justifying various formal calculations (e.g. swapping limits, sums, and integrals with each other). In contrast, with quantitative analysis, one can often restrict to very nice functions (e.g. the continuous compactly supported functions, or perhaps the Schwartz class) in which every formal manipulation is easy to justify; however, the goal is different, namely to obtain an explicit estimate rather than existence or convergence of an expression. Despite these superficial differences, however, there are still many similarities in the two different styles of analysis. For instance, in the qualitative world, a key issue is whether convergence in one topology implies convergence in another; in the quantitative world, the analogous issue is whether control of some norm (or norm-like quantity) implies control of another norm.

Harmonic analysis has long been intertwined with *Fourier analysis*, which is the study of how general functions on symmetric domains (such as Euclidean space, the torus, or the sphere) are decomposed into more symmetric objects (such as plane waves, characters, spherical harmonics, or eigenfunctions). In one direction, in order to justify many of the identities arising in Fourier analysis, one needs the quantitative estimates arising from harmonic analysis. In the converse direction, the Fourier transform enables one to view functions and operators in *frequency space* rather than *physical space*, which can greatly clarify some features of these objects (while greatly obscuring others). More recently, a combined *phase space* viewpoint has proven to be very useful, in which one views all objects in the physical and frequency domain simultaneously, subject of course to the limits given by the uncertainty principle. We will return to these issues much later in this course, and in the next quarter also. However for most of this quarter we shall focus on harmonic analysis (and in particular the art of the estimate) rather than on Fourier analysis.

Historically, both harmonic and Fourier analysis - particularly on the real line $\mathbf{R}$ or circle S^1 - were closely tied to *complex analysis*, which is the study of complex analytic functions and other objects in complex geometry. Complex analysis (and its generalisation to several variables) continues to have a mutually profitable interaction with harmonic analysis today, however the subjects have now moved quite far apart, in that one can now learn much of harmonic analysis without ever having to deal with complex analytic functions (though one always deals with complex numbers, via the fundamental character $x \mapsto e^{2\pi i x}$). Perhaps the key event that separated the two fields was the widespread adoption of the *bump function* (and related cutoff functions) in harmonic analysis, which allowed one to localise many objects in physical space, frequency space, or both. Such functions do not have a fully satisfactory analogue in the complex analytic (or even real analytic) world, as they are incompatible with analytic continuation. Furthermore, the real-variable harmonic analysis methods turned out to extend to higher dimensions much more easily than the complex-variable ones. We shall therefore adopt a modern perspective on harmonic analysis rather than a historical one, and so complex analytic functions will only play a minor role in our presentation.

3. Hardy, Landau, and Vinogradov notation

As we have seen in the above discussion, harmonic analysis will often be concerned with obtaining estimates of the form $X \leq CY$, where X is some quantity measuring size of output, Y is some quantity measuring size of input, and C is a constant. In many cases, the precise value of C is either not important, not interesting, or too difficult to compute exactly[1]. It thus makes great practical sense to adopt notation[2] which allows one to tolerate multiplicative losses of constants without having to do a great deal of book-keeping to track what these constants are exactly. The three main notations for doing this all arose from analytic number theory (which encountered the need for such notation somewhat sooner than the harmonic analysts, who at least had the option of working instead in the qualitative world). They are the *Hardy notation*, the *Landau notation*, and the *Vinogradov notation*. All three are extensively used in the literature; they are essentially equivalent but each has some slight advantages and disadvantages.

In *Hardy notation*, the letter C is used to denote various positive constants between 0 and ∞ (which are typically quite large); the C stands of course[3] for "constant". Generally speaking, these constants C could be evaluated numerically if absolutely necessary[4] but one chooses not to in order not to get bogged down in distractions. The key point to remember is that each different appearance of the letter C can represent a different constant (unless one explictly uses a subscript such as C_1, C_2,

[1]The study of exact inequalities with sharp constants (which typically involve π) is of great interest, and there are important cases in which knowing the sharp inequality can assist with proving the estimate with unspecified constant, particularly when one needs to iterate the inequality repeatedly. For instance, given a sequence of positive numbers $x_1, x_2, \ldots$, a precise inequality $x_n \leq x_{n-1}$ lets one bound x_n uniformly via iteration, whereas an imprecise inequality $x_n \lesssim x_{n-1}$ will not. However, in most cases it is too difficult to obtain sharp constants and so we shall content ourselves with unspecified constants in order to be able to prove more results.

[2]In general, the purpose of good notation is to conceal or deprecate the less important features of a mathematical expression, in order to focus as much attention as possible on the crucial or key features. Of course, the decision as to which features are important and which are not is a subjective one, and depends heavily on the application. Hence it makes sense to adapt notation to a specific field of study, rather than to try to force a uniform one-size-fits-all notational standard across all of mathematics.

[3]In Hardy's original papers, the letter A was used instead.

[4]In number theory there is an interesting phenomenon that some constants are *ineffective* - they are known to be finite, but are not computable with present technology, due for instance to the unknown status of the Riemann hypothesis. This rarely happens in harmonic analysis, except occasionally when using qualitative methods to prove a quantitative result.

etc. to override[5] this convention). Sometimes one needs the constants C to depend on certain parameters, in which case this is denoted by subscripts. For instance, in order for a collection of functions $f_n : \Omega \to \mathbf{R}$ to be *uniformly* bounded, we need $f_n(x) \leq C$ for all n and all $x \in \Omega$, but to be *individually* bounded we need $f_n(x) \leq C_n$ for all n and all $x \in \Omega$. If a parameter stays fixed throughout the entire argument (e.g. the ambient dimension d) then one often omits the explicit dependence of constants C on that parameter, although one should then state this convention explicitly at the start of the argument.

In *Landau notation*[6] the expression $O(X)$ (read: "big-O of X") is used to denote any quantity bounded in magnitude by CX for some finite constant X; thus $Y = O(X)$ is equivalent to the Hardy notation $|Y| \leq CX$, and the constant C is then called the *implied constant* or *implicit constant* in the $O()$ notation. For instance we have $\sin x = O(1)$, $\sin(x) = O(|x|)$, and $\sin(x) = x + O(|x|^3)$ for any real number x. Note that the use of parentheses in the $O()$ notation does not denote a functional relationship: $O(X)$ need not be a function of X. This notation is very convenient (especially in describing expressions such as $X + O(Y)$ with a main term X and an error term $O(Y)$) but there is one major caveat: the notation breaks the symmetry in the equality relation. Basically, when a $O()$ appears on the right of an equality (or any other binary relation), it asserts that the equality is true for *some* choice of function in that class, whereas when it appears on the left, it asserts the equality is true for *all* choices of function in that class. Thus for instance, when n is a positive integer parameter, then $O(n) = O(n^2)$ (i.e. every quantity which is of the form $O(n)$, is automatically also of the form $O(n^2)$), but $O(n^2) \neq O(n)$ (thus a quantity which is of the form $O(n^2)$, is not necessarily of the form $O(n)$). Because of this asymmetry, one generally tries to only place $O()$ notation on the right-hand side of an expression to avoid confusion. Finally, if one wants the implied constant to depend on parameters, this can be done via subscripting; for instance, $X = O_k(Y)$ denotes the estimate $|X| \leq C_k Y$ for some constant C_k depending on a parameter k.

In (modified) *Vinogradov notation*[7], the notation $X \lesssim Y$ (read: X is less than or comparable to Y) or $Y \gtrsim X$ is used synonymously with $|X| \leq CY$ or $X = O(Y)$. We also use $X \sim Y$ to denote $X \lesssim Y \lesssim X$, thus for instance $x + y \sim \max(x, y)$ for all $x, y > 0$. Again, we subscript this to denote dependence on parameters, thus $X \lesssim_k Y$ is synonymous with $X = O_k(Y)$ or $|X| \leq C_k Y$. Note that this notation is transitive if used finitely many times - which is one of the key advantages of this notation - though care should be taken with using it inductively. For instance, we

[5]Of course, an even more unambiguous way to override the convention is simply to use a letter other than C.

[6]Landau notation also includes some other symbols, most notably $o()$, but also the rarer $\Omega()$, $\omega()$, and $\Theta()$; these are useful in analytic number theory, but are not sufficiently relevant in harmonic analysis to be in widespread use.

[7]In the original Vinogradov notation, still in use in analytic number theory, $X \ll Y$ is used for $X = O(Y)$, while $X \lesssim Y$ denotes $|X| \leq (1 + o(1))Y$ (thus X is less than or asymptotic to Y). Asymptotics are not so useful in harmonic analysis and so this notation is not in wide use. Instead, $X \ll Y$ is sometimes used (rather informally) to denote the assertion $X \leq cY$ for a sufficiently *small* constant c, although care should be taken with this notation as it is not fully rigorous.

have $2x \lesssim x$ for any $x > 0$, but one cannot iterate this inductively to conclude that $2^n x \lesssim x$ uniformly in n. Instead, we have $2^n x \lesssim_n x$; the length of the induction depends on n, and so the implied constant will do so also.

Let us use Vinogradov notation for now. A very typical objective is then to upper bound[8] a complicated expression X by a known expression Y, i.e. to establish the bound $X \lesssim Y$. For instance, to prove that a linear operator $T : V \to W$ is bounded one needs to show that $\|Tf\|_W \lesssim \|f\|_V$ for all f in V (or a dense class thereof). One of the most basic techniques in doing so is *divide and conquer* - split X up into two or more pieces, bound each piece separately, and then sum up. For instance, observe that if $X = X_1 + X_2$ for some non-negative X_1, X_2, then $X \lesssim Y$ is true if and only if $X_1 \lesssim Y$ and $X_2 \lesssim Y$ are both true. If the number of pieces is not too large, and one doesn't care about the precise value of the implied constant, then this technique is essentially a "free" reduction, allowing one to replace the task of bounding a large quantity by the easier tasks of bounding several smaller quantities. Besides making the quantity to estimate smaller, the divide-and-conquer technique has the effect of *localising* and thus *isolating* the difficulties of the problem. Suppose for instance one is estimating a quantity Q which contains both an oscillatory component and a singular component; a useful tactic is then to try to cleverly decompose the object $Q = Q_o + Q_s$ into a component Q_o which only has oscillation, and a component which only has singularity Q_s, which will then be easier to estimate[9].

4. Rearrangement-invariant theory - introduction

For the remainder of this week's notes, we shall review the basic theory (in particular, the interpolation theory) of *Lebesgue spaces* L^p, as well as their cousins such as *weak Lebesgue spaces* ($L^{p,\infty}$), *Lorentz spaces* ($L^{p,q}$), and *Orlicz spaces* (such as $L \log L$ and e^L); this entire family of spaces are collectively referred to as the *monotone rearrangement-invariant spaces*. You will already have seen much of this material in 245AB, but since we shall rely on it so much throughout the course, it deserves a thorough review here. As stated above, our emphasis shall be on the quantitative aspects of this theory (in particular, on norms and estimates) as opposed to qualitative aspects (such as measurability, convergence, or integrability). Thus for instance the norms will be more important to us than the function spaces. Also for simplicity we shall restrict attention to real or complex-valued functions; there are important generalisations of the theory here to vector-valued functions,

[8] Lower bounds and asymptotics are of course also desirable, but tend to be significantly harder to prove, requiring tools outside of harmonic analysis. A typical strategy when lower bounding X is to split X into a main term Z, which can be computed by some other means (e.g. by algebraic methods), plus an error E, which one then upper bounds. As long as the upper bound for the magnitude for E is less than the main term Z, one obtains a non-trivial lower bound; if it is significantly less, one can obtain an asymptotic.

[9] More commonly, one performs a *dyadic decomposition* into a countable number of components $Q = \sum_n Q_n$, each of which has some oscillation and some singularity, for instance with Q_n oscillating with wavelength 2^n, but only having amplitude 2^{-n}, so that for large positive n one has little singularity and for large negative n one has little oscillation. By exploiting these quantitative bounds one can hope to obtain bounds on Q_n which decay geometrically as $n \to \pm\infty$, allowing one to easily sum via the triangle inequality.

but rather than develop them here explicitly, it is better to instead present the *techniques* of proof in the theory, so that you can simply apply the techniques yourself whenever one needs to work in a more general setting.

The purpose of rearrangement-invariant norms is to usefully quantify two basic aspects of functions, namely their *height* (i.e. their typical amplitude) and *width* (i.e. the measure of the bulk of their support). The point of using the norms rather than to try to define (the rather fuzzy concepts of) height and width directly is that norms are convex (essentially by definition) and hence stable under many operations.

Throughout the rest of the paper, we fix a measure space $(X, \mathcal{B}, \mu)$, thus X is a set, $\mathcal{B}$ is a σ-algebra of sets in X, and μ is a non-negative measure on X. To avoid irrelevant technicalities we always take μ to be σ-finite. Important examples to keep in mind are the *Euclidean space* $X = \mathbf{R}^d$ (with $\mathcal{B}$ equal to the Borel or Lebesgue σ-algebra, and Lebesgue measure $d\mu = dx$), the *lattice* $X = \mathbf{Z}^d$ (with counting measure $d\mu = d\#$ and the discrete σ-algebra), and the *torus* $X = \mathbf{R}^d/\mathbf{Z}^d$ (with the Borel or Lebesgue σ-algebra, and Lebesgue measure $d\mu = dx$). We should also mention that the finite set $X = \{1, \ldots, N\}$ is also an important example, with μ either equal to counting measure $d\mu = d\#$ or normalised counting measure $d\mu = \frac{1}{N} d\#$).

Unlike these highly structured examples, however, in this week's notes we will not assume any further structures are present on X; thus X will have no topology, no metric, no geometry, and no group structure. (We will of course see these structures in later notes.) On the one hand, this severely limits what we can do with our space X - basically, we can integrate functions and measure sets, and that's about it; thus any result which genuinely exploits more structure than just the measure-theoretic structure will not be provable purely by L^p theory. On the other hand, it will automatically make all of the theory here *rearrangement-invariant*: if we have some measure-isomorphism $\Phi : X \to X$ (i.e. a bijection which is bimeasurable, thus $\Phi(\mathcal{B}) = \mathcal{B}$ and $\Phi^{-1}(\mathcal{B}) = \mathcal{B}$, and measure preserving, thus $\mu(\Phi(E)) = \mu(E)$ for all $E \in \mathcal{B}$) then we can apply this isomorphism to X (thus mapping sets E to $\phi(E)$ and functions f to $f \circ \Phi^{-1}$) without affecting any of the norms or other operations that we shall describe here. Another way of saying this is that, in these notes, the only feature of a set E which is important is its measure; a long thin set and a short round set are viewed equally so long as they have the same measure. Similarly it is the distribution of a function which is relevant, not its precise location in space.

Our norms will also be *monotone*: if $|f| \leq |g|$ pointwise almost everywhere, then g will have a larger norm than f. This means in particular that it is only the magnitude $|f|$ of a function f which will be relevant for all our norms, not the sign.

When we refer here to a *set*, we always mean a measurable subset of X (i.e. an element of $\mathcal{B}$); when we refer here to a *function*, we mean a $\mathcal{B}$-measurable complex-valued function $f : X \to \mathbf{C}$ (of course, the theory for real-valued functions will emerge as a special case). Important examples of functions, in increasing order of generality, are

- *Indicator functions*[10] $f = 1_E$ for some set E, defined by setting $1_E(x) = 1$ when $x \in E$ and $1_E(x) = 0$ otherwise. (If $P(x)$ is a property depending on a point x, we write $1_{P(x)}$ or $1_P(x)$ for $1_{\{x \in X : P(x)\}}$. Thus for instance $1_E(x) = 1_{x \in E}$.)
- *Step functions* $f = c1_E$ for some set E and some $c \in \mathbf{C}$ (i.e. scalar multiples of indicator functions). We refer to $|c|$ as the *height* of the step function, and $\mu(E)$ as the *width*.
- *Simple functions* $f = \sum_{j=1}^J c_j 1_{E_j}$ (i.e. finite linear combinations of indicator functions). Equivalently, simple functions are functions which are measurable with respect to a finite σ-algebra.

As it turns out, simple functions will be dense in every function space considered in these notes, which means that for the purposes of estimates we may restrict attention to simple functions, which is convenient as it means we have essentially no problems in justifying any computation (e.g. swapping a sum and integral). We cannot quite restrict attention in the same way to step functions, however step functions are definitely a major example to consider. Indeed one can gain a lot of intuition about how any given estimate or result here works by first considering how it applies to step functions.

A little later on we shall see some "fuzzier" versions of the above concepts, when we start atomically decomposing L^p functions into components. We shall see in particular the *dyadic pigeonhole principle*, which asserts that general functions only differ from step functions "by a logarithm".

As is usual in measure theory, we identify two functions (or sets) if they agree almost everywhere, and hence every pointwise identity that we assert is understood to only be true outside of sets of measure zero.

Our theory shall mostly focus on the sets E and functions f, and how to decompose or otherwise manipulate these objects. However the space X, the σ-algebra $\mathcal{B}$ and measure μ are not purely passive actors in this theory; we will occasionally see some use in manipulating X, $\mathcal{B}$ or μ in various ways. Indeed this freedom to play with X, $\mathcal{B}$ and μ is a major advantage in working in this abstract setting, rather than staying in a concrete setting such as Euclidean space $\mathbf{R}^d$. (The other major advantage, of course, is that the theory lets one deal with all the concrete examples simultaneously, rather than having to tediously repeat the same theory separately for each concrete case.)

[10]It is also common to see χ_E used instead of 1_E in the literature, though χ has enough other uses (e.g. as a smooth cutoff, or as a character) that it can be better to use 1_E instead to free up some namespace.

5. L^p THEORY - BASICS

We begin with the theory of the L^p norms of a function f, defined for $0 < p < \infty$ as

$$\|f\|_{L^p(X,d\mu)} := (\int_X |f|^p \; d\mu)^{1/p}$$

and for $p = \infty$ as

$$\|f\|_{L^\infty(X,d\mu)} := \operatorname{ess\,sup}_{x \in X} |f(x)|$$

or more usefully, $\|f\|_{L^\infty(X,d\mu)}$ is the least real number for which we have the pointwise bound

$$|f(x)| \leq \|f\|_{L^\infty(X,d\mu)} \text{ for almost every } x \in X.$$

Thus for instance a step function of height H and width W has L^p norm $HW^{1/p}$. We often abbreviate $\|f\|_{L^p(X,d\mu)}$ as $\|f\|_{L^p(X)}$, $\|f\|_{L^p}$, or even just $\|f\|_p$ when there is no chance of confusion.

Problem 5.1. For f a simple function, verify that $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$, and that $\lim_{p\to 0} \|f\|_p^p = \mu(\mathrm{supp}(f))$, where $\mathrm{supp}(f) := \{x : f(x) \neq 0\}$. For this reason, the measure of the support of f is sometimes referred to as the L^0 *norm* of f, though it would be more accurate (though confusing) to refer to it as the 0^{th} power of the L^0 norm.

Example 5.2. *Let $\alpha > 0$. On a Euclidean space $\mathbf{R}^d$, the function $f(x) := |x|^{-\alpha} 1_{|x|>1}$ lies in $L^p(\mathbf{R}^d)$ (with a norm of $O_{p,\alpha}(1)$) if and only if $\alpha > d/p$, while the function $g(x) := |x|^{-\alpha} 1_{|x|\leq 1}$ lies in $L^p(\mathbf{R}^d)$ if and only if $\alpha < d/p$. The function $|x|^{-\alpha}$ does not lie in any $L^p(\mathbf{R}^d)$, although it only fails "logarithmically" to lie in $L^{d/\alpha}$. Thus we see that control in L^p for high p rules out severe local singularities at a point, while control in L^p for low p rules out insufficiently rapid decay at infinity.*

For $1 \leq p \leq \infty$, these norms are indeed norms, in particular we have the triangle inequality

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{1}$$

Proof (Sketch) The case $p = \infty$ is trivial, so take $1 \leq p < \infty$. By homogeneity $\|cf\|_p = |c|\|f\|_p$ we may reduce to the case $\|f\|_p = 1 - \theta$, $\|g\|_p = \theta$ for some $0 \leq \theta \leq 1$. The cases $\theta = 0, 1$ are trivial, so suppose $0 < \theta < 1$. Writing $F := f/(1-\theta)$ and $G := g/\theta$ we reduce to the convexity estimate

$$\|(1-\theta)F + \theta G\|_p \leq 1 \text{ whenever } \|F\|_p, \|G\|_p \leq 1 \text{ and } 0 \leq \theta \leq 1.$$

But since $z \mapsto |z|^p$ is convex for $p \geq 1$, we have the pointwise convexity bound

$$|(1-\theta)F(x) + \theta G(x)|^p \leq (1-\theta)|F(x)|^p + \theta |G(x)|^p.$$

Integrating this we obtain the claim. This proof is a basic example of how one uses a *symmetry* (in this case, homogeneity symmetry) to simplify the task of proving an estimate, by normalising one or more inconvenient factors to equal 1. ■

Problem 5.3. Show that the triangle inequality is sharp if and only if f, g are parallel, thus either $f = 0$, $g = 0$, or $f = cg$ for some real positive c.

We give another proof of the triangle inequality at the end of this section.

Problem 5.3 leads to a useful heuristic: triangle inequalities in general are only expected to be efficient when the functions involved can all "align" or "correlate" in some substantial way. If there is a lot of "orthogonality", "separation" or "cancellation", one expects to improve on triangle inequality bounds somehow. Of course, obtaining this improvement can often be quite difficult. However let us make one remark: if f, g have disjoint supports, then the triangle inequality improves to

$$\|f+g\|_p = (\|f\|_p^p + \|g\|_p^p)^{1/p}$$

for any $0 \leq p < \infty$ (with the usual modification at $p = \infty$).

Iterating the triangle inequality (and using monotone and dominated convergence) we obtain

$$\|\sum_n f_n\|_p \leq \sum_n \|f_n\|_p \tag{2}$$

for either finite or infinite summation. For $0 < p < 1$ we have the pointwise subadditivity property

$$|f(x)+g(x)|^p \leq |f(x)|^p + |g(x)|^p$$

and hence

$$\|f+g\|_p^p \leq \|f\|_p^p + \|g\|_p^p$$

and more generally

$$\|\sum_n f_n\|_p^p \leq \sum_n \|f_n\|_p^p. \tag{3}$$

In particular we have the *quasitriangle inequality*

$$\|f+g\|_p \lesssim_p \|f\|_p + \|g\|_p$$

and more generally (by Hölder's inequality)

$$\|\sum_{n=1}^N f_n\|_p \leq N^{\frac{1}{p}-1} \sum_{n=1}^N \|f_n\|_p. \tag{4}$$

Problem 5.4. Show that the bound $N^{\frac{1}{p}-1}$ here is sharp in general, i.e. it cannot be replaced by any smaller quantity.

Problem 5.5. Show that for any $0 < p < \infty$ and any sequence f_n of functions we have

$$(\sum_{n=1}^\infty \|f_n\|_p^{\max(p,1)})^{1/\max(p,1)} \leq \|\sum_{n=1}^\infty |f_n|\|_p \leq (\sum_{n=1}^\infty \|f_n\|_p^{\min(p,1)})^{1/\min(p,1)}.$$

After the triangle inequality, the next most important inequality in L^p theory is *Hölder's inequality*:

$$\|fg\|_r \leq \|f\|_p \|g\|_q \text{ whenever } 0 < p, q, r \leq \infty \text{ and } \frac{1}{p} + \frac{1}{q} = \frac{1}{r}. \tag{5}$$

Proof (Sketch) We assume $p, q, r < \infty$ as the case $p, q, r = \infty$ is trivial (and can also be obtained by limiting arguments). Using the separate homogeneity symmetry in both f and g we may normalise $\|f\|_p = \|g\|_q = 1$. Writing $F := |f|^p$, $G := |g|^q$, and $\theta = r/q$ (so $1-\theta = r/p$) we reduce to showing that

$$\int_X F^{1-\theta}G^\theta \leq 1 \text{ whenever } \int_X F = \int_X G = 1.$$

But from the convexity of $\theta \mapsto F^{1-\theta}G^\theta$ we have the pointwise estimate

$$F^{1-\theta}(x)G^\theta(x) \leq (1-\theta)F(x) + \theta G(x);$$

integrating this we obtain the claim. ∎

Remark 5.6. Note that Hölder's inequality is not just symmetric under the homogeneities $f \mapsto cf$ and $g \mapsto cg$ of the functions, but also under the homogeneity $\mu \mapsto \lambda\mu$ of the underlying *measure*[11]. This latter symmetry demonstrates why the condition $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ is necessary. (The first two symmetries demonstrate why f appears the same number of times on both sides of the inequality, and similarly for g.)

Remark 5.7. It is instructive to verify Hölder's inequality directly when f and g are step functions.

Hölder's inequality is also equivalent to the *log-convexity of L^p norms*:

$$\|f\|_r \leq \|f\|_p^{1-\theta}\|f\|_q^\theta \text{ whenever } 0 < p < q < \infty, 0 < \theta < 1 \text{ and } \frac{1}{r} = \frac{1-\theta}{p} + \frac{\theta}{q}. \tag{6}$$

Again, it is instructive to verify this inequality for step functions - in fact it is an equality in this case.

Problem 5.8. Derive (6) from Hölder's inequality. Conversely, by manipulating the measure μ appropriately, derive Hölder's inequality from (6). (For technical reasons one needs to first reduce to the case where X has finite measure, and then f and g are everywhere non-vanishing simple functions. Now consider the convexity of $|f|^\alpha|g|^\beta$ with respect to a measure $|f|^\gamma|g|^\delta\mu$ for some suitable exponents $\alpha, \beta, \gamma, \delta$.)

It is instructive to give several additional proofs of this convexity estimate (6), in order to illustrate certain techniques[12] we shall encounter repeatedly in this course. One approach is a direct one:

[11] In Euclidean space $\mathbf{R}^d$, this symmetry is equivalent to the scaling symmetry $x \mapsto ax$ for $a > 0$, as the Jacobian of this map is $\lambda = a^d$. But the point is that by manipulating the measure directly, one still enjoys this symmetry even when no scaling operation is present.

[12] It is always good to have as many proofs of basic results as possible in analysis, because one often needs to generalise these basic results beyond their usual range of applicability, and having many genuinely different proofs increases the chance that one of them will extend to cover the desired generalisation. This is in sharp contrast to algebra, in which theorems can often be used as "black boxes" without knowledge of the proof technique. Algebra draws its power from *modularity*, *abstraction* and *identities*; analysis draws its power from *robustness*, *physical intuition* and *estimates*.

Problem 5.9 (Direct approach to log convexity). Differentiate $\log \|f\|_p$ twice with respect to $\alpha := 1/p$ and show that this is non-negative (take f to be a non-zero simple function with finite measure support to avoid technicalities). This is an example of a *monotonicity formula method* - deriving estimates from a monotonicity property, which in turn follows from the non-negativity of a derivative.

You will see that this approach is surprisingly messy. For all the other ways, observe that (6) enjoys homogeneity symmetry in both f and μ, which lets one normalise *both* $\|f\|_p$ and $\|f\|_q$ to equal one. Thus the task is now to show that if $\|f\|_p = \|f\|_q = 1$, then $\|f\|_r \leq 1$ for all r between p and q. This can be done by the pointwise convexity of $p \mapsto |f(x)|^p$, or more precisely the estimate

$$|f(x)|^r \leq (1-\theta)|f(x)|^p + \theta|f(x)|^q;$$

the observant reader will note that this is merely the proof of Hölder's inequality in disguise.

Let us now give a more unusual proof of the log-convexity which does not appeal to any pointwise convexity estimate, instead combining the "divide and conquer" strategy with an elegant (and rather cheeky) "tensor power trick". Again normalise $\|f\|_p = \|f\|_q = 1$. We split f into a broad flat piece and a narrow tall piece

$$f = f1_{|f|\leq 1} + f1_{|f|>1}$$

which are disjoint, and thus

$$\|f\|_r^r = \int_{|f|\leq 1} |f|^r + \int_{|f|\geq 1} |f|^r.$$

When[13] $|f| \leq 1$, then $|f|^r \leq |f|^p$, and when $|f| \geq 1$, then $|f|^r \leq |f|^q$. Thus we end up with

$$\|f\|_r^r \leq \int_X |f|^p + \int_X |f|^q = 2.$$

The above argument (which is a prototype of the *real interpolation method*) obtained an estimate which is off by a factor of two from what we wanted; this is a typical feature of the method. However we can recover this factor for free by the following *tensor power trick*. Let M be a large integer. We replace the measure space $(X, \mathcal{B}, \mu)$ by its M^{th} power $(X^M, \mathcal{B}^{\oplus M}, \mu^{\oplus M})$ using the product measure construction, and similarly replace f with its tensor power $f^{\oplus M} : X^M \to \mathbf{C}$, defined by

$$f^{\oplus M}(x_1, \dots, x_M) := f(x_1) \dots f(x_M).$$

One then observes that

$$\|f^{\oplus M}\|_{L^p(X^M)} = \|f\|_{L^p(X)}^M = 1;$$
$$\|f^{\oplus M}\|_{L^q(X^M)} = \|f\|_{L^q(X)}^M = 1;$$
$$\|f^{\oplus M}\|_{L^r(X^M)} = \|f\|_{L^r(X)}^M.$$

[13]What we are doing here is exploiting some very basic intuition about L^p norms, namely that L^p bounds for large p tend to exclude tall narrow spikes, whereas L^p bounds for small p tend to exclude short broad tails. Of course, either sort of bound would exclude tall broad functions, and neither excludes narrow short functions. Once again, this intuition can be buttressed by considering the special case of step functions.

Now we apply the preceding arguments to $f^{\oplus M}$ instead of f to deduce that

$$\|f^{\oplus M}\|_{L^r(X^M)}^r \leq 2$$

which on taking M^{th} roots gives

$$\|f\|_{L^r(X)}^r \leq 2^{1/M}.$$

Now the left-hand side is independent of M; take limits as $M \to \infty$ and we obtain $\|f\|_r \leq 1$ as desired.

The tensor power trick can be viewed as another application of symmetry: if an estimate is invariant under raising to a tensor power, then one can automatically replace all absolute constants with 1; thus we obtain the "free lunch" of deducing a bound with an explicit constant 1, from a bound with an unspecified constant[14]. Contrapositively, if an estimate is invariant under tensor power, then a weak counterexample (which shows that the constant must exceed one) can be *amplified* into a strong counterexample (which shows that no finite constant suffices) by tensor powering[15]. The tensor power trick seems like a magical trick at present, but is actually exploiting some basic results in information theory such as the Shannon entropy inequalities and the central limit theorem; it also combines well with virtually any inequality which involves Gaussians. Unfortunately due to lack of time we will not be discussing these beautiful topics further in this course. At any rate one sees the power of abstraction in this tensor power trick. (One could similarly perform this trick in $\mathbf{R}^d$, so long as the constants only grew sub-exponentially in the dimension d.)

The final proof of log-convexity of the norm that we give here proceeds via complex analysis, and the maximum principle - which in many ways is a complex analogue of convexity[16]. We need the following result from complex analysis.

Lemma 5.10 (Three lines lemma)**.** *Let f be a complex-analytic function on the strip $\{0 \leq \mathrm{Re}(z) \leq 1\}$, which is of at most double-exponential growth, or more precisely*[17] *$|f(z)| \lesssim_f e^{O_f(e^{(\pi-\delta)|z|})}$ for some $\delta > 0$. Suppose that we have the bounds $|f(z)| \leq A$ when $\mathrm{Re}(z) = 0$ and $|f(z)| \leq B$ when $\mathrm{Re}(z) = 1$. Then we have $|f(z)| \leq A^{1-\mathrm{Re}(z)} B^{\mathrm{Re}(z)}$ for all z in the strip.*

---

[14]The free lunch is even better; the same trick even allows one to lose logarithmic factors in the constant, the basic point being that $\lim_{M\to\infty}(\log N^M)^{1/M} = 1$ for any N.

[15]The same amplification heuristic also applies, more or less, to other operations which resemble tensor powers, such as Riesz products; we shall return to this point later in the course.

[16]Note that convex functions also obey the maximum principle. Actually, the more accurate analogy is between subharmonic functions of two variables (which include the magnitude of complex analytic functions as special cases) and subharmonic functions of one variable (i.e. convex functions). Subharmonic functions obey the maximum principle in arbitrary dimension.

[17]Because the constants here do not affect the final conclusion of the lemma, we refer to such an estimate as a *qualitative* estimate rather than a quantitative one. In practice, these sorts of qualitative facts are usually easy to establish (especially when compared to quantitative estimates) by restricting, smoothing, or damping to a nice class of functions, or by smoothing out or discretising various operators and domains. See for instance the proof of this lemma in which we upgrade "for free" a weak qualitative bound (sub-double-exponential growth) to a strong qualitative bound (decay at infinity).

The rather strange sub-double-exponential hypothesis here is completely sharp, as the example $f(z) = e^{-ie^{\pi i z}}$ shows.

Proof The hypotheses and conclusion of the lemma are invariant under the operation of multiplying f by a constant (and adjusting A, B appropriately). So we may normalise $A = 1$. Similarly, the hypotheses and conclusion of the lemma are invariant under the operation of multiplying f by an exponential $\exp(cz)$ for some real c. Using this, one can also normalise $B = 1$. So now f is bounded by 1 on both sides of the strip and we want to show it is bounded by 1 inside the strip.

Let us first assume that f is much better than exponential growth, namely that it goes to zero at infinity. Then for all sufficiently large rectangles $\{0 \leq \mathrm{Re}(z) \leq 1; -N \leq \mathrm{Im}(z) \leq N\}$ the complex-analytic function f is bounded by 1 on all four sides of this rectangle, and hence in the interior also by the maximum principle, and we are done by setting $N \to \infty$.

Now let us handle the general case; as is usual when removing a qualitative assumption, we do this by a limiting argument. We replace $f(z)$ by $f(z)\exp(\varepsilon e^{i[(\pi-\varepsilon)z+\varepsilon/2]})$; a little complex arithmetic shows that this converts the almost double-exponentially growing function f to one which is still complex analytic but is now decaying at infinity. It is still bounded by 1 at both sides of the strip, and hence by 1 in the interior also by the previous argument. Now take $\varepsilon \to 0$ to conclude the claim. ■

Problem 5.11. Suppose that f is analytic on the strip $\{0 \leq \mathrm{Re}(z) \leq 1\}$, obeys the sub-double-exponential bound $|f(z)| \lesssim_f e^{O_f(e^{(\pi-\delta)|z|})}$ on the strip, and obeys the polynomial bounds $|f(z)| \lesssim (1+|z|)^{O(1)}$ on the sides of the strip. Show that it obeys the polynomial bound $|f(z)| \lesssim (1+|z|)^{O(1)}$ on the interior of the strip also.

To apply the three-lines lemma to prove (6), take f to be a simple function (with finite measure support) and consider the entire function

$$z \mapsto \int_X |f|^z.$$

This function has exponential growth at most (because of the qualitative assumption that f is simple with finite measure support), and is bounded by 1 on the lines $\mathrm{Re}(z) = p$ and $\mathrm{Re}(z) = q$, and hence (by a trivially rescaled version of the three lines lemma) bounded by 1 on the strip inside the lines. In particular it is bounded by 1 at $z = r$, which gives the claim for simple functions. The claim for more general functions (dropping the qualitative assumption of simpleness and finite measure support) then follows by a standard limiting argument (using for instance monotone convergence) which we leave as an exercise.

The above argument is a prototype of the *complex interpolation method*. As one can see, it can give slightly sharper results than the real interpolation method (though using the tensor power trick the real method can sometimes "catch up"), but on the other hand requires the quantities being studied to depend complex-analytically on a parameter rather than (say) real-analytically.

Having conclusively demonstrated the log-convexity (6) in multiple ways, let us now give some quick applications. It shows that control on two extreme L^p norms implies control of the intermediate L^p norms. Under additional assumptions on the measure space $(X, \mathcal{B}, \mu)$, one of these extremes is not necessary. If the measure space is *finite* in the sense that $\mu(X) < \infty$ (thus prohibiting functions from being arbitrarily broad), then higher L^p norms control lower ones:

$$\|f\|_p \leq \|f\|_q \mu(X)^{\frac{1}{p}-\frac{1}{q}} \text{ whenever } 0 < p \leq q \leq \infty. \tag{7}$$

Indeed this is trivial when $q = \infty$, and the general case then follows by convexity. The bound (7) can also be usefully written in terms of averages: if we write $⨍_X f \, d\mu$ for $\frac{1}{\mu(X)} \int_X f \, d\mu$, then we see that higher L^p averages control lower L^p averages[18]:

$$(⨍_X |f|^p \, d\mu)^{1/p} \leq (⨍_X |f|^q \, d\mu)^{1/q} \text{ whenever } 0 < p \leq q \leq \infty.$$

By restricting X to its support $\mathrm{supp}(f)$ one can refine (7) to

$$\|f\|_p \leq \|f\|_q \mu(\mathrm{supp}(f))^{\frac{1}{p}-\frac{1}{q}} \text{ whenever } 0 < p \leq q \leq \infty.$$

(Note that this is a limiting case of log-convexity at the exponent 0, in view of Problem 5.1). In the converse direction, if the measure space is *granular* in the sense that one has a lower bound $\mu(E) \geq c$ for all sets E of positive measure, then functions are prohibited from being arbitrarily narrow, and lower norms control higher norms:

$$\|f\|_q \leq \|f\|_p c^{\frac{1}{q}-\frac{1}{p}} \text{ whenever } 0 < p \leq q \leq \infty.$$

This can be seen by first checking the $q = \infty$ case, and then using log-convexity to get the remaining cases. In particular, in the l^p spaces (i.e. L^p spaces with counting measure $d\#$) we see that $\|f\|_{l^q} \leq \|f\|_{l^p}$ for $q \geq p$. For l^p spaces on N points, we thus have (non-matching) upper and lower bounds

$$\|f\|_{l^q} \leq \|f\|_{l^p} \leq N^{\frac{1}{p}-\frac{1}{q}} \|f\|_{l^q}, \tag{8}$$

thus the l^p and l^q norms are comparable to some extent, but the comparability gets worse as $N \to \infty$ or as p and q get further apart.

Problem 5.12. When does equality occur for either of the inequalities in (8)? Note how the example that attains the lower bound is in many ways the "opposite extreme" to the example which attains the upper bound.

Once again, one can gain some intuition into the above estimates by specialising everything to the case of step functions, in which it all collapses to high school algebra (and if one takes logarithms, it collapses further, to linear programming).

Lebesgue measure on Euclidean spaces $\mathbf{R}^d$ with the usual Borel or Lebesgue σ-algebra is not granular. However one can create granularity by coarsening the σ-algebra. For instance, if we let $\mathcal{B}_1$ be the σ-algebra generated by the lattice unit cubes $n + [0,1)^d$ for $n \in \mathbf{Z}^d$, then we have granularity with constant $c = 1$, and now lower L^p norms of functions f control higher ones - but only for functions which are *measurable* with respect to this algebra, i.e. only for functions which are constant

[18]One way to view this is that as one lowers the exponent p, the exceptionally large values of f become less important, leaving the small values of f to dominate.

on each lattice unit cube. (This is the first time we have actually manipulated the *σ-algebra* $\mathcal{B}$ to say something non-trivial, as opposed to manipulating f, X, or μ.) Thus we see that local constancy of functions can lead to additional estimates on L^p norms. Later on we shall see that *frequency localisation* achieves a similar effect as local constancy, as quantified by *Bernstein's inequality*; this is a concrete manifestation of the famous *Heisenberg uncertainty principle*.

Finiteness and granularity of the measure space prevent a function from being too broad or too narrow respectively. Similar things happen when a function is being prevented from being too tall or too short; for instance if f is bounded above by a constant M, then we have

$$\|f\|_q \leq \|f\|_p^{p/q} M^{1-p/q} \text{ whenever } 0 \leq p \leq q \leq \infty$$

(this is just log-convexity at the ∞ exponent), while if f is bounded below by M on its support, then we have the reverse inequality

$$\|f\|_p \leq \|f\|_q^{q/p} M^{1-q/p} \text{ whenever } 0 \leq p \leq q \leq \infty.$$

From Hölder's inequality one also obtains the fundamental *duality* property: if $1 \leq p \leq \infty$, and $f \in L^p(X)$, then[19]

$$\|f\|_p := \sup\{|\int_X f\overline{g}| : \|g\|_{p'} \leq 1\} \tag{9}$$

where $1 \leq p' \leq \infty$ is the dual exponent to p, thus $\frac{1}{p} + \frac{1}{p'} = 1$. We leave the proof of this standard result to the reader; it is sometimes referred to as the *converse to Hölder inequality*. But note that (9), which expresses the L^p norm as the supremum of the magnitude of many linear functions, gives an immediate proof of the triangle inequality (1).

Standard density arguments allow one to restrict g in (9) to lie in any dense subclass of $L^{p'}$, and in particular one can restrict g to be a simple function. (When $p = 1$ one needs to argue a little more carefully, taking advantage of the σ-finite hypothesis.)

There are two obvious algebraic identities involving L^p norms which are worth knowing. The first is that one can interchange l^p sums with L^p integrals for any $0 < p < \infty$, in the sense that

$$\|(\sum_n |f_n|^p)^{1/p}\|_{L^p} = (\sum_n \|f_n\|_{L^p}^p)^{1/p};$$

[19] A small subtlety here - one needs f to be in L^p in the first place in order for the integrals $\int_X f\overline{g}$ to be absolutely convergent. Thus this estimate does not quite assert that the dual space of L^p is $L^{p'}$, though that is indeed the case for $1 < p < \infty$ and (for reasonable measures) for $p = 1$ also. However, in practice the trick of reducing to a dense class of functions generally lets one work with L^p functions as a qualitative *a priori* assumption. Also, if one restricts f, g to non-negative functions (not necessarily in any L^p class), then (9) still holds (when $p = 1$ or $p = \infty$ one needs the σ-finite hypothesis to discover sets of arbitrarily large measure inside sets of infinite measure). Thus, even though the dual of L^∞ is usually not L^1, the estimate (9) for $p = \infty$ lets us pretend for the sake of proving estimates that this is indeed the case.

this is just an application of the Fubini-Tonelli theorem. Secondly, exponents can pass through L^p norms by changing the exponent: for any $0 < p, q < \infty$ we have

$$\| |f|^q \|_{L^p} = \|f\|_{L^{pq}}^q.$$

We shall use both of these identities in the sequel without further comment.

6. Lorentz spaces

Recall that the *weak L^p norm* $\|f\|_{L^{p,\infty}(X,d\mu)}$ of a function f is defined for $0 < p < \infty$ as

$$\|f\|_{L^{p,\infty}(X,d\mu)} := \sup_{\lambda > 0} \lambda \mu(\{|f| \geq \lambda\})^{1/p}.$$

Since

$$\|f\|_p^p = \int_X |f|^p \, d\mu \geq \int_X \lambda^p 1_{|f| \geq \lambda} \, d\mu = \lambda^p \mu(\{|f| \geq \lambda\})$$

for any f and λ, we obtain *Chebyshev's inequality*

$$\|f\|_{L^{p,\infty}(X,d\mu)} \leq \|f\|_{L^p(X,d\mu)}$$

(the case $p = 1$ is also known as *Markov's inequality*). When $p = \infty$ we adopt the convention that $L^{\infty,\infty} = L^\infty$.

We define weak $L^p(X, d\mu)$ or $L^{p,\infty}(X, d\mu)$ to be the space of all functions with finite $L^{p,\infty}(X, d\mu)$ norm, with the usual abbreviations. We sometimes refer to L^p as *strong* L^p to distinguish it from weak L^p.

Example 6.1. *On a Euclidean space $\mathbf{R}^d$, the power function $|x|^{-\alpha}$ lies in weak L^p if and only if $\alpha = d/p$. Indeed one can think of a weak L^p function as a function which is pointwise dominated in magnitude by a rearrangement of (a multiple of) $|x|^{-d/\alpha}$.*

Suppose $0 < p < \infty$. From elementary calculus we have

$$|f(x)|^p = p \int_0^\infty 1_{|f(x)| \geq \lambda} \lambda^p \frac{d\lambda}{\lambda}$$

and hence on integration and Fubini's theorem

$$\|f\|_p^p = p \int_0^\infty \mu(\{|f| \geq \lambda\}) \lambda^p \frac{d\lambda}{\lambda}.$$

To summarise, we have

$$\|f\|_{L^{p,\infty}} = \|\lambda \mu(\{|f| \geq \lambda\})^{1/p}\|_{L^\infty(\mathbf{R}^+, \frac{d\lambda}{\lambda})}$$

and

$$\|f\|_{L^p} = p^{1/p} \|\lambda \mu(\{|f| \geq \lambda\})^{1/p}\|_{L^p(\mathbf{R}^+, \frac{d\lambda}{\lambda})}$$

for $0 < p < \infty$. These two identities motivate introducing the *Lorentz (quasi-)norm*[20] $L^{p,q}(X,\mu)$ for $0 < p < \infty$ and $0 < q \leq \infty$ by[21]

$$\|f\|_{L^{p,q}(X,\mu)} := p^{1/q} \|\lambda \mu(\{|f| \geq \lambda\})^{1/p}\|_{L^q(\mathbf{R}^+, \frac{d\lambda}{\lambda})}.$$

Thus for instance L^p norm is identical to the $L^{p,p}$ norm. We shall abbreviate $\|f\|_{L^{p,q}(X,\mu)}$ by $\|f\|_{L^{p,q}(X)}$, $\|f\|_{L^{p,q}}$, or even $\|f\|_{p,q}$ when there is no chance of confusion.

Problem 6.2. If $f : \mathbf{R}^+ \to \mathbf{R}^+$ is a monotone non-increasing function, show that

$$\|f\|_{L^{p,q}(X,\mu)} = \|f(t) t^{1/p}\|_{L^q(\mathbf{R}^+, \frac{dt}{t})}.$$

(For some arguments it may be convenient to first prove this for smoother f, such as diffeomorphisms with strictly negative derivative, in order to apply an inverse function theorem.)

Problem 6.3. Show that a step function of height H and width W has an $L^{p,q}$ norm of $(p/q)^{1/q} H W^{1/p}$ for any $0 < p < \infty$ and $0 < q \leq \infty$.

Problem 6.4. For any $0 < p, r < \infty$ and $0 < q \leq \infty$ show that

$$\|f^r\|_{L^{p,q}(X,\mu)} \sim_{p,q,r} \|f\|^r_{L^{pr,qr}(X,\mu)}.$$

It is obvious that these norms are both rearrangement-invariant and monotone. To get a better intuitive handle on what the $L^{p,q}$ norm represents, we need some more definitions.

- A *sub-step function* of height H and width W is any function f supported on a set E with the bounds $|f(x)| \leq H$ almost everywhere and $\mu(E) \leq W$. (Thus $|f| \leq H 1_E$.)
- A *quasi-step function*[22] of height H and width W is any function f supported on a set E with the bounds $|f(x)| \sim H$ almost everywhere on E, and $\mu(E) \sim W$. (Thus $|f| \sim H 1_E$.)

Remark 6.5. From the binary expansion of the unit interval $[0,1)$ we see that a non-negative sub-step function f of height 1 and width W can always be decomposed as $\sum_{k=1}^\infty 2^{-k} f_k$ where f_k is an actual step function of height 1 and width at most W. By homogeneity we have a similar statement for other heights. Because of this, bounds on step functions tend to automatically extend to sub-step functions (and hence quasi-step functions) without difficulty.

[20]For various reasons it is not worth trying to define Lorentz norms when $p = \infty$, although we will use the convention $L^{\infty,\infty} = L^\infty$. The most important values of q, in descending order, are $q = p$, $q = \infty$, $q = 1$, and $q = 2$; the other cases essentially never occur in applications.

[21]The factor $p^{1/q}$ is inconsequential, but is traditional in order to maintain compatibility with the strong L^p norm. But in practice the exact form of the Lorentz norm is not important; there are many formulations which are equivalent up to constants, and one generally just picks the formulation which is most convenient. The measure $\frac{d\lambda}{\lambda}$ is of course multiplicative Haar measure on $\mathbf{R}^+$. One can interpret the equivalence of (i)-(iii) below by making the change of variables $\lambda = 2^m$, so that the Haar measure just becomes Lebesgue measure in m (modulo an inessential constant) and then passing from continuous m to discrete m.

[22]It is a little dangerous to put fuzzy notation such as $\sim$ within a definition; if multiple quasi-step functions appear in an argument, the question then arises as to whether the implied constants are uniform. In our applications, the implied constants here are true absolute constants (like 1 and 2) so this will not be an issue.

Just like actual step functions, the $L^{p,q}$ norm of sub-step and quasi-step functions are well controlled; a sub-step function of height H and width W has $L^{p,q}$ norm $O_{p,q}(HW^{1/p})$, while a quasi-step function has $L^{p,q}$ norm $\sim_{p,q} HW^{1/p}$ - almost exactly like actual step functions. In the converse direction, it turns out that every $L^{p,q}$ function can be decomposed as an l^q sum of "very different" $L^{p,q}$-normalised sub-step or quasi-step functions.

Theorem 6.6 (Characterisation of $L^{p,q}$)**.** *Let f be a function, let $0 < p < \infty$ and $1 \leq q \leq \infty$, and let $0 < A < \infty$. Then the following are equivalent up to changes in the implied constants:*

(i) *We have $\|f\|_{L^{p,q}} \lesssim_{p,q} A$.*

(ii) *There exists a decomposition $f = \sum_{m \in \mathbf{Z}} f_m$ where each f_m is a quasi-step function of height 2^m and some width $0 < W_m < \infty$, with the f_m having disjoint supports and*

$$\|2^m W_m^{1/p}\|_{l^q_m(\mathbf{Z})} \lesssim_{p,q} A. \tag{10}$$

Here the m subscript in l^q_m denotes the variable that the l^q norm is being taken over.

(iii) *There exists a pointwise bound $|f| \leq \sum_{m \in \mathbf{Z}} 2^m 1_{E_m}$, with*

$$\|2^m \mu(E_m)^{1/p}\|_{l^q_m(\mathbf{Z})} \lesssim_{p,q} A. \tag{11}$$

(iv) *There exists a decomposition $f = \sum_{n \in \mathbf{Z}} f_n$ where each f_n is a sub-step function of width 2^n and some height $0 < H_n < \infty$, with the f_n having disjoint supports, the H_n non-increasing in n, the bounds $H_{n+1} \leq |f_n| \leq H_n$ on the support of f_n, and*

$$\|H_n 2^{n/p}\|_{l^q_n(\mathbf{Z})} \lesssim_{p,q} A. \tag{12}$$

(v) *There exists a pointwise bound $|f| \leq \sum_{n \in \mathbf{Z}} H_n 1_{E_n}$, where $\mu(E_n) \lesssim_{p,q} 2^n$ and (12) holds.*

Remark 6.7. The formulations (ii), (iv) are useful when trying to *use* an $L^{p,q}$ bound on f; the formulations (iii), (v) are useful when trying to *obtain* an $L^{p,q}$ bound on f. Heuristically, the above theorem is trying to say the following. If f is a quasi-step function of height H and width W, then $\|f\|_{L^{p,q}} \sim_{p,q} HW^{1/p}$. But if f is instead the sum $\sum_n f_n$ of quasi-step functions of height H_n and width W_n, and either the heights or the widths are sufficiently variable in n (e.g. one or the other grows like a power of two), then $\|\sum_n f_n\|_{L^{p,q}} \sim_{p,q} \|H_n W_n^{1/p}\|_{l^q_n}$.

Proof We may use homogeneity symmetry to normalise $A = 1$. The implications $(ii) \implies (iii)$ and $(iv) \implies (v)$ are trivial. To see that (i) implies (ii), set $f_m := f 1_{2^{m-1} < |f| \leq 2^m}$ and $W_m := \mu(\{2^{m-1} < |f| \leq 2^m\})$. (This is the "vertically dyadic layer cake decomposition".) The only thing that requires nontrivial verification is (10); but one easily verifies that

$$2^m W_m^{1/p} \lesssim_{p,q} \|\lambda \mu(\{|f| \geq \lambda\})^{1/p}\|_{L^q([2^{m-2}, 2^{m-1}], \frac{d\lambda}{\lambda})}$$

and the claim follows by summing this in l^q.

Similarly, to see that (i) implies (iv), define

$$H_n := \inf\{\lambda : \mu(\{|f| > \lambda\}) \leq 2^{n-1}\};$$

note that this is a non-increasing function of n, which goes to zero as $n \to \infty$ (this comes from the hypothesis that $\|f\|_{L^{p,q}}$ is finite). We then define

$$f_n := f 1_{H_n \geq |f| > H_{n+1}}.$$

(This is the "horizontally dyadic layer cake decomposition".) The only non-trivial thing to verify is (12). But one easily verifies the telescoping estimate

$$\begin{aligned} H_n 2^{n/p} &= (H_n^q 2^{nq/p})^{1/q} \\ &= (\sum_{k=0}^\infty (H_{n+k}^q - H_{n+k+1}^q) 2^{nq/p})^{1/q} \\ &\lesssim_{p,q} (\sum_{k=0}^\infty 2^{-kq/p} \|\lambda 2^{(n+k)/p}\|^q_{L^q([H_{n+k+1}, H_{n+k}], \frac{d\lambda}{\lambda})})^{1/q} \\ &\lesssim_{p,q} (\sum_{k=0}^\infty 2^{-kq/p} \|\lambda \mu(\{|f| \geq \lambda\})^{1/p}\|^q_{L^q([H_{n+k+1}, H_{n+k}], \frac{d\lambda}{\lambda})})^{1/q} \end{aligned}$$

and the claim follows by summing this in l^q and interchanging the summation signs. (We leave to the reader how to modify the above argument to handle the case $q = \infty$.)

It remains to show that (iii) implies (i) and (iv) implies (i). Suppose first that (iii) holds. It is clear that for any m we have

$$\mu(\{|f| > 2^m\}) \leq \sum_{k=0}^\infty \mu(E_{m+k})$$

and hence

$$\|\lambda \mu(\{|f| \geq \lambda\})^{1/p}\|_{L^q((2^m, 2^{m+1}], \frac{d\lambda}{\lambda})} \lesssim_{p,q} 2^m (\sum_{k=0}^\infty \mu(E_{m+k}))^{1/p}$$

and so on taking l^q summation in m it would suffice to show that

$$\|2^m (\sum_{k=0}^\infty \mu(E_{m+k}))^{1/p}\|_{l^q_m} \lesssim_{p,q} 1.$$

Raising to the p^{th} power we rewrite as

$$\|\sum_{k=0}^\infty 2^{pm} \mu(E_{m+k})\|_{l_m^{q/p}} \lesssim_{p,q} 1.$$

But from the hypothesis we have

$$\|2^{pm} \mu(E_m)\|_{l_m^{q/p}} \lesssim_{p,q} 1$$

and hence on shifting m by k

$$\|2^{pm} \mu(E_{m+k})\|_{l_m^{q/p}} \lesssim_{p,q} 2^{-kp}.$$

The claim then follows from (2), (3), or Q1.

Now suppose that (iv) holds. Observe that for any $\lambda > 0$ we have

$$\mu(\{|f| > \lambda\}) \lesssim_{p,q} \sup\{2^n : H'_n \geq \lambda\}$$

where H'_n are the modified heights

$$H'_n := \sum_{k=0}^{\infty} H_{n+k}.$$

Indeed, if $\mu(\{|f| > \lambda\}) > 2^{n-1}$ for some n then one easily verifies that $H_n \geq \lambda$ and hence $H'_n \geq \lambda$. The shifting trick and triangle inequality argument used previously shows that H'_n obeys the same bound (12) as H_n, thus

$$\|H'_n 2^{n/p}\|_{l^q_n(\mathbf{Z})} \lesssim_{p,q} 1.$$

We now compute

$$\begin{aligned}
\|\lambda\mu(\{|f| \geq \lambda\})^{1/p}\|^q_{L^q(\mathbf{R}^+, \frac{d\lambda}{\lambda})} &\lesssim_{p,q} \int_0^\infty \lambda^{q-1} \sup\{2^{nq/p} : H'_n \geq \lambda\}\, d\lambda \\
&\lesssim_{p,q} \sum_n \int_0^\infty \lambda^{q-1} 2^{nq/p} 1_{H'_n \geq \lambda}\, d\lambda \\
&\sim_{p,q} \sum_n 2^{nq/p} (H'_n)^q \\
&\lesssim 1
\end{aligned}$$

as desired. (We leave to the reader how to modify the above argument to handle the case $q = \infty$.) ■

Remark 6.8. Suppose that the ratio between the tallest height and lowest non-zero height of a function f is N (i.e. there exists A such that $A \leq |f(x)| \leq AN$ whenever $f(x)$ is non-zero). Then the above theorem shows that two different Lorentz norms $\|f\|_{L^{p,q_1}}$, $\|f\|_{L^{p,q_2}}$ with the same primary exponent p only differ by multiplicative powers of $\log N$. Similarly if the broadest width and narrowest width of a function differs by N (e.g. if $\mu(X)$ is equal to N times the granularity c of X). What this indicates is that the secondary exponent q in the Lorentz norms only offers "logarithmic correction" to the Lebesgue norms L^p; in contrast, (4) shows that varying the primary exponent p leads to polynomial-strength changes in the norm. So as a first approximation (ignoring logarithms) one can pretend that $L^{p,q} \approx L^p$. Note also that for quasi-step functions, the $L^{p,q}$ norms barely depend on q at all.

One easy corollary of the above theorem is that the $L^{p,q}$ quasi-norm is indeed a quasi-norm; this can be seen for instance by using the equivalence of (i) and (iii). Another easy consequence is that the simple functions are dense in $L^{p,q}$.

A particularly useful consequence of the above theorem is a Hölder inequality for Lorentz spaces, due to O'Neil.

Theorem 6.9 (Hölder's inequality in Lorentz spaces)**.** *If $0 < p_1, p_2, p < \infty$ and $0 < q_1, q_2, q \leq \infty$ obey $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$ and $\frac{1}{q} = \frac{1}{q_1} + \frac{1}{q_2}$ then*

$$\|fg\|_{L^{p,q}} \lesssim_{p_1,p_2,q_1,q_2} \|f\|_{L^{p_1,q_1}} \|g\|_{L^{p_2,q_2}}$$

whenever the right-hand side norms are finite.

Proof We may normalise $\|f\|_{L^{p_1,q_1}} = \|g\|_{L^{p_2,q_2}} = 1$, and drop the dependence of the implied constants on p_1, p_2, q_1, q_2 for brevity. By the equivalence of (i) and (v) in Theorem 6.6 we may dominate $|f| \leq \sum_n H_n 1_{E_n}$ and $|g| \leq \sum_n H'_n 1_{E'_n}$ where $\mu(E_n), \mu(E'_n) \lesssim 2^n$ and

$$\|H_n 2^{n/p_1}\|_{l^{q_1}_n}, \|H'_n 2^{n/p_2}\|_{l^{q_2}_n} \lesssim 1.$$

Then we have

$$|fg| \leq \sum_k \sum_n H_n H'_{n+k} 1_{E_n \cap E'_{n+k}}.$$

By the quasi-triangle inequality and monotonicity it suffices to show that

$$\|\sum_{k \geq 0} \sum_n H_n H'_{n+k} 1_{E_n \cap E'_{n+k}}\|_{L^{p,q}} \lesssim 1$$

and

$$\|\sum_{k < 0} \sum_n H_n H'_{n+k} 1_{E_n \cap E'_{n+k}}\|_{L^{p,q}} \lesssim 1.$$

By symmetry it suffices to consider the $k \geq 0$ component. Here we observe that $E_n \cap E'_{n+k}$ has measure at most 2^n, so by the equivalence of (i) and (v) in Theorem 6.6

$$\|\sum_n H_n H'_{n+k} 1_{E_n \cap E'_{n+k}}\|_{L^{p,q}} \lesssim \|H_n H'_{n+k} 2^{n/p}\|_{l^q_n}.$$

But by the ordinary Hölder inequality

$$\|H_n H'_{n+k} 2^{n/p}\|_{l^q_n} \leq \|H_n 2^{n/p_1}\|_{l^{q_1}_n} \|H'_{n+k} 2^{n/p_2}\|_{l^{q_2}_n};$$

shifting the second n by k we conclude

$$\|\sum_n H_n H'_{n+k} 1_{E_n \cap E'_{n+k}}\|_{L^{p,q}} \lesssim 2^{-k/p_2}.$$

The claim now follows from Q1. ■

One corollary of this Hölder inequality is that $L^{p,q}$ functions are absolutely integrable on sets of finite measure whenever $p > 1$.

Now we consider dual formulations of the $L^{p,q}$ norms. The case $q = \infty$ is fairly straightforward:

Problem 6.10 (Dual formulation of weak L^p). Let $1 < p \leq \infty$. Then for every f in $L^{p,\infty}(X, d\mu)$, we have

$$\|f\|_{L^{p,\infty}(X,d\mu)} \sim_p \sup\{\mu(E)^{-1/p'} |\int_X f 1_E \, d\mu| : 0 < \mu(E) < \infty\} \tag{13}$$

Also show that the hypothesis $f \in L^{p,\infty}(X, d\mu)$ can be dropped if one instead assumes f to be non-negative[23].

[23]What is going on here is that there are two overlapping notions of integral $\int_X \, d\mu$ being used here; the *absolutely convergent integral*, which only makes sense for L^1 functions, and the *non-negative integral*, which only makes sense for non-negative functions. Fortunately, the two concepts of integral agree on their common domain (absolutely integrable non-negative functions), and furthermore dominated convergence and the monotone convergence theorem, coupled with the ability to approximate both L^1 and non-negative functions by simple functions of finite support, generally allows one to translate results from one context to the other.

The right-hand side is clearly a semi-norm at least on f. This leads in particular to a quasi-triangle inequality

$$\|f_1 + \ldots + f_N\|_{L^{p,\infty}(X,d\mu)} \sim_p \|f_1\|_{L^{p,\infty}(X,d\mu)} + \ldots + \|f_N\|_{L^{p,\infty}(X,d\mu)}$$

for any $f_1, \ldots, f_N \in L^{p,\infty}(X, d\mu)$.

Remark 6.11. It is worth comparing (13) to (9). In (9), one takes the inner product of f against all $L^{p'}$-normalised functions, and the worst inner product becomes the L^p norm. In (13), one only takes the inner product of f against the $L^{p'}$-normalised *step functions* $\mu(E)^{-1/p'} 1_E$. This is fully consistent with the fact that the L^p norm is stronger than the weak L^p norm.

Problem 6.10 can be rephrased as follows: if $f \in L^{p,\infty}$ for some $1 < p < \infty$ and $A > 0$, then the following two statements are equivalent (up to changes in the implied constants):

- $\|f\|_{L^{p,\infty}} \lesssim_p A$.
- $\int_X f 1_E \, d\mu = O_p(\mu(E)^{1/p'})$ for all sets E of finite measure.

Unfortunately this equivalence breaks down at $p = 1$ or below (consider for instance the weak L^p function $|x|^{-d/p}$ on $\mathbf{R}^d$, which is not even locally integrable when $p \leq 1$). However, one does have a substitute: see Q5.

For more general $L^{p,q}$ spaces, we have

Theorem 6.12 (Dual characterisation of $L^{p,q}$)**.** *Let $1 < p < \infty$ and $1 \leq q \leq \infty$. Then for any $f \in L^{p,q}$,*

$$\|f\|_{L^{p,q}} \sim_{p,q} \sup\{|\int_X f\overline{g} \, d\mu| : \|g\|_{L^{p',q'}} \leq 1\}.$$

Again, the hypothesis $f \in L^{p,q}$ can be dropped if f is non-negative and g is restricted to also be non-negative.

Thus the $L^{p,q}$ quasi-norm is in fact equivalent to a norm when $1 < p < \infty$ and $q \geq 1$. In particular, weak L^p is equivalent to a normed space when $1 < p < \infty$. (For $p = 1$, weak L^p fails to be normable "by a logarithm"; see Q3.) As with other dual characterisations, one can restrict g to a dense subclass of $L^{p',q'}$, for instance simple functions with finite measure support.

Proof To obtain the $\gtrsim_{p,q}$ part of this theorem, we simply estimate

$$|\int_X f\overline{g} \, d\mu| \leq \|fg\|_{L^1} = \|fg\|_{L^{1,1}}$$

and use Theorem 6.9. To obtain the $\lesssim_{p,q}$ part, we normalise $\|f\|_{L^{p,q}} = 1$. It then suffices by homogeneity to find g with $\|g\|_{L^{p',q'}} \leq 1$ and $\int_X f\overline{g} \, d\mu \gtrsim 1$.

The case $q = \infty$ follows from (13), so let us take $q < \infty$. By the equivalence of (i) and (ii) in Theorem 6.6 may write $f = \sum_m f_m$ where f_m is a quasi-step function of height 2^m and width W_m with disjoint supports such that the sequence $a_m := 2^m W_m^{1/p}$ has an l^q_m norm of $\sim_{p,q} 1$. Now take

$$g := \sum_m g_m$$

where

$$g_m := a_m^{q-p} |f_m|^{p-2} f_m$$

adopting the obvious convention that $|f_m|^{p-2} f_m = 0$ when $f_m = 0$. Then (because of the disjoint supports)

$$|\int_X f\overline{g}| = \sum_m \int_X a_m^{q-p} |f_m|^p.$$

But since f_m has height 2^m and width W_m, $\int_X |f_m|^p \sim_p 2^{mp} W_m = a_m^p$ and so

$$|\int_X f\overline{g}| \sim_p \sum_m a_m^q \sim_{p,q} 1.$$

To conclude it will suffice to show that

$$\|\sum_m g_m\|_{L^{p',q'}} \lesssim_{p,q} 1.$$

If E_m is the support of f_m, then we have the pointwise bound

$$g_m \lesssim_{p,q} a_m^{q-p} 2^{m(p-1)} 1_{E_m}$$

and the measure bound $\mu(E_m) \lesssim_{p,q} W_m = 2^{-mp} a_m^p$.

At this point we would like to apply Theorem 6.6, but neither the height nor width of g_m is necessarily a power of 2. But we can remedy this by introducing the modified heights

$$H_m := \sup_{k \geq 0} a_{m-k}^{q-p} 2^{m(p-1)} 2^{-k(p-1)/2}.$$

We have $H_{m+1} \geq 2^{(p-1)/2} H_m$, and so the H_m increase geometrically. It then suffices to show that

$$\|\sum_m H_m 1_{E_m}\|_{L^{p',q'}} \lesssim_{p,q} 1.$$

By refining the m by a constant factor we can make each H_m at least twice as large as the previous, and so by applying the equivalences of (i) and (iii) in Theorem 6.6 and the triangle inequality it suffices to show that

$$\|H_m \mu(E_m)^{1/p'}\|_{l^{q'}_m} \lesssim_{p,q} 1$$

which we expand using our bound on $\mu(E_m)$ as

$$\|a_m^{p-1} \sup_{k \geq 0} a_{m-k}^{q-p} 2^{-k(p-1)/2}\|_{l^{q'}_m}.$$

But from Hölder's inequality and the l^q bound on a_m we have

$$\|a_m^{p-1} a_{m-k}^{q-p} 2^{-k(p-1)/2}\|_{l^{q'}_m} \lesssim_{p,q} 2^{-k(p-1)/2};$$

summing this using the triangle inequality (and estimating the supremum by a sum) we obtain the claim.

The case when f, g are restricted to be non-negative can be deduced from the above result and a monotone convergence argument (representing f as a monotone limit of simple functions of finite measure support) which we leave as an exercise to the reader. ∎

7. Orlicz spaces (Optional)

So far we have studied the Lebesgue spaces L^p, together with the more general Lorentz spaces $L^{p,q}$, which includes weak L^p as a special case. These spaces are all rearrangement-invariant and monotone. There is a different generalisation of the Lebesgue spaces L^p, the *Orlicz spaces* $\Phi(L)$, which are also rearrangement-invariant and monotone, and which are occasionally useful. (There is a common generalisation of both, the *Lorentz-Orlicz spaces*, but these occur very rarely in applications.)

The motivation for Orlicz spaces starts with the trivial observation that if $1 \leq p < \infty$, then

$$\|f\|_{L^p} \leq 1 \text{ if and only if } \int_X |f|^p \, d\mu \leq 1.$$

Inspired by this, we generalise by letting $\Phi : \mathbf{R}^+ \to \mathbf{R}^+$ be a function (with some additional properties to be selected shortly) and ask if we can find a norm $\|f\|_{\Phi(L)}$ which obeys the property

$$\|f\|_{\Phi(L)} \leq 1 \text{ if and only if } \int_X \Phi(|f|) \, d\mu \leq 1. \tag{14}$$

Since norms need to be homogeneous, this would imply

$$\|f\|_{\Phi(L)} \leq A \text{ if and only if } \int_X \Phi(|f|/A) \, d\mu \leq 1$$

for all $A > 0$. In particular, if $A < A'$, then we need

$$\int_X \Phi(|f|/A) \, d\mu \leq 1 \text{ implies } \int_X \Phi(|f|/A') \, d\mu \leq 1.$$

To ensure this property it is thus very natural to require that Φ be *increasing*. Also to deal with the zero norm case one typically requires $\Phi(0) = 0$.

Next, in order for $\Phi(L)$ to be a norm, the unit ball $\{f : \|f\|_{\Phi(L)} \leq 1\}$ needs to be convex. Looking at (14), we see that this will indeed be the case when Φ is itself convex. (Note that the proof of (1) was a special case of this argument).

We can put all the above discussion together and conclude: if $\Phi : \mathbf{R}^+ \to \mathbf{R}^+$ is increasing and convex with $\Phi(0) = 0$, then the norm

$$\|f\|_{\Phi(L)} := \inf\{A > 0 : \int_X \Phi(|f|/A) \, d\mu \leq 1\}$$

is a norm on the space $\Phi(L) := \{f : \|f\|_{\Phi(L)} < \infty\}$.

As discussed above, the L^p spaces for $1 \leq p < \infty$ are examples of Orlicz spaces with $\Phi(x) := x^p$. The space L^∞ is not really an Orlicz space, but can be viewed the limiting case where $\Phi(x)$ is infinite for $x > 1$ and zero for $x \leq 1$ (or more informally,

$\Phi(x) = x^{+\infty}$). Aside from the Lebesgue spaces, the most common Orlicz spaces which appear are

- The space $L \log L$, defined as the Orlicz space with $\Phi(x) := x \log(2 + x)$;
- The space e^L, defined as the Orlicz space with $\Phi(x) := e^x - 1$;
- The space e^{L^2}, defined as the Orlicz space with $\Phi(x) := e^{x^2} - 1$.

The correction factors of 2 and 1 in the above functions should not be taken too seriously; note that if two functions $\Phi, \tilde{\Phi}$ are comparable then their Orlicz norms are comparable also; a little more generally, if $\Phi \lesssim \tilde{\Phi}$, then $\|f\|_{\Phi(L)} \lesssim \|f\|_{\tilde{\Phi}(L)}$. It is the behaviour of $\Phi(x)$ for large values of x which is the most important, although when X has infinite measure the behaviour at small values of x is also relevant.

Problem 7.1. If X has finite measure, verify the relation

$$\|f\|_{\Phi(L)} \sim_{\Phi,\mu(X)} \|f\|_{1+\Phi(L)}$$

which is another indication of the irrelevance of the low values of Φ in the finite measure case.

The final fact about Orlicz spaces that we give here is the duality relation. Suppose that $\Phi : \mathbf{R}^+ \to \mathbf{R}^+$ is increasing, convex, and is also *superlinear* in the sense that $\lim_{x \to +\infty} \Phi(x)/x = +\infty$. We can then define the *Young dual* $\Psi : \mathbf{R}^+ \to \mathbf{R}^+$ of Φ by the formula

$$\Psi(y) := \sup\{xy - \Phi(x) : x \in \mathbf{R}^+\};$$

the hypothesis that Φ is superlinear ensures that this function is well-defined. We may equivalently define $\Psi(y)$ to be the smallest function for which one has the inequality

$$xy \leq \Phi(x) + \Psi(y) \text{ for all } x, y \in \mathbf{R}^+. \tag{15}$$

Problem 7.2. If $1 < p < \infty$, show that the Young dual of $\Phi(x) = x^p$ is $\Psi(y) = \frac{p^{p'/p}}{p'} y^{p'}$. Show also that the Young dual of $\Phi(x) = x \log(2 + x)$ takes the form $\Psi(y) \sim e^y$ for $y > 1$. What does the Young duals of $e^x - 1$ and $e^{x^2} - 1$ look like?

One can easily verify that Ψ is also increasing, convex, and super-linear, and so the Orlicz norm $\|\|_{\Psi(L)}$ makes sense. From (15) and the triangle inequality it is immediate that

$$|\int_X f\overline{g}\, d\mu| \leq 2 \text{ whenever } \|f\|_{\Phi(L)}, \|g\|_{\Psi(L)} \leq 1$$

and hence by homogeneity we obtain the duality relation

$$|\int_X f\overline{g}\, d\mu| \leq 2\|f\|_{\Phi(L)}\|g\|_{\Psi(L)}$$

whenever $f \in \Phi(L)$ and $g \in \Psi(L)$.

Problem 7.3. Establish the more precise relationship

$$\|f\|_{\Phi(L)} \sim \sup\{|\int_X f\overline{g}\, d\mu| : \|g\|_{\Psi(L)} \leq 1\}.$$

Problem 7.4. Show that if Ψ is the Young dual of Φ, then Φ is the Young dual of Ψ. (It may help to view things geometrically, and in particular understanding Ψ as parameterising the support lines of the graph of the convex function Φ.)

Problem 7.5. When X has finite measure, show that the spaces $L \log L$ and e^L are dual to each other. What is the dual to e^{L^2}?

8. Real interpolation

So far we have only considered functions f on a single measure space $X = (X, \mathcal{B}_X, \mu_X)$. Now we shall consider operators T which take functions on one measure space $X = (X, \mathcal{B}_X, \mu_X)$ to functions on another measure space $Y = (Y, \mathcal{B}_Y, \mu_Y)$; the study of such operators is in fact a major focus of harmonic analysis. Ultimately we want to extend T to a standard normed vector space such as $L^p(X)$, but in practice one has to initially first restrict attention to a dense subspace of functions, such as simple functions or test functions.

We are primarily interested in linear operators, thus $T(cf) = cTf$ and $T(f + g) = Tf + Tg$. But it is also worth considering the more general *sublinear operators*, in which

$$|T(cf)| = |c||Tf|$$

and we have the pointwise estimate

$$|T(f + g)| \leq |Tf| + |Tg|.$$

Apart from the linear operators, the next most important example of a sublinear operator is a *maximal operator*

$$Tf(x) := \sup_n |T_n f(x)|$$

where T_n are a collection (possibly countably or uncountably infinite, though in the latter case one has to take some care in ensuring measurability) of linear or sublinear operators. The third most important example is a *square function* such as

$$Tf(x) := (\sum_n |T_n f(x)|^2)^{1/2}.$$

More generally, one can consider a family T_y of operators indexed by some parameter y, and take $Tf(x)$ to be the norm in the y variable of $T_y f$ in some suitable norm. But the above three examples of linear operators, maximal operators, and square functions already cover the vast majority of applications.

Let $0 < p, q \leq \infty$ be exponents, and let T be sublinear. Let us define the following concepts.

- We say that T is[24] *strong-type* (p, q) (or simply *type* (p, q)) if we have a bound
$$\|Tf\|_{L^q(Y)} \lesssim_{T,p,q} \|f\|_{L^p(X)}$$

[24]This is also called type (L^p, L^q); more generally, type (V, W) means that T is bounded from a (dense subspace of) V to W.

for all f in L^p, or in a dense sub-class thereof. Note that in the latter case there is a unique extension to all of L^p.

- If $q < \infty$, we say that T is *weak-type* (p, q) if we have a bound
$$\|Tf\|_{L^{q,\infty}(Y)} \lesssim_{T,p,q} \|f\|_{L^p(X)}$$
- We say that T is *restricted strong-type* (p, q) if we have a bound
$$\|Tf\|_{L^q(Y)} \lesssim_{T,p,q} HW^{1/p} \tag{16}$$
for all sub-step functions of height H and width W. In particular, we have
$$\|T1_E\|_{L^q(Y)} \lesssim_{T,p,q} \mu(E)^{1/p}. \tag{17}$$
(Conversely, we can deduce (16) from (17) using tricks such as those in Remark 6.5.)
- If $q < \infty$, we say that T is *restricted weak-type* (p, q) if we have a bound
$$\|Tf\|_{L^{q,\infty}(Y)} \lesssim_{T,p,q} HW^{1/p} \tag{18}$$
for all sub-step functions f of height H and width W. In particular, we have
$$\|T1_E\|_{L^{q,\infty}(Y)} \lesssim_{T,p,q} \mu(E)^{1/p}. \tag{19}$$

Clearly, whenever p, q are fixed, strong-type implies weak-type and restricted strong-type, either of which imply restricted weak-type. In most applications, it is the strong-type bounds which are desired; however, we shall see in this section that the *real interpolation method* allows us to deduce strong-type bounds from weak-type, or even restricted weak-type bounds, as long as the strong-type bounds are an interpolant between the restricted weak-type bounds. This can be a very useful strategy, because (as we shall see in next week's notes) weak-type or restricted weak-type bounds are easier to prove than strong-type estimate.

Let us first make a mild (and qualitative) assumption, namely that the form
$$\langle |Tf|, |g| \rangle := \int_Y |Tf||g|\ d\nu \tag{20}$$
is well-defined whenever f, g are simple functions with finite measure support. This is for instance the case if T is of restricted type (p, q) for some $0 < p < \infty$ and $1 \leq q \leq \infty$, or restricted weak-type (p, q) for some $0 < p < \infty$ and $1 < q < \infty$; thus in practice this assumption is easily satisfied. We observe that this form is non-negative, homogeneous and sublinear in both f and g:
$$\langle |Tcf|, |g| \rangle = \langle |Tf|, |cg| \rangle = |c| \langle |Tf|, |g| \rangle$$
$$\langle |T(f_1 + f_2)|, |g| \rangle \leq \langle |Tf_1|, |g| \rangle + \langle |Tf_2|, |g| \rangle$$
$$\langle |Tf|, |g_1 + g_2| \rangle \leq \langle |Tf|, |g_1| \rangle + \langle |Tf|, |g_2| \rangle.$$
The form $\langle |Tf|, |g| \rangle$ turns out to be a convenient way to understand the various types of T, and the ability to decompose both f and g independently is very useful in establishing interpolation type results. There is a near-symmetry between f and g here, if we could somehow take an "adjoint" of the operator T, but we will not explicitly exploit this symmetry here as it is not always available for sublinear operators (though the "duality" or "adjoint" trick is undoubtedly very powerful in the important linear case).

Let us look in particular at the form (20) applied to indicator functions, thus we look at the quantity $\langle |T1_E|, 1_F\rangle$ for $E \subset X$ and $F \subset Y$ of finite measure. Now suppose that T had some strong type (p,q) bound for some $0 < p < \infty$ and $1 \leq q \leq \infty$, say

$$\|Tf\|_{L^q(Y)} \lesssim_{p,q} A\|f\|_{L^p(X)}$$

for some $A > 0$. Then in particular

$$\|T1_E\|_{L^q(Y)} \lesssim_{p,q} A\mu(E)^{1/p}$$

and hence by Hölder's inequality

$$\langle |T1_E|, 1_F\rangle \lesssim_{p,q} \mu(E)^{1/p}\nu(F)^{1/q'}. \tag{21}$$

Actually it is clear that strong type (p,q) is too much of an assumption; restricted type (p,q) would have sufficed for the conclusion. If $q > 1$, we can relax things further to restricted weak-type:

Problem 8.1. Let $0 < p \leq \infty$, $1 < q \leq \infty$, and $A > 0$. Let T be a sublinear operator such that the form (20) is well-defined. Then the following are equivalent up to changes in the implied constant:

- T is restricted weak-type (p,q) with constant A, in the sense that
$$\|Tf\|_{L^{q,\infty}(Y)} \lesssim_{p,q} AHW^{1/p} \tag{22}$$
for all simple sub-step functions f of height H and width W.
- For all $E \subset X$, $F \subset Y$ of finite measure, we have the bound
$$\langle |T1_E|, 1_F\rangle \lesssim_{p,q} A\mu(E)^{1/p}\nu(F)^{1/q'}$$

(Hint: use (13) and Remark 6.5.)

This already gives a simple version of real interpolation:

Corollary 8.2 (Baby real interpolation). *Let T be a sublinear operator such that the form* (20) *is well-defined. Let $0 < p_0, p_1 \leq \infty$, $1 < q_0, q_1 \leq \infty$ and $A_0, A_1 > 0$ be such that T is restricted weak-type (p_i, q_i) with constant A_i (in the sense of* (22)*) for $i = 0, 1$. Then T is also restricted weak-type (p_θ, q_θ) with constant A_θ for $0 \leq \theta \leq 1$, where*

$$\frac{1}{p_\theta} := \frac{1-\theta}{p_0} + \frac{\theta}{p_1}; \quad \frac{1}{q_\theta} := \frac{1-\theta}{q_0} + \frac{\theta}{q_1}; A_\theta := A_0^{1-\theta}A_1^\theta \tag{23}$$

and the implied constant depends on p_0, p_1, q_0, q_1.

Indeed, all we are using here is the obvious algebraic observation that if $X \lesssim Y_0$ and $X \lesssim Y_1$, then $X \lesssim Y_\theta := Y_0^{1-\theta}Y_1^\theta$ for all $0 \leq \theta \leq 1$.

The above corollary has two defects. Firstly, it can only conclude restricted weak-type rather than strong type. Secondly, the restriction $q_0, q_1 > 1$ is inconvenient for many applications, since weak $L^{1,\infty}$ bounds are actually rather common. To address the second concern, we have the following variant of Problem 8.1:

Problem 8.3. Let $0 < p \leq \infty$, $0 < q \leq \infty$, and $A > 0$. Let T be a sublinear operator such that the form (20) is well-defined. Then the following are equivalent up to changes in the implied constant:

- T is restricted weak-type (p, q) with constant A, in the sense that (22) holds.
- For all $E \subset X$, $F \subset Y$ of finite non-zero measure, there exists $F' \subset F$ with $\mu(F') \geq \frac{1}{2}\mu(F)$ such that
$$\langle |T1_E|, 1_{F'}\rangle \lesssim_{p,q} A\mu(E)^{1/p}\nu(F)^{1/q'}$$

Hint: use Q5.

Corollary 8.4. *The hypothesis $q_0, q_1 > 1$ in Corollary 8.2 can be weakened to $q_0, q_1 > 0$.*

Now we can give a significantly more useful real interpolation theorem, which can interpolate between restricted weak-type estimates to obtain strong-type estimates.

Theorem 8.5 (Marcinkeiwicz interpolation theorem)**.** *Let T be a sublinear operator such that the form (20) is well-defined. Let $0 < p_0, p_1 \leq \infty$, $0 < q_0, q_1 \leq \infty$ and $A_0, A_1 > 0$ be such that T is restricted weak-type (p_i, q_i) with constant A_i (in the sense of (22)) for $i = 0, 1$. Suppose also that $p_0 \neq p_1$ and $q_0 \neq q_1$. Then for any $0 < \theta < 1$ and $1 \leq r \leq \infty$ with $q_\theta > 1$ we have*
$$\|Tf\|_{L^{q_\theta,r}(Y)} \lesssim_{p_0,p_1,q_0,q_1,r,\theta} A_\theta \|f\|_{L^{p_\theta,r}(X)}$$
for all simple functions f with finite measure support, where p_θ, q_θ, A were defined in (23). In particular, if $q_\theta \geq p_\theta$, then T is strong-type (p_θ, q_θ) with constant $O_{p_0,p_1,q_1,\theta}(A_\theta)$.

Proof To simplify the notation let us suppress the dependence on $p_0, p_1, q_0, q_1, r, \theta$. Observe that the statement and conclusion of the theorem have several homogeneity symmetries. The most obvious one is that we can multiply T (and the A_θ) by an arbitrary constant, but we may also multiply the measure μ by a constant C (and the A_θ by the constant C^{-1/p_θ}); similarly we may multiply ν by C and A_θ by C^{1/q_θ}). Using these symmetries we may normalise $A_0 = A_1 = 1$ (and hence $A_\theta = 1$ for all θ). Our task is then to show that
$$\|Tf\|_{L^{q_\theta,r}(Y)} \lesssim \|f\|_{L^{p_\theta,r}(X)}$$
for all simple functions f with finite measure support.

Using Theorem 6.12 (and the hypothesis $q_\theta > 1$), this is equivalent to showing that
$$\langle |Tf|, |g|\rangle \lesssim \|f\|_{L^{p_\theta,r}(X)} \|g\|_{L^{q'_\theta,r'}(Y)} \tag{24}$$
for all simple functions f, g of finite measure support.

We currently have $q_0, q_1 > 0$ and $q_\theta > 1$. By using Corollary 8.4 to bring the restricted weak-type exponents (p_0, q_0) and (p_1, q_1) a bit closer to (p_θ, q_θ) we may assume that $q_0, q_1 > 1$ as well. Applying Theorem 8.1 we conclude that

$$\langle |T1_E|, 1_F \rangle \lesssim |E|^{1/p_i} |F|^{1/q_i'}$$

for all sets E, F of finite measure and $i = 0, 1$. From Remark 6.5 and sublinearity we conclude that

$$\langle |Tf|, |g| \rangle \lesssim H W^{1/p_i} H' (W')^{1/q_i'}$$

whenever f, g are sub-step functions of height and widths H, H' and W, W' respectively. We can of course pick the better of the two estimates, leading to

$$\langle |Tf|, |g| \rangle \lesssim H H' \min_{i=0,1} (W^{1/p_i} (W')^{1/q_i'}). \tag{25}$$

Now we can return to proving (24). By homogeneity we may normalise

$$\|f\|_{L^{p_\theta, r}(X)} = \|g\|_{L^{q_\theta', r'}(Y)} = 1.$$

We then apply Theorem 6.6 to decompose $f = \sum_{n \in \mathbf{Z}} f_n$, $g = \sum_{n \in \mathbf{Z}} g_n$ with f_n, g_n sub-step functions of width 2^n and heights H_n, H_n' respectively, with the height bounds

$$\|a\|_{l^r(\mathbf{Z})}, \|b\|_{l^{r'}(\mathbf{Z})} \lesssim 1 \tag{26}$$

where a, b are the sequences

$$a_n := H_n 2^{n/p_\theta}; \quad b_n := H_n' 2^{n/q_\theta'}.$$

Since f, g are simple functions, only finitely many of the f_n and g_n are non-zero[25]. We now use sublinearity to estimate

$$\langle |Tf|, |g| \rangle \leq \sum_{n,m} \langle |Tf_n|, |g_m| \rangle$$

and then use (25) to obtain

$$\langle |Tf|, |g| \rangle \lesssim \sum_{n,m} H_n H_m' \min_{i=0,1} (2^{n/p_i} 2^{m/q_i'}).$$

We can write this in terms of a, b, and reduce to showing that

$$\sum_{n,m} a_n b_m \min_{i=0,1} (2^{n(\frac{1}{p_i} - \frac{1}{p_\theta})} 2^{m(\frac{1}{q_\theta} - \frac{1}{q_i})}) \lesssim 1.$$

Because $p_0 \neq p_1$ and $q_0 \neq q_1$, and because of the definitions of p_θ, q_θ, we can write the left-hand side as

$$\sum_{n,m} a_n b_m \min(2^{\varepsilon(n+\alpha m)}, 2^{-\varepsilon'(n+\alpha m)})$$

for some non-zero $\alpha \in \mathbf{R}$ and some $\varepsilon, \varepsilon' > 0$ depending only on $p_0, p_1, q_0, q_1, \theta$. We substitute $k := \lfloor n + \alpha m \rfloor$ (thus $n = k - \lfloor \alpha m \rfloor$) and estimate this by

$$\sum_k \min(2^{\varepsilon k}, 2^{-\varepsilon k}) \sum_m a_{k - \lfloor \alpha m \rfloor} b_m.$$

[25] This is not explicitly stated in Theorem 6.6, but can be extracted from the proof. Alternately, one can coarsen the sigma-algebras so that there are only finitely many measurable sets in X and Y, while keeping the simple functions f, g measurable, and then apply Theorem 6.6, at which point the finiteness is obvious from the disjoint supports.

By (26) and Hölder we see that the inner sum is $O(1)$ uniformly in k, and the claim follows.

Finally, if we specialise $r = q_\theta$ and recall that the L^{p_θ, q_θ} norm will be dominated by the L^{p_θ} norm for $p_\theta \geq q_\theta$, the last claim of the theorem follows. ∎

There are many other variations on the real interpolation method, for instance an extension to multilinear operators, or to other function spaces. However, the basic method of proof is still the same: dualise, decompose all inputs, estimate each term as optimally as one can, and then sum.

One can illustrate the real interpolation method graphically using *type diagrams*. One plots all points $(\frac{1}{p}, \frac{1}{q})$ where the operator T is strong-type (p, q) or restricted weak-type (p, q). Ignoring some of the technical hypotheses, the above interpolation theorems then essentially assert that the restricted weak-type diagram and strong-type diagrams are both convex, with the latter contained in the former. Furthermore, if two points lie in the former, then the open interval connecting them lies in the latter. Determining the type diagrams of various operators T is a basic task of harmonic analysis, as it conveys a lot of information as to how T transforms the width and height of functions.

There is a different interpolation method, the *complex interpolation method*, which offers similar results to the real interpolation method but with some slight differences. On the plus side, the complex interpolation method gives sharper bounds, and more importantly can handle the case where the operator T itself varies (analytically) with the interpolation parameter. On the minus side, the method cannot upgrade weak or restricted weak-type estimates to strong-type estimates.

In the next set of notes we shall present several applications of the real interpolation method.

9. Exercises

- Q1. Let $\|\|$ be a quasinorm on functions, thus $\|cf\| = |c|\|f\|$ for scalars c, $\|f\| = 0$ if and only if $f = 0$, and we have the quasitriangle inequality.
$$\|f + g\| \lesssim \|f\| + \|g\| \tag{27}$$
for all functions f, g. Let f_n, $n = 1, 2, \ldots, N$ be a sequence of functions obeying the bounds
$$\|f_n\| \lesssim 2^{-\varepsilon n}$$
for some $\varepsilon > 0$. Prove that
$$\|\sum_{n=1}^N f_n\| \lesssim_\varepsilon 1$$
(the point being that the bound is uniform in N.) Hints: Use (27) to reduce to the case where ε is large. Then use (27) to sum *backwards* from N to 1.

- Q2. Let $\|\|$ be a quasinorm, and let f_n, $n = 1, 2, \ldots, N$ obey the bounds
$$\|f_n\| \lesssim n^{-A}$$
for some $A > 0$. Prove that if A is sufficiently large (depending on the implied constant in (27)), then
$$\|\sum_{n=1}^{N} f\| \lesssim_A 1.$$
Hints: dyadically decompose into blocks $2^k \leq n < 2^{k+1}$. Then sum each block using a binary tree - a classic example of divide and conquer.
- Q3. (Stein-Weiss inequality) Let $f_1, \ldots, f_N$ be functions with $N \geq 2$. Show that
$$\|f_1 + \ldots + f_N\|_{L^{1,\infty}} \lesssim \log N(\|f_1\|_{L^{1,\infty}} + \ldots + \|f_N\|_{L^{1,\infty}}).$$
thus the weak $L^{1,\infty}$ quasinorm only fails to be a norm "by a logarithm". Show with an example that the $\log N$ cannot be removed. Hints: first use homogeneity to reduce to showing that
$$\mu(\{|f_1 + \ldots + f_N| \geq 1\}) \lesssim \log N(\|f_1\|_{L^{1,\infty}} + \ldots + \|f_N\|_{L^{1,\infty}}).$$
Then reduce to the case when the f_n are non-negative, bounded above by 1, and bounded below by $1/2N$ (possibly replacing the 1 on the LHS with $1/2$ to achieve this latter reduction). Then compare the $L^{1,\infty}$ norm with the L^1 norm and use the triangle inequality in L^1. Thus $L^{1,\infty}$ is only "a logarithm away" from being a norm; compare with the polynomial-type failure of the triangle inequality in (4). For the counterexample, try to exploit the logarithmic divergence of $\sum_{n=1}^{N} \frac{1}{n}$ or $\int_1^N \frac{dx}{x}$.
- Q4. For each integer n, let f_n be a quasi-step function of height 2^n and width W_n for some $W_n > 0$. Show that
$$\|\sum_n |f_n|\|_p \sim_p \|2^n W_n^{1/p}\|_{l^p_n(\mathbf{Z})}$$
for all $0 < p < \infty$. If instead f_n is a quasi-step function of height H_n and width 2^n for some $H_n > 0$, show that
$$\|\sum_n |f_n|\|_p \sim_p \|H_n 2^{n/p}\|_{l^p_n(\mathbf{Z})}$$
for all $0 < p < \infty$. What goes wrong when we remove the absolute values on the f_n? (This can be repaired by replacing the powers of 2 with powers of a sufficiently large constant (depending on the implied constant in the definition of a quasi-step function) - why?)
- Q5. Let $0 < p < \infty$, $0 < A < \infty$, and $f \in L^{p,\infty}(X, d\mu)$. Show that the following are equivalent (up to changes in the implied constant):
  - $\|f\|_{L^{p,\infty}(X,d\mu)} \lesssim_p A$.
  - For every set E of finite measure, there exists a subset E' of E with $\mu(E') \geq \frac{1}{2}\mu(E)$ such that
  $$\int_X f 1_{E'} \, d\mu = O(\mu(E)^{1/p'})$$
  (in particular, we assert that the integral on the left-hand side is absolutely integrable).

Hint: It may be instructive to work out the example $f(x) = |x|^{-d/p}$ on $\mathbf{R}^d$ by hand to get a sense of what is going on; this should suggest how to prove things in general. The proof is slightly simpler in the case when f is non-negative, so you may want to try that case first. Comment on how this result implies Problem 6.10 (or its equivalent version discussed shortly afterwards) when $p > 1$.

- Q6. Let f be a function on a measure space X of bounded measure $\mu(X) = 0(1)$. Show that the following are equivalent (up to changes in the implied constants):
  (i) $\|f\|_{e^L} = O(1)$.
  (ii) $\|f\|_{L^{p,\infty}} = O(p)$ for all $1 \leq p < \infty$.
  (iii) $\|f\|_{L^p} = O(p)$ for all $1 \leq p < \infty$.
  Hint: You may find the Taylor expansion for e^x, together with the obvious bounds $(k/2)^{k/2-1} \leq k! \leq k^k$ for integer k (or Stirling's formula, if you know what that is), to be useful.
- Q7. Obtain the analogue of Q6 for the Orlicz space e^{L^2}.
- Q8 (Loomis-Whitney inequality). Let $d \geq 2$, let $X_1, \ldots, X_d$ be measure spaces, and for $i = 1, \ldots, d$ let $f_i \in L^p(\prod_{1 \leq j \leq d: j \neq i} X_j)$ for some $0 < p \leq \infty$. Show that the function
$$F(x_1, \ldots, x_d) := \prod_{i=1}^d f_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$$
lies in $L^{p/(d-1)}(\prod_{1 \leq j \leq d} X_j)$ with the *Loomis-Whitney inequality*
$$\|F\|_{L^{p/(d-1)}(\prod_{1 \leq j \leq d} X_j)} \leq \prod_{i=1}^d \|f_i\|_{L^p(\prod_{1 \leq j \leq d: j \neq i} X_j)}.$$
Conclude in particular the *box inequality*
$$\mu_{\prod_{1 \leq j \leq d} X_j}(E) \leq (\prod_{i=1}^d \mu_{\prod_{1 \leq j \leq d: j \neq i}}(\pi_i(E)))^{1/(d-1)}$$
where E is any subset of $\prod_{1 \leq j \leq d} X_j$ and π_i is the canonical projection from $\prod_{1 \leq j \leq d} X_j$ to $\prod_{1 \leq j \leq d: j \neq i} X_j$. From this, deduce the *weak isoperimetric inequality*
$$|E| \lesssim_d |\partial E|^{d/(d-1)}$$
for any $E \subset \mathbf{R}^d$, where $|E|$ is the Lebesgue measure of E and $|\partial E|$ is the $d-1$-dimensional Hausdorff measure of the boundary of E.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: tao@math.ucla.edu