Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

4-16-2015 12:00 AM

Inclusions Among Mixed-Norm Lebesgue Spaces

Wayne R. Grey
The University of Western Ontario

Supervisor
Gord Sinnamon
The University of Western Ontario

Graduate Program in Mathematics
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

INCLUSIONS AMONG MIXED-NORM LEBESGUE SPACES
(Thesis format: Monograph)

by

Wayne Grey

Graduate Program in Mathematics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Acknowledgements

Without my advisor, Gord Sinnamon, this thesis would not have been completed. He never stopped believing that I could do worthwhile mathematics, and has finally convinced me. He led the way in addressing the problem with two atomic spaces, worked hard to refine the presentation for the paper version of our results, and has always been available for consultation during the preparation of the thesis.

The examiners were quite helpful during revision, both in pointing out missed typos and in suggesting clarifications and improvements. In alphabetical order, thanks to Tatyana Barron for spotting various errors and misstatements, as well as recommending clearer and more consistent expectations in Chapter 2. Thanks to Martin Pinsonnault for substantial advice in improving the typography, recommendations to improve readability, and suggesting a one-variable summary. Thanks to Hristo Sendov for many spotted typos, as well as suggested clarifications. Finally, thanks to Javier Soria for locating references on the one-variable problem, pointing out the importance of certain results, and suggesting possible future directions.

Thanks as well to my many instructors in mathematics; often it's taken a while for me to fully appreciate what they've taught, but in the end the lessons have been invaluable.

Abstract

A mixed L^p norm of a function on a product space is the result of successive classical L^p norms in each variable, potentially with a different exponent for each. Conditions to determine when one mixed norm space is contained in another are produced, generalizing the known conditions for embeddings of L^p spaces.

The two-variable problem (with four L^p exponents, two for each mixed norm) is studied extensively. The problem's "unpermuted" case simply reduces to a question of L^p embeddings. The other, "permuted" case further divides, depending on the values of the L^p exponents. Often, they fit the "Minkowski case", when Minkowski's integral inequality provides an easy, complete solution. In the "non-Minkowski case", the solution is determined by the structure of the measures in the component L^p spaces. When no measure is purely atomic, there can be no mixed-norm embedding in the non-Minkowski case, so for such measures the problem is solved.

With at least one purely atomic measure, the non-Minkowski case divides further based on the structure of the measures and the values of the exponents. Various necessary conditions and sufficient conditions are found, solving a number of subcases. Other subcases are shown to be genuinely complicated, with their solutions expressed in terms of an optimization problem known to be computationally difficult.

With some difficult cases already present in the two-variable problem, it is impractical to cover every case of the multivariable problem, but results are presented which fully solve some cases. When no measure is purely atomic, the multivariable problem is solved by a reduction to the Minkowski case of certain two-variable subproblems. The multivariable problem with unweighted ℓ^p spaces has a similar reduction to easy two-variable subproblems. It is conjectured that this applies more generally; that, regardless of the structures of the involved measures, when every permuted two-variable subproblem fits the Minkowski case, the full multivariable mixed norm inclusion must hold.

Keywords: L^p mixed norms, permuted mixed norms, Minkowski's integral inequality, embedding, atomic measure, atomless measure, partition problem

Contents

Chapter 1

Introduction

Every analyst knows at least one theorem regarding mixed norms. Minkowski's integral inequality proves an embedding between mixed-norm spaces. Even Tonelli's theorem tells us that, when every exponent in a mixed norm is the same p, permuting their order is irrelevant, always yielding L^p on the product space. (In case the name sounds unfamiliar, Tonelli's theorem is a version of Fubini's theorem for nonnegative functions, which is valid with extended real values even without requiring integrability.)

Given an n-tuple $P = (p_1, \dots p_n)$, each $p_k \in (0, \infty]$, a mixed L^P norm is a norm applied to functions $f(x_1, \dots, x_n)$ on a product measure space $(X_1, \mu_1) \times \cdots \times (X_n, \mu_n)$, computed by

$$\|f\|_P = \left(\int_{X_n} \left(\cdots \left(\int_{X_1} |f(x_1, \dots, x_n)|^{p_1} \, d\mu_1(x_1) \right)^{p_2/p_1} \cdots \right)^{p_n/p_{n-1}} d\mu_n(x_n) \right)^{1/p_n}, \tag{1.1}$$

successively applying an $L^{p_k}_{\mu_k}(X_k)$ norm in each variable x_k. (Naturally, when some $p_k = \infty$, an essential supremum is used instead. Also, for $p_k < 1$ only quasi-norms are obtained.)

The problem addressed here is the question of when one mixed-norm L^P space is contained in another. As noted in [4], such spaces are Banach function spaces, so whenever this is true there is a continuous inclusion map. The problem then amounts to one of determining whether there is a finite constant, and what the best constant (i.e. least value) is, in the following inequality.

Let $X_1, \dots, X_n$ be measurable spaces, each X_k admitting σ-finite measures μ_k and ν_k. Suppose that $P = (p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_n)$, two n-tuples of exponents drawn from $(0, \infty]$. Take any permutation σ on $\{1, \dots, n\}$. Then the question is one of finding the least constant $C \in [0, \infty]$ (or, at least, conditions when $C < \infty$) such that, for any measurable, complex-valued function f on $X_1 \times \cdots \times X_n$,

$$\|f\|_Q \le C \, \|f\|_{\sigma(P)}, \tag{1.2}$$

where

$$\|f\|_Q = \left(\int_{X_n} \cdots \left(\int_{X_1} |f(x_1, \dots, x_n)|^{q_1} \, d\nu_1(x_1) \right)^{q_2/q_1} \cdots d\nu_n(x_n) \right)^{1/q_n} \quad \text{and}$$

$$\|f\|_{\sigma(P)} = \left(\int_{X_{\sigma(n)}} \cdots \left(\int_{X_{\sigma(1)}} |f(x_1, \dots, x_n)|^{p_{\sigma(1)}} \, d\mu_{\sigma(1)}(x_{\sigma(1)}) \right)^{p_{\sigma(2)}/p_{\sigma(1)}} \cdots d\mu_{\sigma(n)}(x_{\sigma(n)}) \right)^{1/p_{\sigma(n)}}.$$

Although one could permute the mixed norms on both sides of (1.2), only one permutation σ is used, yielding $\|f\|_{\sigma(P)}$. This is because the variables can be relabled, so that without loss of generality the left-hand side is simply $\|f\|_Q$.

Computations involving mixed norms have appeared in the literature at least since Littlewood's famous 4/3 inequality [22] in 1930, generalized by Grothendieck's inequality in 1956. This inequality has proven fundamental in bilinear analysis, and been extended and interpreted in many ways since. (Major results in bilinear and multilinear analysis and their history are discussed by Blei in [6], where Theorem 10 shows a mixed-norm inequality central to the proof of Littlewood's inequality.) However, only in 1961 did Benedek and Panzone define the "L^P spaces with mixed norm", and prove many of their properties, in their paper [4].

Since then, L^P mixed norms have enjoyed a variety of applications in both pure and applied mathematics. Applications in other disciplines seem to be growing in recent years, but the following mention only a handful, not necessarily the most important. In 2009, Kowalski [21] applied mixed norms, as well as the Besov and Triebel-Lizorkin spaces which can be characterized using mixed norms, to sparse methods in signal regression. Various papers since have cited this one, including [30] on dictionary learning and [32], which applies a mixed ℓ_2/ℓ_1 regularization to two learning tasks. In statistics, Zhao, Rocha, and Yu introduced "composite absolute penalties", a variation on mixed norms, in their much-cited [33].

In pure mathematics, the study of Sobolev spaces has come to use mixed norms frequently. In 1987, Fournier [16] found that properties of mixed norms could be used to prove both the approach of Gagliardo [17] and Nirenberg [26] to the Sobolev embedding theorem and Littlewood's inequality [22]. He developed, as an intermediate result, a mixed-norm generalization of Minkowski's integral inequality which is itself a special case of our Theorem 7.2.4. Inspired by this paper and Fournier's subsequent collaboration with Blei in [8] to investigate embeddings of Lorentz spaces into mixed-norm L^p spaces, Milman [25] combined mixed norms with interpolation theory to provide alternate proofs of the major results both in [16] and in Fournier's subsequent collaboration with Blei, [8], which investigated embeddings of Lorentz spaces into mixed-norm L^p spaces.

Those spaces that arise in these investigations of embeddings of the Sobolev spaces $W_1^1(\mathbb{R})$ have attracted recent research. Algervik and Kolyada's 2011 paper [1] proves embeddings of these mixed-norm spaces, dubbed "Fournier-Gagliardo mixed norm spaces" after [16] and [17], into Lorentz-type spaces defined using iterative rearrangements, i.e. rearrangements over one variable at a time. Kolyada's 2012 [20] develops further properties of these Fournier-Gagliardo mixed norm spaces and Lorentz-type spaces with iterative rearrangements, leading to a sharp constant in a Sobolev embedding.

The Bohnenblust-Hille inequality, itself an extension of Littlewood's 4/3 inequality, has recently seen further extension, using the theory of multilinearity and p-summing operators. In 2010, Defant, Popa and Schwarting devised coordinatewise multiple summing operators and applied them to develop vector-valued extensions of Bohnenblust-Hille in [14]. The next year, Defant joined with four other authors in [13] to make great strides toward better estimates of the coefficient in the Bohnenblust-Hille inequality, as well as various applications. Although neither of these papers explicitly discusses mixed norms, some results use expressions involving mixed norms, and certain proofs can be simplified using the mixed-norm generalizations of the Hölder and Minkowski inequalities in this document's Section 7.2. Appendix A provides example applications of these two inequalities, including a quick and easy proof of Lemma 1 in [13], a special case of a result by Blei in [7]. More recently, Popa and Sinnamon [28] established two inequalities which generalize inequalities in [6] and [14]. Some fairly lengthy arguments there can be boiled down to quick applications of the mixed-norm forms of Hölder and Minkowski.

The idea of mixed-norm spaces has been generalized beyond L^p; in [9], Blozinski introduced mixed norm spaces constructed from rearrangement-invariant Banach function spaces. Other than L^p, the specific spaces mentioned for this mixed-norm construction include Orlicz spaces, Lorentz-Zygmund spaces, and Lorentz $\Lambda_\alpha(X)$ and $L(p,q)$ spaces. Several applications are developed, as well as inclusion results involving tensor products, kernels of integral operators, and Lorentz mixed-norm spaces.

Boccuto, Bukhvalov, and Sambucini developed impressive results on general mixed norms in [10]. They primarily sought connections between mixed norms and "similar" norms on product spaces, analogous to the way that an L^p mixed norm with p the same for every variable is simply L^p on the product. Weaker properties are established for Orlicz, Lorentz, and Marcinkiewicz mixed norms, as well as counterexamples refuting certain properties in many Lorentz space cases. An appendix provides a striking generalization of the Kolmogorov-Nagumo theorem. Roughly speaking, the Kolmogorov-Nagumo theorem provides that the embedding given by Minkowski's integral inequality (our Theorem 3.5.1) cannot be reversed unless the exponents are equal, when Minkowski's inequality becomes the equality of Tonelli's theorem. The generalization proves that, outside of special circumstances, this equivalence of the two permutations of a mixed norm ($F[E]$ and $E[F]$ in their notation) is impossible unless both component spaces E and F are L^p with the same p.

To define mixed-norm Lorentz spaces, Barza, Persson, and Soria defined a notion of multidimensional rearrangement in [2]. This turned out to agree with Blozinski's mixed norm construction from [9], so Barza et al., joined by Kamińska, used Blozinski's formulation in their [3], which applied mixed Lorentz norms to find necessary and sufficient conditions for normability of two-dimensional Lorentz spaces, the one-dimensional problem having been solved earlier. Furthermore, they proved embeddings among classical, multidimensional, and mixed norm Lorentz spaces.

Finally, in 2014 Clavero and Soria [12] revisited the work done by Algervik and Kolyada [1] and Fournier [16], generalizing their embeddings. Working with a particular generalization of mixed-norm spaces using rearrangement-invariant spaces, these embeddings are generalized to be between these mixed norm spaces and rerrangement-invariant spaces; optimal domains and ranges are found. Though the definitions may be new, the paper notes that estimates on these quantities have been found as early as the work of Gagliardo [17] and Nirenberg [26].

Yet, even though so much work has been done with mixed norms, over half a century after the term was coined, the embedding problem between L^p mixed norm spaces had not been resolved. Perhaps this is because some simple cases seem unremarkable. Many applications may not need to change the order of the variables, as in amalgam spaces, where the meaning of such a permutation is unclear. The "unpermuted" case where σ is the identity simply reduces to one-variable problems, as treated in Sections 3.4 and 7.5. Even when permutation of mixed norms is considered, when the left and right sides in (1.2) differ only by permutation (i.e. each $p_k = q_k$ and $\mu_k = \nu_k$ for $k = 1, 2$ and σ is not the identity), the Minkowski inequality gives a sufficient condition, and the Kolmogorov-Nagumo theorem shows that this is also, except for fairly trivial measure spaces, necessary. Even the generalization of Minkowski's integral inequality to the "Minkowski condition" for inclusion in Section 3.6 is straightforward, and gives an easily computed best constant. The complexity of certain non-Minkowski cases, and the computationally difficult solutions that arise, are not necessarily expected.

With all the interesting generalizations that have arisen since Benedek and Panzone introduced mixed-norm L^p spaces, one might ask why only L^p mixed norms are treated here. First, the unpermuted case reduces to single-variable problems even for general mixed norms, as noted in Theorem 2.2 of [18]. Therefore, only the permuted case of the problem really teaches anything specifically about mixed norms. In order to develop sufficient conditions for the permuted case, some method to deal with the permutation is required. Fundamentally, the methods here are based on the Minkowski integral inequality and Tonelli's theorem (which can be viewed as an equality case of Minkowski). The generalized Kolmogorov-Nagumo theorem in [10] shows that, for other spaces, an analogue to Tonelli's theorem cannot be expected.

In addressing the problem, first comes a summary of the known one-variable solution in Chapter 2. This is not only necessary for completeness, but important to general solutions, because many cases reduce to one-variable inclusions and the computation of mixed-norm best constants always involves one-variable best constants. Next simplest is the unpermuted case, which in two variables looks like

$$\left(\int_{X_2}\left(\int_{X_1}|f(x_1,x_2)|^{q_1}d\nu_1(x_1)\right)^{\frac{q_2}{q_1}}d\nu_2(x_2)\right)^{\frac{1}{q_2}} \le C\left(\int_{X_2}\left(\int_{X_1}|f(x_1,x_2)|^{p_1}d\mu_1(x_1)\right)^{\frac{p_2}{p_1}}d\mu_2(x_2)\right)^{\frac{1}{p_2}}.$$

Treated for two variables in Section 3.4 and more generally in Section 7.5, this case always reduces to one-variable subproblems. Simply applying the one-variable inclusions successively, when they hold, solves the problem with $C = C_1\cdots C_n$, where each C_k is the best constant over X_k. Even in the permuted case, one-variable inclusions remain necessary for mixed-norm inclusion, as is proven by considering "factorable" functions $f(x_1,x_2) = f_1(x_1)f_2(x_2)$.

The two-variable permuted case looks like

$$\left(\int_{X_2}\left(\int_{X_1}|f(x_1,x_2)|^{q_1}d\nu_1(x_1)\right)^{\frac{q_2}{q_1}}d\nu_2(x_2)\right)^{\frac{1}{q_2}} \le C\left(\int_{X_1}\left(\int_{X_2}|f(x_1,x_2)|^{p_2}d\mu_2(x_2)\right)^{\frac{p_1}{p_2}}d\mu_1(x_1)\right)^{\frac{1}{p_1}}.$$

The so-called "Minkowski case" has $\min(p_1,q_1) \le \max(p_2,q_2)$, i.e. one of the "one" exponents (p_1 and q_1) is no greater than one of the "two" exponents (p_2 and q_2). In this case, solved in Section 3.6, using both single-variable inclusions and applying Minkowski's integral inequality either at the start of this process, in between the inclusions, or at the end, solves the problem. Again, $C = C_1C_2$, the product of single-variable best constants, for Minkowski's integral inequality has the constant 1.

In the non-Minkowski case, when $\max(p_2,q_2) < \min(p_1,q_1)$, such a simple solution is not possible. This more complex situation is first treated in the special case of common measures, i.e. when for each $k \in \{1,2\}$ there is a single measure $\lambda_k = \mu_k = \nu_k$, in section 4.1. In this case, both a proof that the Minkowski condition is necessary for measures which are not purely atomic and a solution for the case where just one of X_1 and X_2 has purely atomic measures are developed, in Sections 4.1.2 and 4.1.3 respectively.

Section 4.2 lifts the restriction of common measures to treat the non-Minkowski case more generally. However, it opens with various results in Section 4.2.1 which can reduce many cases to the common measure situation, demonstrating that this special case is not as limited as it initially seemed. It is also worth noting that the reductions to common measures developed in Propositions 4.2.2 and 4.2.3 produce common measures λ_1 and λ_2 which are connected to the solution of the one-variable problem, where in many cases the solution depends on the properties of a measure λ produced in the same way.

The two-variable problem finishes with its most difficult case, the non-Minkowski case where both X_1 and X_2 have purely atomic measures, treated in Chapter 5. Some basic necessary and sufficient conditions still apply, but to cover nearly all the cases which remain, a variational argument is developed in Section 5.2 to show that we can, in many cases, approach the best constants with simpler functions. In particular, in the non-Minkowski case $\max(p_2, q_2) < \min(p_1, q_1)$, if $p_1 \leq q_1$ or $p_2 \leq q_2$, we need only consider functions analogous to matrices with at most one non-zero entry per column, at most one non-zero entry per row, or both. Section 5.3 uses rearrangements to compute the best constant in the case $p_2 \leq q_2 < p_1 \leq q_1$, when we can use "diagonal" functions with at most one non-zero entry per row and per column. Section 5.5 treats $p_2 \leq q_2 < q_1 < p_1$, using functions with one entry per row, and $q_2 < p_2 < p_1 \leq q_1$, with at most one entry per column. The best constants are given in terms of a genuinely computationally difficult problem described in Section 5.4, of which a special case amounts to the NP-hard optimization version of the partition problem. The very last case, where $q_2 < p_2 < q_1 < p_1$ and both X_1 and X_2 have purely atomic measures, is left unsolved beyond previously established conditions, but is expected to be the most complicated case.

Since certain cases of the two-variable problem are so difficult, a full multivariable solution is not provided, but results covering a fair number of cases are given. First, mixed-norm Hölder and Minkowski inequalities are given in Section 7.2, both for later use and because they're useful inequalities for mixed norms. Section 7.4 shows that, analogously to the previously established necessity of single-variable inclusions, all "subinclusions" are necessary. Such subinclusions are mixed-norm inclusions among a k-variable subset of the full n (or simple L^p inclusions if $k = 1$), ordered according to the relative positions, depending on σ, in which they appear in $\|f\|_Q \leq C\|f\|_{\sigma(P)}$. The proof uses partially factorable functions, where $f(x_1, \ldots, x_n)$ is a product of one function of the selected k variables for a subinclusion and one-variable functions in the other variables.

Section 7.5 proves the unsurprising result that the unpermuted case reduces to single-variable problems however many variables are used in the mixed norms. Fortunately, Section 7.6 establishes a more substantial sufficient condition which brings back the Minkowski case from the two-variable problem. When none of the measure spaces involved is purely atomic, there is a complete solution: permuted mixed-norm inclusion holds if and only if every 2-variable subinclusion is of either unpermuted or fits the Minkowski case. A different argument establishes the same result for unweighted ℓ^p spaces; the connection between these is that one-variable inclusion always has $q \leq p$ for measures which are not purely atomic, while it always has $p \leq q$ for unweighted ℓ^p. The consistent order between each p_k and q_k makes it relatively straightforward to use Minkowski's integral inequality, plus one-variable inclusions, to prove mixed-norm inclusion.

It is conjectured that these are not the only cases in which the multivariable problem can be reduced to two-variable subproblems. Specifically, that to prove the full multivariable mixed norm inclusion, it is sufficient to establish that each permuted two-variable subproblem is in the Minkowski case.

Chapter 2

One-variable review

When is one L^p space contained in another? The known answer is both the simplest possible case of the general problem and an important piece of its solution. Several articles in the American Mathematical Monthly address this problem, which constitutes the one-variable case of the mixed-norm inclusion problem. Miamee [24] addresses the general problem and characterizes it as one of checking the existence of a finite best constant, but does not cover all cases and presents results which do not explicitly specify the best constants. The earlier notes by Subramanian [29] and Villani [31] provide some results, but only treat the case of one measure rather than two. This chapter develops a solution with best constants, where important cases and issues relevant to the more general problem are discussed.

2.1 Problem Statement

Consider a measurable space (X, Σ), i.e. a set X on which Σ is a σ-algebra. Let μ and ν be non-zero, σ-finite measures on Σ and $p, q \in (0, \infty]$. The question is when $L^p_\mu(X) \subset L^q_\nu(X)$, understood in the sense that any function with finite $L^p_\mu(X)$ norm must also have finite $L^q_\nu(X)$ norm. It turns out that the inclusion holds if and only if there is a constant $C < \infty$ such that, for any measurable, complex-valued function f on X,

$$||f||_{L^q_\nu(X)} \leq C||f||_{L^p_\mu(X)}. \tag{2.1}$$

It is not a priori obvious that the inclusion $L^p_\mu(X) \subset L^q_\nu(X)$ is equivalent to the existence of a finite constant C, though it is not difficult to see that $C < \infty$ is sufficient. Proof of this equivalence is provided in Section 2.2.

Definition 2.1.1. Recall that the L^p spaces (or Lebesgue spaces) are a class of function spaces defined by the L^p norms; for any measure space (X, μ) and $0 < p \leq \infty$, define for any measurable function $f : X \to \mathbb{C}$ the quantity

$$||f||_{L^p_\mu(X)} = \begin{cases} \left(\int_X |f(x)|^p \, d\mu(x)\right)^{1/p} & \text{if } p < \infty \\ \operatorname{ess\,sup}_{x \in (X,\mu)} |f(x)| & \text{if } p = \infty \end{cases},$$

which takes values in $[0, \infty]$.

The essential supremum, denoted ess sup, is a variation on the supremum which disregards values on sets of measure zero. Here is a precise definition, along with some related definitions from basic measure theory.

Definition 2.1.2. Given a measure space (X, Σ, μ), a measurable set $E \in \Sigma$ is said to be *μ-null* if and only if $\mu E = 0$. When context makes the measure μ clear, a μ-null set E is called *null*.

Definition 2.1.3. Let (X, μ) be a measure space, and $P(x)$ a property which is either true or false of each point $x \in X$. The property $P(x)$ is said to hold *μ-almost everywhere* (abbreviated *μ-a.e.*) if and only if it fails on a μ-null set, i.e.

$$\mu(\{x \in X : \neg P(x)\}) = 0.$$

When the measure is clear from context, we may simply say that a property holds *almost everywhere*, or *a.e.*.

Definition 2.1.4. Let (X, μ) be a measure space and $f : X \to \mathbb{R}$ be a measurable function. Its *essential supremum* is

$$\operatorname*{ess\,sup}_{x \in (X,\mu)} f(x) = \inf\left\{\alpha \in [-\infty, \infty] : \mu\left(f^{-1}(\alpha, \infty)\right) = 0\right\},$$

in other words, the least α such that $f \leq \alpha$ μ-a.e.

Definition 2.1.5. The vector space $L^p_\mu(X)$ defined by the $\|\cdot\|_{L^p_\mu(X)}$ norm consists of all equivalence classes, identifying functions which agree μ-almost everywhere, of measurable functions $f : X \to \mathbb{C}$ such that $\|f\|_{L^p_\mu(X)} < \infty$. Whenever $p \geq 1$, $L^p_\mu(X)$ is a Banach space, i.e. a complete normed vector space. When $0 < p < 1$, $L^p_\mu(X)$ is still a vector space, but is not normed since the triangle inequality fails. When the measure μ or the underlying set X is clear from context, the simpler notations $L^p(X)$ or L^p may be used, with the corresponding norm (or quasi-norm for $p < 1$) denoted $\|\cdot\|_{L^p(X)}$ or $\|\cdot\|_p$.

The standard notation ℓ^p is used instead of L^p when weighted or unweighted counting measure is involved, so that integrals are sums. Given an at most countable index set I, a weight $w : I \to (0, \infty)$, a measurable, complex-valued function f on I, and $p \in (0, \infty]$,

$$\|f\|_{\ell^p_w(I)} = \begin{cases} \left(\sum_{i \in I} |f(i)|^p\, w(i)\right)^{1/p} & \text{if } p < \infty \\ \sup_{i \in I} |f(i)| & \text{if } p = \infty \end{cases},$$

where every term is nonnegative, so the sum converges absolutely, if at all. (Therefore the order of elements in I need not be specified.) The corresponding space $\ell^p_w(I)$ consists of those functions $f(i)$ (equivalently, sequences $(a_i)_{i \in I}$, since I is a countable index set) such that $\|f\|_{\ell^p_w(I)}$. When the space is unweighted (i.e. w is a constant 1), the notation $\ell^p(I)$ is also used for the space and norm. By itself, ℓ^p is understood to use unweighted counting measure on $\mathbb{N}$. For any of these, the short form $\|\cdot\|_p$ may be used for the norm.

Since the values of $\|f\|_{L^p_\mu(X)}$ and $\|f\|_{L^q_\nu(X)}$ depend only on the modulus $|f(x)|$ and the following results will not involve the addition of functions, it suffices to consider only nonnegative functions. The notation $L^+(X)$ denotes the space of all measurable (with respect to Σ) functions on X which take values in $[0, \infty]$.

Definition 2.1.6. Given measures μ and ν on a space X, the measure ν is *absolutely continuous* with respect to μ if and only if every μ-null set is ν-null. This is denoted by $\nu \ll \mu$.

Note that, if $\nu \not\ll \mu$, there is some measurable set $E \subset X$ such that $\mu E = 0$ and $\nu E > 0$. Since $\|f\|_{L^p_\mu(X)} = 0$ while $\|f\|_{L^q_\nu(X)} > 0$, there is no constant $C < \infty$ such that

$$\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)}$$

for all $f \in L^+(X)$. Furthermore, consider $\infty\chi_E$, for which $\|\infty\chi_E\|_{L^p_\mu(X)} = 0$ and $\|\infty\chi_E\|_{L^q_\nu(X)} = \chi\nu(E)^{1/q} = \infty$. Therefore $L^p_\mu(X) \not\subset L^q_\nu(X)$.

(The example $\infty\chi_E$ may seem contrived, but when $\mu E = 0$, it is μ-a.e. equal to the constant zero function. A meaning of inclusion $L^p_\mu(X) \subset L^q_\nu(X)$ which does not respect a.e. equality would be rather strange.)

Assumption For the rest of Chapter 2, assume that $\nu \ll \mu$. (Otherwise, inclusion always fails.)

When $\nu \ll \mu$, the Radon-Nikodym Theorem guarantees that there is a μ-measurable function, denoted by $\frac{d\nu}{d\mu}$, such that $\nu(E) = \int_E \frac{d\nu}{d\mu} d\mu$ for any measurable set E. The more general Lebesgue-Radon-Nikodym Theorem is described for signed measures in Section 3.2 of Folland's [15], where it is Theorem 3.8. Because only unsigned measures are used here, only the following corollary is needed.

Corollary 2.1.7 (Radon-Nikodym Theorem for Unsigned Measures)**.** *Let μ and ν be σ-finite (positive) measures on a common measurable space (X, Σ) such that $\nu \ll \mu$. Then there is a μ-measurable function $f : X \to [0, \infty)$ such that, for each $E \in \Sigma$,*

$$\nu(E) = \int_E f(x)d\mu(x).$$

Furthermore, any two such functions f must agree μ-a.e.

Because the Radon-Nikodym derivative $\frac{d\nu}{d\mu}$ is only defined by μ and ν uniquely up to modification on μ-null sets, it is most appropriate to think of $\frac{d\nu}{d\mu}$ as representing an equivalence class of functions, rather than a function. However, this is the same identification made throughout and whenever working with L^p norms. The following theorem provides the solution to the one-variable inclusion problem, in terms of the Radon-Nikodym derivative, to be proven in Section 2.4.

Theorem 2.1.8. *The least constant $C \in [0, \infty]$ such that $\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)}$ for any measurable function f on X is as specified below, separated by case.*

$$\begin{aligned}
q = p = \infty :& \qquad C = 1. \\
0 < q < p = \infty :& \qquad C = \nu(X)^{1/q}. \\
0 < q \leq p < \infty :& \qquad C = \left(\int_X \left(\frac{d\nu}{d\mu}(x)\right)^{\frac{p}{p-q}} d\mu(x)\right)^{\frac{p-q}{pq}}. \\
0 < p < q \leq \infty :& \qquad C = \varepsilon^{1/q-1/p},
\end{aligned}$$

where $\varepsilon \geq 0$ denotes the infimum of all strictly positive values of $\int_E \left(\frac{d\nu}{d\mu}\right)^{p/(p-q)} d\mu$, for $E \in \Sigma$.

Observe that the second case is the limit of the third case as $p \to \infty$, so they could be combined into one case with $C = \left\| \frac{d\nu}{d\mu} \right\|_{p/(p-q)}^{1/q}$. This means that when $0 < q \le p \le \infty$ (and not both ∞), assuming $\nu \ll \mu$, the necessary and sufficient condition for $L^p_\mu(X) \subset L^q_\nu(X)$ is that $\frac{d\nu}{d\mu}$ be in $L^{p/(p-q)}_\mu(X)$. In the final case $p < q$, again assuming $\nu \ll \mu$, the necessary and sufficient condition for inclusion is that the positive values of $\int_E \left(\frac{d\nu}{d\mu}\right)^{p/(p-q)} d\mu$ be bounded away from zero. Otherwise, $\varepsilon = 0$ and inclusion fails with $C = 0^{1/q-1/p} = \infty$, since $1/q - 1/p < 0$ in this case. Corollary 2.5.7 gives a condition on measures that implies $\varepsilon = 0$, showing that this holds for many measures.

A slightly more condensed version of this theorem is given as part of the summary of the one-variable solution in Section 2.7.

2.2 Preliminaries

Intuitively, we can think of $L^p_\mu(X) \subset L^q_\nu(X)$ as saying that any function $f \in L^+(X)$ must be in $L^q_\nu(X)$ if it is in $L^p_\mu(X)$, i.e. that $\|f\|_{L^p_\mu(X)} < \infty$ implies $\|f\|_{L^q_\nu(X)} < \infty$. Of course, technically, the elements of $L^p_\mu(X)$ are not functions, but equivalence classes of functions, identified by agreement μ-almost everywhere. Similarly, the elements of $L^q_\nu(X)$ are equivalence classes for agreement ν-almost everywhere. The following describes the connection between this functionwise notion of inclusion and what it means in terms of equivalence classes.

Definition 2.2.1. The inclusion $L^p_\mu(X) \subset L^q_\nu(X)$ is understood here in a functionwise sense; that is, for each $f \in L^+(X)$, if $\|f\|_{L^p_\mu(X)} < \infty$ then $\|f\|_{L^q_\nu(X)}$.

Definition 2.2.2. Given a measure space (X, μ), let $[f]_\mu$ denote the equivalence class of f for agreement μ-a.e., i.e.

$$[f]_\mu = \{g \in L^+(X) : f = g \ \mu\text{-a.e.}\}.$$

Proposition 2.2.3. *The formula* $\iota\left([f]_\mu\right) = [f]_\nu$ *specifies a well-defined map*

$$\iota : L^p_\mu(X) \to L^q_\nu(X)$$

if and only if $\|f\|_{L^p_\mu(X)} < \infty$ *implies* $\|f\|_{L^q_\nu(X)} < \infty$ *for any* $f \in L^+(X)$.

Proof. Assume that $\|f\|_{L^p_\mu(X)} < \infty$ implies $\|f\|_{L^q_\nu(X)}$ for any $f \in L^+(X)$. To see that ι is well-defined, consider any $f_1, f_2 \in L^+(X)$ which are in the same class, in the sense that $[f_1]_\mu = [f_2]_\mu$. This means that $f_1 = f_2$ μ-a.e. Recall that we assume $\nu \ll \mu$, so because the set where $f_1 \neq f_2$ is μ-null, it is also ν-null; that is, $f_1 = f_2$ ν-a.e. Therefore $[f_1]_\nu = [f_2]_\nu$.

The space $L^q_\nu(X)$ is valid as the codomain because, as long as $[f]_\mu \in L^p_\mu(X)$, the function $f \in L^q_\nu(X)$. For any $g \in [f]_\nu$, i.e. such that $f = g$ ν-a.e., $\|g\|_{L^q_\nu(X)} = \|f\|_{L^q_\nu(X)} < \infty$ and so the class $\iota\left([f]_\mu\right) = [f]_\nu$ is in the space $L^q_\nu(X)$.

Conversely, suppose that $\iota : L^p_\mu(X) \to L^q_\nu(X)$ as a transformation of equivalence classes. Suppose that $f \in L^+(X)$ is such that $\|f\|_{L^p_\mu(X)} < \infty$. Then $[f]_\mu \in L^p_\mu(X)$, so $\iota([f]_\mu) = [f]_\nu \in L^q_\nu(X)$ and $\|f\|_{L^q_\nu(X)} < \infty$. □

So, the functionwise inclusion $L^p_\mu(X) \subset L^q_\nu(X)$ is equivalent to a natural embedding of function spaces which consist of equivalence classes. For simplicity, we will from now on work functionwise, making such statements as $f \in L^p_\mu(X)$ rather than $[f]_\mu \in L^p_\mu(X)$.

Next, we address why the inclusion problem $L^p_\mu(X) \subset L^q_\nu(X)$ can be formulated in terms of the existence of a finite best constant C in $\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)}$. This is based on a property of inclusions among Banach function spaces, which is quite helpful when $p, q \geq 1$. If either exponent is less than 1, there is still a method to convert the problem to one where the exponents are at least 1, as the following result shows.

Lemma 2.2.4. *Fix an arbitrary real number $t > 0$. Let $C \in [0, \infty]$ denote the least constant such that*

$$\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)} \tag{2.2}$$

for all $f \in L^+(X)$, and let $D \in [0, \infty]$ denote the least constant such that

$$\|h\|_{L^{tq}_\nu(X)} \leq D\,\|h\|_{L^{tp}_\mu(X)} \tag{2.3}$$

for each $h \in L^+(X)$. Then $C = D^t$.

Furthermore, $L^p_\mu(X) \subset L^q_\nu(X)$ if and only if $L^{tp}_\mu(X) \subset L^{tq}_\nu(X)$.

Proof. Given any $h \in L^+(X)$, let $f = h^t$. Then

$$\|h\|_{L^{tq}_\nu(X)} = \|f\|^{1/t}_{L^q_\nu(X)} \qquad \text{and} \qquad \|f\|_{L^{tp}_\mu(X)} = \|h\|^t_{L^p_\mu(X)}\,.$$

Therefore

$$\|h\|_{L^{tq}_\nu(X)} = \|f\|^{1/t}_{L^q_\nu(X)} \leq (C\,\|f\|_{L^p_\mu(X)})^{1/t} = C^{1/t}\,\|h\|_{L^p_\mu(X)}\,.$$

Because D is the least constant for (2.3), $D \leq C^{1/t}$, so $D^t \leq C$.

To obtain the reverse inequality, for any $f \in L^+(X)$, let $h = f^{1/t}$. Of course, this implies that $f = h^t$, as above, so the same equalities apply, and

$$\|f\|_{L^q_\nu(X)} = \|h\|^t_{L^{tq}_\nu(X)} \leq \left(D\,\|h\|_{L^{tp}_\mu(X)}\right)^t = D^t\,\|f\|_{L^p_\mu(X)}\,.$$

Because C is the least constant for (7.7), $C \leq D^t$.

For equivalence of the inclusion problems, suppose that $L^p_\mu(X) \subset L^q_\nu(X)$. For any $h \in L^+(X)$, let $f = h^t$ and note that if $h \in L^{tp}_\mu(X)$, then $\|f\|_{L^p_\mu(X)} = \|h\|^t_{L^{tp}_\mu(X)} < \infty$, so $\|h\|_{L^{tq}_\nu(X)} = \|f\|^{1/t}_{L^q_\nu(X)} < \infty$. This means that $L^{tp}_\mu(X) \subset L^{tq}_\nu(X)$. The converse is proven similarly. □

With this, it is possible to convert the original inclusion problem to one involving Banach function spaces (defined in such sources as [5]; see Definitions 1.1 and 1.3), if this is not already the case. Recall the standard fact that any linear operator between normed vector spaces is continuous if and only if it is bounded.

Definition 2.2.5. Where V and W are normed vector spaces, a linear map $T : V \to W$ is *bounded* if and only if there is some constant $C \in [0, \infty)$ such that, for all $v \in V$,

$$\|Tv\|_W \leq C\,\|v\|_V\,.$$

Remark A linear map $T : V \to W$ between normed vector spaces is bounded if and only if it is continuous.

This equivalence is standard and given here without proof. See Folland's [15], where it is Proposition 5.2. The above notion of boundedness for linear maps is defined immediately preceding the proposition. The following proof rests on the more powerful result that every inclusion between Banach function spaces is continuous, found for example in [5] as Theorem 1.8.

Proposition 2.2.6. *The inclusion $L^p_\mu(X) \subset L^q_\nu(X)$ holds if and only if there is a constant $C < \infty$ such that*

$$\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)}$$

for all $f \in L^+(X)$.

Proof. If either p or q is strictly less than 1, let

$$t = \max(p^{-1}, q^{-1})$$

and use this t in Lemma 2.2.4 to convert the inclusion problem $L^p_\mu(X) \subset L^q_\nu(X)$ to the equivalent $L^{tp}_\mu(X) \subset L^{tq}_\nu(X)$, with both exponents $tp, tq \geq 1$. This transformation also takes the problem of finding the least constant C such that

$$\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)}$$

for all $f \in L^+(X)$ to the problem of finding the least constant D such that

$$\|h\|_{L^{tq}_\nu(X)} \leq D\,\|h\|_{L^{tp}_\mu(X)}$$

for all $h \in L^+(X)$. If either constant exists and is finite, so is the other, and $C = D^t$, by Lemma 2.2.4. This also means that, if one fails to exist (as a finite constant), then so does the other. Replace p by tp and q by tq for the remainder, so that $p, q \geq 1$.

With exponents at least 1, the Lebesgue spaces are Banach function spaces, as defined in [5]. Therefore, if $L^p_\mu(X) \subset L^q_\nu(X)$, then by Theorem 1.8 in [5], there is a constant $C < \infty$ such that $\|f\|_{L^q_\nu(X)} \leq C\,\|f\|_{L^p_\mu(X)}$ for all $f \in L^+(X)$. The converse is simple, as the existence of $C < \infty$ implies that there is a continous inclusion map $L^p_\mu(X) \hookrightarrow L^q_\nu(X)$. □

This result is why the inclusion problem is addressed by finding the least constant C; the inclusion holds if and only if $C < \infty$, in which case C provides the operator norm of the inclusion map.

2.3 Hölder's inequality

Recall Hölder's inequality, a fundamental result in the theory of L^p spaces. It involves the concept of conjugate exponents, defined here along with one standard notation, the prime notation (as in p') used for it. This notation is used throughout.

Definition 2.3.1. Given any $p \in [1, \infty]$, its *conjugate* exponent $p' \in [1, \infty]$ is defined by

$$\frac{1}{p} + \frac{1}{p'} = 1. \tag{2.4}$$

Because this definition is symmetric in the roles of p and p', any $(p')' = p$. Note that although this computation can be applied when $0 < p < 1$, in this case $-\infty < p' < 0$, and vice versa with p and p' exchanged. Furthermore, Hölder's inequality, in the form described below, is not valid for such exponents.

Theorem 2.3.2 (Hölder's inequality)**.** *For any measure space (X, μ), any μ-measurable, complex-valued functions f and g on X, any $p \in [1, \infty]$ and its conjugate exponent $p' \in [1, \infty]$,*

$$\int_X |fg| d\mu \leq \|f\|_{L^p_\mu} \|g\|_{L^{p'}_\mu}.$$

If $1 < p < \infty$ and $f \in L^p_\mu$, $g \in L^{p'}_\mu$, then $fg \in L^1$, with equality if and only if f^p and $g^{p'}$ are linearly dependent, i.e. there are constants α, β not both zero such that $\alpha f^p = \beta g^{p'}$ μ-a.e.

Hölder's inequality is sharp, which can be expressed in the following way. (A similar result is true even for measures which are not σ-finite, but with the additional requirement that f be in $L^p_\mu(X)$. Since this work is only concerned with σ-finite measures anyway, this version is more convenient.)

Corollary 2.3.3. *Suppose that (X, μ) is a σ-finite measure space. For any particular $g \in L^+(X)$, the least constant $0 \leq C \leq \infty$ such that, for any $f \in L^+(X)$,*

$$\int_X fg d\mu \leq C \|f\|_{L^p_\mu(X)}$$

is $C = \|g\|_{L^{p'}_\mu(X)}$. Naturally, this means that it is possible to have $C < \infty$ if and only if $g \in L^{p'}_\mu(X)$.

Proof. By Hölder's inequality, this inequality is valid with $C = \|g\|_{L^{p'}_\mu(X)} \in [0, \infty]$, so it remains only to establish that this is the least possible constant. Note that both sides are zero (and $C = 0 = \|g\|_{L^{p'}_\mu(X)}$) when μ is the zero measure on X, so assume that μ is nonzero.

Since μ is σ-finite and nonzero, there is an increasing sequence $(E_n)_{n \geq 1}$ of measurable subsets of X such that each $\mu E_n \in (0, \infty)$ and $X = \bigcup_{n \geq 1} E_n$.

If $p = \infty$, define for each $n \geq 1$ $f_n = \chi_{E_n}$, so that each $\|f_n\|_{L^\infty_\mu(X)} = 1$ since $\mu(E_n) > 0$. Also, $\int_X f_n g d\mu = \int_{E_n} g d\mu \to \int_X g d\mu = \|g\|_{L^1_\mu(X)}$ by the Monotone Convergence Theorem. Therefore no constant less than $\|g\|_{L^1_\mu(X)}$ is satisfactory.

If $p = 1$, observe that for any $c < \|g\|_{L^\infty_\mu(X)}$, there is a measurable $S_c \subset X$ such that $\mu S_c \in (0, \infty)$ and $g > c$ on S_c. Therefore $\int_X \chi_{S_c} g d\mu = \int_{S_c} g d\mu \geq c\mu(S_c) = c\|\chi_{S_c}\|_{L^1_\mu(X)}$, so $C \geq c$. Because this applies to any $c < \|g\|_{L^\infty_\mu(X)}$, the least possible value of C is $\|g\|_{L^\infty_\mu(X)}$.

Now suppose that $1 < p < \infty$. For each $n \geq 1$, let $g_n = \min(g, n)\chi_{E_n}$ and $f_n = \min(g, n)^{p'/p}\chi_{E_n}$. Note that $f_n^p = g_n^{p'} = \min(g, n)^{p'}\chi_{E_n}$, which has an integral bounded above by $\int_{E_n} n^{p'} d\mu \leq n^{p'}\mu(E_n) < \infty$, so $f_n \in L^p_\mu(X)$ and $g_n \in L^{p'}_\mu(X)$. There is now equality in Hölder's inequality, so $\int_X f_n g_n d\mu = \|f_n\|_{L^p_\mu(X)} \|g_n\|_{L^{p'}_\mu(X)}$, implying that $C \geq \|g_n\|_{L^{p'}_\mu(X)}$. By the Monotone Convergence Theorem, $\lim_{n \to \infty} \|g_n\|_{L^{p'}_\mu(X)} = \left(\lim_{n \to \infty} \int_X \min(g, n)^{p'} \chi_{E_n} d\mu\right)^{1/p'} = \left(\int_X g^{p'} d\mu\right)^{1/p'} = \|g\|_{L^{p'}_\mu(X)}$, so $\|g\|_{L^{p'}_\mu(X)}$ is the least possible value of C. □

Simply rearranging the above inequality then provides the following result.

Corollary 2.3.4. *For any* $g \in L^+(X)$*,*

$$\sup_f \frac{\int_X fg d\mu}{\|f\|_{L^p_\mu(X)}} = \|g\|_{L^{p'}_\mu(X)},$$

where this supremum is taken over $f \in L^+(X)$*, not almost everywhere zero.*

Corollary 2.3.3 to Hölder's inequality easily provides answers to some cases of the Lebesgue space inclusion problem, including the following familiar examples.

Example 2.3.5. Whenever $1 \le q \le p \le \infty$, $L^p[0,1] \subset L^q[0,1]$ (with Lebesgue measure), where the inclusion map has norm 1. (This is a slight generalization of the fact that any function on $[0,1]$ which is p^{th} power integrable ($p \ge 1$), e.g. square-integrable, must also be integrable.)

Proof. This can be established by showing that the least constant C so that the following inequality holds for all $f \in L^+(X)$ is 1.

$$\|f\|_q \le C\,\|f\|_p$$

If $q = \infty$, then $p = \infty$ and clearly the least constant is $C = 1$. Now, if $q < p = \infty$, then $C = 1$ works since $f \le \|f\|_\infty$ a.e., and therefore $\int_0^1 |f(x)|^q dx \le \|f\|_\infty^q (1-0) = \|f\|_\infty^q$; it is the least constant because, for any constant $c \ge 0$, $(\int_0^1 c^q dx)^{1/q} = c$. On the other hand, for $p < \infty$, this inequality has the form

$$\left(\int_0^1 f(x)^q dx\right)^{1/q} \le C\left(\int_0^1 f(x)^p dx\right)^{1/p}.$$

Substitute $h = f^q$ and take the q^{th} power of each side for

$$\int_0^1 h(x)dx \le C^q\left(\int_0^1 h(x)^{p/q} dx\right)^{q/p}.$$

Corollary 2.3.3, applied with g as the constant function 1, shows that the least value of C^q (and therefore C itself) is $\|g\|_{(p/q)'} = \left(\int_0^1 1dx\right)^{1/(p/q)'} = 1$. □

A similar argument shows that $L^p_\mu(X) \subset L^q_\mu(X)$ whenever μ is a finite measure on X, though the constant will not necessarily be one. In fact, this is a special case of Proposition 2.4.1 with the Radon-Nikodym derivative $\frac{d\mu}{d\mu} = 1$, giving $C = (\mu(X))^{1/q}$. On the other hand, when there is just one measure μ and $\mu(X) = \infty$, there is no inclusion with $q < p$, as illustrated in the following special case.

Example 2.3.6. When $1 \le q < p \le \infty$, $\ell^p \not\subset \ell^q$, with the customary counting measure. (This includes the fact that square-summable or other p^{th}-power summable series need not be summable.)

Proof. The case $p = \infty$ is trivial since, when $f(n)$ is the constant sequence 1, $\|f\|_\infty = 1$ while $\|f\|_q = (\sum_n 1)^{1/q} = \infty$, so the least constant in $\|f\|_q \le C\,\|f\|_\infty$ is $C = \infty$. When $p < \infty$, the result after substituting $h = f^q$ and taking q^{th} powers is

$$\sum_n h(n) \le C^q \left(\sum_n h(n)^{p/q} \right)^{q/p}.$$

Again, use $g = 1$ in Corollary 2.3.3, this time to find that the least $C^q = (\sum_n 1)^{1/(p/q)'} = \infty$, so $C = \infty$ as well. □

However, because no non-empty set can have a counting measure less than one, when $p < q$ there is an argument which shows that $\ell^p \subset \ell^q$. The exponents are not suitable for Hölder's inequality, so it is not applied, but the general argument is developed in Proposition 2.4.3.

2.4 Proof for one-variable case

The following results, taken together, establish Theorem 2.1.8, giving necessary and sufficient conditions that $L^p_\mu(X) \subset L^q_\nu(X)$, i.e. the existence of $C < \infty$ such that $\|f\|_{L^q_\nu(X)} \le C\|f\|_{L^p_\mu(X)}$.

Proposition 2.4.1 (Hölder condition)**.** *When $0 < q \le p \le \infty$, there is a constant $C < \infty$ such that $\|f\|_{L^q_\nu(X)} \le C\|f\|_{L^p_\mu(X)}$ for any $f \in L^+(X)$ if and only if the appropriate condition holds:*

If $p < \infty$, $\quad \dfrac{d\nu}{d\mu} \in L^{p/(p-q)}_\mu(X)$ $\quad$ *with* $C = \left(\left\| \dfrac{d\nu}{d\mu} \right\|_{L^{\frac{p}{p-q}}_\mu(X)} \right)^{1/q}.$

If $q < p = \infty$, $\quad \dfrac{d\nu}{d\mu} \in L^1_\mu(X)$ $\quad$ *with* $C = \left(\left\| \dfrac{d\nu}{d\mu} \right\|_{L^1_\mu(X)} \right)^{1/q}.$

If $q = p = \infty$, $\quad C = 1.$

where, as noted above, $\frac{d\nu}{d\mu}$ denotes the Radon-Nikodym derivative of ν with respect to μ.

Proof. If $q < \infty$, Hölder's inequality gives

$$\|f\|^q_{L^q_\nu(X)} = \int_X |f|^q \frac{d\nu}{d\mu} d\mu \le \||f|^q\|_{L^{p/q}_\mu(X)} \left\| \frac{d\nu}{d\mu} \right\|_{L^{(p/q)'}_\mu(X)} = \left\| \frac{d\nu}{d\mu} \right\|_{L^{(p/q)'}_\mu(X)} \|f\|^q_{L^p_\mu(X)}.$$

Because the inequality is sharp, as found in Corollary 2.3.3, the least possible constant $C \in [0, \infty]$ in the inequality is $\left\| \frac{d\nu}{d\mu} \right\|^{1/q}_{L^{(p/q)'}_\mu(X)}$. If furthermore $p < \infty$, then $(p/q)' = \left(1 - \frac{q}{p}\right)^{-1} = \frac{p}{p-q}$, while if $p = \infty$, then $p/q = \infty$ and $(p/q)' = 1$. In either case, $C < \infty$ if and only if $\frac{d\nu}{d\mu}$ is in the appropriate Lebesgue space.

Now suppose $p = q = \infty$. For any $f \in L^+(X)$ such that $\|f\|_{L^\infty_\mu(X)} = \infty$, the inequality holds regardless of C. When $\|f\|_{L^\infty_\mu(X)} < \infty$, let $E = f^{-1}(\|f\|_{L^\infty_\mu(X)}, \infty]$. Because $\mu E = 0$ and $\nu \ll \mu$, $\nu E = 0$, so $\|f\|_{L^\infty_\nu(X)} \le 1 \cdot \|f\|_{L^\infty_\mu(X)}$. To see that 1 is the least possible constant, consider any constant function, for which $\|f\|_{L^\infty_\nu(X)} = \|f\|_{L^\infty_\mu(X)}$. □

Proposition 2.4.2. *When* $0 < p < q \le \infty$*, define a measure*

$$\lambda(S) = \int_{S \cap \{\frac{d\nu}{d\mu} > 0\}} \left(\frac{d\nu}{d\mu}\right)^{p/(p-q)} d\mu$$

(with the natural interpretation $\lambda = \mu$ *when* $q = \infty$*). Then* λ *is* σ*-finite and* $\|f\|_{L^q_\nu(X)} \le C\,\|f\|_{L^p_\mu(X)}$ *for all* $f \in L^+(X)$ *if and only if, for all* $h \in L^+(X)$*,*

$$\|h\|_{L^q_\lambda(X)} \le C\|h\|_{L^p_\lambda(X)}. \tag{2.5}$$

Proof. Because μ is σ-finite and the function

$$\chi_{\{\frac{d\nu}{d\mu}>0\}} \left(\frac{d\nu}{d\mu}\right)^{p/(p-q)}$$

is measurable and finite μ-a.e., λ is also σ-finite. Assuming $\|f\|_{L^q_\nu(X)} \le C\,\|f\|_{L^p_\mu(X)}$, for any $h \in L^+(X)$, let

$$f = h\chi_{\{\frac{d\nu}{d\mu}>0\}} \left(\frac{d\nu}{d\mu}\right)^{1/(p-q)}$$

and substitute to derive (2.5).

Conversely, given (2.5), for any $f \in L^+(X)$ let $h = f\left(\frac{d\nu}{d\mu}\right)^{1/(q-p)}$ and compute

$$\|h\|_{L^q_\lambda(X)} = \left\| f\left(\frac{d\nu}{d\mu}\right)^{1/q} \chi_{\{\frac{d\nu}{d\mu}>0\}} \right\|_{L^q_\mu(X)} = \left\| f\left(\frac{d\nu}{d\mu}\right)^{1/q} \right\|_{L^q_\mu(X)} = \|f\|_{L^q_\nu(X)},$$

$$\|h\|_{L^p_\lambda(X)} = \left\| f\chi_{\{\frac{d\nu}{d\mu}>0\}} \right\|_{L^p_\mu(X)} \le \|f\|_{L^p_\mu(X)},$$

from which $\|f\|_{L^q_\nu(X)} \le C\,\|f\|_{L^p_\mu(X)}$ follows. □

Proposition 2.4.3. *Let* $0 < p < q \le \infty$*. If there exists* $\varepsilon > 0$ *such that every measurable* $E \subset X$ *with* $\lambda(E) < \varepsilon$ *is* λ*-null, then for any* $C \ge \varepsilon^{1/q-1/p}$*,*

$$\|h\|_{L^q_\lambda(X)} \le C\,\|h\|_{L^p_\lambda(X)}$$

for any $h \in L^+(X)$*.*

Proof. If $h \notin L^p_\lambda(X)$ or $h = 0$ λ-a.e., then the inequality is clearly true, so suppose neither of these is true. Then $\lambda_h(t) := \lambda\left(h^{-1}(t,\infty]\right)$ is a non-increasing function of $t > 0$, everywhere finite and not everywhere zero. Let $T = \inf\{t > 0 \mid \lambda_h(t) = 0\} \in (0,\infty]$ and note that $h(x) \le T$ for λ-a.e. $x \in X$. Whenever $0 < t < T$, $\lambda_h(t) > 0$, so $\lambda_h(t) \ge \varepsilon$. Thus,

$$t^p\varepsilon \le t^p\lambda_h(t) = \int_{\{h(x)>t\}} t^p d\lambda(x) \le \int_{\{h(x)>t\}} h(x)^p d\lambda(x) \le \int_X h(x)^p d\lambda(x) = \|h\|^p_{L^p_\lambda(X)}.$$

Taking p^{th} roots and letting $t \to T$, this yields $T \le \varepsilon^{-1/p}\|h\|_{L^p_\lambda(X)}$. This is exactly what is required when $q = \infty$. For $q < \infty$,

$$\|h\|^q_{L^q_\lambda(X)} = \int h(x)^q d\lambda(x) \le T^{q-p} \int h(x)^p d\lambda(x) \le \left(\varepsilon^{-1/p}\|h\|_{L^p_\lambda(X)}\right)^{q-p} \|h\|^p_{L^p_\lambda(X)}.$$

With q^{th} roots, the proof is complete. □

Corollary 2.4.4. *Suppose that* $0 < p < q \leq \infty$ *and let* ε *denote the infimum of* $\lambda(E)$ *for all measurable* $E \subset X$ *such that* $\lambda(E) > 0$. *If* $\varepsilon > 0$, *then the least value of* C *such that*

$$\|h\|_{L^q_\lambda(X)} \leq C\,\|h\|_{L^p_\lambda(X)}$$

for any $h \in L^+(X)$ *is* $C = \varepsilon^{1/q-1/p}$.

Proof. Proposition 2.4.3 has established that the least $C \leq \varepsilon^{1/q-1/p}$, so it remains only to show that no lesser value works. By the definition of ε, for any $n \geq 1$ there is some measurable $E_n \subset X$ such that $\varepsilon \leq \lambda(E_n) < \varepsilon + \frac{1}{n}$. Let $h_n = \chi_{E_n}$ and observe that

$$\frac{\|h_n\|_{L^q_\lambda(X)}}{\|h_n\|_{L^p_\lambda(X)}} = \frac{\left(\int_X \chi^q_{E_n}\, d\lambda\right)^{1/q}}{\left(\int_X \chi^p_{E_n}\, d\lambda\right)^{1/p}} = \frac{\lambda(E_n)^{1/q}}{\lambda(E_n)^{1/p}} = \lambda(E_n)^{1/q-1/p}.$$

(Because $\lambda(E_n) > 0$, $\|h_n\|_{L^\infty_\lambda(X)} = 1$, so the numerator is still $\lambda(E_n)^{1/q}$ when $q = \infty$, with the natural interpretation $\frac{1}{\infty} = 0$.) Since the minimum value of C is $\sup_{f\neq 0} \frac{\|f\|_{L^q_\lambda(X)}}{\|f\|_{L^p_\lambda(X)}}$ and, because $\frac{1}{q} - \frac{1}{p} < 0$, $\lambda(E_n)^{1/q-1/p} > \left(\varepsilon + \frac{1}{n}\right)^{1/q-1/p}$, $C \geq \left(\varepsilon + \frac{1}{n}\right)^{1/q-1/p}$ for any n, and therefore $C \geq \varepsilon^{1/q-1/p}$. □

Lemma 2.4.5. *For* $0 < p < q \leq \infty$, *if there are measurable sets of arbitrarily small positive* λ *measure, then there is no finite constant* C *such that, for any* $h \in L^+(X)$,

$$\|h\|_{L^q_\lambda(X)} \leq C\,\|h\|_{L^p_\lambda(X)}\,.$$

Proof. Substituting $h = \chi_E$ into (2.5) produces $\lambda(E)^{1/q} \leq C\lambda(E)^{1/p}$, which implies that $C \geq \lambda(E)^{1/q-1/p}$ when $\lambda(E) > 0$. Because the exponent is negative, choosing E of sufficiently small positive λ measure makes the right-hand side arbitrarily large. □

Taken together, these propositions yield Theorem 2.1.8.

2.5 Atomless measures only allow Hölder inclusion

Of the two types of conditions for inclusion between Lebesgue spaces, for many measures inclusion is only possible with the Hölder condition, Proposition 2.4.1. The relevant properties of measures are described below.

Definition 2.5.1. A measurable subset A of a measure space (X, μ) is called an *atom* when $\mu A > 0$ and, for any measurable $F \subset X$, either $\mu(A \cap F) = 0$ or $\mu(A \setminus F) = 0$.

Naturally, this means that the other piece, $A \setminus F$ or $A \cap F$ respectively, must have the full measure of A.

Proposition 2.5.2. *Equivalently, we can define an atom* A *as a measurable set which, when decomposed into any disjoint union* $A = E \cup F$ *of measurable sets* E *and* F, *must have one of* $\mu E = 0$ *or* $\mu F = 0$.

Proof. Were A not an atom by the second definition, there would be a disjoint union $A = E \cup F$ with both $\mu E > 0$ and $\mu F > 0$. Then, since (by disjointness) $\mu E + \mu F = \mu A$, $0 < \mu E = \mu A - \mu F < \mu A$. Consequently, A cannot be an atom by the first definition. Conversely, were A not an atom by the first definition, there would be a subset $E \subset A$ with $0 < \mu F < \mu A$ and so, letting $E = A \setminus F$, we'd also have $\mu E = \mu A - \mu F > 0$, so A could not be an atom by the second definition. □

As the name suggests, an atom is essentially an indivisible set. (Null sets might as well be empty for many purposes, including the functional analytic operations here, based on the integral and essential supremum.) This indivisibility makes atoms effectively much like singletons, as suggested by results in Section 2.6, especially the finding in Lemma 2.6.7 that any measurable function on a measure space must be almost constant (constant up to null sets) on any atom in the space. But of more immediate importance are two properties of measures involving atoms, which will also be important in later chapters to separate cases of the problem for more than one variable.

Definition 2.5.3. A measure is called *purely atomic* when every measurable set of positive measure contains an atom. At the other extreme, it is *atomless* when it has no atom, i.e. every subset of positive measure can be expressed as a disjoint union of two sets of positive measure.

Example 2.5.4. Counting measure on any set X is purely atomic, as is any weighted counting measure (a measure μ on X where, for any $S \subset X$, $\mu(S) = \sum_{x \in S} m(x)$, where $m : X \to [0, \infty]$).

Example 2.5.5. Lebesgue measure on any Euclidean space is atomless. So is any weighted Lebesgue measure, i.e. a measure μ defined by $\mu(E) = \int_E w(x)dx$, for any Lebesgue measurable $E \subset X$, where w is a fixed nonnegative measurable function. (By the Radon-Nikodym theorem, every measure which is absolutely continuous with respect to Lebesgue measure is a weighted Lebesgue measure.)

These terms are the same as those used by Bogachev [11], though other authors differ; for example, sometimes "non-atomic" is used instead of "atomless". The following simple lemma shows that any atomless space must have sets of arbitrarily small measure. Although a theorem with a rather stronger result is cited as Theorem 4.1.4, that much won't be needed to show that inclusion among Lebesgue spaces with atomless measures is only possible through the Hölder condition, Proposition 2.4.1.

Lemma 2.5.6. *If (X, μ) is an atomless measure space and $E \subset X$ a set with $\mu E > 0$, then for any $\varepsilon > 0$ there is a measurable subset $F \subset E$ with $0 < \mu F < \varepsilon$.*

Proof. The goal is to prove by induction that, for any $n \geq 0$, there is some $F_n \subset E$ such that $0 < \mu F_n \leq 2^{-n} \mu E$. (This is sufficient since, for any $\varepsilon > 0$, there is some $n \geq 0$ such that $2^{-n} \mu E < \varepsilon$.) When $n = 0$, $F_0 = E$ suffices. For the inductive step, suppose that $F_n \subset E$ has $0 < \mu F_n \leq 2^{-n} \mu E$. Because μ is atomless, by Proposition 2.5.2 there are disjoint subsets A and B of F_n such that $F_n = A \cup B$ and both $\mu A > 0$ and $\mu B > 0$. Because $\mu A + \mu B = \mu F_n$, let F_{n+1} be whichever of A and B has the lesser measure and then $0 < \mu F_{n+1} \leq \frac{1}{2} \mu F_n \leq 2^{-(n+1)} \mu E$. □

Corollary 2.5.7. *Let μ and ν be σ-finite measures on X and $p, q \in (0, \infty]$. When ν is not purely atomic, $\|f\|_{L^q_\nu(X)} \leq C \|f\|_{L^p_\mu(X)}$ cannot hold with a constant $C < \infty$ unless $q \leq p$.*

Proof. Assume that ν is not purely atomic, so there is some measurable set $Y' \subset X$, with $\nu Y' > 0$, which contains no atom for ν. Let g denote the Radon-Nikodym derivative of ν with respect to μ, which must exist if inclusion is to be possible, and let $Y = Y' \cap g^{-1}(0, \infty)$, i.e. the subset of Y' where $g > 0$. Since $\nu(Y' \setminus Y) = \int_{Y' \setminus Y} g(x) d\mu(x) = \int_{Y' \setminus Y} 0 d\mu(x) = 0$, $\nu(Y) = \nu(Y') > 0$. Therefore $\mu(Y) > 0$ as well.

Furthermore, Y cannot contain any atom for μ. For any measurable $A \subset Y$ there must be disjoint union $A = E \cup F$ such that $\nu E > 0$ and $\nu F > 0$, by Proposition 2.5.2, since A cannot be an atom for ν. Of course, this means that $\mu E > 0$ and $\mu F > 0$, so A cannot be an atom for μ. Restricting μ and ν to Y then yields atomless measures with the Radon-Nikodym derivative $g > 0$ μ-almost everywhere. Since $L^p_\mu(X) \subset L^q_\nu(X)$ would also apply to functions supported on Y, it can be disproven by refuting $L^p_\mu(Y) \subset L^q_\nu(Y)$.

For any $p < q$, define λ as in Proposition 2.4.2, so that $d\lambda = g^{p/(p-q)} d\mu$. By Lemma 2.4.5, it suffices to show that there are subsets of Y with arbitrarily small λ measure, which Lemma 2.5.6 establishes must be true as long as λ is atomless on Y. For any $A \subset Y$ with $\lambda A > 0$, also $\mu A > 0$, so using Proposition 2.5.2 again there is a disjoint union $A = E \cup F$ such that $\mu E > 0$ and $\mu F > 0$. But, since $g > 0$ μ-a.e., also $g^{p/(p-q)} > 0$ μ-a.e., and therefore $\lambda E > 0$ and $\lambda F > 0$. This means that A cannot be an atom for λ, which therefore is atomless on Y. □

2.6 Integrals over σ-finite purely atomic spaces are weighted series

Having found that, unless ν is purely atomic, the inclusion $L^p_\mu(X) \subset L^q_\nu(X)$ is only possible when $q \leq p$, and only with the Hölder condition given in Proposition 2.4.1, we now investigate the purely atomic case. Although there is more flexibility in the exponents, as p and q need not be in any particular order, Lebesgue spaces with purely atomic measures turn out to be quite constrained. Not only are the sequence spaces ℓ^p the archetypal example, but weighted ℓ^p is essentially the only kind of Lebesgue space over a purely atomic, σ-finite space.

The following arguments establish that any σ-finite purely atomic measure space can be decomposed into a countable disjoint collection of atoms and a leftover null set, and ultimately that integration over any σ-finite, purely atomic measure space can be computed as a weighted series. There's nothing new here, but since later treatment will freely represent such measure spaces as weighted counting measure on the natural numbers, an explanation follows.

One technical difficulty in making sense of counting atoms and even a basic notion of distinct atoms is that, given any atom, adding or removing a null set will result in another atom. As in the familiar notion of equivalence of functions which agree almost everywhere, we must disregard such differences of measure zero, as made precise in the following equivalence relation.

Definition 2.6.1. Let (X, Σ, μ) be a measure space. Atoms $A_1, A_2 \in \Sigma$ are *equivalent*, denoted $A_1 \sim A_2$, if and only if their symmetric difference is a null set, i.e. $\mu(A_1 \triangle A_2) = 0$.

By additivity, any equivalent atoms $A_1 \sim A_2$ have equal measures, so for any equivalence class $\mathcal{A}$, $\mu(\mathcal{A})$ is well defined. It is also useful to have a notion of disjointness applicable to these equivalence classes, for which the measure-theoretic definition of "almost disjoint" is suitable. (Naturally, given any atoms A_1 and A_2 and a non-empty null set N, $A_1 \cup N$ and $A_2 \cup N$ cannot be disjoint, though each $A_k \cup N \sim A_k$.)

Definition 2.6.2. Two atoms A_1 and A_2 are said to be almost disjoint when $A_1 \cap A_2$ has zero measure. Similarly, we can call two equivalence classes $\mathcal{A}_1$ and $\mathcal{A}_2$ almost disjoint when there exist representatives $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ which are almost disjoint.

Note that this apparently weak demand that there exist an almost disjoint pair of representatives nonetheless implies that every pair of representatives is almost disjoint. Furthermore, it turns out that any pair of inequivalent atoms must, in this sense, be almost disjoint.

Lemma 2.6.3. *Any two inequivalent atoms are almost disjoint.*

Proof. Suppose that two atoms A_1 and A_2 are not almost disjoint. Then $A_1 \cap A_2$ is a measurable subset of A_1 with positive measure. Since A_1 is an atom, $A_1 \setminus (A_1 \cap A_2) = A_1 \setminus A_2$ is a null set. Similarly, $A_2 \setminus A_1$ is a null set. Therefore, so is $A_1 \triangle A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1)$, and $A_1 \sim A_2$. □

The following corollary restates this result in terms of equivalence classes.

Corollary 2.6.4. *Any two distinct equivalence classes of atoms are almost disjoint.*

The countable additivity property of measures applies not only to disjoint sets, but also to almost disjoint sets.

Lemma 2.6.5. *For any sequence (E_n) of pairwise almost disjoint measurable sets in a space with measure μ, $\mu(\bigcup_n E_n) = \sum_n \mu E_n$.*

Proof. Let $F_1 = E_1$ and, for $n > 1$, define $F_n = E_n \setminus (E_{n-1} \cup \cdots \cup E_1)$. Because $E_n = F_n \cup (E_n \cap E_{n-1}) \cup \cdots \cup (E_n \cap E_1)$, $\mu E_n \leq \mu F_n + \mu(E_n \cap E_{n-1}) + \cdots + \mu(E_n \cap E_1) = \mu F_n + 0 + \cdots + 0$. But $F_n \subset E_n$, so $\mu F_n \leq \mu E_n$, and therefore $\mu F_n = \mu E_n$. The sets (F_n) are disjoint and have the same union as the (E_n), so by countable additivity $\mu(\bigcup_n E_n) = \mu(\bigcup_n F_n) = \sum_n \mu F_n = \sum_n \mu E_n$. □

Proposition 2.6.6. *Let μ be a σ-finite measure on a space X. Then there are only countably many equivalence classes of atoms under $\sim$.*

Proof. Consider any measurable $E \subset X$ and natural number m. If there are infinitely many equivalence classes of atoms with measure at least $\frac{1}{m}$ and a representative contained in E, then there is a sequence $(\mathcal{A}_n)$ of such classes. Taking representatives $A_n \subset \mathcal{A}_n$ with each $A_n \subset E$, by Lemma 2.6.3 the sequence (A_n) is almost disjoint. Because $\bigcup_n A_n \subset E$, the countable additivity provided by Lemma 2.6.5 gives $\mu E \geq \mu(\bigcup_n A_n) = \sum_n \mu A_n \geq \sum_n \frac{1}{m} = \infty$.

Conversely, if $\mu E < \infty$, then the collection C_m of classes of atoms with measure at least $\frac{1}{m}$ and a representative contained in E is finite. Since every atom has positive measure, the set of all classes of atoms with a representative contained in E is $\bigcup_m C_m$, which must be countable.

Because μ is σ-finite, there is a sequence $E_1 \subset E_2 \subset E_3 \cdots$ of measurable sets such that each $\mu E_k < \infty$ and $X = \bigcup_k E_k$. For any atom A, were $\mu(A \cap E_k) = 0$ for every k, then by subadditivity $\mu(A) = \mu(\bigcup_k A \cap E_k) \leq \sum_k \mu(A \cap E_k) = 0$; of course, atoms have positive measure, so this means that there must be some k such that $\mu(A \cap E_k) > 0$. Because A is an atom, this implies that $\mu(A \cap E_k) = \mu A$, and A is equivalent to the atom $A \cap E_k$ contained in E_k. Every equivalence class of atoms has a representative contained in some E_k. Since there are at most countably many equivalence classes of atoms contained in each E_k, this implies that, as a countable union of countable sets, there collection of all equivalence classes of atoms is countable. □

The following fact suggests that, for the purpose of working with functions on measure spaces, atoms might as well be singletons.

Lemma 2.6.7. *Any measurable function on a measure space must be almost constant on any atom in the space, in the sense of having a single value almost everywhere on the atom.*

Proof. Let f be a function on a measure space containing an atom A. Let y be the supremum of all t such that $\mu\{x \in A : f(x) < t\} = 0$. For any $y' > y$, $\mu\{x \in A : y \leq f(x) < y'\} > 0$; were this measure zero, then the supremum would have to have been at least y'. Because A is an atom, then, $\{x \in A : y \leq f(x) < y'\}$ must have the same measure as A. Therefore, $\mu\{x \in A : f(x) > y'\} = 0$. Taking a countable union using $y' = y + \frac{1}{n}$, we find that $\mu\{x \in A : f(x) > y\} = 0$. Consequently, $\mu\{x \in A : f(x) \neq y\} = \mu\{x \in A : f(x) < y\} + \mu\{x \in A : f(x) > y\} = 0$, so $f(x) = y$ almost everywhere on A. □

Finally, we have the following result, as a consequence of which we can replace functions defined on σ-finite purely atomic measure spaces by sequences of values. Since the integral and essential supremum are preserved by mapping each function to the corresponding sequence, all L^p norms are kept intact.

Theorem 2.6.8. *For any σ-finite purely atomic measure space (X, μ), there is a sequence (A_n), finite or infinite, of atoms such that every atom in X is equivalent to some A_n. For any $f \in L^+(X)$, for each n let $f(A_n)$ denote the value which f takes μ-a.e. on A_n. Then $L^+(X)$ is in one-to-one correspondence with sequences defined by $a_n = f(A_n)$; furthermore, this equivalence gives*

$$\int_X f d\mu = \sum_n a_n w_n,$$

with weights defined by $w_n = \mu(A_n)$, and $\operatorname{ess\,sup}_\mu f = \sup_n a_n$.

Proof. By Proposition 2.6.6, we can enumerate the classes of atoms in a (finite or infinite) sequence $(\mathcal{A}_n)$. For each n, let $w_n = \mu(A_n)$ for any $A_n \in \mathcal{A}_n$, noting that the measure is independent of the choice of representative. For each n, Lemma 2.6.7 tells us that f takes on one value $a_n = f(A_n)$ almost everywhere on the atom A_n. Because μ is purely atomic, the complement $X \setminus \bigcup_n A_n$ of all the atoms has measure zero, for it contains no atom. Therefore the integral

$$\int_X f d\mu = \sum_n \int_{A_n} f d\mu = \sum_n a_n \mu(A_n) = \sum_n a_n w_n.$$

Each value a_n is attained on A_n (which must have positive measure), so $\operatorname{ess\,sup} f \geq \sup_n a_n$. On the other hand, $f \leq \sup_n a_n$ on $\bigcup_n A_n$, which means that $f \leq \sup_n a_n$ almost everywhere, so $\operatorname{ess\,sup} f \leq \sup_n a_n$. □

Finally, the solution to the inclusion problem is specialized to purely atomic measures in the following corollaries. Suppose that μ and ν are purely atomic measures, the necessary

$\nu \ll \mu$ assumed, with the μ-atoms of X enumerated as $(E_i)_{i\in I}$ for some index set I which is at most countable, typically a subset of $\mathbb{N}$. For each $i \in I$, let $u_i = \mu(E_i) > 0$ and $v_i = \nu(E_i)$. Note that each E_i is either a ν-atom, if $v_i > 0$, or ν-null (when, of course, $v_i = 0$). The problem can then be described in terms of weighted ℓ^p, asking when $\ell^p_u(I) \subset \ell^q_v(I)$.

Corollary 2.6.9. *Let I be an at most countable set and, for each $i \in I$, let $u_i > 0$ and $v_i \geq 0$. The least constant $C < \infty$ such that, for any sequence $(c_i)_{i\in I}$,*

$$\|c_i\|_{\ell^q_v(I))} \leq C\,\|c_i\|_{\ell^p_u(I)}\,,$$

is as follows in each case, observing the conventions that $1/\infty = 0$ *and* $1/0 = \infty$.

$$\begin{aligned}
q = p = \infty : &\qquad C = 1\\
0 < q \leq p \leq \infty : &\qquad C = \left\| v_i^{1/q} u_i^{-1/p} \right\|_{\ell^{(q^{-1}-p^{-1})^{-1}}(I)}\\
0 < p < q \leq \infty : &\qquad C = \sup_{i\in I}\left\{ \left(v_i^{\frac{p}{p-q}} u_i^{-\frac{q}{p-q}} \right)^{q^{-1}-p^{-1}} < \infty \right\}
\end{aligned}$$

Proof. Based on Proposition 2.4.1 and Corollary 2.4.4,

$$\begin{aligned}
q = p = \infty : &\qquad C = 1\\
0 < q < p = \infty : &\qquad C = \left(\sum_{i\in I} v_i \right)^{1/q} = \nu(X)^{q^{-1}}\\
0 < q \leq p < \infty : &\qquad C = \left\| \left(\frac{v_i}{u_i} \right) \right\|^{1/q}_{\ell_u^{\frac{p}{p-q}}(I)} = \left(\sum_{i\in I} v_i^{\frac{p}{p-q}} u_i^{-\frac{q}{p-q}} \right)^{q^{-1}-p^{-1}}\\
0 < p < q \leq \infty : &\qquad C = \left(\inf_{S\subset I} \left\{ \sum_{i\in S} v_i^{\frac{p}{p-q}} u_i^{-\frac{q}{p-q}} > 0 \right\} \right)^{q^{-1}-p^{-1}}
\end{aligned}$$

The second and third cases combine neatly in the above formula. In the last case, $q^{-1} - p^{-1} < 0$, so the overall expression achieves greater values when the inner sum is smaller; since zero values (when $v_i = 0$) are excluded from the infimum, so the supremum excludes ∞. □

The following simple fact about inclusion among unweighted ℓ^p sequence spaces is presented for easy reference later; one half appeared before as an example.

Corollary 2.6.10. *The inclusion $\ell^p(\mathbb{N}) \subset \ell^q(\mathbb{N})$ occurs if and only if $p \leq q$, in which case the least constant $C = 1$.*

Proof. Simply use $u_i = v_i = 1$ in the above. As noted in an example at the end of Section 2.3, since $\sum_i 1 = \infty$, the Hölder criterion for inclusion (applicable when $q \leq p$) cannot possibly be satsified, except of course when $p = q$ and the spaces are identical. (Clearly, $C = 1$, which agrees with $\|1\|_\infty$.) On the other hand, when $p < q$ there is always inclusion, with $C = 1$, the supremum of $\left(1^{\frac{p}{p-q}} 1^{-\frac{q}{p-q}} \right)^{q^{-1}-p^{-1}}$. □

2.7 Summary

This chapter's solution to the one-variable problem is summarized in two following results. First, a helpful piece of notation which will be used more extensively later. For exponents $p, q \in (0, \infty]$,

$$(p:q)^{-1} = q^{-1} - p^{-1},$$

as defined in Definition 4.2.8, generalizing the standard Hölder conjugate $p' = p:1$. It is defined with the conventions $\infty^{-1} = 0$ and $0^{-1} = \infty$; for example, $\infty : \infty = 0^{-1} = \infty$.

First, a condensed version of Theorem 2.1.8. It is still assumed that $\nu \ll \mu$, for otherwise the best constant is $C = \infty$.

Theorem 2.7.1. *The least constant $C \in [0, \infty]$ such that $\|f\|_{L^q_\nu(X)} \leq C \|f\|_{L^p_\mu(X)}$ for any measurable function f on X is as specified below, separated by case.*

$$0 < q \leq p < \infty : \qquad C = \left\| \left(\frac{d\nu}{d\mu} \right)^{1/q} \right\|_{p:q} \qquad \textit{(Hölder condition)}$$

$$0 < p < q \leq \infty : \qquad C = \sup_E \left\| \left(\frac{d\nu}{d\mu} \right)^{1/q} \chi_{E \cap \left\{ \frac{d\nu}{d\mu} > 0 \right\}} \right\|_{p:q}$$

$$p = q = \infty : \qquad C = 1,$$

where the supremum in the second case is taken over all measurable $E \subset X$.

Recall that $L^p_\mu(X) \subset L^q_\nu(X)$ if and only if $C < \infty$, so this provides the following characterization of when inclusion holds, in terms of the measure λ defined in Proposition 2.4.2.

Corollary 2.7.2. *Define a measure $\lambda(E) = \int_{E \cap \left\{ \frac{d\nu}{d\mu} > 0 \right\}} \left(\frac{d\nu}{d\mu} \right)^{\frac{p}{p-q}} d\mu$ on the measurable space (X, Σ). Depending on p and q, the indicated condition on λ is both necessary and sufficient for inclusion.*

$0 < q \leq p < \infty : L^p_\mu(X) \subset L^q_\nu(X)$ *if and only if* $\{\lambda(E) > 0 : E \in \Sigma\}$ *is bounded away from* ∞.

$0 < p < q \leq \infty : L^p_\mu(X) \subset L^q_\nu(X)$ *if and only if* $\{\lambda(E) > 0 : E \in \Sigma\}$ *is bounded away from zero.*

$p = q = \infty : L^p_\mu(X) \subset L^q_\nu(X).$

A simpler way to state the first condition is $\lambda(X) < \infty$, but the formulation above is chosen to emphasize the similarity between these two conditions.

Chapter 3

Two-variable basics

Because of its length and the numerous cases and subcases it features, the solution to the two-variable mixed-norm inclusion problem is spread among Chapters 3, 4, and 5. (Chapter 6 summarizes the major results in these chapters, describing the major results and how they fit together to ssolve the two-variable inclusion problem.) As cases are solved, the remainder of the solution focuses on the complements of these solved cases. Note that there are two assumptions which hold without loss of generality, and should be treated as given beginning from the point where they appear.

Specifically, on page 26, an assumption of absolute continuity is introduced. This is shown to be a necessary condition, so there is no point in proceeding without it, for inclusion is certainly impossible unless it holds. This allows free use of the Radon-Nikodym derivative, which as in the one-variable case is important to the solution. Proposition 4.2.1 provides a reduction to problems involving only positive Radon-Nikodym derivatives, $\frac{d\nu_k}{d\mu_k} > 0$ for $k = 1, 2$. From 42 on, then, we assume this.

(These assumptions are based on ideas which work throughout, but the reader only needs to keep them in mind beginning where they occur, until the end of the two-variable problem.)

3.1 Definitions

Let (X_1, Σ_1) be a measurable space (i.e. Σ_1 is a σ-algebra on the set X_1) with a σ-finite measure μ_1, and let (X_2, Σ_2) be a measurable space with a σ-finite measure μ_2. Given, as well, exponents $p_1, p_2 \in (0, \infty]$, we let $P = (p_1, p_2)$, following the notation from Benedek and Panzone [4], and define a mixed norm by computing, for any measurable function f on $X_1 \times X_2$,

$$\|f\|_P = \left\| \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{p_2}_{\mu_2}(X_2)} = \left(\int_{X_2} \left(\int_{X_1} |f(x_1, x_2)|^{p_1} \, d\mu_1(x_1) \right)^{p_2/p_1} d\mu_2(x_2) \right)^{1/p_2} .$$

That is, first compute the L^{p_1} norm of the function $f(\cdot, x_2)$ on X_1 using the measure μ_1, then compute the L^{p_2} norm of the result (which depends on x_2). If either p_1 or p_2 is ∞, the appropriate integral is replaced by an essential supremum. The corresponding function space, the mixed norm space L^P, consists of the equivalence classes (under the standard identification of a.e. equal functions) of all measurable functions $f : X_1 \times X_2 \to \mathbb{C}$ such that $\|f\|_P < \infty$. When

$p_1 \geq 1$ and $p_2 \geq 1$, as shown in [4], this space is a Banach function space. For exponents less than one, the triangle inequality fails, so we do not have a true norm.

This mixed-norm computation is well-defined since the function $\|f(x_1, x_2)\|_{L^{p_1}_{\mu_1}(X_1)}$ of x_2 is itself measurable, by Tonelli's theorem if $p_1 < \infty$. In the case $p_1 = \infty$, the measurability of the intermediate function $\operatorname{ess\,sup}_{x_1} |f(x_1, \cdot)|$ is a consequence of the Luxemburg-Gribanov theorem [23], which applies for any Banach function norms with the Fatou property.

We also wish to discuss reversing the order in which the mixed norm is computed, with $L^{p_2}_{\mu_2}(X_2)$ before $L^{p_1}_{\mu_1}(X_1)$. The exponent vector $P = (p_1, p_2)$ notation does not really describe this concept, since writing, say, $(3, 2)$ instead of $(2, 3)$ would be understood as changing which exponent is associated with each variable, but not the order of integration. This is not surprising, since although Benedek and Panzone describe many properties of mixed-norm Lebesgue spaces in [4], they do not cover what Fournier calls "permuted mixed-norm spaces" in [16]. (This paper's main topics are Sobolev spaces and bilinearity, but both the concept of "permuted mixed-norm spaces" and a generalization of Minkowski's integral inequality to mixed-norm spaces are described.)

Since the order of integration is essentially arbitrary, and there is no fundamental reason to prefer X_1 before X_2 (or why one space should be "one" and the other "two"), all mixed norms here are assumed to be "permuted". For now, given a permutation σ (of $\{1, 2\}$, in this two-variable chapter) and $P = (p_1, p_2)$, we define the (permuted) mixed norm

$$\|f\|_{\sigma(P)} = \left(\int_{X_{\sigma(2)}} \left(\int_{X_{\sigma(1)}} |f(x_1, x_2)|^{p_{\sigma(1)}} \, d\mu_{\sigma(1)}\left(x_{\sigma(1)}\right) \right)^{p_{\sigma(2)}/p_{\sigma(1)}} d\mu_{\sigma(2)}\left(x_{\sigma(2)}\right) \right)^{1/p_{\sigma(2)}}$$

where, again, integration is replaced by the essential supremum for any ∞ exponent. Naturally, the mixed-norm space $L^{\sigma(P)}$ consists of (the equivalence classes of) those functions f for which $\|f\|_{\sigma(P)} < \infty$. If σ is the identity, then this is identical to $\|f\|_P$ from the previous definition, while if σ is the transposition (12), then $\|f\|_{\sigma(P)}$ is what would be obtained from $\|f\|_P$ by reversing the order in which the integrations over μ_1 and μ_2, as well as the exponentiations and roots p_1 and p_2, are applied. (Basically, every "one" object, such as X_1, μ_1, and p_1, becomes a "two" and vice versa.)

With that in mind, though, it's worth asking why there is any need to introduce the permutation. For discussing a single mixed norm, there really isn't, since relabeling would work as well. However, this is useful for discussing multiple mixed norms with different orders of integration. As described in Section 3.5, Minkowski's integral inequality provides a familiar example of comparing differently permuted mixed norms, especially in its Corollary 3.5.2.

Note that when $p_1 = p_2$, regardless of the permutation σ, $L^{\sigma(P)} = L^p$, the classical Lebesgue space, with $p = p_1 = p_2$. (Tonelli's theorem can be used to verify the result's independence of the order of integration for $p < \infty$.) However, for unequal exponents, the permutation makes a big difference. Not only can the numeric result of computing $\|f\|_{\sigma(P)}$ vary with σ, but the two norms need not be equivalent, as the following example shows. (However, Corollary 3.5.2 to Minkowski's integral inequality shows that one of them consistently gives greater values than the other, and consequently one of the spaces $L^{\sigma(P)}$ is always contained in the other.)

Example Using $X_1 = X_2 = \mathbb{N}$ with counting measure, $P = (1, 2)$, and σ the non-identity permutation in S_2, $L^P \not\subset L^{\sigma(P)}$.

Proof. Let

$$c_{i,j} = \begin{cases} i^{-1} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

The function $f(i,j) = c_{i,j}$ is in L^P but not in $L^{\sigma(P)}$.

$$\|f\|_P = \left(\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} c_{i,j}\right)^2\right)^{1/2} = \left(\sum_{j=1}^{\infty} j^{-2}\right)^{1/2} = \frac{\pi}{\sqrt{6}} < \infty,$$
$$\|f\|_{\sigma(P)} = \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} c_{i,j}^2\right)^{1/2} = \sum_{i=1}^{\infty} i^{-1} = \infty.$$

□

As we'll see later, because the exponent $p_1 = 1 < 2 = p_2$, in this example (by Corollary 3.5.2) $\|\cdot\|_P \le \|\cdot\|_{\sigma(P)}$, so $L^{\sigma(P)} \subset L^P$.

3.2 Problem statement

Given two mixed-norm spaces in two variables, we seek necessary and sufficient conditions for one to be a subset of another. Because mixed-norm spaces are Banach function spaces, whenever this happens the inclusion map must be continuous. That is, there is a mixed-norm inclusion if and only if there is a finite best constant. (This is demonstrated in multiple variables in Proposition 7.3.2.)

Let (X_1, Σ_1) be a measurable space which admits σ-finite measures μ_1 and ν_1 and (X_2, Σ_2) have σ-finite measures μ_2 and ν_2. Also let $P = (p_1, p_2)$ and $Q = (q_1, q_2)$, where $p_1, p_2, q_1, q_2 \in (0, \infty]$, and $\sigma, \tau \in S_2$. When is there a constant $C < \infty$ such that the below inequality holds for any measurable function f on $X_1 \times X_2$? And, in this case, what is the least value of C in $\|f\|_{\tau(Q)} \le C\,\|f\|_{\sigma(P)}$, i.e. the norm of the inclusion map $L^{\sigma(P)} \hookrightarrow L^{\tau(Q)}$?

If τ is not the identity, then relabeling so as to swap every "one" (p_1, q_1, $x_1 \in X_1$, μ_1, ν_1) with its corresponding "two" reduces to the case where τ is the identity and $\sigma \circ \tau^{-1}$ replaces σ, so that we only need consider

$$\|f\|_Q \le C\,\|f\|_{\sigma(P)}.$$

(It is evident that the existence of $C < \infty$ implies that $L^{\sigma(P)} \subset L^Q$. For the converse, see Proposition 7.3.2 which covers the multi-variable case, relying on Lemma 7.3.1 to reduce to the case where each $p_k, q_k \ge 1$ for $k = 1, 2$. Then, as observed by Benedek and Panzone in [4], our mixed-norm spaces are Banach function spaces, as defined in Bennett and Sharpley [5]. As proven in Theorem 1.8 of [5], any inclusion between Banach function spaces must be a bounded map, i.e. there must be such a $C < \infty$.)

Concretely, this breaks down into two cases, of which the first is the "unpermuted case" with τ as the identity and where the order of integration is the same on each side,

$$\left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} \le C\left\|\|f\|_{L^{p_1}_{\mu_1}(X_1)}\right\|_{L^{p_2}_{\mu_2}(X_2)}. \tag{3.1}$$

The other case is the "permuted case", with non-identity τ, where the inner and outer variables are reversed between the two sides,

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}. \tag{3.2}$$

3.3 One-variable inclusions are necessary

For either inequality (3.1) or (3.2) to hold for all measurable functions, it is necessary that both $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$ and $L^{p_2}_{\mu_2}(X_2) \subset L^{q_2}_{\nu_2}(X_2)$. This is most readily seen by considering factorable functions, i.e. those of the form $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. From these, it follows that each $\nu_k \ll \mu_k$, but proving this consequence first makes it easier to establish the one-variable inclusions.

Lemma 3.3.1. *When either (3.1) or (3.2) holds with $C < \infty$, each $\nu_k \ll \mu_k$ for $k = 1, 2$.*

Proof. If $\nu_1 \not\ll \mu_1$, then there must be a measurable $S_1 \subset X_1$ with $\mu_1 S_1 = 0$ and $\nu_1 S_1 > 0$. Let $f(x_1, x_2) = \chi_{S_1}(x_1)$ and observe that

$$\xi(f) = \left\|\chi_{S_1}\right\|_{L^{q_1}_{\nu_1}(X_1)} \|1\|_{L^{q_2}_{\nu_2}(X_2)} = (\nu_1 S_1)^{1/q_1} (\nu_2 X_2)^{1/q_2} > 0,$$

with the convention $(\nu_2 X_2)^0 = 1$. However,

$$\rho(f) = (\mu_1 S_1)^{1/p_1} (\mu_2 X_2)^{1/p_2} = 0,$$

since $\mu_1 S_1 = 0$. This means that there is no suitable constant $C < \infty$.

Similarly, if $\nu_2 \not\ll \mu_2$, then we can replace S_1 by X_1 and X_2 by some S_2 with $\mu_2 S_2 = 0$ and $\nu_2 S_2 > 0$, and proceed as above to prove that $C = \infty$. □

Again, since absolute continuity is necessary, it can be assumed.

Assumption For the rest of the two-variable problem, assume that $\nu_k \ll \mu_k$ for $k = 1, 2$.

Lemma 3.3.2. *When either (3.1) or (3.2) holds with $C < \infty$, there must be finite constants C_1 and C_2 such that, for each $k \in \{1, 2\}$ and any measurable function f_k on X_k,*

$$\|f_k\|_{L^{q_k}_{\nu_k}(X)} \leq C_k \|f_k\|_{L^{p_k}_{\mu_k}(X)}. \tag{3.3}$$

Proof. Suppose that $f(x_1, x_2) = f_1(x_1)f_2(x_2)$ for some measurable functions f_1 on X_1 and f_2 on X_2. For any such function and any exponent $p > 0$, $\left(\int_{X_1} |f(x_1, x_2)|^p \, d\mu_1(x_1)\right)^{1/p} = |f_2(x_2)| \left(\int_{X_1} |f(x_1)|^p \, d\mu_1(x_1)\right)^{1/p}$ and $\operatorname{ess\,sup}_{x_1} |f(x_1, x_2)| = |f_2(x_2)| \operatorname{ess\,sup}_{x_1} |f_1(x_1)|$. Similarly, $|f_1|$ factors out of any L^p norm computation over X_2. Therefore, both inequalities (3.1) and (3.2) reduce, for these factorable functions, to

$$\|f_1\|_{L^{q_1}_{\nu_1}(X_1)} \|f_2\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \|f_1\|_{L^{p_1}_{\mu_1}(X_1)} \|f_2\|_{L^{p_2}_{\mu_2}(X_2)}. \tag{3.4}$$

Because ν_1 is a non-zero, σ-finite measure, there is a measurable $E_1 \subset X_1$ such that $0 < \nu_1 E_1 < \infty$. As $\nu_1 \ll \mu_1$ by assumption, $\mu_1 E_1 > 0$, and by σ-finiteness there is a measurable $S_1 \subset E_1$ so that $0 < \mu_1 S_1 < \infty$, while still $0 < \nu_1 S_1 < \infty$. (Were there no subset $S_1 \subset E_1$ with $\mu_1 S_1 < \infty$ and $\nu_1 S_1 > 0$, then ν_1 would have to be zero on E_1.)

Following a similar procedure to that which established the existence of $S_1 \subset X_1$ with $0 < \mu_1(S_1) < \infty$ and $0 < \nu_1(S_1) < \infty$ proves that there is some $S_2 \subset X_2$ with $0 < \mu_2(S_2) < \infty$ and $0 < \nu_2(S_2) < \infty$.

For any measurable function f_1 on X_1, let $f_2 = \chi_{S_2}$ and inequality (3.4) gives

$$\left(\int_{X_1} |f_1(x_1)|^{q_1} d\nu_1(x_1)\right)^{1/q_1} \leq \frac{(\mu_2 S_2)^{1/p_2}}{(\nu_2 S_2)^{1/q_2}} C \left(\int_{X_1} |f_1(x_1)|^{p_1} d\mu_1(x_1)\right)^{1/p_1},$$

which establishes that the minimal constant C_1 in (3.3) satisfies $C_1 \leq (\mu_2 S_2)^{1/p_2}(\nu_2 S_2)^{-1/q_2} C < \infty$. Similarly, letting $f_1 = \chi_{S_1}$, for any measurable function f_2 on X_2,

$$\left(\int_{X_2} |f_2(x_2)|^{q_2} d\nu_2(x_2)\right)^{1/q_2} \leq \frac{(\mu_1 S_1)^{1/p_1}}{(\nu_1 S_1)^{1/q_1}} C \left(\int_{X_2} |f_2(x_2)|^{p_2} d\mu_2(x_2)\right)^{1/p_2},$$

thus the minimal constant $C_2 \leq (\mu_1 S_1)^{1/p_1}(\nu_1 S_1)^{-1/q_1} C$ is also finite. □

The following result includes the necessity of single-variable inclusions, but also provides a quantitative lower bound on the mixed-norm best constant, connecting it to the single-variable constants.

Proposition 3.3.3. *Let $C \in [0, \infty]$ denote the least constant for which either (3.1) or (3.2) is satisfied for any measurable f, and for $k = 1, 2$ let $C_k \in [0, \infty]$ denote the least constant so that for any measurable f_k on X_k*

$$\|f_k\|_{L^{q_k}_{\nu_k}(X_k)} \leq C_k \|f_k\|_{L^{p_k}_{\mu_k}(X_k)}. \tag{3.5}$$

Then $C \geq C_1 C_2$.

Proof. From Lemma 3.3.2, we know that if $C < \infty$, then each $C_k < \infty$. (Because no measure is zero, it is impossible for either C_k to be zero.) As in that lemma, for any measurable functions f_1 on X_1 and f_2 on X_2, let $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ and then, for such f, either inequality (3.1) or (3.2) reduces to

$$\|f_1\|_{L^{q_1}_{\nu_1}(X_1)} \|f_2\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \|f_1\|_{L^{p_1}_{\mu_1}(X_1)} \|f_2\|_{L^{p_2}_{\mu_2}(X_2)}.$$

When both C_1 and C_2 are finite, for any $\varepsilon > 0$ it is possible to choose functions f_k so that each

$$C_k - \varepsilon < \frac{\|f_k\|_{L^{q_k}_{\nu_k}(X_k)}}{\|f_k\|_{L^{p_k}_{\mu_k}(X_k)}} \leq C_k,$$

because C_k is the supremum of all such ratios when the denominator is nonzero. For such functions, inequality (3.2) yields

$$(C_1 - \varepsilon)(C_2 - \varepsilon) \leq C,$$

so $C_1 C_2 \leq C$ because this is true whenever $\varepsilon > 0$. □

3.4 Unpermuted case

In the unpermuted case described by inequality (3.1), the necessary condition that both single-variable inclusions $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$ hold turns out to be sufficient, as well, and the lower bound C_1C_2 is the best constant C.

Theorem 3.4.1. *For $k \in \{1,2\}$, let $C_k \in [0,\infty]$ denote the least constant such that, for any measurable function f_k on X_k,*

$$\|f_k\|_{L^{q_k}_{\nu_k}(X_k)} \le C_k \|f_k\|_{L^{p_k}_{\mu_k}(X_k)},$$

and let $C \in [0,\infty]$ denote the least constant such that, for any measurable function $f(x_1,x_2)$ on $X_1 \times X_2$,

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \le C \left\| \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{p_2}_{\mu_2}(X_2)}.$$

Then $C = C_1C_2$ (in particular, $C < \infty$ if and only if both $C_1 < \infty$ and $C_2 < \infty$.)

Proof. To see that $C \le C_1C_2$, i.e. that the two-variable inequality is valid with C_1C_2 in place of C, take an arbitrary measurable $f(x_1,x_2)$. For each particular value $x_2 \in X_2$, $|f(\cdot,x_2)|$ is a measurable function on X_1, so the $k=1$ inequality yields

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \le \left\| C_1 \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} = C_1 \left\| \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)}.$$

Furthermore, the inner $L^{p_1}_{\mu_1}(X_1)$ norm produces a measurable function of x_2, so the $k=2$ inequality applies.

$$C_1 \left\| \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \le C_1C_2 \left\| \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{p_2}_{\mu_2}(X_2)},$$

so $C \le C_1C_2$.

Of course, $C \ge C_1C_2$ regardless of permutation, from Proposition 3.3.3. □

Example Let $X_1 = [0,1]$ with μ_1 Lebesgue measure, $\nu_1 = x\mu_1$, $p_1 = 2$, and $q_1 = 1$. Let $X_2 = \mathbb{N}$ with $\mu_2 = \nu_2$ being counting measure, $p_2 = 1$ and $q_2 = 2$. Then the least constant C such that, for any measurable function $f(x,n)$ on $[0,1] \times \mathbb{N}$, $\|f\|_Q \le C\|f\|_P$, i.e.

$$\left(\sum_n \left(\int_0^1 |f(x,n)|\, x dx \right)^2 \right)^{1/2} \le C \sum_n \left(\int_0^1 |f(x,n)|^2\, dx \right)^{1/2}$$

is $C = \frac{1}{\sqrt{3}}$.

Proof. By Hölder's inequality, for any Lebesgue measurable function g on $[0,1]$ (since $\frac{1}{2} + \frac{1}{2} = 1$),

$$\int_0^1 |g(x)|\, x dx \le \left(\int_0^1 |g(x)|^2\, dx \right)^{1/2} \left(\int_0^1 x^2 dx \right)^{1/2} = \frac{1}{\sqrt{3}} \left(\int_0^1 |g(x)|^2\, dx \right)^{1/2}.$$

Since Hölder's inequality is sharp (Corollary 2.3.3), the least constant C_1 in $\|g\|_{L^1_{\nu_1}([0,1])} \leq C_1 \|g\|_{L^2_{\mu_1}([0,1])}$ is $\frac{1}{\sqrt{3}}$. Because $p_2 = 1 < 2 = q_2$, Corollary 2.4.4 applies using the common measure $\lambda = \mu_2 = \nu_2$, counting measure on $\mathbb{N}$. The least non-zero counting measure of any set is 1, which is the value of ε in that result, so $C_2 = 1^{1/q_2 - 1/p_2} = 1$. By Theorem 3.4.1, the least value of $C = C_1 C_2 = \frac{1}{\sqrt{3}}$.

Furthermore, in this case we can confirm that $C = \frac{1}{\sqrt{3}}$ is achieved as $\frac{\|f\|_Q}{\|f\|_P}$, for example with

$$f(x,n) = \begin{cases} x & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases}$$

For this function,

$$\frac{\|f\|_Q}{\|f\|_P} = \frac{\left(\sum_n \left(\int_0^1 |f(x,n)|\, x dx\right)^2\right)^{1/2}}{\sum_n \left(\int_0^1 |f(x,n)|^2\, dx\right)^{1/2}} = \frac{\int_0^1 x^2 dx}{\left(\int_0^1 x^2 dx\right)^{1/2}} = \frac{\frac{1}{3}}{\frac{1}{\sqrt{3}}} = \frac{1}{\sqrt{3}}$$

□

Because the unpermuted case has such a simple solution, the rest of this chapter covers the permuted case, described by inequality (3.2). The example at the end of Section 3.1 already shows that, for the permuted case, having both single-variable inclusions is not sufficient. (Recall that, in that example, the one-variable norms are identical on each side; however, the mixed norms differ in the order of summation and turn out to be inequivalent.)

3.5 Minkowski integral inequality

To understand the permuted case, Minkowski's integral inequality is an essential tool. Recall that it generalizes the triangle inequality for Lebesgue spaces, known as Minkowski's inequality, which itself is derived from Hölder's inequality. A statement of this well-known theorem follows here, without proof. Details can be found in many references, such as Folland's [15], where this is part (a) of Theorem 6.19, "Minkowski's Inequality for Integrals".

Theorem 3.5.1 (Minkowski integral inequality)**.** *Let (X,μ) and (Y,ν) be σ-finite measure spaces. For any $f \in L^+(X \times Y)$ and any $p \in [1,\infty)$,*

$$\left(\int_Y \left(\int_X f(x,y) d\mu(x)\right)^p d\nu(y)\right)^{1/p} \leq \int_X \left(\int_Y f(x,y)^p d\nu(y)\right)^{1/p} d\mu(x).$$

Each side of the inequality can be regarded as computing a mixed norm, with an implicit 1 exponent associated with the integral in x and the explicit p exponent for the integral in y, where the difference between sides is that the y integral moves to the inside on the right, its exponent coming with it. This is because $p \geq 1$, as we can see in the following simple generalization which replaces $p \geq 1$ and 1 by two exponents $p \geq q$; moving the integral with the greater exponent, p, to the inside yields a norm which always produces greater values.

Corollary 3.5.2. *Let (X,μ) and (Y,ν) be σ-finite measure spaces. For any $f \in L^+(X \times Y)$ and any $0 < q \le p \le \infty$,*

$$\left\| \|f\|_{L^q_\mu(X)} \right\|_{L^p_\nu(Y)} \le \left\| \|f\|_{L^p_\nu(Y)} \right\|_{L^q_\mu(X)}$$

Proof. Suppose that $p < \infty$. Let $g(x,y) = f(x,y)^q$ and $r = p/q$, which must be at least 1. Then we can apply the Minkowski integral inequality in

$$\begin{aligned}
\left(\int_Y \left(\int_X f(x,y)^q d\mu(x) \right)^{p/q} d\nu(y) \right)^{q/p} &= \left(\int_Y \left(\int_X g(x,y) d\mu(x) \right)^{r} d\nu(y) \right)^{1/r} \\
&\le \int_X \left(\int_Y g(x,y)^r d\nu(y) \right)^{1/r} d\mu(x) \\
&= \left(\int_X \left(\int_Y f(x,y)^p d\nu(y) \right)^{q/p} d\mu(x) \right)^{q/q}
\end{aligned}$$

and, by taking q^{th} roots, obtain the desired result.

When $q < p = \infty$, for ν-a.e. $y \in Y$

$$\left(\int_X f(x,y)^q d\mu(x) \right)^{1/q} \le \left(\int_X \left(\operatorname*{ess\,sup}_y f(x,y) \right)^q d\mu(x) \right)^{1/q},$$

so this inequality holds for the essential supremum in y,

$$\left\| \|f\|_{L^q_\mu(X)} \right\|_{L^\infty_\nu(Y)} \le \left\| \|f\|_{L^\infty_\nu(Y)} \right\|_{L^q_\mu(X)}.$$

Finally, in the case $p = q = \infty$, both sides are equal, simply the essential supremum of f on $(X,\mu) \times (Y,\nu)$. □

As an immediate consequence, if $P = (p,q)$ and σ is the non-identity permutation of two elements, then so long as $q \le p$, $L^P \subset L^{\sigma(P)}$, with an inclusion map of norm 1. (The inequality shows that $C = 1$ works, and it is easily verified to be the least constant by considering factorable functions $f = f_1(x)f_2(y)$, i.e. products of one-variable functions. For such f, each side of the inequality equals $\|f_1\|_{L^q_\mu(X)} \|f_2\|_{L^p_\nu(Y)}$.) Since the one-variable Lebesgue norms on each side match, the one-variable constants are $C_1 = C_2 = 1$, so along with the unpermuted case, this is another case where $C = C_1C_2$, the lower bound on C given in Proposition 3.3.3. However, the next section establishes a much more general sufficient condition, based on Minkowski's integral inequality, for mixed-norm inclusion with $C = C_1C_2$.

3.6 Minkowski's inequality sufficient condition

Recall that here and for the rest of the chapter we're studying the permuted two-variable case, seeking the least constant $C \in [0,\infty]$ in (3.2), which is repeated here:

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \le C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

Observe that the least $C \in [0, \infty]$ satisfying (3.2) can be characterized as a supremum of ratios over all measurable functions f on $X_1 \times X_2$ which are not almost everywhere zero,

$$C = \sup_{f \neq 0} \frac{\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)}}{\left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}}. \tag{3.6}$$

(As long as f is not $\mu_1 \times \mu_2$-a.e. zero, the denominator cannot be zero.)

The mixed-norm form of Minkowski's integral inequality immediately provides certain sufficient conditions for $C < \infty$ (that is, for the mixed-norm inclusion $L^{p_1}_{\mu_1}(L^{p_2}_{\mu_2}) \subset L^{q_2}_{\nu_2}(L^{q_1}_{\nu_1})$), depending (outside of the necessary per-variable inclusions) only on the exponents.

Theorem 3.6.1 (Minkowski sufficient condition)**.** *Assume the necessary one-variable inclusions, i.e. that for each $k \in \{1, 2\}$ there is a least possible constant C_k such that, for any measurable f_k on X_k, $\|f_k\|_{L^{q_k}_{\nu_k}(X_k)} \leq C_k \|f_k\|_{L^{p_k}_{\mu_k}(X_k)}$. If $\min(p_1, q_1) \leq \max(p_2, q_2)$, then inequality (3.2) is satisfied with the least possible $C = C_1 C_2$.*

Proof. Considering factorable functions $f(x_1, x_2) = g(x_1)h(x_2)$, we find (as in Theorem 3.4.1) that $C \geq C_1 C_2$. For the reverse inequality, there are four cases in which to prove inequality (3.2) is satisfied with C replaced by $C_1 C_2$.

- Case 1: $p_1 \leq p_2$

  The one-variable inclusions, followed by Minkowski's integral inequality (Theorem 3.5.2), establish that

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq C_2 \left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{p_2}_{\mu_2}(X_2)} \leq C_2 \left\| C_1 \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{p_2}_{\mu_2}(X_2)} \\ &\leq C_1 C_2 \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.\end{aligned}$$

- Case 2: $q_1 \leq p_2$

  Using the one-variable inclusion $\|g\|_{L^{q_2}_{\nu_2}(X_2)} \leq C_2 \|g\|_{L^{p_2}_{\mu_2}(X_2)}$ with $g = \|f\|_{L^{q_1}_{\nu_1}(X_1)}$, followed by Minkowski's integral inequality,

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq C_2 \left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{p_2}_{\mu_2}(X_2)} \\ &\leq C_2 \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{q_1}_{\nu_1}(X_1)}.\end{aligned}$$

  Finally, applying the one-variable inclusion in X_1 yields

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C_1 C_2 \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

- Case 3: $p_1 \leq q_2$

This time, it's inclusion in the first variable, Minkowski, and finally inclusion in the second variable.

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq \left\| C_1 \|f\|_{L^{p_1}_{\mu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\ &\leq C_1 \left\| \|f\|_{L^{q_2}_{\nu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &\leq C_1 C_2 \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}\end{aligned}$$

- Case 4: $q_1 \leq q_2$

  Minkowski's integral inequality provides that $\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq \left\| \|f\|_{L^{q_2}_{\nu_2}(X_2)} \right\|_{L^{q_1}_{\nu_1}(X_1)}$. The rest reduces to the unpermuted case, where both one-variable inclusions take this to the conclusion

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C_1 C_2 \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)} .$$

□

Chapter 4

Two-variable non-Minkowski case

4.1 Non-Minkowski case with common measures

4.1.1 Measure summability condition on blocks

Having found in Theorem 3.6.1 that the Minkowski condition $\min(p_1, q_1) \leq \max(p_2, q_2)$ is sufficient for mixed norm inclusion with $C = C_1C_2$ (assuming the necessary one-variable inclusions), it remains only to consider the non-Minkowski permuted case, Inequality (3.2) with $\max(p_2, q_2) < \min(p_1, q_1)$.

Although this non-Minkowski case splits into several subcases depending on properties of the measures involved, this section treats the simplest situation, where all exponents are finite (for which it is sufficient that $\max(p_1, q_1) < \infty$) and, rather than separate measures μ_k and ν_k on each X_k, they are replaced by common measures λ_k. Although this seems rather restrictive, in most cases there are methods to reduce the problem to common measures, as seen in Section 4.2.1. Where this does not work, the basic arguments presented here will be generalized in Section 4.2.

For this introductory section, the case where both measures λ_k are purely atomic is also omitted (although one may be purely atomic), as this is a particularly complex case which merits its own section.

Definition 4.1.1. Whenever $\max(p_2, q_2) < \min(p_1, q_1)$, define

$$\alpha = \frac{q_1^{-1} - p_1^{-1}}{q_2^{-1} - p_1^{-1}} \qquad \text{and} \qquad \beta = \frac{q_2^{-1} - p_2^{-1}}{q_2^{-1} - p_1^{-1}}, \tag{4.1}$$

with the convention that $\infty^{-1} = 0$.

Observe that, when $p_1 < \infty$ and $q_1 < \infty$,

$$\alpha = \frac{q_2(p_1 - q_1)}{q_1(p_1 - q_2)} \qquad \text{and} \qquad \beta = \frac{p_1(p_2 - q_2)}{p_2(p_1 - q_2)},$$

though this also applies with $p_1 = \infty$ or $q_1 = \infty$, if understood with appropriate limits.

Proposition 4.1.2. *Suppose that* $\max(p_2, q_2) < \min(p_1, q_1)$ *and each* $L^{p_k}_{\mu_k} \subset L^{q_k}_{\nu_k}$. *Then*

1. $\alpha < 1, \beta < 1,$ *and* $\alpha + \beta < 1$.

2. If ν_1 *is not purely atomic then* $\alpha \geq 0$.

3. If ν_2 *is not purely atomic then* $\beta \geq 0$.

Proof. Because $q_2 < p_1$, the denominator $q_2^{-1} - p_1^{-1}$ is always positive. Additionally, $q_2 < q_1$, so $q_1^{-1} - p_1^{-1} < q_2^{-1} - p_1^{-1}$; dividing both sides by $q_2^{-1} - p_1^{-1}$ then yields $\alpha < 1$. A similar argument based on $p_2 < p_1$ shows that $\beta < 1$. Finally,

$$\alpha + \beta = \frac{q_1^{-1} + q_2^{-1} - p_1^{-1} - p_2^{-1}}{q_2^{-1} - p_1^{-1}} = 1 + \frac{q_1^{-1} - p_2^{-1}}{q_2^{-1} - p_1^{-1}}.$$

Of course, though the denominator is positive, $q_1^{-1} - p_2^{-1} < 0$, since $p_2 < q_1$.

Statements 2 and 3 are consequences of Corollary 2.5.7. If ν_1 is not purely atomic, then $p_1 \geq q_1$, so $q_1^{-1} - p_1^{-1} \geq 0$. Similarly, if ν_2 is not purely atomic, then $p_2 \geq q_2$, so $q_2^{-1} - p_2^{-1} \geq 0$. □

The first use of these exponents is in the following necessary condition based on the properties of the measures λ_k, which also gives a lower bound on the least constant C.

Proposition 4.1.3. *If max*$(p_2, q_2) < \min(p_1, q_1)$ *(i.e. the Minkowski sufficient condition does not apply),* $\max(p_1, q_1) < \infty$*, and there is some* C *such that for all measurable functions* f *on* $X_1 \times X_2$

$$\left\| \|f\|_{L^{q_1}_{\lambda_1}(X_1)} \right\|_{L^{q_2}_{\lambda_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\lambda_2}(X_2)} \right\|_{L^{p_1}_{\lambda_1}(X_1)}$$

then, for any disjoint sequence (A_i) *of measurable subsets of* X_1 *and any disjoint sequence* (B_i) *of measurable subsets of* X_2*,*

$$\left(\sum_i a_i^{\alpha} b_i^{\beta} \right)^{\frac{1}{q_2} - \frac{1}{p_1}} \leq C$$

where $a_i = \lambda_1(A_i)$ *and* $b_i = \lambda_2(B_i)$*. Because* $\frac{1}{q_2} - \frac{1}{p_1} > 0$ *in the non-Minkowski case, as a consequence it is necessary for* $C < \infty$ *not only that every such* $\sum_i a_i^{\alpha} b_i^{\beta}$ *converge, but that the collection of such series be bounded above.*

Proof. The idea is to use combinations $\sum_i c_i \chi_{E_i \times F_i}$ of block characteristic functions. Consider any such disjoint sequences (A_i) and (B_i). Define $c_i = a_i^{\frac{q_2 - q_1}{q_1(p_1 - q_2)}} b_i^{\frac{p_2 - p_1}{p_2(p_1 - q_2)}}$ and, for each N, a function $f_N(x_1, x_2) = \sum_{i\le N} c_i \chi_{A_i}(x_1)\chi_{B_i}(x_2)$. Then

$$\begin{aligned}
\frac{\left\| \|f_N\|_{L^{q_1}_{\lambda_1}(X_1)} \right\|_{L^{q_2}_{\lambda_2}(X_2)}}{\left\| \|f_N\|_{L^{p_2}_{\lambda_2}(X_2)} \right\|_{L^{p_1}_{\lambda_1}(X_1)}} &= \frac{\left(\int_{X_2}\left(\int_{X_1}\sum_{i\le N} c_i^{q_1}\chi_{A_i}(x_1)d\lambda_1(x_1)\right)^{\frac{q_2}{q_1}}\chi_{B_i}(x_2)d\lambda_2(x_2)\right)^{\frac{1}{q_2}}}{\left(\int_{X_1}\left(\int_{X_2}\sum_{i\le N} c_i^{p_2}\chi_{B_i}(x_2)d\lambda_2(x_2)\right)^{\frac{p_1}{p_2}}\chi_{A_i}(x_1)d\lambda_1(x_1)\right)^{\frac{1}{p_1}}} \\
&= \frac{\left(\sum_{i\le N} c_i^{q_2}\lambda_1(A_i)^{\frac{q_2}{q_1}}\lambda_2(B_i)\right)^{\frac{1}{q_2}}}{\left(\sum_{i\le N} c_i^{p_1}\lambda_2(B_i)^{\frac{p_1}{p_2}}\lambda_1(A_i)\right)^{\frac{1}{p_1}}} \\
&= \frac{\left(\sum_{i\le N} a_i^{\frac{q_2(q_2-q_1)}{q_1(p_1-q_2)}} b_i^{\frac{q_2(p_2-p_1)}{p_2(p_1-q_2)}} a_i^{\frac{q_2}{q_1}} b_i\right)^{\frac{1}{q_2}}}{\left(\sum_{i\le N} a_i^{\frac{p_1(q_2-q_1)}{q_1(p_1-q_2)}} b_i^{\frac{p_1(p_2-p_1)}{p_2(p_1-q_2)}} a_i b_i^{\frac{p_1}{p_2}}\right)^{\frac{1}{p_1}}} \\
&= \left(\sum_{i\le N} a_i^{\alpha} b_i^{\beta}\right)^{\frac{1}{q_2}-\frac{1}{p_1}}.
\end{aligned} \tag{4.2}$$

Naturally, the original left-hand side is at most C, while the supremum over N of the last expression is $\left(\sum_i a_i^{\alpha} b_i^{\beta}\right)^{\frac{1}{q_2}-\frac{1}{p_1}}$. □

Although the proof is complete as given, if one is curious about the source of the coefficients c_i, they are computed so as to obtain equality in Hölder's inequality. Specifically, the greatest possible value of the q_2 power of the expression (4.2),

$$\frac{\sum_{i\le N} c_i^{q_2}\lambda_1(A_i)^{\frac{q_2}{q_1}}\lambda_2(B_i)}{\left(\sum_{i\le N} c_i^{p_1}\lambda_1(A_i)\lambda_2(B_i)^{\frac{p_1}{p_2}}\right)^{\frac{q_2}{p_1}}} = \frac{\sum_{i\le N}\left(c_i^{q_2} a_i^{\frac{q_2}{p_1}} b_i^{\frac{q_2}{p_2}}\right) a_i^{q_2(q_1^{-1}-p_1^{-1})} b_i^{q_2(q_2^{-1}-p_2^{-1})}}{\left\| c_i^{q_2} a_i^{\frac{q_2}{p_1}} b_i^{\frac{q_2}{p_2}} \right\|_{\ell^{\frac{p_1}{q_2}}}},$$

is the $\ell^{\frac{p_1}{p_1-q_2}}$ norm (the conjugate $\left(\frac{p_1}{q_2}\right)' = \frac{p_1}{p_1-q_2}$) of the other factor,

$$\left\| a_i^{q_2(q_1^{-1}-p_1^{-1})} b_i^{q_2(q_2^{-1}-p_2^{-1})} \right\|_{\ell^{\frac{p_1}{p_1-q_2}}} = \left(\sum_{i\le N} a_i^{\alpha} b_i^{\beta}\right)^{q_2\left(\frac{1}{q_2}-\frac{1}{p_1}\right)},$$

by Corollary 2.3.4 applied with $f(i) = c_i^{q_2} a_i^{\frac{q_2}{p_1}} b_i^{\frac{q_2}{p_2}}$ and $g(i) = a_i^{q_2(q_1^{-1}-p_1^{-1})} b_i^{q_2(q_2^{-1}-p_2^{-1})}$. As noted in Hölder's inequality, Theorem 2.3.2, this value is achieved when $f(i)^{\frac{p_1}{q_2}}$ is a scalar multiple of $g(i)^{\frac{p_1}{p_1-q_2}}$. Solving $f(i)^{\frac{p_1}{q_2}} = kg(i)^{\frac{p_1}{p_1-q_2}}$ for c_i produces the formula for coefficients given above, $c_i = a_i^{\frac{q_2-q_1}{q_1(p_1-q_2)}} b_i^{\frac{p_2-p_1}{p_2(p_1-q_2)}}$ (with $k = 1$), or a constant multiple of it. (Any extra constant factor is cancelled in the ratio, because norms are homogeneous, i.e. $\|cf\| = |c|\,\|f\|$.)

4.1.2 Necessity of the Minkowski criterion for non-atomic measures

Although in general the criterion based on Minkowski's inequality and described in Section 3.6 is only a sufficient condition, it turns out to be necessary, as well, when neither λ_1 nor λ_2 is a purely atomic measure. The flexibility afforded by measures which are not atomic allows the construction of a counterexample in any non-Minkowski permuted case. A review of facts from measure theory precedes the main result. First, recall these definitions from Section 2.5.

Definition 2.5.1 A measurable subset A of a measure space (X, μ) is called an *atom* when $\mu A > 0$ and, for any measurable $F \subset X$, either $\mu(A \cap F) = 0$ or $\mu(A \setminus F) = 0$.

(From Proposition 2.5.2, we know this is equivalent to the requirement that $\mu A > 0$ and, for any disjoint decomposition $A = E \cup F$, either $\mu E = 0$ or $\mu F = 0$.)

Definition 2.5.3 A measure is called *purely atomic* when every measurable set of positive measure contains an atom. At the other extreme, it is *atomless* when it has no atom, i.e. every subset of positive measure can be expressed as a disjoint union of two sets of positive measure.

The following theorem regarding atomless measures was originally proven by Sierpiński in 1922. It is given without proof, but can be found in texts on measure theory, for example Corollary 1.12.10 in Bogachev [11].

Theorem 4.1.4. *If (X, μ) is an atomless measure space, then for any real number $\alpha \in [0, \mu X]$, there is some measurable $F \subset X$ with $\mu F = \alpha$.*

It has the following corollary, helpful to construct counterexamples.

Corollary 4.1.5. *If (X, μ) is a measure space with a measurable subset E, $\mu E > 0$ which contains no μ-atom, and (c_k) is a sequence (finite or infinite) of nonnegative real numbers with $\sum_k c_k = \mu E$, then X can be partitioned into a pairwise disjoint sequence (F_k), with as many terms as (c_k), of measurable subsets of E such that each $\mu F_k = c_k$.*

Proof. First, observe that the restriction $\mu|_E$ of μ to E, which applies to the σ-algebra of measurable subsets of E, is an atomless measure. Were this not so, there would be an atom A of $\mu|_E$, i.e. a measurable subset $A \subset E$ such that, for any measurable $F \subset E$, either $\mu(A \cap F) = 0$ or $\mu(A \setminus F) = 0$. Therefore this set A would also be an atom for the original measure μ, on X. (For any measurable $S \subset X$, note that $E \cap S$ is a measurable subset of E. Because A would be an atom for $\mu|_E$, and recalling that $A \subset E$, either $0 = \mu(A \cap E \cap S) = \mu(A \cap S)$ or $0 = \mu(A \setminus (E \cap S)) = \mu((A \setminus E) \cup (A \setminus S)) = \mu(A \setminus S)$.) By hypothesis, however, E contains no μ-atom, so $\mu|_E$ must be an atomless measure; work with it rather than μ.

Now, find sets F_k one-by-one, establishing by induction on n that, when we have just added F_n, the sets $F_1, \ldots, F_n$ are pairwise disjoint and, for each $1 \leq k \leq n$, $\mu F_k = c_k$. Because $c_1 \leq \sum_k c_k = \mu E$, Sierpiński's theorem (Theorem 4.1.4) guarantees that there is some $F_1 \subset E$ with $\mu F_1 = c_1$.

For the inductive step, suppose that $F_1, \ldots, F_n$ are pairwise disjoint subsets of E, each $\mu F_k = c_k$. If the sequence (c_k) has only n terms, then this process is finished. Otherwise, because

$$c_{n+1} \leq \sum_{k>n} c_k = \mu E - \sum_{k=1}^{n} c_k = \mu\left(E \setminus \bigcup_{k=1}^{n} F_k\right),$$

there is a measurable $F_{n+1} \subset E \setminus (F_1 \cup \cdots \cup F_n)$ with $\mu F_{n+1} = c_{n+1}$, so that we now have pairwise disjoint subsets $F_1, \ldots, F_{n+1}$ of E with each $\mu F_k = c_k$ for $k \in \{1, \ldots, n+1\}$.

Thus we reach a sequence (F_k) of measurable subsets of E with each $\mu F_k = c_k$. These sets are pairwise disjoint; it is sufficient to prove that arbitrarily finitely many of them are pairwise disjoint, as the induction shows. Finally, to obtain a partition of X, note that by countable additivity,

$$\mu\left(\bigcup_k F_k\right) = \sum_k \mu F_k = \sum_k c_k = \mu X.$$

Therefore $\mu(X \setminus \bigcup_k F_k) = 0$. Add the null set $X \setminus \bigcup_k F_k$ to any one of the sets F_k, not changing its measure. This preserves pairwise disjointness and the property that each $\mu F_k = c_k$, and now provides that $X = \bigcup_k F_k$, in a disjoint union. □

Now, this is enough to prove that, when neither λ_k measure is purely atomic, the permuted mixed-norm inclusion $L^{\sigma(P)} \subset L^Q$ is true if and only if the exponents satisfy the Minkowski condition $\min(p_1, q_1) \leq \max(p_2, q_2)$ and both one-variable inclusions $L^{p_k}_{\lambda_k}(X_k) \subset L^{q_k}_{\lambda_k}(X_k)$ ($k \in \{1, 2\}$) are true. (Recall that this mixed-norm inclusion is equivalent to the existence of a $C < \infty$ such that inequality (3.2) is true, with $\mu_k = \nu_k = \lambda_k$, for any measurable function f on $X_1 \times X_2$.)

Having already established in Section 3.6 that one-variable inclusions plus the Minkowski condition on exponents are sufficient for permuted mixed-norm inclusion, it now remains to refute mixed-norm inclusion when the Minkowski condition fails. Given the non-Minkowski case $\max(p_2, q_2) < \min(p_1, q_1)$, the (always necessary) one-variable inclusions, and that neither λ_k measure is purely atomic, the aim is to produce a counterexample where the mixed-norm inclusion fails. Note that, as with the rest of the section, the hypotheses include $\max(p_1, q_1) < \infty$, but Theorem 4.2.16 shows that this result holds more generally.

The key to this argument is working on subsets of each X_k which contain no atom for λ_k, and therefore where the measures' restrictions are atomless. On these sets, Corollary 4.1.5 allows the production of subsets of various desired measures, from which a function in $L^{\sigma(P)}$ but not in L^Q can be obtained.

Theorem 4.1.6. *Assume that neither λ_1 nor λ_2 is purely atomic. If* $\max(p_2, q_2) < \min(p_1, q_1)$ *and* $\max(p_1, q_1) < \infty$, *then there is no constant $C < \infty$ such that, for every measurable function f on $X_1 \times X_2$,*

$$\left\| \|f\|_{L^{q_1}_{\lambda_1}(X_1)} \right\|_{L^{q_2}_{\lambda_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\lambda_2}(X_2)} \right\|_{L^{p_1}_{\lambda_1}(X_1)}.$$

Proof. Since neither λ_k is purely atomic, for each there is a measurable subset $E_k \subset X_k$ with positive λ_k measure which contains no λ_k atom. Because λ_k is σ-finite, we can choose E_k such that $0 < \lambda_k(E_k) < \infty$.

When considering functions f supported on $E_1 \times E_2$, inequality (3.2) with $\mu_k = \nu_k = \lambda_k$ implies that

$$\left\| \|f\|_{L^{q_1}_{\lambda_1}(E_1)} \right\|_{L^{q_2}_{\lambda_2}(E_2)} \leq C \left\| \|f\|_{L^{p_2}_{\lambda_2}(E_2)} \right\|_{L^{p_1}_{\lambda_1}(E_1)}.$$

So, to disprove mixed-norm inclusion, we need only consider functions on $E_1 \times E_2$, where the restricted measures $\lambda_k|_{E_k}$ are atomless. Proposition 4.1.3 shows that it suffices to find disjoint sequences (A_i) of measurable subsets of E_1 and (B_i) of measurable subsets of E_2 such that, letting each $a_i = \lambda_1(A_i)$ and $b_i = \lambda_2(B_i)$, $\sum_i a_i^\alpha b_i^\beta = \infty$.

Proposition 4.1.2 establishes that, in this non-Minkowski case of $\max(p_2, q_2) < \min(p_1, q_1)$, $\alpha + \beta < 1$. Furthermore, because neither λ_1 nor λ_2 is purely atomic, $\alpha \geq 0$ and $\beta \geq 0$, so $0 \leq \alpha + \beta < 1$.

When $0 < \alpha + \beta < 1$, the series $\sum_{m\geq 1} m^{-1/(\alpha+\beta)}$ converges; let $M = \sum_{m\geq 1} m^{-1/(\alpha+\beta)}$. For each $i \geq 1$, let $a_i = M^{-1}\lambda_1(E_1)m^{-1/(\alpha+\beta)}$ and $b_i = M^{-1}\lambda_2(E_2)m^{-1/(\alpha+\beta)}$, so that $\sum_i a_i = \lambda_1(E_1)$ and $\sum_i b_i = \lambda_2(E_2)$. Because each E_k contains no atom of λ_k, by Corollary 4.1.5 there is a pairwise disjoint sequence (A_i) of measurable subsets of E_1 with each $\lambda_1 A_i = a_i$ and there is a pairwise disjoint sequence (B_i) of measurable subsets of E_2 with each $\lambda_2 B_i = b_i$. As desired, $\sum_{i\geq 1} a_i^\alpha b_i^\beta = M^{-\alpha-\beta} (\lambda_1 E_1)^\alpha (\lambda_2 E_2)^\beta \sum_{i\geq 1} m^{-1} = \infty$.

Finally, when $\alpha + \beta = 0$ (i.e. $\alpha = \beta = 0$), let $a_i = \lambda_1(E_1)2^{-i}$ and $b_i = \lambda_2(E_2)2^{-i}$, in which case again $\sum_{i\geq 1} a_i = \lambda_1(E_1)$ and $\sum_{i\geq 1} b_i = \lambda_2(E_2)$, so there are again disjoint sequences (A_i), (B_i) with each $A_i \subset E_1$, $a_i = \lambda_1(A_i)$ and $B_i \subset E_2$, $b_i = \lambda_2(B_i)$. Now, $\sum_{i\geq 1} a_i^\alpha b_i^\beta = \sum_{i\geq 1} 1 = \infty$. □

4.1.3 Two-variable permuted case, one purely atomic measure

When at least one space is purely atomic, permuted mixed-norm inclusion is possible even outside of the Minkowski case. We're still considering common measures, where a single λ_k takes the place of μ_k and ν_k, for each $k \in \{1, 2\}$, the non-Minkowski case $\max(p_2, q_2) < \min(p_1, q_1)$, and supposing that $\max(p_1, q_1) < \infty$. (In the non-Minkowski case, this also implies that $p_2 < \infty$ and $q_2 < \infty$.) The question is whether $L^{\sigma(P)} \subset L^Q$, where σ denotes the unique non-identity permutation of $\{1, 2\}$, and, if so, finding the norm of the inclusion map. Based on the results in Section 2.6, the purely atomic factor will be taken to be an at most countable index set I, with weighted counting measure. (The weights are the measures of atoms in the original purely atomic space.)

Proposition 4.1.7. *Suppose that* $\max(p_1, q_1) < \infty$. *If* λ_2 *is purely atomic but* λ_1 *is not, then* $L^{\sigma(P)} \subset L^Q$ *(with non-identity* σ*) if and only if each* $L^{p_k}_{\lambda_k}(X_k) \subset L^{q_k}_{\lambda_k}(X_k)$ *for* $k = 1, 2$ *and one of these conditions is true:*

1. $\min(p_1, q_1) \leq \max(p_2, q_2)$
2. *the measures of all* λ_2 *atoms are* $\frac{\beta}{1-\alpha}$*-summable*

In the first case, the least constant $C = C_1C_2$*, the product of the one-variable constants. In the second,* $C_1C_2 \leq C \leq C_1 \|(v_i)\|_{\ell^{\frac{\beta}{1-\alpha}}}^{q_2^{-1}-p_2^{-1}}$*, where* (v_i) *denotes the sequence of measures of atoms of* λ_2.

Recall that, as defined in Section 4.1.1, $\alpha = \left(q_1^{-1} - p_1^{-1}\right) / \left(q_2^{-1} - p_1^{-1}\right)$ and $\beta = \left(q_2^{-1} - p_2^{-1}\right) / \left(q_2^{-1} - p_1^{-1}\right)$, so that $\frac{\beta}{1-\alpha} = \left(q_2^{-1} - p_2^{-1}\right) / \left(q_2^{-1} - q_1^{-1}\right)$.

Proof. Factorable functions provide the lower bound $C_1C_2 \leq C$, as shown in Proposition 3.3.3. In the first case, the Minkowski sufficient condition for inclusion, demonstrated in Section 3.6 gives $C \leq C_1C_2$. For the rest of the proof, assume this condition fails, so that $\max(p_2, q_2) < \min(p_1, q_1)$. Because (X_2, λ_2) is a σ-finite, purely atomic measure space, as shown in Section 2.6 it can be replaced by the natural numbers, with weighted counting measure. Let these weights be denoted by (v_i), each v_i the measure of an atom in X_2. If there are only finitely many atoms, the tail of the sequence (v_i) can be filled with zeroes.

The following computation proves inclusion when (v_i) is $\frac{\beta}{1-\alpha}$-summable, i.e. $\sum_i v_i^{\frac{\beta}{1-\alpha}} < \infty$. For any measurable function $f(x,i) \geq 0$ of $x \in X_1$ and $i \geq 1$,

$$\|f\|_Q^{q_2} = \sum_i \left(\int_{X_1} f(x,i)^{q_1} d\lambda_1(x) \right)^{\frac{q_2}{q_1}} v_i = \sum_i \left(\int_{X_1} f(x,i)^{q_1} d\lambda_1(x) \right)^{\frac{q_2}{q_1}} v_i^{\frac{q_2}{p_2}} v_i^{1-\frac{q_2}{p_2}}.$$

By Hölder's inequality using the conjugate exponents $\frac{q_1}{q_2}$ and $\frac{q_1}{q_1-q_2}$, recalling that $q_2 < q_1$ in the non-Minkowski case,

$$\begin{aligned} \|f\|_Q^{q_2} &\leq \left(\sum_i \left(\int_{X_1} f(x,i)^{q_1} d\lambda_1(x) \right) v_i^{\frac{q_1}{p_2}} \right)^{\frac{q_2}{q_1}} \left(\sum_i v_i^{\frac{p_2-q_2}{p_2} \cdot \frac{q_1}{q_1-q_2}} \right)^{\frac{q_1-q_2}{q_1}} \\ &= \left(\int_{X_1} \sum_i f(x,i)^{q_1} v_i^{\frac{q_1}{p_2}} d\lambda_1(x) \right)^{\frac{q_2}{q_1}} \left(\sum_i v_i^{\frac{p_2-q_2}{p_2} \cdot \frac{q_1}{q_1-q_2}} \right)^{\frac{q_1-q_2}{q_1}}, \end{aligned}$$

where the equality is due to Tonelli's theorem, allowing the order of integration and summation to be reversed. Next, because $\frac{q_1}{p_2} > 1$ in the non-Minkowski case, for any sequence (c_i) the series $\sum_i c_i^{q_1/p_2} \leq (\sum_i c_i)^{q_1/p_2}$. Apply that here with $c_i = f(x,i)^{p_2} v_i$:

$$\|f\|_Q^{q_2} \leq \left(\int_{X_1} \left(\sum_i f(x,i)^{p_2} v_i \right)^{\frac{q_1}{p_2}} d\lambda_1(x) \right)^{\frac{q_2}{q_1}} \left(\sum_i v_i^{\frac{p_2-q_2}{p_2} \cdot \frac{q_1}{q_1-q_2}} \right)^{q_2\left(q_2^{-1}-q_1^{-1}\right)}.$$

Taking the q_2 root, this is

$$\|f\|_Q \leq \left(\int_{X_1} \left(\sum_i f(x,i)^{p_2} v_i \right)^{\frac{q_1}{p_2}} d\lambda_1(x) \right)^{\frac{1}{q_1}} \left(\sum_i v_i^{\frac{\beta}{1-\alpha}} \right)^{q_2^{-1}-q_1^{-1}}.$$

The final step uses the one-variable inclusion $L_{\lambda_1}^{p_1}(X_1) \subset L_{\lambda_1}^{q_1}(X_1)$, with the norm of that inclusion represented as usual by C_1,

$$\begin{aligned} \|f\|_Q &\leq C_1 \left(\int_{X_1} \left(\sum_i f(x,i)^{p_2} v_i \right)^{\frac{p_1}{p_2}} d\lambda_1(x) \right)^{\frac{1}{p_1}} \left(\sum_i v_i^{\frac{\beta}{1-\alpha}} \right)^{q_2^{-1}-q_1^{-1}} \\ &= C_1 \left(\sum_i v_i^{\frac{\beta}{1-\alpha}} \right)^{q_2^{-1}-q_1^{-1}} \|f\|_{\sigma(P)} = C_1 \|(v_i)\|_{\ell^{\frac{\beta}{1-\alpha}}}^{q_2^{-1}-p_2^{-1}} \|f\|_{\sigma(P)}. \end{aligned}$$

On the other hand, if (v_i) is not $\frac{\beta}{1-\alpha}$-summable, we can disprove inclusion with Proposition 4.1.3, by obtaining appropriate sets so that $\sum_n a_n^\alpha b_n^\beta$ is arbitrarily large. The divergence of the series $\sum_{i \in I} v_i^{\beta/(1-\alpha)}$ implies that the index set I is infinite; being countable, it has an enumeration $(i_n)_{n \geq 1}$. In the purely atomic factor, we simply let each B_n consist of exactly one atom, $b_n = v_{i_n}$.

Since λ_1 is not purely atomic, it has a subset A with $\lambda_1(A) > 0$ which contains no atom. For each $N \geq 1$, define a sequence $a_N(1)$ to $a_N(N)$ by

$$a_N(n) = \frac{v_{i_n}^{\frac{\beta}{1-\alpha}}}{\sum_{n=1}^{N} v_{i_n}^{\frac{\beta}{1-\alpha}}} \lambda_1(A).$$

Using Corollary 4.1.5, because λ_1 restricted to A is atomless and $\sum_{n=1}^{N} a_N(n) = \lambda_1(A)$, there are disjoint subsets $A_N(1), \ldots, A_N(n)$ of A with each $\lambda_1(A_N(n)) = a_N(n)$.

$$\begin{aligned}\sum_{n=1}^{N} a_N(n)^\alpha b_n^\beta &= \frac{\sum_{n=1}^{N} v_{i_n}^{\frac{\alpha\beta}{1-\alpha}} v_{i_n}^{\beta}}{\left(\sum_{i=1}^{I} v_{i_n}^{\frac{\beta}{1-\alpha}}\right)^{\alpha}} \\ &= \left(\sum_{n=1}^{N} v_{i_n}^{\frac{\beta}{1-\alpha}}\right)^{1-\alpha}\end{aligned}$$

Recalling that $\alpha < 1$, for large N this sum becomes arbitrarily large. Proposition 4.1.3 then shows that the least $C = \infty$, i.e. $L^{\sigma(P)} \not\subset L^Q$. □

Proposition 4.1.8. *Suppose that* $\max(p_1, q_1) < \infty$. *If* λ_1 *is purely atomic but* λ_2 *is not, then* $L^{\sigma(P)} \subset L^Q$ *(with non-identity* σ*) if and only if each* $L_{\lambda_k}^{p_k}(X_k) \subset L_{\lambda_k}^{q_k}(X_k)$, *for* $k = 1, 2$, *and one of these conditions is true:*

1. $\min(p_1, q_1) \leq \max(p_2, q_2)$
2. *the measures of all* λ_1 *atoms are* $\frac{\alpha}{1-\beta}$*-summable*

In the first case, the least constant $C = C_1C_2$, *the product of the one-variable constants. In the second,* $C_1C_2 \leq C \leq C_2 \|(u_i)\|_{\ell^{\frac{\alpha}{1-\beta}}}^{q_1^{-1}-p_1^{-1}}$, *where* (u_i) *denotes the sequence of measures of atoms of* λ_1.

With $\alpha = \left(q_1^{-1} - p_1^{-1}\right) / \left(q_2^{-1} - p_1^{-1}\right)$ and $\beta = \left(q_2^{-1} - p_2^{-1}\right) / \left(q_2^{-1} - p_1^{-1}\right)$, $\frac{\alpha}{1-\beta} = \left(q_1^{-1} - p_1^{-1}\right) / \left(p_1^{-1} - p_2^{-1}\right)$.

Proof. Factorable functions provide the lower bound $C_1C_2 \leq C$, as shown in Proposition 3.3.3. In the first case, the Minkowski sufficient condition for inclusion, demonstrated in Section 3.6 gives $C \leq C_1C_2$. For the rest of the proof, assume this condition fails and that $\max(p_2, q_2) < \min(p_1, q_1)$.

As a σ-finite purely atomic measure space, (X_1, λ_1) has countably many atoms, with measures to be enumerated as (u_i). Represent functions on X_1 by functions on $\{1, 2, \ldots\}$, with integrals as series weighted by (u_i). In these terms, if (u_i) is $\frac{\alpha}{1-\beta}$-summable, for any measurable function $f(i, y)$ with $y \in X_2$,

$$\|f\|_Q = \left(\int_{X_2} \left(\sum_i f(i,y)^{q_1} u_i\right)^{\frac{q_2}{q_1}} d\lambda_2(y)\right)^{\frac{1}{q_2}} \leq C_2 \left(\int_{X_2} \left(\sum_i f(i,y)^{q_1} u_i\right)^{\frac{p_2}{q_1}} d\lambda_2(y)\right)^{\frac{1}{p_2}},$$

using the inclusion $L^{p_2}_{\lambda_2}(X_2) \subset L^{q_2}_{\lambda_2}(X_2)$. In the non-Minkowski case, $p_2 < q_1$, so for any non-negative terms c_i, $\sum_i c_i^{\frac{q_1}{p_2}} \le (\sum_i c_i)^{\frac{q_1}{p_2}}$. Applied to the above with $c_i = f(i,y)^{p_2} u_i^{\frac{p_2}{q_1}}$,

$$\|f\|_Q \le C_2 \left(\int_{X_2} \left(\sum_i f(i,y)^{p_2} u_i^{\frac{p_2}{q_1}} \right) d\lambda_2(y) \right)^{\frac{1}{p_2}} = C_2 \left(\sum_i \int_{X_2} f(i,y)^{p_2} d\lambda_2(y) u_i^{\frac{p_2}{q_1}} \right)^{\frac{1}{p_2}},$$

where the final equality is due to Tonelli's theorem. By Hölder's inequality with the conjugate exponents $\frac{p_1}{p_2}$ and $\frac{p_1}{p_1-p_2}$,

$$\begin{aligned}
\sum_i \int_{X_2} f(i,y)^{p_2} d\lambda_2(y) u_i^{\frac{p_2}{q_1}} &= \sum_i \left(\int_{X_2} f(i,y)^{p_2} d\lambda_2(y) u_i^{\frac{p_2}{p_1}} \right) u_i^{p_2\left(\frac{1}{q_1}-\frac{1}{p_1}\right)} \\
&\le \left(\sum_i \left(\int_{X_2} f(i,y)^{p_2} d\lambda_2(y) \right)^{\frac{p_1}{p_2}} u_i \right)^{\frac{p_2}{p_1}} \left(\sum_i u_i^{p_2\left(\frac{1}{q_1}-\frac{1}{p_1}\right)\cdot\frac{p_1}{p_1-p_2}} \right)^{\frac{p_1-p_2}{p_1}} \\
&= \|f\|^{p_2}_{\sigma(P)} \left(\sum_i u_i^{\frac{\alpha}{1-\beta}} \right)^{p_2\left(\frac{1}{p_2}-\frac{1}{p_1}\right)}.
\end{aligned}$$

Therefore

$$\|f\|_Q \le C_2 \left(\sum_i u_i^{\frac{\alpha}{1-\beta}} \right)^{\frac{1}{p_2}-\frac{1}{p_1}} \|f\|_{\sigma(P)} = C_2 \, \|(u_i)\|^{q_1^{-1}-p_1^{-1}}_{\ell^{\frac{\alpha}{1-\beta}}} \, \|f\|_{\sigma(P)}.$$

However, if (u_i) is not $\frac{\alpha}{1-\beta}$-summable, take inequivalent atoms $(A_i)_{i\in I}$ of X_1, where I must be infinite (but countable, by Proposition 2.6.6), so it has an infinite enumeration $(i_n)_{n\ge1}$. Let each $a_n = u_{i_n}$. Let $B \subset X_2$ be a subset with $\lambda_2(B) > 0$ which contains no atom. For each $N \ge 1$, define for $1 \le n \le N$

$$b_N(n) = \frac{u_{i_n}^{\frac{\alpha}{1-\beta}}}{\sum_{n=1}^N u_{i_n}^{\frac{\beta}{1-\alpha}}} \lambda_2(B).$$

Because λ_2 is atomless on B and $\sum_{n=1}^N b_N(n) \le \lambda_2(B)$, there are disjoint subsets $B_N(1), \ldots, B_N(N)$ of B such that $\lambda_2(B_N(n)) = b_N(n)$.

$$\begin{aligned}
\sum_{n=1}^N a_n^\alpha b_N(n)^\beta &= \frac{\sum_{n=1}^N u_{i_n}^\alpha u_{i_n}^{\frac{\alpha\beta}{1-\beta}}}{\left(\sum_{n=1}^N u_{i_n}^{\frac{\alpha}{1-\beta}}\right)^\beta} \\
&= \left(\sum_{n=1}^N u_{i_n}^{\frac{\alpha}{1-\beta}} \right)^{1-\beta}
\end{aligned}$$

Since $\beta < 1$, for large N this sum is arbitrarily large. As in the previous proposition, by Proposition 4.1.3 $L^{\sigma(P)} \not\subset L^Q$. □

4.2 Non-Minkowski case, in general

The previous section's restrictions of considering common measures and finite exponents are here lifted. One-variable inclusions remain necessary, and the Minkowski case $\min(p_1, q_1) \leq \max(p_2, q_2)$ remains, with such inclusions, sufficient for permuted mixed-norm inclusion, with $C = C_1C_2$. Therefore, the focus remains on the non-Minkowski $\max(p_2, q_2) < \min(p_1, q_1)$ case, but with the possibility that one or both of p_1 and q_1 may be ∞. Furthermore, rather than studying only common measures λ_k, different measures μ_k and ν_k are allowed.

Absolute continuity, $\nu_k \ll \mu_k$, remains necessary and is therefore still assumed. Methods using the Radon-Nikodym derivative turn out, in most cases, to allow a reduction of the problem with different measures to one with common measures, as we are about to see. Thus, the previous simplified arguments remain relevant, not only in inspiring the ideas behind somewhat more complicated approaches, but in solving some problems after reduction. Although there are limitations and cases better handled otherwise, the idea of reducing a problem in μ and ν to one in λ, which first showed up in the one-variable problem in Proposition 2.4.2 (and could also be applied for Proposition 2.4.1 in many cases), still has its uses.

4.2.1 Reducing to common measures

Recall from Theorem 2.1.8 that each necessary one-variable inclusion $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\mu_k}(X_k)$ implies that $\nu_k \ll \mu_k$, so each Radon-Nikodym derivative $\frac{d\nu_k}{d\mu_k}$ exists. For this problem, we can further suppose that each $\frac{d\nu_k}{d\mu_k} > 0$ μ_k-a.e.

Assumption For the rest of the two-variable problem, assume that $\frac{d\nu_k}{d\mu_k} > 0$, μ_k-a.e., for $k = 1, 2$.

The justification comes from the following proposition.

Proposition 4.2.1. *For each $k \in \{1, 2\}$, let $Y_k = \left\{x \in X_k : \frac{d\nu_k}{d\mu_k}(x) > 0\right\}$. Any constant $C \geq 0$ satisfies, for any measurable function f on $X_1 \times X_2$,*

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)},$$

if and only if it satisfies, for any measurable function g on $Y_1 \times Y_2$,

$$\left\| \|g\|_{L^{q_1}_{\nu_1}(Y_1)} \right\|_{L^{q_2}_{\nu_2}(Y_2)} \leq C \left\| \|g\|_{L^{p_2}_{\mu_2}(Y_2)} \right\|_{L^{p_1}_{\mu_1}(Y_1)}.$$

(Equivalently, for any measurable function g on $X_1 \times X_2$ which is supported on $Y_1 \times Y_2$.)

Proof. Naturally, if C works in the first inequality, it works in the second, since $Y_1 \times Y_2 \subset X_1 \times X_2$. (Any function g can be extended to $f = g$ on $Y_1 \times Y_2$, $f = 0$ off $Y_1 \times Y_2$, to which the first inequality applies.) It remains only to show that, if the second inequality is always valid for a particular C, the first inequality also holds with it.

Assuming the second inequality for a particular C, take any measurable function f on $X_1 \times X_2$. Let g be the restriction $f|_{Y_1 \times Y_2}$ of f to $Y_1 \times Y_2$. Because each $\nu_k(X_k \setminus Y_k) = 0$,

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} = \left\| \|g\|_{L^{q_1}_{\nu_1}(Y_1)} \right\|_{L^{q_2}_{\nu_2}(Y_2)},$$

while because each $Y_k \subset X_k$,

$$\left\| \|g\|_{L^{p_2}_{\mu_2}(Y_2)} \right\|_{L^{p_1}_{\mu_1}(Y_1)} \le \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

Therefore, as desired,

$$\begin{aligned} \left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left\| \|g\|_{L^{q_1}_{\nu_1}(Y_1)} \right\|_{L^{q_2}_{\nu_2}(Y_2)} \\ &\le C \left\| \|g\|_{L^{p_2}_{\mu_2}(Y_2)} \right\|_{L^{p_1}_{\mu_1}(Y_1)} \\ &\le C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}. \end{aligned}$$

□

Replacing each X_k by Y_k to work on this $Y_1 \times Y_2$, where each $\frac{d\nu_k}{d\mu_k} > 0$, we get the same best constant C with the convenience that each $\frac{d\nu_k}{d\mu_k} > 0$, μ_k-almost everywhere. (If desired, it can be defined and positive everywhere.)

The next two results reduce to a common measure in a particular variable; first in X_1, then in X_2. When both are applicable, they can be applied together to reduce the problem entirely to common measures λ_1 and λ_2. Note that the constant remains the same in this reduction, so not only does the qualitative question of whether inclusion holds have the same answer afterward, but the quantitative best value of the constant C is preserved.

(The assumption that $\frac{d\nu_1}{d\mu_1} > 0$ is helpful because it means that $\left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{p_1}{p_1-q_1}}$ makes sense whether $p_1 > q_1$ or $p_1 < q_1$, although even without that assumption we could use it times the characteristic function of the set where $\frac{d\nu_1}{d\mu_1} > 0$.)

Proposition 4.2.2. *Suppose that $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$ and either $p_1 = q_1 = \infty$ or $p_1 \ne q_1$. Define a measure λ_1 as follows.*

$$\text{If } p_1 = q_1 = \infty \text{ then } \lambda_1 = \mu_1$$

$$\text{If } p_1 \ne q_1 \text{ and } p_1, q_1 < \infty \text{ then } \lambda_1 = \left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{p_1}{p_1-q_1}} \mu_1$$

Then the following are equivalent:

$$\text{For each measurable } f \text{ on } X_1 \times X_2, \left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \le C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}$$

$$\text{For each measurable } h \text{ on } X_1 \times X_2, \left\| \|h\|_{L^{q_1}_{\lambda_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \le C \left\| \|h\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\lambda_1}(X_1)}$$

Proof. First, suppose that $p_1 = q_1 = \infty$. Because $\nu_1 \ll \mu_1$, any subset $E \subset X_1$ with $\nu_1(E) > 0$ must also have $\mu_1(E) > 0$. With the further assumption that $\frac{d\nu_1}{d\mu_1} > 0$ μ_1-a.e., explained earlier in this section, if $\mu_1(E) > 0$ then $\nu_1(E) > 0$. Consequently, the essential supremum norms $\|\cdot\|_{L^\infty_{\mu_1}(X_1)}$ and $\|\cdot\|_{L^\infty_{\nu_1}(X_1)}$ are identical, so we can simply use either μ_1 or ν_1 as the common measure λ_1 and obtain the same problem.

For the rest of the argument, assume that $p_1 \neq q_1$ and both are finite. If the first inequality holds, then for any measurable h let $f = |h| \left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{1}{p_1-q_1}}$.

$$
\begin{aligned}
\left\| \|h\|_{L^{q_1}_{\lambda_1}} \right\|_{L^{q_2}_{\nu_2}} &= \left\| \left(\int_{X_1} |h|^{q_1} \left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{p_1}{p_1-q_1}} d\mu_1 \right)^{\frac{1}{q_1}} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\
&= \left\| \left(\int_{X_1} f^{q_1} d\nu_1 \right)^{\frac{1}{q_1}} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\
&\leq C \left(\int_{X_1} \|f\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} d\mu_1 \right)^{\frac{1}{p_1}} \\
&= C \left(\int_{X_1} \|h\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} \left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{p_1}{p_1-q_1}} d\mu_1 \right)^{\frac{1}{p_1}} \\
&= C \left\| \|h\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\lambda_1}(X_1)}
\end{aligned}
$$

If the second inequality holds, then for any measurable f let $h = |f| \left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{-1}{p_1-q_1}}$.

$$
\begin{aligned}
\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left\| \left(\int_{X_1} |f|^{q_1} d\nu_1 \right)^{\frac{1}{q_1}} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\
&= \left\| \left(\int_{X_1} h^{q_1} d\lambda_1 \right)^{\frac{1}{q_1}} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\
&\leq C \left(\int_{X_1} \|h\|^{q_1}_{L^{p_2}_{\mu_2}(X_2)} d\lambda_1 \right)^{\frac{1}{p_1}} \\
&= C \left(\int_{X_1} \|f\|_{L^{p_2}_{\mu_2}(X_2)} d\mu_1 \right)^{\frac{1}{p_1}} \\
&= C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}
\end{aligned}
$$

□

(Note that in this next result, $p_2 = q_2 = \infty$ is impossible in the non-Minkowski case. This is left in since the result could be applied in the Minkowski case, even though that case is simple enough as it is.)

Proposition 4.2.3. *Suppose that $L^{p_2}_{\mu_2}(X_2) \subset L^{q_2}_{\nu_2}(X_2)$ and either $p_2 = q_2 = \infty$ or $p_2 \neq q_2$. Define a measure λ_2 as follows.*

$$\textit{If } p_2 = q_2 = \infty \textit{ then } \lambda_2 = \mu_2$$

$$\textit{If } p_2 \neq q_2 \textit{ and } p_2, q_2 < \infty \textit{ then } \lambda_2 = \left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{p_2}{p_2-q_2}} \mu_2$$

Then the following are equivalent:

$$\text{For each measurable } f \text{ on } X_1 \times X_2,\ \left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}$$

$$\text{For each measurable } h \text{ on } X_1 \times X_2,\ \left\| \|h\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\lambda_2}(X_2)} \leq C \left\| \|h\|_{L^{p_2}_{\lambda_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}$$

Proof. The case $p_2 = q_2 = \infty$ works as in the preceding result, where the one-variable inclusion from the hypothesis and $\frac{d\nu_2}{d\mu_2} > 0$ provide that $\|\cdot\|_{L^{\infty}_{\mu_2}(X_2)} = \|\cdot\|_{L^{\infty}_{\nu_2}(X_2)}$. Again, for the rest assume that $p_2 \neq q_2$ and both are finite.

If the first inequality holds, then for any measurable h let $f = |h| \left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{1}{p_2 - q_2}}$.

$$\begin{aligned}
\left\| \|h\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\lambda_2}} &= \left(\int_{X_2} \|h\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} \left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{p_2}{p_2 - q_2}} d\mu_2 \right)^{\frac{1}{q_2}} \\
&= \left(\int_{X_2} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} d\nu_2 \right)^{\frac{1}{q_2}} \\
&\leq C \left\| \left(\int_{X_2} f^{p_2} d\mu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\
&= C \left\| \left(\int_{X_2} |h|^{p_2} \left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{p_2}{p_2 - q_2}} d\mu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\
&= C \left\| \|h\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}
\end{aligned}$$

If the second inequality holds, then for any measurable f let $h = |f| \left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{-1}{p_2 - q_2}}$.

$$\begin{aligned}
\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left(\int_{X_2} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} d\nu_2 \right)^{\frac{1}{q_2}} \\
&= \left(\int_{X_2} \|h\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} d\lambda_2 \right)^{\frac{1}{q_2}} \\
&\leq C \left\| \left(\int_{X_2} h^{p_2} d\lambda_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\
&= C \left\| \left(\int_{X_2} |f|^{p_2} d\mu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\
&= C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}
\end{aligned}$$

□

In the remaining cases where $p_1 = q_1 < \infty$ or $p_2 = q_2 < \infty$, inclusions with common measures turn out to be sufficient, but may not be necessary. Also, unlike the previous results, the best constants may differ between the original $L^{\sigma(P)} \subset L^Q$ problems and the versions with common measures.

For example, the following result shows that proving an inclusion using either μ_1 or ν_1, alone, as a common measure on X_1 implies the original inclusion using both μ_1 and ν_1. The converse is true if the Radon-Nikodym derivative is bounded away from zero. Because of the limitations on these results, alternative methods are instead used to solve the problem with general measures, but they are given here in case they are of interest.

Proposition 4.2.4. *Suppose that $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$, $p_1 = q_1 < \infty$, and that there is some constant $C < \infty$ such that at least one of the following is true for every measurable function f on $X_1 \times X_2$.*

$$\left\| \|f\|_{L^{q_1}_{\mu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}$$

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\nu_1}(X_1)}$$

Then there is a constant $C' < \infty$ such that, for every measurable function f on $X_1 \times X_2$,

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C' \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

If additionally the Radon-Nikodym derivative $\frac{d\nu_1}{d\mu_1}$ has a strictly positive a.e. lower bound, then the existence of such a constant C' implies that there is a constant C satisfying both of the first two inequalities.

Proof. Recall from Theorem 2.1.8 that the inclusion $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$ with $p_1 = q_1 < \infty$ implies that $\frac{d\nu_1}{d\mu_1} \in L^{\infty}_{\mu_1}(X_1)$, so there is some $M \geq 0$ such that $\frac{d\nu_1}{d\mu_1} \leq M$ μ_1-a.e.

Suppose that the first inequality holds with some $C < \infty$.

$$\begin{aligned}
\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left\| \left(\int_{X_1} |f|^{q_1} d\nu_1 \right)^{\frac{1}{q_1}} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\
&\leq \left\| M^{1/q_1} \left(\int_{X_1} |f|^{q_1} d\mu_1 \right)^{\frac{1}{q_1}} \right\|_{L^{q_2}_{\nu_2}(X_2)} \\
&\leq M^{1/q_1} C \left(\int_{X_1} \|f\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} \, d\mu_1 \right)^{\frac{1}{p_1}} \\
&= M^{1/q_1} C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}
\end{aligned}$$

That is, the final inequality is then true with the (not necessarily least) constant $C' = M^{1/q_1} C$. Next, assume that the second statement is true with some $C < \infty$.

$$\begin{aligned}
\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq C \left(\int_{X_1} \|f\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} \, d\nu_1 \right)^{\frac{1}{p_1}} \\
&\leq C \left(\int_{X_1} M \, \|f\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} \, d\mu_1 \right)^{\frac{1}{p_1}} \\
&= M^{1/p_1} C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}
\end{aligned}$$

The final inequality is true with $C' = M^{1/p_1}C$.

Now suppose that there is some $m > 0$ such that $\frac{d\nu_1}{d\mu_1} \geq m$ μ_1-a.e. and that there is some $C' < \infty$ satisfying the last inequality. Then we can verify both of the first two inequalities for different constants C, as follows.

$$\begin{aligned}\left\|\|f\|_{L^{q_1}_{\mu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left\|\left(\int_{X_1} |f|^{q_1}\, d\mu_1\right)^{\frac{1}{q_1}}\right\|_{L^{q_2}_{\nu_2}(X_2)} \\ &\leq m^{1/q_1}\left\|\left(\int_{X_1} |f|^{q_1}\, d\nu_1\right)^{\frac{1}{q_1}}\right\|_{L^{q_2}_{\nu_2}(X_2)} \\ &\leq m^{1/q_1}C'\left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{L^{p_1}_{\mu_1}(X_1)}\end{aligned}$$

$$\begin{aligned}\left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq C'\left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &= C'\left(\int_{X_1} \|f\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)}\, d\mu_1\right)^{\frac{1}{p_1}} \\ &\leq m^{1/p_1}C'\left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{L^{p_1}_{\nu_1}(X_1)}\end{aligned}$$

□

Unsurprisingly, there is a similar result involving μ_2 and ν_2.

Proposition 4.2.5. *Suppose that $L^{p_2}_{\mu_2}(X_2) \subset L^{q_2}_{\nu_2}(X_2)$, $p_2 = q_2 < \infty$, and that there is some constant $C < \infty$ such that at least one of the following is true for every measurable function f on $X_1 \times X_2$.*

$$\left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\mu_2}(X_2)} \leq C\left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{L^{p_1}_{\mu_1}(X_1)}$$
$$\left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C\left\|\|f\|_{L^{p_2}_{\nu_2}(X_2)}\right\|_{L^{p_1}_{\mu_1}(X_1)}$$

Then there is a constant $C' < \infty$ such that, for every measurable function f on $X_1 \times X_2$,

$$\left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C'\left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

If additionally the Radon-Nikodym derivative $\frac{d\nu_2}{d\mu_2}$ has a strictly positive a.e. lower bound, then the existence of such a constant C' implies that there is a constant C satisfying both of the first two inequalities.

Proof. As in the preceding result, the one-variable inclusion implies that $\frac{d\nu_2}{d\mu_2} \in L^{\infty}_{\mu_2}(X_2)$, so there is some $M \geq 0$ such that $\frac{d\nu_2}{d\mu_2} \leq M$ μ_2-a.e.

Assuming that there is some $C < \infty$ for which the first inequality holds,

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left(\int_{X_2} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} \, d\nu_2 \right)^{\frac{1}{q_2}} \\ &\leq M^{1/q_2} \left(\int_{X_2} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} \, d\mu_2 \right)^{\frac{1}{q_2}} \\ &\leq M^{1/q_2} C \left\| \left(\int_{X_2} |f|^{p_2} d\mu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &= M^{1/q_2} C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)} .\end{aligned}$$

This is the last inequality with $C' = M^{1/q_2} C$. Now suppose that the second inequality is true with $C < \infty$.

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq C \left\| \left(\int_{X_2} |f|^{p_2} d\nu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &\leq M^{1/p_2} C \left\| \left(\int_{X_2} |f|^{p_2} d\mu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &= M^{1/p_2} C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}\end{aligned}$$

This time, $C' = M^{1/p_2} C$ works.

Now, suppose that for some $m > 0$, $\frac{d\nu_1}{d\mu_1} \geq m \ \mu_1$-a.e. and that the last inequality holds for some $C' < \infty$. The first two inequalities are established below.

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\mu_2}(X_2)} &= \left(\int_{X_2} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} \, d\mu_2 \right)^{\frac{1}{q_2}} \\ &\leq m^{1/q_2} \left(\int_{X_2} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} \, d\nu_2 \right)^{\frac{1}{q_2}} \\ &\leq m^{1/q_1} C' \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}\end{aligned}$$

$$\begin{aligned}\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} &\leq C' \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &= C' \left\| \left(\int_{X_2} |f|^{p_2} d\mu_2 \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &\leq m^{1/p_2} C' \left\| \|f\|_{L^{p_2}_{\nu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}\end{aligned}$$

□

4.2.2 Block factorable function necessary condition

The following results generalize the main result, Proposition 4.1.3, of Section 4.1.1. Although they cover the cases $p_1 = \infty$ and $q_1 = \infty$, they also serve to generalize the combinations of characteristic functions of blocks, used in that section, to a more flexible notion of block factorable functions. Factorable functions, i.e. those of the form $f(x_1, x_2) = f_1(x_1)f_2(x_2)$, established the basic lower bound $C_1C_2 \le C$, which is refined by considering block factorable functions (which are so named since they are locally factorable on blocks, although they are generally not globally factorable functions) as described below.

Definition 4.2.6. Let I be an (at most) countable index set, with a corresponding disjoint collection $(A_i)_{i\in I}$ of measurable subsets of X_1 and another disjoint collection $(B_i)_{i\in I}$ of measurable subsets of X_2. For each $i \in I$, let $C_1(A_i)$ and $C_2(B_i)$ be the least nonnegative constants such that, for any $g_i \in L^+(A_i)$ and $h_i \in L^+(B_i)$,

$$\begin{aligned} \|g_i\|_{L^{q_1}_{\nu_1}(A_i)} &\le C_1(A_i)\,\|g_i\|_{L^{p_1}_{\mu_1}(A_i)} \text{ and} \\ \|h_i\|_{L^{q_2}_{\nu_2}(B_i)} &\le C_2(B_i)\,\|h_i\|_{L^{p_2}_{\mu_2}(B_i)} . \end{aligned} \tag{4.3}$$

(Equivalently, $C_1(A_i)$ is the least constant such that $\|g_i\|_{L^{q_i}_{\nu_1}(X_1)} \le C_1(A_i)\,\|g_i\|_{L^{p_i}_{\mu_1}(X_1)}$ for any $g_i \in L^+(X_1)$ supported on A_i, and similarly for $C_2(B_i)$.)

Although the index set I would normally be either $\mathbb{N}$ or, for finite collections, $\{1, \dots n\}$, any (at most) countable set, being in bijection with one of these, would do. The constants $C_1(A_i)$ are so named because they are local versions of C_1, and $C_2(B_i)$ of C_2, applicable to functions supported on the sets A_i and B_i, respectively. (In this sense, $C_1 = C_1(X_1)$ and $C_2 = C_2(X_2)$.) Of course, any function on A_i is a function on X_1, and similarly for B_i and X_2, so each $C_1(A_i) \le C_1$ and $C_2(B_i) \le C_2$.

Definition 4.2.7. Given disjoint collections $(A_i)_{i\in I}$ and $(B_i)_{i\in I}$ as above, a block factorable function f supported on $(A_i \times B_i)_{i\in I}$ is one of the form

$$f(x_1, x_2) = \sum_{i\in I} g_i(x_1)h_i(x_2),$$

where each g_i is supported on A_i and each h_i supported on B_i. (Naturally, the order of summation here is irrelevant since each point (x_1, x_2) is in at most one block $A_i \times B_i$, and therefore at most one non-zero term contributes to $f(x_1, x_2)$.)

Definition 4.2.8. Given two exponents $p, q \in (0, \infty]$, define the *relative conjugate* of p with respect to q as $p : q \in (-\infty, \infty]$, computed using the conventions that $\infty^{-1} = 0$ and $0^{-1} = \infty$, by

$$(p : q)^{-1} = q^{-1} - p^{-1}. \tag{4.4}$$

Note that the standard Hölder conjugate is recovered by $p : 1 = p'$. Although this idea is hardly new, the term "relative conjugate" is improvised and probably not standard. This name for the exponent $p : q$ is chosen because of its role in the following generalization of Corollary 2.3.3, a sharp form of Hölder's inequality.

Corollary 4.2.9. *Let (X,μ) be a σ-finite measure space and suppose that $0 < q \le p \le \infty$. For fixed $g \in L^+(X)$, the least constant $0 \le C_g \le \infty$ such that, for any $f \in L^+(X)$,*

$$\|fg\|_q \le C_g \|f\|_p$$

is $C_g = \|g\|_{p:q}$. Consequently, $C < \infty$ if and only if $g \in L^{p:q}_\mu(X)$.

Proof. If $q = \infty$, then $p = \infty$ and $p:q = 0^{-1} = \infty$ as well. Because $|fg| \le \|f\|_\infty \|g\|_\infty$ μ-a.e., in the essential supremum $\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty$. Therefore the least constant $C_g \le \|g\|_\infty$.

To find that $\|g\|_\infty$ is the least constant C_g, note that for any $\epsilon > 0$, there is a set $E_\epsilon \subset X$ on which $|g| > \|g\|_\infty - \epsilon$ and such that $0 < \mu(E_\epsilon) < \infty$. (Use σ-finiteness of μ to produce E_ϵ of finite measure, if necessary.) Let $f_\epsilon = \chi_{E_\epsilon}$ and then observe that, on E_ϵ itself, $|f_\epsilon g| = |g| > \|g\|_\infty - \epsilon$. Therefore

$$\|f_\epsilon g\|_\infty \ge \|g\|_\infty - \epsilon = \|f_\epsilon\|_\infty \left(\|g\|_\infty - \epsilon\right),$$

and in the limit as $\epsilon \to \infty$, $C_g \ge \|g\|_\infty$.

Now suppose that $q < \infty$. Simply let $\tilde{f} = f^q$ and $\tilde{g} = g^q$ and apply Corollary 2.3.3:

$$\|fg\|_q^q = \int_X \tilde{f}\tilde{g}\,d\mu \le C\|\tilde{f}\|_{p/q} = C\|f\|_p^q \tag{4.5}$$

has the least constant $C = \|\tilde{g}\|_{(p/q)'} = \|\tilde{g}\|_{p/(p-q)} = \|g\|^q_{p:q}$. Taking q^{th} roots,

$$\|fg\|_q \le C_g \|f\|_p \tag{4.6}$$

is valid with $C_g = \|g\|_{p:q}$; this is also the least constant C_g since, if (4.6) holds with some C_g, then (4.5) must hold with $C = C_g^q$, and the least value of C there is $C = \|g\|^q_{p:q}$. □

Many following results will be easier to write in terms of relative conjugates. The complicated exponents which appeared in the case of common measures, dubbed α, β, $\frac{\alpha}{1-\beta}$, and $\frac{\beta}{1-\alpha}$ and defined in (4.1), can be easily written using this notation.

$$\alpha = \frac{p_1 : q_2}{p_1 : q_1} \qquad \frac{\alpha}{1-\beta} = \frac{p_1 : p_2}{p_1 : q_1}$$
$$\beta = \frac{p_1 : q_2}{p_2 : q_2} \qquad \frac{\beta}{1-\alpha} = \frac{q_1 : q_2}{p_2 : q_2}$$

It is simply computed, and convenient to note, that the sign of $p : q$ corresponds to the order of p and q, as expressed below.

$p : q > 0$	if and only if	$p > q$
$p : q < 0$	if and only if	$p < q$

Proposition 4.2.10. *Suppose that* $\max(p_2, q_2) < \min(p_1, q_1)$ *(the Minkowski sufficient condition for inclusion does not apply) and let* $C \in [0, \infty]$. *For any (at most) countable I and disjoint collections $(A_i)_{i\in I}$ of measurable subsets of X_1, (B_i) of measurable subsets of X_2, let $C_1(A_i)$ and*

$C_2(B_i)$ denote the least constants in (4.3). Then the least constant $C_{BF}(A, B)$ such that, for any block factorable function f supported on $(A_i, B_i)_{i\in I}$,

$$\|f\|_Q \le C_{BF}(A, B)\,\|f\|_{\sigma(P)}$$

is computed by

$$C_{BF}(A, B) = \|C_1(A_i)C_2(B_i)\|_{\ell^{p_1:q_2}(I)}\,.$$

Proof. For any block factorable function $f = \sum_{i\in I} g_i h_i$ supported on $(A_i \times B_i)_{i\in I}$, define ratios

$$r_1(i) = \frac{\|g_i\|_{L^{q_1}_{\nu_1}(X_1)}}{\|g_i\|_{L^{p_1}_{\mu_1}(X_1)}}, \qquad r_2(i) = \frac{\|h_i\|_{L^{q_2}_{\nu_2}(X_2)}}{\|h_i\|_{L^{p_2}_{\mu_2}(X_2)}}.$$

For this block-factorable f, when $q_1 < \infty$ (so $q_2 < q_1 < \infty$ as well), since the collections (g_i) and (h_i) are disjointly supported,

$$\begin{aligned}
\|f\|_Q = \left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left(\int_{X_2}\left(\int_{X_1}\sum_{i\in I}|g_i(x_1)h_i(x_2)|^{q_1}\,d\nu_1(x_1)\right)^{\frac{q_2}{q_1}} d\nu_2(x_2)\right)^{\frac{1}{q_2}}\\
&= \left(\int_{X_2}\sum_i\left(\int_{X_1}|g_i(x_1)|^{q_1}\,d\nu_1(x_1)\right)^{\frac{q_2}{q_1}}|h_i(x_2)|^{q_2}\,d\nu_2(x_2)\right)^{\frac{1}{q_2}}\\
&= \left(\sum_i\left(\int_{X_1}|g_i(x_1)|^{q_1}\,d\nu_1(x_1)\right)^{\frac{q_2}{q_1}}\int_{X_2}|h_i(x_2)|^{q_2}\,d\nu_2(x_2)\right)^{\frac{1}{q_2}}\\
&= \left(\sum_i\|g_i\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)}\,\|h_i\|^{q_2}_{L^{q_2}_{\nu_2}(X_2)}\right)^{\frac{1}{q_2}}.
\end{aligned}$$

When $q_1 = \infty$, the same conclusion is true, computed as follows.

$$\begin{aligned}
\|f\|_Q = \left\|\|f\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{L^{q_2}_{\nu_2}(X_2)} &= \left(\int_{X_2}\left(\operatorname*{ess\,sup}_{x_1\in X_1}\sum_{i\in I}|g_i(x_1)h_i(x_2)|^{q_1}\right)^{q_2} d\nu_2(x_2)\right)^{\frac{1}{q_2}}\\
&= \left(\sum_i\left(\operatorname*{ess\,sup}_{x_1\in X_1}|g_i(x_1)|\right)^{q_2}\int_{X_2}|h_i(x_2)|^{q_2}\,d\nu_2(x_2)\right)^{\frac{1}{q_2}}\\
&= \left(\sum_i\|g_i\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)}\,\|h_i\|^{q_2}_{L^{q_2}_{\nu_2}(X_2)}\right)^{\frac{1}{q_2}}.
\end{aligned}$$

Similarly,

$$\|f\|_{\sigma(P)} = \left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{L^{p_1}_{\mu_1}(X_1)} = \left(\sum_i\|g_i\|^{p_1}_{L^{p_1}_{\mu_1}(X_1)}\,\|h_i\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)}\right)^{\frac{1}{p_1}},$$

understood as $\sup_i \|g_i\|_{L^{p_1}_{\mu_1}(X_1)} \|h_i\|_{L^{p_2}_{\mu_2}(X_2)}$ when $p_1 = \infty$. These can be used together with the inequality which defines $C_{BF}(A, B)$ to produce

$$\begin{aligned}\sum_i \|g_i\|^{q_2}_{L^{p_1}_{\mu_1}(X_1)} \|h_i\|^{q_2}_{L^{p_2}_{\mu_2}(X_2)} r_1(i)^{q_2} r_2(i)^{q_2} &= \left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|^{q_2}_{L^{q_2}_{\nu_2}(X_2)} \\ &\le C_{BF}(A, B)^{q_2} \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|^{q_2}_{L^{p_1}_{\mu_1}(X_1)} \\ &= C_{BF}(A, B)^{q_2} \left(\sum_i \|g_i\|^{p_1}_{L^{p_1}_{\mu_1}(X_1)} \|h_i\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} \right)^{\frac{q_2}{p_1}}.\end{aligned}$$

By Hölder's inequality (Corollary 2.3.3) applied with the exponent $\frac{p_1}{q_2}$ to $\|g_i\|^{q_2}_{L^{p_1}_{\mu_1}(X_1)} \|h_i\|^{q_2}_{L^{p_2}_{\mu_2}(X_2)}$ and $\left(\frac{p_1}{q_2}\right)' = \frac{p_1}{p_1 - q_2}$ to $r_1(i)^{q_2} r_2(i)^{q_2}$, the least constant K in

$$\sum_i \|g_i\|^{q_2}_{L^{p_1}_{\mu_1}(X_1)} \|h_i\|^{q_2}_{L^{p_2}_{\mu_2}(X_2)} r_1(i)^{q_2} r_2(i)^{q_2} \le K \left(\sum_i \|g_i\|^{p_1}_{L^{p_1}_{\mu_1}(X_1)} \|h_i\|^{p_1}_{L^{p_2}_{\mu_2}(X_2)} \right)^{\frac{q_2}{p_1}} \tag{4.7}$$

is $K = \left\| r_1^{q_2} r_2^{q_2} \right\|_{\ell^{\frac{p_1}{p_1 - q_2}}}$. Since $C_{BF}(A, B)^{q_2}$ is a viable constant above,

$$C_{BF}(A, B)^{q_2} \ge K = \left(\sum_{i \in I} (r_1(i) r_2(i))^{\frac{p_1 q_2}{p_1 - q_2}} \right)^{\frac{p_1 - q_2}{p_1}} = \left(\sum_{i \in I} (r_1(i) r_2(i))^{p_1 : q_2} \right)^{\frac{q_2}{p_1 : q_2}},$$

for the right-hand side here is the least constant given by Hölder. Consequently, $C_{BF}(A, B) \ge \|r_1 r_2\|_{\ell^{p_1 : q_2}(I)}$. (When $p_1 = \infty$, $\left(\frac{p_1}{q_2}\right)' = 1$, so this is an ℓ^{q_2} norm, as desired.) Finally, since each $C_1(A_i)$ is the supremum of $r_1(i)$ over all functions g_i supported on A_i and $C_2(B_i)$ is the supremum of $r_2(i)$ over h_i supported on B_i, taking these suprema we have $C_{BF}(A, B) \ge \|C_1(A_i) C_2(B_i)\|_{\ell^{p_1 : q_2}(I)}$.

On the other hand, $K = \|C_1(A_i)^{q_2} C_2(B_i)^{q_2}\|_{\ell^{\frac{p_1}{p_1 - q_2}}(I)}$ satisfies (4.7) for any functions g_i supported on A_i and h_i supported on B_i, so taking the q_2 root implies that

$$\|f\|_Q \le \|C_1(A_i) C_2(B_i)\|_{\ell^{p_1 : q_2}(I)} \|f\|_{\sigma(P)}$$

for any block-factorable f supported on $(A_i \times B_i)$. Because $C_{BF}(A, B)$ is the least such constant, $C_{BF} \le \|C_1(A_i) C_2(B_i)\|_{\ell^{p_1 : q_2}(I)}$. □

Theorem 4.2.11. *Suppose that* $\max(p_2, q_2) < \min(p_1, q_1)$ *(the non-Minkowski case) and let* $C \in [0, \infty]$ *be the least constant so that inequality (3.2),*

$$\|f\|_Q \le C \|f\|_{\sigma(P)},$$

holds for all $f \in L^+(X_1 \times X_2)$.

For any (at most) countable I *and disjoint collections* $(A_i)_{i \in I}$ *of measurable subsets of* X_1, $(B_i)_{i \in I}$ *of measurable subsets of* X_2, *let* $C_1(A_i)$ *and* $C_2(B_i)$ *denote the least constants as above. Then*

$$C \ge \|C_1(A_i) C_2(B_i)\|_{\ell^{p_1 : q_2}(I)}.$$

Proof. By Proposition 4.2.10, the best constant for block-factorable f supported on $(A_i \times B_i)_{i\in I}$ is $C_{BF}(A, B) = \|C_1(A_i)C_2(B_i)\|_{\ell^{p_1:q_2}(I)}$. No lesser constant works for this special class of functions, so certainly no lesser constant will work for the more general $f \in L^+(X_1 \times X_2)$. Therefore $C \geq C_{BF}(A, B) = \|C_1(A_i)C_2(B_i)\|_{\ell^{p_1:q_2}(I)}$. □

This is easily seen to refine the established $C \geq C_1C_2$ since, with only one $A_1 = X_1$ and $B_1 = X_2$, $C_1(A_1) = C_1$, $C_2(B_1) = C_2$, and $\|C_1(A_1)C_2(A_2)\|_{\ell^{p_1:q_2}(I)} = C_1C_2$. More generally, assuming the mixed-norm inclusion $L^{\sigma(P)} \hookrightarrow L^Q$ on $X_1 \times X_2$, the same inclusion holds locally on blocks $A_i \times B_i$. On each block, the product $C_1(A_i)C_2(B_i)$ of local best constants $C_1(A_i)$ (for $L^{p_1}_{\mu_1}(A_i) \hookrightarrow L^{q_1}_{\nu_1}(A_i)$) and $C_2(B_i)$ (for $L^{p_2}_{\mu_2}(B_i) \hookrightarrow L^{q_2}_{\nu_2}(B_i)$) gives a lower bound on the norm of inclusion $L^{\sigma(P)}(A_i \times B_i) \hookrightarrow L^Q(A_i \times B_i)$. These products combine through the $\ell^{p_1:q_2}$ norm to give a lower bound on the best constant C in the global mixed-norm inclusion $L^{\sigma(P)}(X_1 \times X_2) \hookrightarrow L^Q(X_1 \times X_2)$.

The resulting necessary condition lets us rule out mixed-norm inclusion if there is any way to form disjoint sequences (A_i) in X_1 and (B_i) in X_2 such that the resulting $\|C_1(A_i)C_2(B_i)\|_{p_1:q_2} = \infty$. This can be established by finding functions g_i and h_i such that $\|r_1r_2\|_{p_1:q_2} = \infty$ or is arbitrarily large (for various choices of functions with fixed (A_i) and (B_i)), where $r_1(i)$ and $r_2(i)$ are as defined in the preceding proof.

One possible application is in considering characteristic functions, as a quick way to obtain a generalization of the necessary measure summability condition from Proposition 4.1.3 of $\sum_i a_i^\alpha b_i^\beta < \infty$, i.e.

$$\sum_i a_i^{\frac{p_1:q_2}{p_1:q_1}} b_i^{\frac{p_1:q_2}{p_2:q_2}} < \infty,$$

beyond the case of common measures. The below result can also be obtained directly by computing the ratio of mixed norms $\|f\|_Q / \|f\|_{\sigma(P)}$ where f is a combination of characteristic functions of blocks $(\sum_i c_i\chi_{A_i\times B_i})$ and choosing coefficients c_i for which Hölder's inequality is sharp.

Corollary 4.2.12. *Whenever* $\max(p_2, q_2) < \min(p_1, q_1)$*, for any disjoint sequences* (A_i) *of measurable subsets of* X_1 *and* (B_i) *of measurable subsets of* X_2*, any constant* C *satisfying inequality (3.2) is bounded below,*

$$C \geq \left\| \frac{\nu_1(A_i)^{1/q_1}}{\mu_1(A_i)^{1/p_1}} \frac{\nu_2(B_i)^{1/q_2}}{\mu_2(B_i)^{1/p_2}} \right\|_{\ell^{p_1:q_2}(I)}.$$

(Naturally, this means that the above sequence norm must converge, regardless of the sets A_i *and* B_i*, if* $C < \infty$*; furthermore, in this case its values must be bounded above.)*

Proof. Let $g_i(x_1) = \chi_{E_i}(x_1)$ and $h_i(x_2) = \chi_{F_i}(x_2)$ and apply Theorem 4.2.11. The local best constants

$$C_1(A_i) \geq \frac{\|g_i\|_{L^{q_1}_{\nu_1}(A_i)}}{\|g_i\|_{L^{p_1}_{\mu_1}(A_i)}} = \frac{\nu_1(A_i)^{1/q_1}}{\mu_1(A_i)^{1/p_1}}$$

$$C_2(B_i) \geq \frac{\|h_i\|_{L^{q_2}_{\nu_2}(B_i)}}{\|h_i\|_{L^{p_2}_{\mu_2}(B_i)}} = \frac{\nu_2(B_i)^{1/q_2}}{\mu_2(B_i)^{1/p_2}},$$

since each g_i is supported on A_i and h_i on B_i. Therefore this is an immediate consequence of $C \geq \|C_1(A_i)C_2(B_i)\|_{p_1:q_2}$. □

In the case of common measures $\lambda_k = \mu_k = \nu_k$, Proposition 4.1.3 that $\left(\sum_i a_i^\alpha b_i^\beta\right)^{\frac{1}{q_2}-\frac{1}{p_1}} \leq C$ is an easy consequence, generalized by Corollary 4.2.12. When each $\lambda_1(A_i) = a_i$ and $\lambda_2(B_i) = b_i$,

$$\left\|\frac{\lambda_1(A_i)^{1/q_1}}{\lambda_1(A_i)^{1/p_1}}\frac{\lambda_2(B_i)^{1/q_2}}{\lambda_2(B_i)^{1/p_2}}\right\|_{\ell^{p_1:q_2}(I)} = \left\|a_i^{\frac{1}{p_1:q_1}}b_i^{\frac{1}{p_2:q_2}}\right\|_{p_1:q_2} = \left(\sum_i a_i^\alpha b_i^\beta\right)^{q_2^{-1}-p_1^{-1}}.$$

4.2.3 Necessity of the Minkowski criterion for non-atomic measures

The result here generalizes Theorem 4.1.6 beyond the special case of common measures, and allows exponents to be ∞. This establishes that, in the case of measures which are not purely atomic, the Minkowski sufficient condition is always necessary. Relevant definitions and properties from Section 2.6 and Subsection 4.1.2 are used, and a few more added here to deal with distinct measures μ_k and ν_k.

First, recall that Proposition 4.2.1 reduced the problem to the case where each $\frac{d\nu_k}{d\mu_k} > 0$, for $k = 1, 2$. This means that μ_k and ν_k have the same atoms, are either both atomless or neither atomless, and both purely atomic or neither purely atomic. (Although this assumption remains in place for the mixed-norm inclusion problem, for these more general measure-theoretic issues, the hypotheses are stated explicitly.)

Lemma 4.2.13. *Whenever $\nu \ll \mu$ and $\frac{d\nu}{d\mu} > 0$ μ-a.e., any measurable set has positive μ measure if and only if it has positive ν measure. (Equivalently, the null sets are also the same.)*

Proof. For any measurable set F with $\mu(F) > 0$, since $\frac{d\nu}{d\mu} > 0$ μ-a.e., the nonnegative function $\chi_F \frac{d\nu}{d\mu}$ is not μ-a.e zero. Consequently, $\nu(F) = \int_F d\nu = \int \chi_F \frac{d\nu}{d\mu} d\mu > 0$. On the other hand, if $\mu(F) = 0$, then of course $\nu(F) = \int_F \frac{d\nu}{d\mu} d\mu = 0$. □

Lemma 4.2.14. *Whenever $\nu \ll \mu$ and $\frac{d\nu}{d\mu} > 0$ μ-a.e., the measures μ and ν have the same atoms.*

Proof. Let A be any μ-atom. By definition, for any measurable $E \subset A$, either $\mu(E) = 0$ or $\mu(A \setminus E) = 0$. Because $\nu \ll \mu$, in the first case $\nu(E) = 0$, while in the second $\nu(A \setminus E) = 0$. Therefore, as long as $\nu(A) > 0$ (i.e. A is not ν-null), A is a ν-atom.

Conversely, let A be any ν-atom. Since $\nu(A) > 0$, of course $\mu(A) > 0$ as well. For any measurable $E \subset A$, either $\nu(E) = 0$ or $\nu(A \setminus E) = 0$. By Lemma 4.2.13 applied to $F = E$ or $F = A \setminus E$, as appropriate, if $\nu(E) = 0$, then $\mu(E) = 0$; also, if $\nu(A \setminus E) = 0$, then $\mu(A \setminus E) = 0$. One or the other must be true for any E, so A must be a μ-atom. □

Lemma 4.2.15. *Whenever $\nu \ll \mu$ and $\frac{d\nu}{d\mu} > 0$ μ-a.e., μ is purely atomic if and only if ν is purely atomic. Similarly, μ is atomless if and only if ν is atomless.*

Proof. Lemmas 4.2.13 and 4.2.14 show that μ and ν have the same atoms, the same null sets, and the same sets of positive measure. Because the definitions of "purely atomic" and "atomless" are dependent only on which measurable sets have these properties, clearly either both of μ and ν are purely atomic or neither is, and either both of μ and ν are atomless or neither is. □

The next theorem generalizes Theorem 4.1.6 to cover distinct measures μ_k and ν_k, as well as ∞ as a possible exponent. The basic idea remains the same, though.

Theorem 4.2.16. *Suppose that neither μ_1 (equivalently, ν_1) nor μ_2 (equivalently, ν_2) is purely atomic. If $\max(p_2, q_2) < \min(p_1, q_1)$, then there is no constant $C < \infty$ such that, for every measurable function f on $X_1 \times X_2$,*

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(X_1)} \right\|_{L^{q_2}_{\nu_2}(X_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

Proof. Since neither ν_k is purely atomic, for each there is a measurable subset $\tilde{E}_k \subset X_k$ with positive ν_k measure which contains no ν_k atom. Because μ_k is σ-finite, we can also have $0 < \nu_k(\tilde{E}_k) < \infty$. Decompose $\tilde{E}_k = \cup_{n\geq 1} \left\{ x \in \tilde{E}_k : \frac{d\nu_k}{d\mu_k} \geq \frac{1}{n} \right\}$, possible since $\frac{d\nu_k}{d\mu_k} > 0$. By subadditivity, at least one of these sets must have positive measure, since $\nu_k(\tilde{E}_k) > 0$; that is, there must be some n_k such that the subset $E_k \subset \tilde{E}_k$ where $\frac{d\nu_k}{d\mu_k} \geq \frac{1}{n_k}$ has positive measure. Now we have sets E_k with $0 < \nu_k(E_k) < \infty$, where $\frac{d\nu_k}{d\mu_k} \geq \frac{1}{n_k}$, and which contain no ν_k atom. Additionally, $0 < \mu_k(E_k) < \infty$ and each E_k contains no μ_k atom. (By Lemma 4.2.13, also $0 < \mu_k(E_k)$. And $\nu_k(E_k) = \int_{E_k} \frac{d\nu_k}{d\mu_k} d\mu_k \geq \int_{E_k} \frac{1}{n_k} d\mu_k = \frac{1}{n_k}\mu_k(E_k)$, so $\mu_k(E_k) < \infty$ because $\nu_k(E_k) < \infty$.) Furthermore, Lemma 4.2.14 shows that E_k contains no μ_k atom because it contains no ν_k atom.)

For any function f supported on $E_1 \times E_2$, inequality (3.2) implies that

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(E_1)} \right\|_{L^{q_2}_{\nu_2}(E_2)} \leq C \left\| \|f\|_{L^{p_2}_{\mu_2}(E_2)} \right\|_{L^{p_1}_{\mu_1}(E_1)}.$$

Because each $\frac{d\nu_k}{d\mu_k} \geq \frac{1}{n_k}$ on E_k, by direct computation we find that

$$\left\| \|f\|_{L^{q_1}_{\nu_1}(E_1)} \right\|_{L^{q_2}_{\nu_2}(E_2)} \geq \left(\frac{1}{n_1}\right)^{1/q_1} \left(\frac{1}{n_2}\right)^{1/q_2} \left\| \|f\|_{L^{q_1}_{\mu_1}(E_1)} \right\|_{L^{q_2}_{\mu_2}(E_2)},$$

with the natural convention $1/\infty = 0$ in case $q_1 = \infty$. This implies that

$$\begin{aligned} \left\| \|f\|_{L^{q_1}_{\mu_1}(E_1)} \right\|_{L^{q_2}_{\mu_2}(E_2)} &\leq n_1^{1/q_1} n_2^{1/q_2} \left\| \|f\|_{L^{q_1}_{\nu_1}(E_1)} \right\|_{L^{q_2}_{\nu_2}(E_2)} \\ &\leq n_1^{1/q_1} n_2^{1/q_2} C \left\| \|f\|_{L^{p_2}_{\mu_2}(E_2)} \right\|_{L^{p_1}_{\mu_1}(E_1)}. \end{aligned}$$

So, to disprove mixed-norm inclusion, we need only consider functions on $E_1 \times E_2$ with, rather than μ_k and ν_k, the common measure $\mu_k|_{E_k}$ on each E_k. Proposition 4.1.3 shows that it suffices to find disjoint sequences $(A_m)_{m\geq 1}$ of measurable subsets of E_1 and $(B_m)_{m\geq 1}$ of measurable subsets of E_2 such that, letting each $a_m = \mu_1(A_m)$ and $b_m = \mu_2(B_m)$, $\sum_m a_m^\alpha b_m^\beta = \infty$, with the exponents α and β defined in (4.1).

Proposition 4.1.2 establishes that, in this non-Minkowski case of $\max(p_2, q_2) < \min(p_1, q_1)$, $\alpha + \beta < 1$. Furthermore, because neither ν_1 nor ν_2 is purely atomic, $\alpha \geq 0$ and $\beta \geq 0$, so $0 \leq \alpha + \beta < 1$.

When $0 < \alpha + \beta < 1$, the series $\sum_{m\geq 1} m^{-\frac{1}{\alpha+\beta}}$ converges; let $M = \sum_{m\geq 1} m^{-\frac{1}{\alpha+\beta}}$. For each $i \geq 1$, let $a_m = \frac{\mu_1(E_1)}{M} m^{-\frac{1}{\alpha+\beta}}$ and $b_m = \frac{\mu_2(E_2)}{M} m^{-\frac{1}{\alpha+\beta}}$, so that $\sum_m a_m = \mu_1(E_1)$ and $\sum_m b_m = \mu_2(E_2)$. Because each E_k contains no atom of μ_k, by Corollary 4.1.5 there is a pairwise disjoint sequence (A_m) of measurable subsets of E_1 with each $\mu_1 A_m = a_m$ and there is a pairwise disjoint sequence (B_m) of measurable subsets of E_2 with each $\mu_2 B_m = b_m$. As desired, $\sum_{m\geq 1} a_m^\alpha b_m^\beta = M^{-\alpha-\beta}(\mu_1 E_1)^\alpha (\mu_2 E_2)^\beta \sum_{m\geq 1} m^{-1} = \infty$.

Finally, when $\alpha + \beta = 0$ (i.e. $\alpha = \beta = 0$), let $a_m = \mu_1(E_1)2^{-m}$ and $b_m = \mu_2(E_2)2^{-m}$, in which case again $\sum_{m\geq 1} a_m = \mu_1(E_1)$ and $\sum_{m\geq 1} b_m = \mu_2(E_2)$, so there are again disjoint sequences (A_m), (B_m) with each $A_m \subset E_1$, $a_m = \mu_1(A_m)$ and $B_m \subset E_2$, $b_m = \mu_2(B_m)$. Now, $\sum_{m\geq 1} a_m^\alpha b_m^\beta = \sum_{m\geq 1} 1 = \infty$. □

4.2.4 Two-variable permuted case, one purely atomic measure

Now, we address the final topic which was introduced in the setting of common measures with finite exponents, that of Section 4.1.3: the situation where one measure is purely atomic. Fortunately, with the assumption that $\frac{d\nu_k}{d\mu_k} > 0$ (justified by the reduction in Proposition 4.2.1), Lemma 4.2.14 shows that μ_k is purely atomic if and only if ν_k is. As before, the purely atomic space is represented by the natural numbers, with different weights for the measures μ_k and ν_k. The first step, however, is to reduce to a simpler formulation of the problem.

Recall that Proposition 4.2.1 reduces the problem to considering only the case where each Radon-Nikodym derivative $\frac{d\nu_k}{d\mu_k} > 0$ (since the region where $\frac{d\nu_k}{d\mu_k} = 0$ can be discarded), and Lemma 4.2.15 shows that, in this case, μ_k is purely atomic if and only if ν_k is purely atomic. Therefore, in this case it makes sense to say either that X_k is a purely atomic space, or it is not. Furthermore, Lemma 4.2.14 establishes that each μ_k and ν_k have the same atoms.

Suppose that X_1 has purely atomic measures, with atoms enumerated by $(E_i)_{i\in I}$; the meaning of $\frac{d\nu_k}{d\mu_k} > 0$ is that each $\nu_1(E_i) > 0$, as well as $\mu_1(E_i) > 0$. (This is achieved by tossing out any atoms with zero ν_k measure, in accordance with Proposition 4.2.1.) As explained in Section 2.6, any measurable function f on X_1 can be represented by a sequence c_i, such that $\int_{X_1} f d\mu_1 = \sum_{i\in I} c_i \mu_1(E_i)$ and $\int_{X_1} f d\nu_1 = \sum_{i\in I} c_i \nu_1(E_i)$. Similarly, when X_2 has purely atomic measures, enumerate the atoms (for both μ_2 and ν_2) by $(F_j)_{j\in J}$, so that integrals of measurable functions over X_2 become sums $\sum_{j\in J} c_j \mu_2(F_j)$ and $\sum_{j\in J} c_j \nu_2(F_j)$.

Since it is possible that measure spaces may be neither purely atomic nor atomless, the standard measure-theoretic decomposition into purely atomic and atomless parts is used. An appropriate notation, as well as that for the measures of atoms, is described below.

Definition 4.2.17.

$$X_1 = E_0 \,\dot\cup \left(\dot\bigcup_{i\in I} E_i \right), \qquad X_2 = F_0 \,\dot\cup \left(\dot\bigcup_{j\in J} F_j \right),$$

where E_0 and F_0 are atomless, each E_i is an atom for μ_1 and ν_1, and each F_j is an atom for μ_2 and ν_2.

Note that E_0 is a null set if and only if X_1 is purely atomic, while I is empty if and only if X_1 is atomless. Similarly, F_0 is null if and only if X_2 is purely atomic, and J is empty if and only if X_2 is atomless.

As noted in Lemma 2.6.7, measurable functions are almost constant on atoms. For convenience, then, functions on a purely atomic X_k will be represented as functions on the index set I and J. The same principle applies to functions $f(x_1, x_2)$, which will be written as $f(i, x_2)$ when X_1 has purely atomic measures, $f(x_1, j)$ when X_2 has purely atomic measures, or $f(i, j)$ with both.

Definition 4.2.18. Weights in the resulting sums are represented with the following brief forms.

$$u_1(i) = \mu_1(E_i) \qquad u_2(j) = \mu_2(F_j)$$
$$v_1(i) = \nu_1(E_i) \qquad v_2(j) = \nu_2(F_j)$$

The mixed norms $\sigma(P)$ and Q implicitly use weighted one-variable norms for the atomic factors, here to be denoted $\ell^{p_1}(u_1)$, $\ell^{p_2}(u_2)$, $\ell^{q_1}(v_1)$, and $\ell^{q_2}(v_2)$, as appropriate. Arguments will also include unweighted sequence norms, denoted $\ell^p(I)$ and $\ell^p(J)$, for various values $0 < p \le \infty$. The colon notation $p:q$ is used extensively in the following, as well. To repeat the definition,

$$(p:q)^{-1} = q^{-1} - p^{-1}.$$

For another notational convenience, note that a sequence (finite or infinite, depending on the index set) $(a_i)_{i\in I}$ can be referred to as a, and an ℓ^p norm $\|a_i\|_{\ell^p(I)}$ can be written as $\|a\|_{\ell^p(I)}$. Similarly, we can use the abbreviation b to refer to a sequence $(b_j)_{j\in J}$, and write $\|b\|_{\ell^p(J)} = \|b_j\|_{\ell^p(J)}$. This is used, for example, in the norm $\left\|v_2^{1/q_2}u_2^{-1/p_2}\right\|_{\ell^{p_1:q_2}(J)}$ of $\left(v_2(j)^{1/q_2}u_2(j)^{-1/p_2}\right)_{j\in J}$.

Proposition 4.2.19. *Suppose that* $0 < \max(p_2, q_2) < \min(p_1, q_1) \le \infty$ *and the two necessary one-variable inclusions* $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$ *(for* $k = 1, 2$*) hold, with best constants* C_k*. Furthermore, assume that neither* μ_1 *nor* ν_1 *is purely atomic, while* μ_2 *and* ν_2 *are purely atomic. Represent* X_2 *by a countable index set* J *with weights* u_2 *and* v_2*.*

Then $L^{\sigma(P)} \subset L^Q$ *if and only if the sequence* $(v_2(j)^{1/q_2}u_2(j)^{-1/p_2})_{j\in J}$ *is* $(q_1:q_2)$*-summable, in which case* $C_1(E_0)\left\|v_2^{1/q_2}u_2^{-1/p_2}\right\|_{\ell^{p_1:q_2}(J)} \le C \le C_1\left\|v_2^{1/q_2}u_2^{-1/p_2}\right\|_{\ell^{q_1:q_2}(J)}$*, where* E_0 *is the atomless part of* X_1*.*

Naturally, when X_1 *is atomless, this simplifies to* $C = C_1\left\|v_2^{1/q_2}u_2^{-1/p_2}\right\|_{\ell^{q_1:q_2}(J)}$*.*

Proof. First, a computation to verify that $(q_1:q_2)$-summability of $v_2^{1/q_2}u_2^{-1/p_2}$ is sufficient for inclusion, establishing the upper bound on C.

$$\begin{aligned}
\|f(x_1, j)\|_Q &= \left(\sum_{j\in J} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)}\, u_2(j)^{\frac{q_2}{p_2}}\, \frac{v_2(j)}{u_2(j)^{\frac{q_2}{p_2}}}\right)^{\frac{1}{q_2}} \\
&\le \left(\sum_{j\in J} \|f\|^{q_1}_{L^{q_1}_{\nu_1}(X_1)}\, u_2(j)^{\frac{q_1}{p_2}}\right)^{\frac{1}{q_1}} \left(\sum_{j\in J}\left(\frac{v_2(j)}{u_2(j)^{\frac{q_2}{p_2}}}\right)^{\frac{q_1}{q_1-q_2}}\right)^{\frac{q_1-q_2}{q_1q_2}} \\
&= \left\|\left\|f(x_1, j)u_2(j)^{\frac{1}{p_2}}\right\|_{L^{q_1}_{\nu_1}(X_1)}\right\|_{\ell^{q_1}(J)} \left\|\frac{v_2^{1/q_2}}{u_2^{1/p_2}}\right\|_{\ell^{q_1:q_2}(J)},
\end{aligned}$$

by Hölder's inequality with conjugate exponents $\frac{q_1}{q_2}$ and $\frac{q_1}{q_1-q_2}$. The first factor's value is independent of the order of the one-variable q_1 norms (by Tonelli's theorem if $q_1 < \infty$), so it can be reversed. Next, recall from Corollary 2.6.10 that, since $p_2 < q_1$ in the non-Minkowski case, $\|\cdot\|_{\ell^{q_1}(J)} \leq \|\cdot\|_{\ell^{p_2}(J)}$.

$$\begin{aligned}
\|f\|_Q &\leq \left\| \left\| \left(f u_2^{\frac{1}{p_2}} \right) \right\|_{\ell^{q_1}(J)} \right\|_{L^{q_1}_{\nu_1}(X_1)} \left\| v_2^{1/q_2} u_2^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \\
&\leq \left\| \left\| \left(f u_2^{\frac{1}{p_2}} \right) \right\|_{\ell^{p_2}(J)} \right\|_{L^{q_1}_{\nu_1}(X_1)} \left\| v_2^{1/q_2} u_2^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \\
&= \left\| \left(\sum_{j\in J} f(x_1, j)^{p_2} u_2(j) \right)^{\frac{1}{p_2}} \right\|_{L^{q_1}_{\nu_1}(X_1)} \left\| v_2^{1/q_2} u_2^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \\
&\leq C_1 \left\| \left(\sum_{j\in J} f(x_1, j)^{p_2} u_2(j) \right)^{\frac{1}{p_2}} \right\|_{L^{p_1}_{\mu_1}(X_1)} \left\| v_2^{1/q_2} u_2^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \\
&= C_1 \left\| v_2^{1/q_2} u_2^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \|f\|_{\sigma(P)},
\end{aligned}$$

where the final step is applying the $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$ inclusion.

Notice that, because any measurable function on a single atom F_j is constant, the local best constant $C_2(F_j)$ on each atom is

$$C_2(F_j) = v_2(j)^{1/q_2} u_2(j)^{-1/p_2}.$$

(Regardless of the values of p_2 and q_2, the best constant given in Corollary 2.6.9 reduces to that simple value.)

For any finite subset $J_0 \subset J$, let $S(J_0)$ denote the partial sum $\sum_{j\in J_0} \left(v_2(j)^{1/q_2} u_2(j)^{-1/p_2} \right)^{q_1:q_2}$. Define, for $j \in J_0$,

$$\gamma_j = S(J_0)^{-1} \left(v_2(j)^{1/q_2} u_2(j)^{-1/p_2} \right)^{q_1:q_2} = S(J_0)^{-1} C_2(F_j)^{q_1:q_2}. \tag{4.8}$$

Observe that $\sum_j \gamma_j = 1$.

Because $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$ with measures which are not purely atomic, by Corollary 2.5.7 we know that $q_1 \leq p_1$. The atomless part E_0 of X_1 is then not null, and on E_0 as a subspace $L^{p_1}_{\mu_1}(E_0) \subset L^{q_1}_{\nu_1}(E_0)$. The case $p_1 = q_1$ will be postponed, but otherwise Proposition 2.4.1 shows that the Radon-Nikodym derivative $\frac{d\nu_1}{d\mu_1}$ is in $L^{\frac{p_1}{p_1-q_1}}_{\mu_1}(E_0)$, and that the best constant on E_0 is

$$C_1(E_0) = \left(\int_{E_0} \left(\frac{d\nu_1}{d\mu_1}(x_1) \right)^{\frac{p_1}{p_1-q_1}} d\mu_1(x_1) \right)^{q_1^{-1}-p_1^{-1}} \leq C_1(X_1) < \infty.$$

If we define the measure $\lambda_1 = \left(\frac{d\nu_1}{d\mu_1} \right)^{\frac{p_1}{p_1-q_1}} \mu_1$, this means that $\lambda_1(E_0) = C_1(E_0)^{p_1:q_1}$.

Corollary 4.1.5 provides a disjoint sequence $\left(A_j \right)_{j\in J_0}$ of measurable subsets which partitions E_0, each

$$\lambda_1(A_j) = \gamma_j \lambda_1(E_0) = \gamma_j C_1(E_0)^{p_1:q_1}.$$

(Since $\sum_{j\in J_0}\gamma_j = 1$, the total $\sum_{j\in J_0}\lambda_1(A_j) = \lambda_1(E_0)$.)

Proposition 2.4.1 can be applied to each A_j to show that $C_1(A_j) = \left(\int_{A_j}\left(\frac{d\nu_1}{d\mu_1}\right)^{\frac{p_1}{p_1-q_1}}\,d\mu_1\right)^{q_1^{-1}-p_1^{-1}}$; in other words, $\lambda_1(A_j) = \int_{A_j}\left(\frac{d\nu_1}{d\mu_1}(x_1)\right)^{\frac{p_1}{p_1-q_1}}\,d\mu_1(x_1) = C_1(A_j)^{p_1\,:\,q_1}$, and therefore each

$$C_1(A_j) = \gamma_j^{\frac{1}{p_1\,:\,q_1}}C_1(E_0).$$

The necessary condition using block factorable functions, Theorem 4.2.11, yields

$$C \geq \left\|C_1(A_j)C_2(B_j)\right\|_{\ell^{p_1\,:\,q_2}(J_0)}.$$

Applied with the A_j above and $B_j = F_j$, and using (4.8), this means that

$$\begin{aligned}
C &\geq \left\|\gamma_j^{\frac{1}{p_1\,:\,q_1}}C_1(E_0)C_2(F_j)\right\|_{\ell^{p_1\,:\,q_2}(J_0)}\\
&= \frac{C_1(E_0)}{S(J_0)^{p_1\,:\,q_1}}\left\|C_2(F_j)^{\frac{q_1\,:\,q_2}{p_1\,:\,q_2}}\right\|_{\ell^{p_1\,:\,q_2}(J_0)}\\
&= C_1(E_0)\frac{\left(\sum_{j\in J_0}C_2(F_j)^{q_1\,:\,q_2}\right)^{\frac{1}{p_1\,:\,q_2}}}{\left(\sum_{j\in J_0}C_2(F_j)^{q_1\,:\,q_2}\right)^{\frac{1}{p_1\,:\,q_1}}}\\
&= C_1(E_0)\left(\sum_{j\in J_0}C_2(F_j)^{q_1\,:\,q_2}\right)^{\frac{1}{q_1\,:\,q_2}}\\
&= C_1(E_0)\left\|\nu_2(j)^{1/q_2}u_2(j)^{-1/p_2}\right\|_{\ell^{q_1\,:\,q_2}(J_0)}.
\end{aligned}$$

This provides the lower bound on C, by taking the supremum over the various finite $J_0 \subset J$. As a consequence, if $\left(\nu_2^{1/q_2}u_2^{-1/p_2}\right)$ is not $(q_1 : q_2)$-summable, this lower bound can be made arbitrarily large for various $J_0 \subset J$. In this case, $C = \infty$.

The case $p_1 = q_1 = \infty$ is simple. For any measurable $E \subset X_1$ with $\nu_1 E > 0$, the local best constant $C_1(E) = 1$. (Of course, the best constant on a ν_1-null set would be zero, but the hypothesis that both measures be non-zero in the one-variable inclusion problem $L^p_\mu(X) \subset L^q_\nu(X)$ would not be satisfied on such a set.) We need only obtain any disjoint sequence $\left(A_j\right)_{j\in J}$ of measurable $A_j \subset X_1$ such that each $\nu_1 A_j > 0$, which is easy to do within E_0 since ν_1 is atomless there. (With $j \geq 1$, say, use Corollary 4.1.5 to obtain sets with $\nu_1 A_j = 2^{-j}\nu_1 E_0$, shrinking E_0 to a subset with finite ν_1 measure if necessary.) Then each $C_1(A_j) = 1$, so applying Theorem 4.2.11 with $B_j = F_j$ yields (since $p_1 = q_1$)

$$C \geq \left\|C_2(F_j)\right\|_{\ell^{p_1\,:\,q_2}(J)} = \left\|\nu_2^{1/q_2}u_2^{-1/p_1}\right\|_{\ell^{q_1\,:\,q_2}(J)} = C_1\left\|\nu_2^{1/q_2}u_2^{-1/p_1}\right\|_{\ell^{q_1\,:\,q_2}(J)}.$$

Finally, suppose that $p_1 = q_1 < \infty$. In this case, $C_1(E) = \operatorname{ess\,sup}_E\left(\frac{d\nu_1}{d\mu_1}\right)^{1/q_1}$ for any measurable $E \subset X_1$, including $C_1 = C_1(X_1)$. (For the following, recall that the assumption $\frac{d\nu_k}{d\mu_k} > 0$ means that any subset of each X_k has positive μ_k measure if and only if it has positive ν_k measure. Such sets will simply be said to "have positive measure".) Fix an arbitrary $\varepsilon > 0$ and note that the subset of E_0 where $\left(\frac{d\nu_1}{d\mu_1}\right)^{1/q_1} \geq C_1(E_0) - \varepsilon$ has positive measure. Take a collection $\left(A_j\right)_{j\in J}$ of disjoint measurable subsets of that set, each with positive measure, possible since the measures μ_1 and ν_1 are atomless on E_0. Observe that each $C_1(A_j) = \operatorname{ess\,sup}_{A_j}\left(\frac{d\nu_1}{d\mu_1}\right)^{1/q_1} \geq C_1(E_0) - \varepsilon$. Apply Theorem 4.2.11 with such A_j and let each B_j be the atom F_j, again with $C_2(F_j) = \nu_2(j)^{1/q_2}u_2(j)^{-1/p_2}$.

$$
\begin{aligned}
C &\geq \left\|C_1(A_j)C_2(B_j)\right\|_{\ell^{p_1:q_2}(J)} \geq \left\|(C_1(E_0)-\varepsilon)\,C_2(F_j)\right\|_{\ell^{p_1:q_2}(J)} \\
&= (C_1(E_0)-\varepsilon)\left\|\nu_2^{1/q_2}u_2^{-1/p_2}\right\|_{\ell^{q_1:q_2}(J)}
\end{aligned}
$$

This applies with arbitrary $\varepsilon > 0$, so with $\varepsilon \to 0^+$, $C \geq C_1\left\|\nu_2^{1/q_2}u_2^{-1/p_2}\right\|_{\ell^{q_1:q_2}(J)}$. □

Proposition 4.2.20. *Suppose that* $0 < \max(p_2, q_2) < \min(p_1, q_1) \leq \infty$ *and the two necessary one-variable inclusions* $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$ *(for* $k = 1, 2$*) hold, with best constants* C_k*. Furthermore, assume that* μ_1 *and* ν_1 *are purely atomic, while neither* μ_2 *nor* ν_2 *is purely atomic. Represent* X_1 *by a countable index set* I *with weights* u_1 *and* ν_1*.*

Then $L^{\sigma(P)} \subset L^Q$ *if and only if the sequence* $\left(\nu_1(i)^{1/q_1}u_1(i)^{-1/p_1}\right)_{i\in I}$ *is* $(p_1 : p_2)$*-summable, in which case* $C_2(F_0)\left\|\nu_1^{1/q_1}u_1^{-1/p_1}\right\|_{\ell^{p_1:p_2}(I)} \leq C \leq C_2\left\|\nu_1^{1/q_1}u_1^{-1/p_1}\right\|_{\ell^{p_1:p_2}(I)}$*, where* F_0 *is the atomless part of* X_2*.*

Naturally, when X_2 *is atomless, this simplifies to* $C = C_2\left\|\nu_1^{1/q_1}u_1^{-1/p_1}\right\|_{\ell^{p_1:p_2}(I)}$*.*

Proof. First, a computation to verify that $(p_1 : p_2)$-summability of $\nu_1^{1/q_1}u_1^{-1/p_1}$ is sufficient for inclusion. Using $L^{p_2}_{\mu_2}(X_2) \subset L^{q_2}_{\nu_2}(X_2)$,

$$
\|f(i,x_2)\|_Q = \left(\sum_{j\in J}\|f\|^{q_2}_{\ell^{q_1}(\nu_1)}\right)^{\frac{1}{q_2}} \leq C_2\left(\int_{X_2}\left\|f\nu_1^{1/q_1}\right\|^{p_2}_{\ell^{q_1}(I)}d\mu_2(x_2)\right)^{\frac{1}{p_2}}.
$$

In the non-Minkowski case, $p_2 < q_1$, so by Corollary 2.6.10, $\|\cdot\|_{\ell^{q_1}(I)} \leq \|\cdot\|_{\ell^{p_2}(I)}$. Together with Tonelli's inequality (to exchange the order of the L^{p_2} one-variable norms),

$$
\begin{aligned}
\|f\|_Q &\leq C_2\left(\int_{X_2}\left\|f\nu_1^{1/q_1}\right\|^{p_2}_{\ell^{p_2}(I)}d\mu_2(x_2)\right)^{\frac{1}{p_2}} = C_2\left\|\left\|f\nu_1^{1/q_1}\right\|_{L^{p_2}_{\mu_2}(X_2)}\right\|_{\ell^{p_2}(I)} \\
&= C_2\left(\sum_{i\in I}\|f\|^{p_2}_{L^{p_2}_{\mu_2}(X_2)}u_1(i)^{p_2/p_1}\left(\nu_1(i)^{1/q_1}u_1(i)^{-1/p_1}\right)^{p_2}\right)^{\frac{1}{p_2}}.
\end{aligned}
$$

By Hölder's inequality with the conjugate exponents $\frac{p_1}{p_2}$ and $\frac{p_1}{p_1-p_2}$,

$$
\begin{aligned}
\|f\|_Q &\leq C_2\left\|\|f\|_{L^{p_2}_{\mu_2}(X_2)}u_1^{1/p_1}\right\|_{\ell^{p_1}(I)}\left\|\nu_1^{1/q_1}u_1^{-1/p_1}\right\|_{\ell^{\frac{p_1p_2}{p_1-p_2}}(I)} \\
&= C_2\left\|\nu_1^{1/q_1}u_1^{-1/p_1}\right\|_{\ell^{p_1:p_2}(I)}\|f\|_{\sigma(P)}.
\end{aligned}
$$

The lower bound on C proceeds much as in the previous proposition. Because any measurable function on a single atom E_i is constant, the local best constant $C_1(E_i)$ is

$$
C_1(E_i) = \nu_1(i)^{1/q_1}u_1(i)^{-1/p_1}.
$$

For any finite subset $I_0 \subset I$, let $S(I_0)$ denote the partial sum $\sum_{i\in I_0}\left(v_1(i)^{1/q_1}u_1(i)^{-1/p_1}\right)^{p_1\,:\,p_2}$. Define, for $i \in I_0$,

$$\gamma_i = S(I_0)^{-1}\left(v_1(i)^{1/q_1}u_1(i)^{-1/p_1}\right)^{p_1\,:\,p_2} = S(I_0)^{-1}C_1(E_i)^{p_1\,:\,p_2}, \tag{4.9}$$

so that $\sum_i \gamma_i = 1$.

Because $L^{p_2}_{\mu_2}(X_2) \subset L^{q_2}_{\nu_2}(X_2)$ with measures which are not purely atomic, by Corollary 2.5.7 we know that $q_2 \le p_2$. The atomless part F_0 of X_2 is then not null, and on F_0 as a subspace $L^{p_2}_{\mu_2}(F_0) \subset L^{q_2}_{\nu_2}(F_0)$. Postponing the case $p_2 = q_2$, otherwise Proposition 2.4.1 shows that the Radon-Nikodym derivative $\frac{d\nu_2}{d\mu_2}$ is in $L^{\frac{p_2}{p_2-q_2}}_{\nu_2}(F_0)$, and that the best constant in F_0 is

$$C_2(F_0) = \left(\int_{F_0}\left(\frac{d\nu_2}{d\mu_2}(x_2)\right)^{\frac{p_2}{p_2-q_2}} d\mu_2(x_2)\right)^{q_2^{-1}-p_2^{-1}} \le C_2(X_2) < \infty.$$

Define $\lambda_2 = \left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{p_2}{p_2-q_2}}\mu_2$, so that $\lambda_2(F_0) = C_2(F_0)^{p_1\,:\,q_2}$.

Partition F_0 into a disjoint sequence $(B_i)_{i\in I_0}$ of measurable subsets, by Corollary 4.1.5, with each

$$\lambda_2(B_i) = \gamma_i\lambda_2(F_0) = \gamma_i C_2(F_0)^{p_2\,:\,q_2}.$$

Proposition 2.4.1 shows that each $C_2(B_i) = \left(\int_{B_i}\left(\frac{d\nu_2}{d\mu_2}\right)^{\frac{p_2}{p_2-q_2}} d\mu_2\right)^{q_2^{-1}-p_2^{-1}}$. Then

$$C_2(B_i) = \gamma_i^{\frac{1}{p_2\,:\,q_2}} C_2(F_0).$$

Use $A_i = E_i$ and the B_i above, plus (4.9), in the necessary condition for block factorable functions, Theorem 4.2.11.

$$\begin{aligned}
C &\ge \|C_1(A_i)C_2(B_i)\|_{\ell^{p_1\,:\,q_2}(I_0)}\\
&= \frac{C_2(F_0)}{S(I_0)^{\frac{1}{p_2\,:\,q_2}}}\left\|C_1(E_i)^{\frac{p_1\,:\,p_2}{p_1\,:\,q_2}}\right\|_{\ell^{p_1\,:\,q_2}(I_0)}\\
&= C_2(F_0)\frac{\left(\sum_{i\in I_0} C_1(E_i)^{p_1\,:\,p_2}\right)^{\frac{1}{p_1\,:\,q_2}}}{\left(\sum_{i\in I_0}\left(v_1^{1/q_1}u_1^{-1/p_1}\right)^{p_1\,:\,p_2}\right)^{\frac{1}{p_2\,:\,q_2}}}\\
&= C_2(F_0)\left(\sum_{i\in I_0}\left(v_1^{1/q_1}u_1^{-1/p_1}\right)^{p_1\,:\,p_2}\right)^{\frac{1}{p_1\,:\,p_2}}\\
&= C_2(F_0)\left\|v_1^{1/q_1}u_1^{-1/p_1}\right\|_{\ell^{p_1\,:\,p_2}(I_0)}.
\end{aligned}$$

When $\left(v_1^{1/q_1}u_1^{-1/p_1}\right)_{i\in I}$ is not $(p_1 : p_2)$-summable, this lower bound can be arbitrarily large for different choices of I_0, so $C = \infty$.

In the non-Minkowski case, neither p_2 nor q_2 can be ∞, so all that remains is $p_2 = q_2 < \infty$. In this case, for any measurable $F \subset X_2$, $C_2(F) = \operatorname{ess\,sup}_F\left(\frac{d\nu_2}{d\mu_2}\right)^{1/q_2}$. Fix an arbitrary $\varepsilon > 0$ and note that $\left(\frac{d\nu_2}{d\mu_2}\right)^{1/q_2} \ge C_2(F_0)-\varepsilon$ on a subset of F_0 of positive measure. Since the measures on F_0

are atomless, we can take a collection $(B_i)_{i\in I}$ of disjoint measurable subsets, each with positive measure and $C_2(B_i) = \operatorname{ess\,sup}_{B_i} \left(\frac{d\nu_2}{d\mu_2}\right)^{1/q_2} \geq C_2(F_0) - \varepsilon$. By Theorem 4.2.11 with each $A_i = E_i$, recalling that $C_1(E_i) = v_1(i)^{1/q_1} u_1(i)^{-1/p_1}$,

$$\begin{aligned} C &\geq \|C_1(E_i)\,(C_2(F_0) - \varepsilon)\|_{\ell^{p_1:q_2}(I)} \\ &= (C_2(F_0) - \varepsilon)\left\|v_1^{1/q_1} u_1^{-1/p_1}\right\|_{\ell^{p_1:p_2}(I)}. \end{aligned}$$

As $\varepsilon \to 0^+$, $C \geq C_2(F_0)\left\|v_1^{1/q_1} u_1^{-1/p_1}\right\|_{\ell^{p_1:p_2}(I)}$. Once more, $v_1^{1/q_1} u_1^{-1/p_1}$ must be $(p_1 : p_2)$-summable to have $C < \infty$. □

Chapter 5

Non-Minkowski case with two atomic measures

5.1 Weight summability sufficient conditions

When all measures are purely atomic, we represent integrals in either space as weighted series, with weights $(u_1(i))_{i\in I}$ for μ_1, $(v_1(i))_{i\in I}$ for v_1, $(u_2(j))_{j\in J}$ for μ_2, and $(v_2(j))_{j\in J}$ for v_2. The previous arguments that $(q_1 : q_2)$-summability of $v_2^{1/q_2}u_2^{-1/p_2}$ and $(p_1 : p_2)$-summability of $v_1^{1/q_1}u_1^{-1/p_1}$ are sufficient conditions for $L^{\sigma(P)} \subset L^Q$ still apply, just as when one space has purely atomic measures. However, they might not now be necessary, as it may not be possible to produce similar counterexamples when both spaces are purely atomic.

The following results establish that $(p_1 : q_2)$-summability of $v_2^{1/q_2}u_2^{-1/p_2}$ or $v_1^{1/q_1}u_1^{-1/p_1}$ is also sufficient for $L^{\sigma(P)} \subset L^Q$. Although these conditions are also valid for the case of one purely atomic space, they are less useful there.

Specifically, when X_1 has measures which are not purely atomic, by Corollary 2.5.7 the necessary inclusion $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{v_1}(X_1)$ implies that $q_1 \le p_1$. Therefore $q_2^{-1} - q_1^{-1} \le q_2^{-1} - p_1^{-1}$, so $\left(q_2^{-1} - p_1^{-1}\right)^{-1} = p_1 : q_2 \le q_1 : q_2 = \left(q_2^{-1} - q_1^{-1}\right)^{-1}$. Corollary 2.6.10 then establishes that $(p_1 : q_2)$-summability implies $(q_1 : q_2)$-summability. Any theorem should be given with the weakest hypothesis possible, so when X_1 does not have purely atomic measures, the weaker $(q_1 : q_2)$-summability condition on $\left(v_2^{1/q_2}u_2^{-1/p_2}\right)$ (which, in that case, is also necessary) is preferred. Similarly, when X_2 does not have purely atomic measures, the weaker $(p_1 : p_2)$-summability condition on $\left(v_1^{1/q_1}u_1^{-1/p_1}\right)$ is used.

When both X_1 and X_2 have purely atomic measures, the preferred condition may vary. When $p_1 > q_1$, $(q_1 : q_2)$-summability of $\left(v_2^{1/q_2}u_2^{-1/p_2}\right)$ is the preferred sufficient condition for inclusion, but it is also possible that $p_1 < q_1$, so that $(p_1 : q_2)$-summability is preferred. Similarly, when $p_2 > q_2$, the weaker inclusion condition is $(p_1 : p_2)$-summability of $\left(u_1^{1/q_1}u_1^{-1/p_1}\right)$, while when $p_2 < q_2$, we would look for $(p_1 : q_2)$-summability.

Proposition 5.1.1. *Suppose that* $0 < \max(p_2, q_2) < \min(p_1, q_1) \le \infty$ *and that* μ_2 *and* v_2 *are purely atomic. Represent* X_2 *by a countable index set* J *with weights* u_2 *and* v_2.

If the sequence $\left(v_2(j)^{1/q_2}u_2(j)^{-1/p_2}\right)_{j\in J}$ *is* $(p_1 : q_2)$*-summable, then* $L^{\sigma(P)} \subset L^Q$*, with*

$$C \le C_1 \left\| v_2^{1/q_2}u_2^{-1/p_2} \right\|_{\ell^{p_1 : q_2}(J)}.$$

Proof. Consider any $f(x_1, j) \in L^+(X_1 \times J)$. Using $L^{p_1}_{\mu_1}(X_1) \subset L^{q_1}_{\nu_1}(X_1)$,

$$\begin{aligned}\|f(x_1, j)\|_Q &= \left(\sum_{j\in J} \|f\|^{q_2}_{L^{q_1}_{\nu_1}(X_1)} u_2(j)^{\frac{q_2}{p_2}} \left(v_2(j)^{1/q_2} u_2(j)^{-1/p_2}\right)^{q_2}\right)^{\frac{1}{q_2}} \\ &\leq C_1 \left(\sum_{j\in J} \|f\|^{q_2}_{L^{p_1}_{\mu_1}(X_1)} u_2(j)^{\frac{q_2}{p_2}} \left(v_2(j)^{1/q_2} u_2(j)^{-1/p_2}\right)^{q_2}\right)^{\frac{1}{q_2}}.\end{aligned}$$

Next, Hölder's inequality with the conjugate exponents $\frac{p_1}{q_2}$ and $\frac{p_1}{p_1-q_2}$ yields

$$\|f\|_Q \leq C_1 \left\| \|f\|_{L^{p_1}_{\mu_1}(X_1)} u_2(j)^{\frac{1}{p_2}} \right\|_{\ell^{p_1}(J)} \left(\sum_{j\in J} \left(v_2(j)^{1/q_2} u_2(j)^{-1/p_2}\right)^{\frac{p_1 q_2}{p_1-q_2}}\right)^{\frac{p_1-q_2}{p_1 q_2}}.$$

Reversing the order of the L^{p_1} norms (by Tonelli's theorem if $p_1 < \infty$),

$$\|f\|_Q \leq C_1 \left\| v_2(j)^{1/q_2} u_2(j)^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \left\| \left\| f u_2^{1/p_2} \right\|_{\ell^{p_1}(J)} \right\|_{L^{p_1}_{\mu_1}(X_1)}.$$

Corollary 2.6.10 shows that $\|\cdot\|_{\ell^{p_1}(J)} \leq \|\cdot\|_{\ell^{p_2}(J)}$, so

$$\begin{aligned}\|f\|_Q &\leq C_1 \left\| v_2(j)^{1/q_2} u_2(j)^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \left\| \left\| f u_2^{1/p_2} \right\|_{\ell^{p_2}(J)} \right\|_{L^{p_1}_{\mu_1}(X_1)} \\ &= C_1 \left\| v_2(j)^{1/q_2} u_2(j)^{-1/p_2} \right\|_{\ell^{q_1:q_2}(J)} \|f\|_{\sigma(P)}.\end{aligned}$$

□

Proposition 5.1.2. *Suppose that* $0 < \max(p_2, q_2) < \min(p_1, q_1) \leq \infty$ *and that* μ_1 *and* ν_1 *are purely atomic. Represent* X_1 *by a countable index set* I *with weights* u_1 *and* v_1.

If the sequence $\left(v_1(i)^{1/q_1} u_1(i)^{1/p_1}\right)_{i\in I}$ *is* $(p_1 : q_2)$*-summable, then* $L^{\sigma(P)} \subset L^Q$*, with* $C \leq C_2 \left\| v_1^{1/q_1} u_1^{-1/p_1} \right\|_{\ell^{p_1:q_2}(I)}$.

Proof. Consider any $f(i, x_2) \in L^+(I \times X_2)$. Because $q_2 < q_1$, Corollary 2.6.10 provides $\|\cdot\|_{\ell^{q_1}(I)} \leq \|\cdot\|_{\ell^{q_2}(I)}$.

$$\begin{aligned}\|f(i, x_2)\|_Q &= \left(\int_{X_1} \|f(i, x_2)\|^{q_2}_{\ell^{q_1}(I)} v_1(i)^{\frac{q_2}{q_1}} d\nu_2(x_2)\right)^{\frac{1}{q_2}} \\ &\leq \left(\int_{X_1} \left(\sum_{i\in I} f(i, x_2)^{q_2}\right) v_1(i)^{\frac{q_2}{q_1}} d\nu_2(x_2)\right)^{\frac{1}{q_2}} \\ &= \left(\sum_{i\in I} \left(\int_{X_1} f(i, x_2)^{q_2} d\nu_2(x_2)\right) v_1(i)^{\frac{q_2}{q_1}}\right)^{\frac{1}{q_2}}\end{aligned}$$

with the order of integration reversed by Tonelli's theorem. After some rewriting, apply Hölder's inequality with the conjugate exponents $\frac{p_1}{q_2}$ and $\frac{p_1}{p_1-q_2}$.

$$\|f\|_Q \le \left(\sum_{i\in I} \left(\int_{X_2} f(i,x_2)^{q_2} d\nu_2(x_2) \right) u_1(i)^{\frac{q_2}{p_1}} \left(v_1(i)^{1/q_1} u_1(i)^{-1/p_1} \right)^{q_2} \right)^{\frac{1}{q_2}}$$
$$\le \left\| \left(\int_{X_2} f(i,x_2)^{q_2} d\nu_2(x_2) \right)^{\frac{1}{q_2}} u_1(i)^{\frac{1}{p_1}} \right\|_{\ell^{p_1}(I)} \left(\sum_{i\in I} \left(v_1(i)^{1/q_1} u_1(i)^{-1/p_1} \right)^{\frac{p_1 q_2}{p_1-q_2}} \right)^{\frac{p_1-q_2}{p_1 q_2}} .$$

The inclusion $L^{p_2}_{\mu_2}(X_2) \subset L^{q_2}_{\nu_2}(X_2)$ then yields

$$\|f\|_Q \le C_2 \left\| \|f\|_{L^{p_2}_{\mu_2}(X_2)} \right\|_{\ell^{p_1}(u_1)} \left\| v_1^{1/q_1} u_1^{-1/p_1} \right\|_{\ell^{p_1 : q_2}(I)}$$
$$= C_2 \left\| v_1^{1/q_1} u_1^{-1/p_1} \right\|_{\ell^{p_1 : q_2}(I)} \|f\|_{\sigma(P)} .$$

□

5.2 Classifying extremal functions

With both measures purely atomic, we identify functions on $X_1 \times X_2$ with functions $f(i,j)$ on $I \times J$, for some (at most) countable index sets I and J. As noted earlier, the measures of atoms become weights on the elements of I and J, so that, where $(E_i)_{i\in I}$ enumerates the atoms on X_1 and $\left(F_j\right)_{j\in J}$ enumerates the atoms on X_2,

$$u_1(i) = \mu_1(E_i) \qquad u_2(j) = \mu_2(F_j)$$
$$v_1(i) = \nu_1(E_i) \qquad v_2(j) = \nu_2(F_j).$$

Recall that Proposition 4.2.1 reduces to the case where each $\frac{d\nu_k}{d\mu_k} > 0$ μ_k-a.e., which is assumed. This way, each pair μ_k and ν_k has the same atoms, so the phrase "the atoms on X_k" is not ambiguous. This also means that the u_k and v_k sequences are strictly positive.

The functions $f(i,j)$ can be thought of as possibly infinite matrices $\left(f_{i,j}\right)$. Matrices of specific forms may prove particularly easy to analyze; Corollary 4.2.12, with its combinations of characteristic functions of blocks, in this purely atomic case is about considering matrices which, up to reordering of rows and columns, are block diagonal and constant on each block. The expression it gives as a lower bound on C is computed as a best constant for all such functions with a particular division into blocks; it gives an $\ell^{p_1 : q_2}$ norm of terms each of which is a best constant for constant functions on the block $A_i \times B_i$.

Measurable functions are always (almost everywhere) constant on atoms, so by using the atoms E_i and F_j as blocks (corresponding to singletons $i \in I$ and $j \in J$ for $f(i,j)$) the special case of blockwise constant functions actually becomes general; in terms of matrices, it is only natural that each entry contains a single value. Given finite or infinite sequences (i_n) in I and (j_n) in J, the inequality from Corollary 4.2.12 becomes

$$C \ge \left(\sum_n \left(\frac{v_1(i_n)^{1/q_1}}{u_1(i_n)^{1/p_1}} \frac{v_2(j_n)^{1/q_2}}{u_2(i_n)^{1/p_2}} \right)^{p_1 : q_2} \right)^{\frac{1}{p_1 : q_2}} ,$$

obtained by considering functions which correspond to matrices $(f_{i,j})$ with at most one entry in each row and in each column, and which can thus be rearranged to be diagonal.

(To clarify, lest we be tempted to start thinking about diagonalizability, this matrix conceptualization is only a convenience, to provide a simple way to imagine these purely atomic functions and to describe certain special types which are easy to work with. The term "matrix" naturally brings to mind linear algebra and the properties of well-behaved linear operators, but none of that is actually used here. While it is possible that matrix multiplication could provide a meaningful operation which would enrich these ideas somehow, it is not clear that there is a reasonable application of linear algebra here.)

Although special classes of functions give lower bounds on C, to obtain upper bounds we must either consider more general functions or prove that the general best constant C can be achieved, or at least approached, by the ratio $\|f\|_Q \,/\, \|f\|_{\sigma(P)}$ for functions f from a special class. Ideally, we want conditions when it's sufficient to consider functions representable by diagonal matrices (for some order of I and J), i.e. when for each $i \in I$ there is at most one j such that $f(i, j)$ is nonzero, and for each $j \in J$ there is at most one i such that $f(i, j)$ is nonzero. If not, it still may be possible to allow at most one non-zero entry per row, or at most one non-zero entry per column.

First, a demonstration that it is sufficient to work with functions represented by finite-dimensional matrices.

Proposition 5.2.1. *Let C denote the least nonnegative constant such that, for any $f \in L^+(I\times J)$, $\|f\|_Q \le C\,\|f\|_{\sigma(P)}$, and recall that $C = \sup_{f\neq 0} \|f\|_Q/\|f\|_{\sigma(P)}$. Then there is a sequence (f_k) such that:*

- *Each f_k has a corresponding integer N_k such that $f_k(i, j) = 0$ except on $I_0(k) \times J_0(k)$, for particular finite sets $I_0(k) \subset I$ and $J_0(k) \subset J$ with cardinality at most N_k.*
- $\lim_{k\to\infty} \frac{\|f_k\|_Q}{\|f_k\|_{\sigma(P)}} = C.$

Proof. Fix any enumeration (i_m) of I and any enumeration (j_m) of J. For each $k \ge 1$, there is some g_k such that

$$C - \frac{1}{k} < \frac{\|g_k\|_Q}{\|g_k\|_{\sigma(P)}} \le C.$$

(Take $g_k = |g_k| \ge 0$, valid since this doesn't change the norms at all.) For $N \ge 1$, let $\chi_N(i_m, j_{\bar m}) = 1$ if $m \le N$ and $\bar m \le N$, and $\chi_N(i, j) = 0$ otherwise. Every L^p norm has the Fatou property, so mixed L^p norms do as well (as stated in Part I of [4] immediately following the mixed-norm triangle inequality); therefore,

$$\lim_{N\to\infty} \frac{\|g_k\chi_N\|_Q}{\|g_k\chi_N\|_{\sigma(P)}} = \frac{\|g_k\|_Q}{\|g_k\|_{\sigma(P)}}$$

because $\lim_{N\to\infty} g_k\chi_N = g_k$ pointwise.

(Recall that a Banach function norm ρ is said to have the Fatou property when $0 \le h_n \nearrow h$ implies $\rho(h_n) \uparrow \rho(h)$. The Monotone Convergence Theorem establishes this for L^p with $p < \infty$, and it's fairly straightforward for L^∞ as well.)

Therefore there is some N_k such that $\frac{\|g_k\chi_{N_k}\|_Q}{\|g_k\chi_{N_k}\|_{\sigma(P)}}$ is within $\frac{1}{k}$ of $\frac{\|g_k\|_Q}{\|g_k\|_{\sigma(P)}}$. Let $f_k = g_k\chi_{N_k}$ and the triangle inequality gives

$$C - \frac{2}{k} < \frac{\|f_k\|_Q}{\|f_k\|_{\sigma(P)}} \le C.$$

Of course, each f_k is supported on the set $I_0(k) \times J_0(k)$, where $I_0(k) = \{i_m : m \le N_k\}$ and $J_0(k) = \{j_m : m \le N_k\}$. Each factor has cardinality at most N_k. □

Corollary 5.2.2. *Let C denote the least nonnegative constant such that, for any $f \in L^+(I \times J)$, $\|f\|_Q \le C\,\|f\|_{\sigma(P)}$. For any particular finite subsets $I_0 \subset I$ and $J_0 \subset J$, let $C(I_0, J_0)$ denote the least nonnegative constant such that, for any $f \in L^+(I, J)$ supported on $I_0 \times J_0$ (that is, zero off $I_0 \times J_0$), $\|f\|_Q \le C(I_0, J_0)\,\|f\|_{\sigma(P)}$.*

Then C is the supremum of $C(I_0, J_0)$ over all finite subsets $I_0 \subset I$ and $J_0 \subset J$.

Proof. By Proposition 5.2.1, there is a sequence (f_k) of functions, each $f_k \in L^+(I \times J)$ supported on $I_0(k) \times J_0(k)$ for some finite subsets $I_0(k) \subset I$ and $J_0(k) \subset J$, such that $\lim_{k\to\infty} \|f_k\|_Q \,/\, \|f_k\|_{\sigma(P)} = C$. Any $C(I_0, J_0)$ is the best constant for a subset of $L^+(I \times J)$, so $C(I_0, J_0) \le C$. Therefore

$$C = \sup_k \frac{\|f_k\|_Q}{\|f_k\|_{\sigma(P)}} \le \sup_k C\left(I_0(k), J_0(k)\right) \le \sup_{I_0, J_0} C\left(I_0, J_0\right) \le C.$$

□

Proposition 5.2.3. *Suppose that $0 < p_2 \le q_2 < \min(p_1, q_1) \le \infty$ (the non-Minkowski case with the additional constraint $p_2 \le q_2$) and let $I_0 \subset I$ and $J_0 \subset J$ be any finite subsets of the index sets for $f(i, j)$. Let $C(I_0, J_0)$ be the least constant such that, for any $f(i, j)$ for which all nonzero values have $i \in I_0$ and $j \in J_0$,*

$$\|f\|_Q \le C(I_0, J_0)\,\|f\|_{\sigma(P)}\,.$$

Then $C(I_0, J_0)$ can be achieved as $\|f\|_Q \,/\, \|f\|_{\sigma(P)}$ where f has at most one non-zero entry per row, i.e. for each fixed $i \in I_0$, there is at most one $j \in J_0$ such that $f(i, j) \ne 0$.

Proof. Let $f(i, j) \in L^+(I \times J)$ be supported on $I_0 \times J_0$, i.e. zero except for $(i, j) \in I_0 \times J_0$, with $\|f\|_{\sigma(P)} = 1$, $\|f\|_Q = C(I_0, J_0)$, and among such functions the fewest possible non-zero values. (The maximum is achieved since $L^+(I_0 \times J_0)$ is finite-dimensional, so the unit sphere of $L^{\sigma(P)}$ there is compact, and any continuous function achieves its optima on a compact domain. Well-ordering of the natural numbers provides that among those functions achieving the maximum, there is a minimum number of non-zero values.) Suppose, in order to produce a contradiction, that there is some $r \in I_0$ with distinct $s_1, s_2 \in J_0$ such that both $f(r, s_1) > 0$ and $f(r, s_2) > 0$.

First suppose that $q_1 = \infty$. In this case, define $g(i, j)$ to match $f(i, j)$, except replacing whichever of $f(r, s_1)$ and $f(r, s_2)$ is lesser by zero. Because $\max_{j\in J_0} f(r, j) = \max_{j\in J_0} g(r, j)$, $\|f\|_Q = \|g\|_Q$. Then, since $g \le f$, $\|g\|_{\sigma(P)} \le \|f\|_{\sigma(P)} = 1$. Define $h(i, j) = \|g\|_{\sigma(P)}^{-1}\, g(i, j)$ and observe that

$$\|h\|_{\sigma(P)} = \frac{\|g\|_{\sigma(P)}}{\|g\|_{\sigma(P)}} = 1 \qquad \text{while} \qquad \|h\|_Q = \frac{\|g\|_Q}{\|g\|_{\sigma(P)}} \ge \|g\|_Q = C(I_0, J_0).$$

This means that h is in the $L^{\sigma(P)}$ unit sphere and achieves the maximum value $C(I_0, J_0)$ of the L^Q mixed norm on it. However, h has one fewer non-zero value than f, contradicting the minimality of the number of non-zero values in f.

For the remainder, suppose that $q_1 < \infty$. Let T denote the open interval

$$T = (-f(r, s_1)^{p_2} u_2(s_1), f(r, s_2)^{p_2} u_2(s_2))$$

and define

$$a(\theta) = \left(f(r, s_1)^{p_2} + \frac{\theta}{u_2(s_1)} \right)^{\frac{1}{p_2}} \quad \text{and} \quad b(\theta) = \left(f(r, s_2)^{p_2} - \frac{\theta}{u_2(s_2)} \right)^{\frac{1}{p_2}}$$

for θ in the closed interval $\overline{T} = [-f(r, s_1)^{p_2} u_2(s_1), f(r, s_2)^{p_2} u_2(s_2)]$. Note that $a(0) = f(r, s_1)$, $b(0) = f(r, s_2)$, and $a^{p_2} u_2(s_1) + b^{p_2} u_2(s_2)$ is constant. For $\theta \in T$, $a'(\theta) > 0$ and $b'(\theta) < 0$. On this open interval, differentiating the constant $a(\theta)^{p_2} u_2(s_1) + b(\theta)^{p_2} u_2(s_2)$ shows that

$$a^{p_2 - 1} a' u_2(s_1) = b^{p_2 - 1} (-b') \, u_2(s_2).$$

Take the logarithm and differentiate to get

$$\begin{aligned} (p_2 - 1) \log a + \log a' + \log u_2(s_1) &= (p_2 - 1) \log b + \log(-b') + \log u_2(s_2) \\ (p_2 - 1) \frac{a'}{a} + \frac{a''}{a'} &= (p_2 - 1) \frac{b'}{b} + \frac{b''}{b'} \end{aligned} \tag{5.1}$$

For each value $\theta \in T$, define a modified version f_θ of f by

$$f_\theta(i, j) = \begin{cases} a(\theta) & \text{if } i = r, j = s_1 \\ b(\theta) & \text{if } i = r, j = s_2 \\ f(i, j) & \text{otherwise} \end{cases}$$

and observe that $f_0 = f$, and for all $\theta \in \overline{T}$

$$\|f_\theta\|_{\sigma(P)}^{p_1} = \sum_{i \in I_0} \left(\sum_{j \in J_0} f_\theta(i, j)^{p_2} u_2(j) \right)^{\frac{p_1}{p_2}} u_1(i) = 1,$$

i.e. each such f_θ is on the unit sphere of $L^{\sigma(P)}$. (When $p_1 = \infty$, instead observe that, for each $\theta \in T$, $\|f_\theta\|_{\sigma(P)} = \max_{i \in I_0} \left(\sum_{j \in J_0} f_\theta(i, j)^{p_2} u_2(j) \right)^{\frac{1}{p_2}} = 1$, because the inner sum is always the same as in f.) Because $\|f\|_Q$ achieves the maximum value, $C(I_0, J_0)$, for f supported on $I_0 \times J_0$ in the $L^{\sigma(P)}$ unit sphere, the function

$$\|f_\theta\|_Q^{q_2} = \sum_{j \in J_0} \left(\sum_{i \in I_0} f_\theta(i, j)^{q_1} v_1(i) \right)^{\frac{q_2}{q_1}} v_2(j) \tag{5.2}$$

of θ has a maximum at zero. Only two terms vary with θ, so define

$$A = \sum_{i \in I_0} f(i, s_1)^{q_1} v_1(i) \quad \text{and} \quad B = \sum_{i \in I_0} f(i, s_2)^{q_1} v_1(i)$$

and this means that the function F defined by

$$\begin{aligned} F(\theta) = {} & (A - f(r, s_1)^{q_1} v_1(r) + a(\theta)^{q_1} v_1(r))^{\frac{q_2}{q_1}} \, v_2(s_1) \\ & + (B - f(r, s_2)^{q_1} v_1(r) + b(\theta)^{q_1} v_1(r))^{\frac{q_2}{q_1}} \, v_2(s_2) \end{aligned} \tag{5.3}$$

has a maximum at $\theta = 0$. Therefore $F'(0) = 0$ and $F''(0) \leq 0$.

$$\begin{aligned} F'(\theta) =& \frac{q_2}{q_1}\left(A - f(r,s_1)^{q_1} v_1(r) + a^{q_1} v_1(r)\right)^{\frac{q_2}{q_1}-1} v_2(s_1)\left(q_1 a^{q_1-1} a' v_1(r)\right) \\ &- \frac{q_2}{q_1}\left(B - f(r,s_2)^{q_1} v_1(r) + b^{q_1} v_1(r)\right)^{\frac{q_2}{q_1}-1} v_2(s_2)\left(q_1 b^{q_1-1}\left(-b'\right) v_1(r)\right) \\ =& x(\theta) - y(\theta), \end{aligned}$$

where

$$\begin{aligned} x(\theta) &= q_2\left(A - f(r,s_1)^{q_1} v_1(r) + a^{q_1} v_1(r)\right)^{\frac{q_2}{q_1}-1} a^{q_1-1} a' v_1(r) v_2(s_1), \\ y(\theta) &= q_2\left(B - f(r,s_2)^{q_1} v_1(r) + b^{q_1} v_1(r)\right)^{\frac{q_2}{q_1}-1} b^{q_1-1}\left(-b'\right) v_1(r) v_2(s_2), \end{aligned}$$

and both x and y are strictly positive functions with $x(0) = y(0)$ and $x'(0) \leq y'(0)$.

($F'(0) = 0$ gives $x(0) = y(0)$ and $x'(0) \leq y'(0)$ because $F''(0) \leq 0$. Strict positivity of $A - f(r,s)^{q_1} u_1(r) + a^{q_1} u_1(r)$ and $B_f(r,t)^{q_1} u_1(r) + b^{q_1} u_1(r)$ is because $a' > 0$, $b' < 0$, and having $a(\theta)$ and $b(\theta)$ replace terms in A and B still leaves a nonnegative sum, with the replaced terms strictly positive within the open interval T.)

This means that

$$(\log x)'(0) = \frac{x'(0)}{x(0)} \leq \frac{y'(0)}{y(0)} = (\log y)'(0). \tag{5.4}$$

Differentiating $\log x$ and $\log y$,

$$\begin{aligned} (\log x)'(\theta) &= (q_2 - q_1)\frac{a^{q_1-1} a' v_1(r)}{A - f(r,s_1)^{q_1} v_1(r) + a^{q_1} v_1(r)} + (q_1 - 1)\frac{a'}{a} + \frac{a''}{a'} \\ (\log y)'(\theta) &= (q_2 - q_1)\frac{b^{q_1-1} b' v_1(r)}{B - f(r,s_2)^{q_1} v_1(r) + b^{q_1} v_1(r)} + (q_1 - 1)\frac{b'}{b} + \frac{b''}{b'} \end{aligned}$$

and, since the denominators are A and B, respectively, when $\theta = 0$, (5.4) gives

$$\begin{aligned} &(q_2 - q_1)\frac{a(0)^{q_1}}{A}\frac{a'(0)}{a(0)} v_1(r) + (q_1 - 1)\frac{a'(0)}{a(0)} + \frac{a''(0)}{a'(0)} \\ &\leq (q_2 - q_1)\frac{b(0)^{q_1}}{B}\frac{b'(0)}{b(0)} v_1(r) + (q_1 - 1)\frac{b'(0)}{b(0)} + \frac{b''(0)}{b'(0)}. \end{aligned}$$

Subtracting (5.1) evaluated at $\theta = 0$ from both sides,

$$\begin{aligned} &(q_2 - q_1)\frac{a(0)^{q_1}}{A}\frac{a'(0)}{a(0)} v_1(r) + (q_1 - p_2)\frac{a'(0)}{a(0)} \\ &\leq (q_2 - q_1)\frac{b(0)^{q_1}}{B}\frac{b'(0)}{b(0)} v_1(r) + (q_1 - p_2)\frac{b'(0)}{b(0)}. \end{aligned} \tag{5.5}$$

Divide by $a(0)^{p_2-1} a'(0) u_2(s_1) = b(0)^{p_2-1}\left(-b'(0)\right) u_2(s_2) > 0$:

$$\frac{(q_2 - q_1)\, a(0)^{q_1-p_2} v_1(r)}{A u_2(s_1)} + \frac{q_1 - p_2}{a(0)^{p_2} u_2(s_1)} \leq -\frac{(q_2 - q_1)\, b(0)^{q_1-p_2} v_1(r)}{B u_2(s_2)} - \frac{q_1 - p_2}{b(0)^{p_2} u_2(s_2)}.$$

That is,

$$\frac{(q_2-p_2)\,a(0)^{q_1}v_1(r)-(q_1-p_2)\,a(0)^{q_1}v_1(r)}{Aa(0)^{p_2}u_2(s_1)}+\frac{(q_1-p_2)\,A}{Aa(0)^{p_2}u_2(s_1)}$$
$$\le -\frac{(q_2-q_1)\,b(0)^{q_1}v_1(r)-(q_1-p_2)\,b(0)^{q_1}v_1(r)}{Bb(0)^{p_2}u_2(s_2)}-\frac{(q_1-p_2)\,B}{Bb(0)^{p_2}u_2(s_2)},$$

so

$$\frac{(q_2-p_2)\,a(0)^{q_1}v_1(r)+(q_1-p_2)\,(A-a(0)^{q_1}v_1(r))}{Aa(0)^{p_2}u_2(s_1)}$$
$$\le -\frac{(q_2-p_2)\,b(0)^{q_1}v_1(r)+(q_1-p_2)\,(B-b(0)^{q_1}v_1(r))}{Bb(0)^{p_2}u_2(s_2)}.$$

However, $0<a(0)^{q_1}v_1(r)=f(r,s_1)^{q_1}v_1(r)\le A$ and $0<b(0)^{q_1}v_1(r)=f(r,s_2)^{q_1}v_1(r)\le B$, so with $p_2\le q_2<\min(p_1,q_1)$, the left-hand side is at least zero while the right-hand side is at most zero. Therefore, both sides must be zero, so that $p_2=q_2$, $A=a(0)^{q_1}v_1(r)$, and $B=b(0)^{q_1}v_1(r)$. These equations, plus the definitions of $a(\theta)$ and $b(\theta)$, allow the simplification of $F(\theta)$ from the formula in (5.3) to

$$\begin{aligned}F(\theta)&=a^{q_2}v_1(r)^{\frac{q_2}{q_1}}v_2(s_1)+b^{q_2}v_1(r)^{\frac{q_2}{q_1}}v_2(s_2)\\&=a^{p_2}v_1(r)^{\frac{p_2}{q_1}}v_2(s_1)+b^{p_2}v_1(r)^{\frac{p_2}{q_1}}v_2(s_2)\\&=v_1(r)^{\frac{p_2}{q_1}}\left(f(r,s_1)^{p_2}v_2(s_1)+f(r,s_2)^{p_2}v_2(s_2)+\left(\frac{v_2(s_1)}{u_2(s_1)}-\frac{v_2(s_2)}{u_2(s_2)}\right)\theta\right).\end{aligned}$$

As noted before, F has a maximum at the interior point zero, so since it is an affine function of θ, it must be constant. Therefore the expression (5.2) is constant with respect to $\theta\in T$ and, by continuity, $\theta\in\overline{T}$. Consequently, $\|f_\theta\|_Q=C(I_0,J_0)$ for any $\theta\in\overline{T}$, including at either end. But there is one fewer non-zero value of f_θ for endpoint θ; when $\theta=-f(r,s_1)^{p_2}u_2(s_1)$, $a(\theta)=f_\theta(r,s_1)=0$, while when $\theta=f(r,s_2)^{p_2}u_2(s_2)$, $b(\theta)=f_\theta(r,s_2)=0$. The minimality of the number of non-zero values in f is therefore contradicted. □

For the next argument, it is convenient to note that, by the characterizaation

$$C=\sup_{f\neq 0}\frac{\|f\|_Q}{\|f\|_{\sigma(P)}}$$

and by homogeneity,

$$C=\sup\left\{\|g\|_Q:\|g\|_{\sigma(P)}=1\right\}=\left(\inf\left\{\|h\|_{\sigma(P)}:\|h\|_Q=1\right\}\right)^{-1}.$$

In detail, given any f which is not almost everywhere zero, define $g=\|f\|_{\sigma(P)}^{-1}f$. Observe that

$$\|g\|_{\sigma(P)}=1\quad\text{and}\quad\|g\|_Q=\frac{\|f\|_Q}{\|f\|_{\sigma(P)}},$$

so $\|g\|_Q\le C$ and can be brought arbitrarily close to C, which must be the supremum. Similarly, let $h=\|f\|_Q^{-1}f$.

$$\|h\|_{\sigma(P)}=\frac{\|f\|_{\sigma(P)}}{\|f\|_Q}\quad\text{and}\quad\|h\|_Q=1.$$

Therefore $\|h\|_{\sigma(P)}^{-1} \leq C$, so $\|h\|_{\sigma(P)} \geq C^{-1}$, and can be brought arbitrarily close to C^{-1}, which therefore is the infimum.

These same remarks apply, of course, to the best constants $C(I_0, J_0)$ for functions supported on the finite sets $I_0 \times J_0$, where compactness of the norms' unit spheres means that

$$\max\left\{\|f\|_Q : \|f\|_{\sigma(P)} = 1\right\} = C(I_0, J_0)$$
$$\min\left\{\|f\|_{\sigma(P)} : \|f\|_Q = 1\right\} = C(I_0, J_0)^{-1}.$$

Proposition 5.2.4. *Suppose that* $0 < \max(p_2, q_2) < p_1 \leq q_1 \leq \infty$ *(the non-Minkowski case with the additional constraint* $p_1 \leq q_1$*) and let* $I_0 \subset I$ *and* $J_0 \subset J$ *be any finite subsets of the index sets for* $f(i, j)$*. Let* $C(I_0, J_0)$ *be the least constant such that, for any* $f(i, j)$ *for which all nonzero values have* $i \in I_0$ *and* $j \in J_0$*,*

$$\|f\|_Q \leq C(I_0, J_0)\, \|f\|_{\sigma(P)}\,.$$

Then $C(I_0, J_0)$ *can be achieved as* $\|f\|_Q \,/\, \|f\|_{\sigma(P)}$ *where* f *has at most one non-zero entry per column, i.e. for each fixed* $j \in J_0$*, there is at most one* $i \in I_0$ *such that* $f(i, j) \neq 0$*.*

Proof. Let $F(i, j) \in L^+(I \times J)$ be supported on $I_0 \times J_0$, i.e. zero except for $(i, j) \in I_0 \times J_0$, with $\|f\|_{\sigma(P)} = C(I_0, J_0)^{-1}$, $\|f\|_Q = 1$, and among such functions the fewest possible non-zero values. (The minimum $C(I_0, J_0)^{-1}$ is achieved since $L^+(I_0 \times J_0)$ is finite-dimensional, so the unit sphere of L^Q there is compact.) Suppose, in order to produce a contradiction, that there is some $s \in J_0$ with distinct $r_1, r_2 \in I_0$ such that both $f(r_1, s) > 0$ and $f(r_2, s) > 0$.

First suppose that $q_1 = \infty$, in which case

$$\|f\|_Q = \left(\sum_{j \in J_0} \left(\max_{i \in I_0} f(i, j)\right)^{q_2} v_2(j)\right)^{\frac{1}{q_2}}.$$

Changing f to a new function g by reducing the lesser of $f(r_1, s)$ and $f(r_2, s)$ to zero will not change $\max_{i \in I_0} f(i, s)$, so $\|g\|_Q = \|f\|_Q = 1$. However, because $g \leq f$, $\|g\|_{\sigma(P)} \leq \|f\|_{\sigma(P)} = C(I_0, J_0)^{-1}$, so it must be the minimum, $C(I_0, J_0)^{-1}$. This contradicts the minimal number of non-zero entries in f.

For the remainder, suppose that $q_1 < \infty$ and therefore, by hypothesis, $p_1 < \infty$ as well. Let T denote the open interval $(-f(r_1, s)^{q_1} v_1(r_1), f(r_2, s)^{q_1} v_1(r_2))$ and define

$$a(\theta) = \left(f(r_1, s)^{q_1} + \frac{\theta}{v_1(r_1)}\right)^{\frac{1}{q_1}} \qquad\qquad b(\theta) = \left(f(r_2, s)^{q_1} - \frac{\theta}{v_1(r_2)}\right)^{\frac{1}{q_1}}$$

for θ in the closed interval $\overline{T}$. Because $a(\theta)^{q_1} v_1(r_1) + b(\theta)^{q_1} v_1(r_2)$ is constant,

$$a^{q_1 - 1} a' v_1(r_1) = b^{q_1 - 1} \,(-b')\, v_1(r_2).$$

Differentiate the logarithm to obtain

$$(q_1 - 1)\,\frac{a'}{a} + \frac{a''}{a'} = (q_1 - 1)\,\frac{b'}{b} + \frac{b''}{b'}. \tag{5.6}$$

For any $\theta \in \overline{T}$,

$$\|f_\theta\|_Q^{q_2} = \sum_{j\in J_0}\left(\sum_{i\in I_0} f_\theta(i,j)^{q_1} v_1(i)\right)^{\frac{q_2}{q_1}} v_2(j) = 1,$$

so $\|f_\theta\|_Q = 1$. Because $\|f\|_{\sigma(P)}$ achieves its minimum value, $C(I_0, J_0)^{-1}$, for f supported on $I_0 \times J_0$ in the L^Q unit sphere, the function

$$\|f_\theta\|_{\sigma(P)}^{p_1} = \sum_{i\in I_0}\left(\sum_{j\in J_0} f_\theta(i,j)^{p_2} v_1(i)\right)^{\frac{p_1}{p_2}} u_1(i) \tag{5.7}$$

has a minimum at $\theta = 0$. Define

$$A = \sum_{j\in J_0} f_\theta(r_1, j)^{p_2} u_2(j) \quad \text{and} \quad B = \sum_{j\in J_0} f_\theta(r_2, j)^{p_2} u_2(j).$$

The function

$$\begin{aligned} F(\theta) =& (A - f(r_1,s)^{p_2}u_2(s) + a^{p_2}u_2(s))^{\frac{p_1}{p_2}} u_1(r_1) \\ &+ (B - f(r_2,s)^{p_2}u_2(s) + b^{p_2}u_2(s))^{\frac{p_1}{p_2}} u_1(r_2) \end{aligned} \tag{5.8}$$

has a minimum at zero. Therefore $F'(0) = 0$ and $F''(0) \geq 0$.

$$\begin{aligned} F'(\theta) =& \frac{p_1}{p_2}(A - f(r_1,s)^{p_2}u_2(s) + a^{p_2}u_2(s))^{\frac{p_1}{p_2}-1} u_1(r_1)\left(p_2 a^{p_2-1}a' u_2(s)\right) \\ &- \frac{p_1}{p_2}(B - f(r_2,s)^{p_2}u_2(s) + b^{p_2}u_2(s))^{\frac{p_1}{p_2}-1} u_1(r_1)\left(p_2 b^{p_2-1}(-b')\, u_2(s)\right) \\ =& x(\theta) - y(\theta) \end{aligned}$$

where

$$\begin{aligned} x(\theta) &= p_1 (A - f(r_1,s)^{p_2}u_2(s) + a^{p_2}u_2(s))^{\frac{p_1}{p_2}-1} a^{p_2-1}a' u_1(r_1)u_2(s) \\ y(\theta) &= p_1 (B - f(r_2,s)^{p_2}u_2(s) + b^{p_2}u_2(s))^{\frac{p_1}{p_2}-1} b^{p_2-1}(-b')\, u_1(r_2)u_2(s) \end{aligned}$$

and both x and y are strictly positive functions with $x(0) = y(0)$, since $F'(0) = 0$ and $x'(0) \geq y'(0)$, because $F''(0) \geq 0$. Therefore

$$(\log y)'(0) = \frac{y'(0)}{y(0)} \leq \frac{x'(0)}{x(0)} = (\log x)'(0).$$

Differentiate $\log x$ and $\log y$ to find that

$$\begin{aligned} (\log y)'(\theta) &= (p_1 - p_2)\frac{b^{p_2-1}b' u_2(s)}{B - f(r_2,s)^{p_2}u_2(s) + a^{p_2}u_2(s)} + (p_2 - 1)\frac{b'}{b} + \frac{b''}{b'} \\ (\log x)'(\theta) &= (p_1 - p_2)\frac{a^{p_2-1}a' u_2(s)}{A - f(r_1,s)^{p_2}u_2(s) + a^{p_2}u_2(s)} + (p_2 - 1)\frac{a'}{a} + \frac{a''}{a'} \end{aligned}$$

and, when $\theta = 0$,

$$\begin{aligned}(p_1 - p_2)\,\frac{b(0)^{p_2}}{B}\frac{b'(0)}{b(0)}u_2(s) + (p_2 - 1)\,\frac{b'(0)}{b(0)} + \frac{b''(0)}{b'(0)} \\ \le (p_1 - p_2)\,\frac{a(0)^{p_2}}{A}\frac{a'(0)}{a(0)}u_2(s) + (p_2 - 1)\,\frac{a'(0)}{a(0)} + \frac{a''(0)}{a'(0)}.\end{aligned}$$

Subtract the value of (5.6) at $\theta = 0$ from both sides to obtain

$$(p_1 - p_2)\,\frac{b(0)^{p_2}}{B}\frac{b'(0)}{b(0)}u_2(s) + (p_2 - q_1)\,\frac{b'(0)}{b(0)} \le (p_1 - p_2)\,\frac{a(0)^{p_2}}{A}\frac{a'(0)}{a(0)}u_2(s) + (p_2 - q_1)\,\frac{a'(0)}{a(0)}$$

Divide by $a(0)^{q_1-1}a'(0)v_1(r_1) = b(0)^{q_1-1}\,(-b'(0))\,v_1(r_2) > 0$:

$$\begin{aligned}-\frac{(p_1 - p_2)b(0)^{p_2-q_1}u_2(s)}{Bv_1(r_2)} - \frac{p_2 - q_1}{b(0)^{q_1}v_1(r_2)} &\le \frac{(p_1 - p_2)a(0)^{p_2-q_1}u_2(s)}{Av_1(r_1)} + \frac{p_2 - q_1}{a(0)^{q_1}v_1(r_1)} \\ -\frac{(p_1 - p_2)b(0)^{p_2}u_2(s)}{Bb(0)^{q_1}v_1(r_2)} - \frac{(p_2 - q_1)\,B}{Bb(0)^{q_1}v_1(r_2)} &\le \frac{(p_1 - p_2)a(0)^{p_2}u_2(s)}{Aa(0)^{q_1}v_1(r_1)} + \frac{(p_2 - q_1)\,A}{Aa(0)^{q_1}v_1(r_1)}\end{aligned}$$

$$\begin{aligned}&\frac{(q_1 - p_2)\,(B - b(0)^{p_2}u_2(s)) + (q_1 - p_1)b(0)^{p_2}u_2(s)}{Bb(0)^{q_1}v_1(r_2)} \\ &\le -\frac{(q_1 - p_2)\,(A - a(0)^{p_2}u_2(s)) + (q_1 - p_1)a(0)^{p_2}u_2(s)}{Aa(0)^{q_1}v_1(r_1)}\end{aligned}$$

Because $p_1 \le q_1$, the left-hand side is at least zero, while the right-hand side is at most zero. Therefore each side is zero, $p_1 = q_1$, $A = a(0)^{p_2}u_2(s)$, and $B = b(0)^{p_2}u_2(s)$. Together with the definitions of $a(\theta)$ and $b(\theta)$, these simplify the formula in (5.8) to

$$\begin{aligned}F(\theta) &= a^{p_1}u_2(s)^{\frac{p_1}{p_2}}u_1(r_1) + b^{p_1}u_2(s)^{\frac{p_1}{p_2}}u_1(r_2) \\ &= u_2(s)^{\frac{q_1}{p_2}}\left(\left(f(r_1,s)^{q_1} + \frac{\theta}{v_1(r_1)}\right)u_1(r_1) + \left(f(r_2,s)^{q_1} - \frac{\theta}{v_1(r_2)}\right)u_1(r_2)\right) \\ &= u_2(s)^{\frac{q_1}{p_2}}\left(f(r_1,s)^{q_1}u_1(r_1) + f(r_2,s)^{q_1}u_1(r_2) + \left(\frac{u_1(r_1)}{v_1(r_1)} - \frac{u_1(r_2)}{v_1(r_2)}\right)\theta\right).\end{aligned}$$

Because this affine function has a minimum at the interior point $\theta = 0$, it must be constant. Therefore, $\|f_\theta\|_{\sigma(P)} = C(I_0, J_0)^{-1}$ for any $\theta \in \overline{T}$, including the endpoints, where f_θ has one fewer non-zero value than f. This contradicts the minimality of the number of non-zero values in f among those functions in the L^Q unit sphere with $\|f\|_{\sigma(P)} = C(I_0, J_0)^{-1}$. □

5.3 Best constants for diagonal case, $p_2 \le q_2 < p_1 \le q_1$

When the non-Minkowski condition $\max(p_2, q_2) < \min(p_1, q_1)$ is combined with the necessary condition $p_2 \le q_2$ for Proposition 5.2.3 and $p_1 \le q_1$ for Proposition 5.2.4, the best constant is obtained by considering functions which are, up to rearrangement of the indices, diagonal, in the sense that each row ($\{f(i_0, j) : j \in J\}$ for fixed $i_0 \in I$) and each column ($\{f(i, j_0) : i \in I\}$ for fixed $j_0 \in J$) contains at most one non-zero entry. So far, this is only established over finite subsets $I_0 \subset I$ and $J_0 \subset J$, but the following results show that this is generally true, and develop a formula for the resulting best constant.

Lemma 5.3.1. *For any $i \in I$, let $C_1(i)$ denote the least constant such that, for any function $f_1 \in L^+(I)$ supported on $\{i\}$, $\|f_1\|_{\ell^{q_1}_{v_1}(I)} \le C_1 \|f_1\|_{L^{p_1}_{u_1}(I)}$. For $j \in J$, let $C_2(j)$ denote the least constant such that, for any $f_2^+(J)$ supported on $\{j\}$, $\|f_2\|_{L^{q_2}_{v_2}(J)} \le C_2(j)\|f_2\|_{L^{p_2}_{u_2}(J)}$. Then*

$$C_1(i) = v_1(i)^{1/q_1} u_1(i)^{-1/p_1} \quad \text{and} \quad C_2(j) = v_2(j)^{1/q_2} u_2(j)^{-1/p_2}.$$

Proof. Any function f_1 supported on $\{i\}$ has the form $f = c\chi_{\{i\}}$ for some constant c. By homogeneity,

$$\frac{\|f\|_{L^{q_1}_{v_1}(I)}}{\|f\|_{L^{p_1}_{u_1}(I)}} = \frac{|c| \left\|\chi_{\{i\}}\right\|_{L^{q_1}_{v_1}(X_1)}}{|c| \left\|\chi_{\{j\}}\right\|_{L^{p_1}_{\mu_1}(X_1)}} = \frac{v_1(i)^{1/q_1}}{u_1(i)^{1/p_1}},$$

and a similar computation works for J. □

Lemma 5.3.2. *For any $i_0 \in I$, $j_0 \in J$, let $C(i_0, j_0)$ denote the least constant such that*

$$\|f\|_Q \le C(i_0, j_0) \|f\|_{\sigma(P)}$$

for any function $f \in L^+(I \times J)$ which is zero everywhere except (i_0, j_0). Then

$$C(i_0, j_0) = C_1(i_0)C_2(j_0) = \frac{v_1(i_0)^{1/q_1}}{u_1(i_0)^{1/p_1}} \frac{v_2(j_0)^{1/q_2}}{u_2(j_0)^{1/p_2}}.$$

Proof. Any such function has the form $f(i, j) = c\chi_{\{i_0\}}(i)\chi_{\{j_0\}}(j)$. Using homogeneity and the fact that mixed norms computed of factorable functions are products of Lebesgue space norms of the factors, so long as $c \neq 0$

$$\frac{\|f\|_Q}{\|f\|_{\sigma(P)}} = \frac{|c| \left\|\chi_{\{i_0\}}\right\|_{\ell^{q_1}_{v_1}(I)} \left\|\chi_{\{j_0\}}\right\|_{\ell^{q_2}_{v_2}(J)}}{|c| \left\|\chi_{\{i_0\}}\right\|_{\ell^{p_1}_{u_1}(I)} \left\|\chi_{\{j_0\}}\right\|_{\ell^{p_2}_{u_2}(J)}} \le C_1(i_0)C_2(j_0).$$

Therefore $C(i_0, j_0) \le C_1(i_0)C_2(j_0)$, while Proposition 3.3.3 (also based on factorable functions), applied with I and J replaced by the singleton subspaces $\{i_0\}$ and $\{j_0\}$ respectively, yields $C(i_0, j_0) \ge C_1(i_0)C_2(j_0)$. □

Lemma 5.3.3. *Suppose that $p_2 \le q_2 < p_1 \le q_1$. Let $I_0 \subset I$ and $J_0 \subset J$ be finite subsets, each with cardinality $|I_0| = |J_0| = N$, where of course $N \le \min(|I|, |J|)$. Let $\left(C_1^*(m)\right)_{1 \le m \le N}$ be a nonincreasing enumeration, i.e. $C_1^*(1) \ge \cdots \ge C_1^*(N)$, of $\{C_1(i) : i \in I_0\}$, corresponding to an enumeration $(i_m^*)_{1 \le m \le N}$ according to $C_1^*(m) = C_1(i_m^*)$. Similarly, let $\left(C_2^*(m)\right)_{1 \le m \le N}$ be a nonincreasing enumeration of $(C_2(j))_{j \in J_0}$, corresponding to an enumeration $(j_m^*)_{1 \le m \le N}$ by $C_2^*(m) = C_2(j_m^*)$. Then the best constant $C(I_0, J_0)$ such that*

$$\|f\|_Q \le C(I_0, J_0) \|f\|_{\sigma(P)}$$

for all functions $f \in L^+(I \times J)$ supported on $I_0 \times J_0$ is

$$C(I_0, J_0) = \left\|C_1^*(m)C_2^*(m))\right\|_{\ell^{p_1 : q_2}(\{1, \ldots, N\})} = \left\|\frac{v_1(i_m^*)^{1/q_1} v_2(j_m^*)^{1/q_2}}{u_1(i_m^*)^{1/p_1} u_2(j_m^*)^{1/p_2}}\right\|_{p_1 : q_2}.$$

Proof. The order of exponents provides, using Propositions 5.2.3 and 5.2.4, that the best constant $C(I_0, J_0)$ can be achieved with some function $f \in L^+(I_0 \times J_0)$ which has at most one entry in each row, and at most one entry in each column. Any such function has the form

$$f(i,j) = \sum_{m=1}^{N} c_m \chi_{\{i_m^*\}}(i) \chi_{\left\{j_{\sigma(m)}^*\right\}}(j)$$

for some coefficients $c_1, \ldots, c_N \ge 0$ and some permutation $\sigma \in S_N$. (There is at most one non-zero value in the "row" identified with each i_m^*; its "column" is given by $j_{\sigma(m)}^*$, where injectivity is because there is at most one non-zero value in each column. The value is given by c_m; if there is no non-zero value in that row, then $c_m = 0$ and the value of $\sigma(m)$ can be any not reserved for a non-zero value.)

Such a function is clearly block-factorable, supported on blocks $\left(\{i_m^*\} \times \left\{j_{\sigma(m)}^*\right\}\right)_{1 \le m \le N}$. For those particular blocks, dependent on σ, the best constant for block-factorable functions is

$$C_{BF}\left(\{i_m^*\}, \left\{j_{\sigma(m)}^*\right\}\right) = \left\| C_1(i_m^*) C_2(j_{\sigma(m)}^*) \right\|_{\ell^{p_1 : q_2}(\{1,\ldots,N\})},$$

according to Proposition 4.2.10. The choice of σ which maximizes this constant is the one which maximizes its $p_1 : q_2$ power,

$$\sum_{m=1}^{N} a_m b_{\sigma(m)} \tag{5.9}$$

where each $a_m = C_1(i_m^*)^{p_1 : q_2}$ and $b_{\sigma(m)} = C_2(j_{\sigma(m)}^*)^{p_1 : q_2}$. Recall that the enumeration (i_m^*) of I_0 is chosen so that $(C_1(i_m^*))$ is nonincreasing, and therefore so is (a_m), i.e. $a_1 \ge \cdots \ge a_N$. The rearrangement inequality, given by Hardy, Littlewood, and Pólya as Theorem 368 in [19], establishes that $\sum a_m b_{\sigma(m)}$ is greatest when (a_m) and $(b_{\sigma(m)})$ are similarly ordered, i.e. when $b_1 \ge \cdots \ge b_N$, which occurs when σ is the identity and

$$C_{BF}\left(\{i_m^*\}, \{j_m^*\}\right) = \left\| C_1(i_m^*) C_2(j_m^*) \right\|_{\ell^{p_1 : q_2}(\{1,\ldots,N\})}.$$

Therefore $C(I_0, J_0)$ is equal to this greatest value, $\left\| C_1^*(m) C_2^*(m) \right\|_{\ell^{p_1 : q_2}(\{1,\ldots,N\})}$. □

To solve the case with infinite index sets, we need the notion of decreasing rearrangement of sequences, a special case of the decreasing rearrangement of functions described in such references as [5] and [27]. When the definition there is applied to a sequence $K = (K_1, K_2, \ldots)$, finite or infinite, its rearrangement $K^* = \left(K_1^*, K_2^*, \ldots,\right)$ is a function on $(0, \infty)$,

$$K^*(t) = \inf\left\{\alpha \in (0, \infty] : \{n : |K_n| > \alpha\} \text{ has at most } t \text{ elements}\right\}.$$

But the function K^* is a step function, constant on intervals $[n, n+1)$. We identify each such function with a sequence, and let K^* refer to that sequence.

For convenience, the rearrangement K_n^* of a nonnegative sequence $K = (K_n)$ can be characterized as follows. Let N denote the possibly infinite number of elements of K which strictly exceed $\limsup_i K_i$. Then

$$\begin{aligned} &\text{if } 1 \le n \le N, \text{ then } K_n^* \text{ is the } n^{th} \text{ greatest value in } K, \\ &\text{but if } n > N, \text{ then } K_n^* = \limsup_i K_i. \end{aligned}$$

(Given any $y > \limsup_i K_i$, there are only finitely many values in K which are at least y; this is why the n^{th} largest value is well-defined for $n \leq N$.)

For any nonnegative sequence K which is either finite or convergent to zero, K^* is a permutation of K, sorting its entries into nonincreasing order.

The superscript asterisk notation used for the finite case in Lemma 5.3.3 is consistent with this definition, and used in this sense in $C_1^*(m)$ and $C_2^*(m)$. The is not the case for i_m^* and j_m^*, since I and J need not have any inherent order, and the values in the index sets are irrelevant; the notations i_m^* and j_m^* are only used due to their connection to the rearrangements of the atomic best constants $C_1(i)$ and $C_2(j)$.

The well-known Hardy-Littlewood inequality gives an upper bound in terms of rearrangements,

$$\int_X |fg| \leq d\mu \leq \int_0^\infty f^*(t)g^*(t)dt.$$

This also applies when f is replaced by a function equimeasurable to itself, and the same for g. (Note that equimeasurable functions have identical rearrangements, so the right-hand side remains fixed.) The measure μ is said to be resonant, as in Definition 2.2.3 from [5], if and only if

$$\sup \int_X |f\tilde{g}| d\mu = \int_0^\infty f^*(t)g^*(t)dt,$$

where the supremum is over all $\tilde{g}$ equimeasurable to g. By symmetry in the roles of f and g, we could just as well use functions $\tilde{f}$ equimeasurable to f, and the supremum replacing both f and g with equimeasurable functions must then be $\int f^* g^*$.

Theorem 2.2.7 in [5] establishes that counting measure is resonant. In terms of sequences, let A and B be countably infinite collections of nonnegative numbers. Then

$$\sup \sum_n a_n b_n = \sum a_n^* b_n^*,$$

where the supremum is taken over all enumerations (a_n) of A and (b_n) of B.

Theorem 5.3.4. *Suppose that $p_2 \leq q_2 < p_1 \leq q_1$ and each X_k has purely atomic measures. Let the atoms of X_1 be $(E_i)_{i \in I}$ and the atoms of X_2 be $\left(F_j\right)_{j \in J}$, where I and J are (at most) countable index sets. For each $i \in I$ and each $j \in J$, let*

$$C_1(i) = \nu_1(E_i)^{1/q_1} \mu_1(E_i)^{-1/p_1} \qquad \text{and} \qquad C_2(j) = \nu_2(F_j)^{1/q_2} \mu_2(F_j)^{-1/p_2}.$$

Then the least constant C such that

$$\|f\|_Q \leq C \|f\|_{\sigma(P)}$$

for any $f \in L^+(X_1 \times X_2)$ can be computed by

$$C = \left\| C_1^*(m) C_2^*(m) \right\|_{p_1 : q_2},$$

where C_1^* and C_2^* denote the decreasing rearrangements of these sequences. (If I, J, or both are finite, pad out C_1, C_2, or both with zeroes.)

Proof. For convenience, identify functions on $X_1 \times X_2$ with functions on $I \times J$, using weights u_k and v_k, as described in Definition 4.2.18 on page 57. For any finite subsets $I_0 \subset I$ and $J_0 \subset J$, let $C(I_0, J_0)$ denote the least constant such that, for any $f \in L^+(I \times J)$ supported on the subset $I_0 \times J_0$ of $I \times J$,

$$\|f\|_Q \le C(I_0, J_0)\, \|f\|_{\sigma(P)}\,.$$

The order of exponents provides, by Propositions 5.2.3 and 5.2.4, that the constant can be achieved by some f with at most one entry in each row and at most one entry in each column. Any such function has the form

$$f(i, j) = \sum_{m=1}^{N} c_m \chi_{\{i_m\}}(i) \chi_{\{j_m\}}(j)$$

for coefficients $c_1, \ldots, c_N \ge 0$ and distinct elements $i_1, \ldots, i_N \in I_0$ and $j_1, \ldots, j_N \in J$. (There is at most one non-zero value in the "row" identified with each i_m; its column is given by j_m; if there is no non-zero value in that row, $c_m = 0$.)

Such a function is clearly block-factorable, supported on blocks $(\{i_m\} \times \{j_m\})_{1 \le m \le N}$. For those particular blocks, the best constant for block-factorable functions is

$$C_{BF}(\{i_m\} \times \{j_m\}) = \|C_1(i_m) C_2(j_m)\|_{\ell^{p_1 : q_2}(\{1,\ldots,N\})}\,,$$

according to Proposition 4.2.10. Therefore C is the supremum of

$$\left(\sum_m C_1(i_m)^{p_1 : q_2} C_2(j_m)^{p_1 : q_2} \right)^{1/(p_1 : q_2)}, \tag{5.10}$$

over all enumerations (i_m) of I and (j_m) of J, padded with zeroes if either set is finite. As noted earlier, counting measure is resonant (as described and proven in such sources as [27] and [5]), which means that the supremum of (5.10) over all enumerations (i_m) and (j_m) is

$$C = \left(\sum_m (C_1^*(m) C_2^*(m))^{p_1 : q_2} \right)^{1/(p_1 : q_2)},$$

as desired. □

This is enough to solve the two-variable permuted inclusion problem where every measure is counting measure on a countably infinite set, i.e. where every one-variable space is an unweighted ℓ^p space.

Proposition 5.3.5. *Let $C \in [0, \infty]$ be the least constant such that, for any function $f(i, j)$ on $\mathbb{N}^2$,*

$$\|f\|_Q \le C\, \|f\|_{\sigma(P)}\,.$$

If $p_1 \le q_1$. $p_2 \le q_2$, and $p_1 \le q_2$, then $C = 1$. Otherwise, $C = \infty$.

Proof. If, indeed, $p_1 \le q_1 \le p_2 \le q_2$, then the fact that each $p_k \le q_k$ means that $\ell^{p_k} \subset \ell^{q_k}$ with $C_k = 1$ by Corollary 2.6.10. Also, this is in the Minkowski case, so Theorem 3.6.1 proves that $C = C_1 C_2 = 1$.

The rest of the argument is devoted to refuting inclusion in every other case. If either $p_k > q_k$, then by Corollary 2.6.10, $\ell^{p_k} \not\subset \ell^{q_k}$; that is, $C_k = \infty$. By Proposition 3.3.3, $C \geq C_1 C_2 = \infty$. When $p_1 \leq q_1$ and $p_2 \leq q_2$, the only case left is $p_2 \leq q_2 < p_1 \leq q_1$. But this case is the one handled by Theorem 5.3.4; unweighted counting measure with $p_k \leq q_k$ gives the local best constant $C_k(n) = 1$ on each atom $n \in \mathbb{N}$. This constant sequence is its own decreasing rearrangement, $C_k^* = 1$. Theorem 5.3.4 then shows that

$$C = \left(\sum_{n \in \mathbb{N}} 1^{p_1 : q_2} \right)^{1/(p_1 : q_2)} = \infty.$$

□

5.4 Generalized partition problem

Solutions to the remaining non-Minkowski cases involve a generalization of a known problem, the optimization version of the partition problem, known to be NP-hard. Considering this, a full solution to the problem is not presented, just the definition and certain properties.

Definition 5.4.1 (Generalized Partition Problem)**.** Let $\gamma \in (0, 1)$ and two nonnegative sequences $(a_i)_{i \in I}$ and $\left(b_j\right)_{j \in J}$. Define

$$GPP_\gamma(a, b) = \sup \sum_{i \in I} a_i \left(\sum_{j \in J(i)} b_j \right)^\gamma ,$$

a supremum over all partitions $\{J(i) : i \in I\}$ of J into pairwise disjoint, possibly empty subsets indexed by I.

The partition problem itself is obtained when $I = \{1, 2\}$, with weights $a_1 = a_2 = 1$, and a finite set $J = \{1, \ldots, N\}$. The problem is one of maximizing, over all subsets $J' \subset J$,

$$\left(\sum_{j \in J'} b_j \right)^\gamma + \left(\sum_{j \in J \setminus J'} b_j \right)^\gamma .$$

Regardless of the partition, the two sums always add up to the same total, $\sum_{j \in J} b_j$. Because the exponent γ is less than one, the maximum is achieved when these sums are as close to equal as possible. Trying to partition a finite set into subsets with as close to equal totals as possible is the optimization version of the partition problem.

1. If $GPP_\gamma(a, b) < \infty$ then $b \in \ell^1$ and $a \in \ell^\infty$. To see this, fix any $i_0 \in I$ and consider the trivial partition with $J(i_0) = J$, every other $J(i) = \emptyset$. With this,

$$\sum_{i \in I} a_i \left(\sum_{j \in J(i)} b_j \right)^\gamma = a_{i_0} \left(\sum_{j \in J} b_j \right)^\gamma \leq GPP_\gamma(a, b) < \infty.$$

   This implies that $\|b\|_1 = \sum_{j \in J} b_j < \infty$, and furthermore that every $a_{i_0} \leq \|b\|_1^{-\gamma} GPP_\gamma(a, b)$, so $\|a\|_\infty \leq \|b\|_1^{-\gamma} GPP_\gamma(a, b) < \infty$.

2. If $b \in \ell^1$ and $a \in \ell^{1/(1-\gamma)}$, then $GPP_\gamma(a,b) < \infty$. Regardless of the partition, apply Hölder's inequality with conjugate exponents $1/(1-\gamma)$ and $1/\gamma$:

$$\sum_{i\in I} a_i \left(\sum_{j\in J(i)} b_j\right)^\gamma \le \left(\sum_{i\in I} a_i^{\frac{1}{1-\gamma}}\right)^\gamma \left(\sum_{i\in I}\sum_{j\in J(i)} b_j\right)^\gamma$$
$$= \|a\|_{1/(1-\gamma)} \|b\|_1^\gamma$$

   This also applies to the supremum, $GPP_\gamma(a,b) \le \|a\|_{1/(1-\gamma)} \|b\|_1^\gamma$.

3. If $a \in \ell^\infty$ and $b \in \ell^\gamma$ then $GPP_\gamma(a,b) < \infty$.

$$\sum_{i\in I} a_i \left(\sum_{j\in J(i)} b_j\right)^\gamma \le \sum_{i\in I} \|a\|_\infty \left(\sum_{j\in J(i)} b_j\right)^\gamma \le \|a\|_\infty \sum_{i\in I}\sum_{j\in J(i)} b_j^\gamma = \|a\|_\infty \|b\|_\gamma^\gamma$$

   In the supremum, $GPP_\gamma(a,b) \le \|a\|_\infty \|b\|_\gamma^\gamma$.

4. Let $N = \min(|I|, |J|)$, possibly infinite, and define, for $1 \le m \le N$, $a^*(m)$ and $b^*(m)$ to enumerate the N greatest values in a_i and b_i in nonincreasing order. If $GPP_\gamma(a,b) < \infty$ then $a^*(b^*)^\gamma \in \ell^1$.

   For each $1 \le m \le N$, let $i^*(m)$ be the element of I such that $a_{i^*(m)} = a^*(m)$, and let $j^*(m)$ be the element of J such that $b_{j^*(m)} = b^*(m)$. Then partition J into singletons, each $J(i^*(m)) = \{j^*(m)\}$. (If $|J| > |I|$, there will be leftover, lesser elements of j, but inserting those in any set in the partition preserves the following inequality.)

$$GPP_\gamma(a,b) \ge \sum_{m=1}^{N} a^*(m)(b^*(m))^\gamma = \|a^*(b^*)^\gamma\|_1$$

Although the simple kinds of partitions used in the above examples shed some light on the problem, this GPP is too complex to be solved by considering only such partitions. Example 5.7 in [18] establishes that, even when all these partition types give bounded values, the supremum defining $GPP_\gamma(a,b)$ can be infinite.

5.5 Best constants for cases with one entry per row or per column

Next, consider the case $p_2 \le q_2 < q_1 < p_1$, where Proposition 5.2.3 guarantees that there is at most one non-zero entry per row, but Proposition 5.2.4 does not apply, so this need not be the case for columns.

Theorem 5.5.1. *Suppose that $p_2 \le q_2 < q_1 < p_1$ and each X_k has purely atomic measures. Let the atoms of X_1 be $(E_i)_{i\in I}$ and the atoms of X_2 be $\left(F_j\right)_{j\in J}$, where I and J are (at most) countable index sets. For each $i \in I$ and each $j \in J$, let*

$$C_1(i) = \nu_1(E_i)^{1/q_1}\mu_1(E_i)^{-1/p_1} \quad \text{and} \quad C_2(j) = \nu_2(F_j)^{1/q_2}\mu_2(F_j)^{-1/p_2}.$$

Then the least constant C such that

$$\|f\|_Q \leq C \, \|f\|_{\sigma(P)}$$

for any $f \in L^+(X_1 \times X_2)$ is

$$C = \left(GPP_{\frac{p_1 : q_2}{p_1 : q_1}} \left(C_2(j)^{p_1 : q_2}, C_1(i)^{p_1 : q_1} \right) \right)^{\frac{1}{p_1 : q_2}},$$

in terms of a solution to a generalized partition problem.

Proof. Fix finite subsets $I_0 \subset I$ and $J_0 \subset J$. By Proposition 5.2.3, the best constant $C(I_0, J_0)$ for functions supported on $I_0 \times J_0$ can be achieved with a function $f \in L^+(I \times J)$, supported on $I_0 \times J_0$, with at most one non-zero entry in each row, i.e. for each $i \in I_0$, there is at most one $j \in J_0$ such that $f(i, j) > 0$.

This means that there are pairwise disjoint subsets $(I(j))_{j \in J_0}$ of I_0, each $I(j) = \{i \in I_0 : f(i, j) > 0\}$, the places in column j where f takes positive values. For each fixed $j \in J_0$, define $g_j(i) = f(i, j)$. (Note that then $I(j)$ is the support of g_j.) With these,

$$\|f\|_Q = \left(\sum_{j \in J_0} \left\| g_j \right\|^{q_2}_{\ell^{q_1}_{v_1}(I(j))} v_2(j) \right)^{1/q_2}.$$

Also, if $p_1 < \infty$, then since for each $i \in I_0$ there is at most one $j \in J_0$ where $f(i, j) > 0$,

$$\begin{aligned} \|f\|_{\sigma(P)} &= \left(\sum_{i \in I_0} \left(\sum_{j \in J_0} f(i, j)^{p_2} u_2(j) \right)^{\frac{p_1}{p_2}} u_1(i) \right)^{\frac{1}{p_1}} = \left(\sum_{j \in J_0} \sum_{i \in I(j)} f(i, j)^{p_1} u_1(i) u_2(j)^{\frac{p_1}{p_2}} \right)^{\frac{1}{p_1}} \\ &= \left\| \left\| g_j \right\|_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{1/p_2} \right\|_{\ell^{p_1}(J_0)}. \end{aligned}$$

If $p_1 = \infty$, the same result holds:

$$\begin{aligned} \|f\|_{\sigma(P)} &= \sup_{i \in I_0} \left(\sum_{j \in J_0} f(i, j)^{p_2} u_2(j) \right)^{\frac{1}{p_2}} = \sup_{i \in I_0} \left(\sup_{j \in J_0} f(i, j)^{p_2} u_2(j) \right)^{\frac{1}{p_2}} \\ &= \sup_{j \in J_0} \sup_{i \in I(j)} \left(f(i, j) u_2(j)^{\frac{1}{p_2}} \right)^{p_2} = \left\| \left\| g_j \right\|_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{\frac{1}{p_2}} \right\|_{\ell^{p_1}(J_0)}. \end{aligned}$$

Therefore, $\|f\|_Q \leq C(I_0, J_0) \, \|f\|_{\sigma(P)}$ can be rewritten as

$$\left(\sum_{j \in J_0} \left\| g_j \right\|^{q_2}_{\ell^{q_1}_{v_1}(I(j))} v_2(j) \right)^{\frac{1}{q_2}} \leq C(I_0, J_0) \left\| \left\| g_j \right\|_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{\frac{1}{p_2}} \right\|_{\ell^{p_1}(J_0)},$$

where $C(I_0, J_0)$ is the least constant so that this holds for all $f \in L^+(I, J)$ supported on $I_0 \times J_0$.

By Hölder's inequality with the conjugate exponents p_1/q_2 and $(p_1 : q_2)/q_2$,

$$\begin{aligned}
\|f\|_Q^{q_2} &= \sum_{j\in J_0} \left\|g_j\right\|^{q_2}_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{\frac{q_2}{p_2}} \frac{\left\|g_j\right\|^{q_2}_{\ell^{q_1}_{v_1}(I(j))} v_2(j)}{\left\|g_j\right\|^{q_2}_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{\frac{q_2}{p_2}}} \\
&\le \left\| \left\|g_j\right\|_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{\frac{1}{p_2}} \right\|^{q_2}_{\ell^{p_1}(J_0)} \left(\sum_{j\in J_0} \left(\frac{\left\|g_j\right\|_{\ell^{q_1}_{v_1}(I(j))} v_2(j)^{\frac{1}{q_2}}}{\left\|g_j\right\|_{\ell^{p_1}_{u_1}(I(j))} u_2(j)^{\frac{1}{p_2}}} \right)^{p_1 : q_2} \right)^{\frac{q_2}{p_1 : q_2}} \\
&= \left\| \frac{\left\|g_j\right\|_{\ell^{q_1}_{v_1}(I(j))}}{\left\|g_j\right\|_{\ell^{p_1}_{u_1}(I(j))}} C_2(j) \right\|^{q_2}_{\ell^{p_1 : q_2}(J_0)} \|f\|^{q_2}_{\sigma(P)} .
\end{aligned}$$

Therefore, the supremum of $\|f\|_Q / \|f\|_{\sigma(P)}$ over $f \in L^+(I \times J)$, supported on $I_0 \times J_0$ and not almost everywhere zero is

$$C(I_0, J_0) = \sup \frac{\|f\|_Q}{\|f\|_{\sigma(P)}} = \sup_{\{I(j)\}} \|C_1(I(j))C_2(j)\|_{\ell^{p_1 : q_2}(J_0)} ,$$

where the rightmost supremum denotes one taken over all disjoint collections $\{I(j) : j \in J_0\}$ of subsets of I. (The ratio $\|g_j\|_{\ell^{q_1}_{v_1}(I(j))}\|g_j\|^{-1}_{\ell^{p_1}_{u_1}(I(j))}$ can approach $C_1(I(j))$ arbitrarily closely for various g_j.) Appropriate choices of f can select both any desired disjoint subsets $I(j)$ of I and the functions g_j on them. Because $q_1 < p_1$, Corollary 2.6.9 gives the formula

$$C_1(I(j)) = \left\| v_1(i)^{\frac{1}{q_1}} u_1(i)^{-\frac{1}{p_1}} \right\|_{\ell^{p_1 : q_1}(I(j))} = \|C_1(i)\|_{\ell^{p_1 : q_1}(I(j))}$$

so

$$\begin{aligned}
C(I_0, J_0) &= \sup_{\{I(j)\}} \left\| \|C_1(i)\|_{\ell^{p_1 : q_1}(I(j))} C_2(j) \right\|_{\ell^{p_1 : q_2}(J_0)} \\
&= \sup_{\{I(j)\}} \left(\sum_{j\in J_0} \left(\sum_{i\in I(j)} C_1(i)^{p_1 : q_1} \right)^{\frac{p_1 : q_2}{p_1 : q_1}} C_2(j)^{p_1 : q_2} \right)^{\frac{1}{p_1 : q_2}} . \qquad (5.11)
\end{aligned}$$

By Corollary 5.2.2, the supremum of all $C(I_0, J_0)$ over finite subsets I_0 and J_0 is C itself. Furthermore, since the expression for $C(I_0, J_0)$ in (5.11) grows with more terms, as J_0 is brought closer to J and the collection $\{I(j) : j \in J\}$ comes closer to partitioning I, the supremum

$$\begin{aligned}
C^{p_1 : q_2} &= \sup_{I_0, J_0} C(I_0, J_0)^{p_1 : q_2} = \sup_{\{I(j)\}} \sum_{j\in J} C_2(j)^{p_1 : q_2} \left(\sum_{i\in I(j)} C_1(i)^{p_1 : q_1} \right)^{\frac{p_1 : q_2}{p_1 : q_1}} \\
&= GPP_{\frac{p_1 : q_2}{p_1 : q_1}} \left(C_2(j)^{p_1 : q_2}, C_1(i)^{p_1 : q_1}\right) ,
\end{aligned}$$

where it is worth noting that $0 < (p_1 : q_2) / (p_1 : q_1) < 1$. Finally, the formula for best constants on singletons is given by Lemma 5.3.1. □

Theorem 5.5.2. *Suppose that $q_2 < p_2 < p_1 \leq q_1$ and each X_k has purely atomic measures. Let the atoms of X_1 be $(E_i)_{i\in I}$ and the atoms of X_2 be $(F_j)_{j\in J}$, where I and J are (at most) countable index sets. For each $i \in I$ and each $j \in J$, let*

$$C_1(i) = \nu_1(E_i)^{1/q_1}\mu_1(E_i)^{-1/p_1} \qquad \text{and} \qquad C_2(j) = \nu_2(F_j)^{1/q_2}\mu_2(F_j)^{-1/p_2}.$$

Then the least constant C such that

$$\|f\|_Q \leq C\,\|f\|_{\sigma(P)}$$

for any $f \in L^+(X_1 \times X_2)$ is

$$C = \left(GPP_{\frac{p_1:q_2}{p_2:q_2}}\left(C_1(i)^{p_1:q_2}, C_2(j)^{p_2:q_2}\right)\right)^{\frac{1}{p_1:q_2}},$$

in terms of a solution to a generalized partition problem.

Proof. Fix finite subsets $I_0 \subset I$ and $J_0 \subset J$. By Proposition 5.2.4, the best constant $C(I_0, J_0)$ for functions supported on $I_0 \times J_0$ can be achieved with a function $f \in L^+(I \times J)$, supported on $I_0 \times J_0$, with at most one non-zero entry in each column, i.e. for each $j \in J_0$ there is at most one $i \in I_0$ such that $f(i, j) > 0$.

This means that there are pairwise disjoint subsets $(J(i))_{i\in I_0}$ of J_0, each $\{j \in J_0 : f(i, j) > 0\}$, the places in row i where f takes positive values. For each fixed $i \in I_0$, define $h_i(j) = f(i, j)$. (Note that then $J(i)$ is the support of h_i.) With these facts,

$$\|fi\|_Q = \left\| \|h_i\|_{\ell^{q_2}_{\nu_2}(I(j))}\, \nu_1(i)^{1/q_1} \right\|_{\ell^{q_2}(I_0)}. \tag{5.12}$$

This is proven by one of two computations, depending on the value of q_1. If $q_1 < \infty$,

$$\begin{aligned}
\|f\|_Q &= \left(\sum_{j\in J_0}\left(\sum_{i\in I_0} f(i,j)\right)^{q_2} \nu_2(j)\right)^{\frac{1}{q_2}} \\
&= \left(\sum_{i\in I_0}\sum_{j\in J(i)} f(i,j)^{q_2}\nu_1(i)^{\frac{q_2}{q_1}}\nu_2(j)\right)^{\frac{1}{q_2}} \\
&= \left(\sum_{i\in I_0} \|h_i\|^{q_2}_{\ell^{q_2}_{\nu_2}(J(i))}\, \nu_1(i)^{\frac{q_2}{q_1}}\right)^{\frac{1}{q_2}}.
\end{aligned}$$

If $q_1 = \infty$,

$$\|f\|_Q = \left(\sum_{j\in J_0}\left(\sup_{i\in I_0} f(i,j)\right)^{q_2} \nu_2(j)\right)^{\frac{1}{q_2}} = \left(\sum_{j\in J_0}\sum_{i\in I_0} f(i,j)^{q_2}\nu_2(j)\right)^{\frac{1}{q_2}},$$

because the supremum and sum over a unique term are the same. Next, because $\nu_1(i)^{q_2}q_1 = 1$ when $q_1 = \infty$,

$$\|f\|_Q = \left(\sum_{i\in I_0}\left(\sum_{j\in J(i)} f(i,j)^{q_2}\nu_2(j)\right)\nu_1(i)^{\frac{q_2}{q_1}}\right)^{\frac{1}{q_2}} = \left(\sum_{i\in I_0}\|h_i\|^{q_2}_{\ell^{q_2}_{\nu_2}(J(i))}\,\nu_1(i)^{\frac{q_2}{q_1}}\right)^{\frac{1}{q_2}}.$$

It is easily seen that

$$\|f\|_{\sigma(P)} = \left\| \|h_i\|_{\ell^{p_2}_{u_2}(J(i))} \, u_1(i)^{\frac{1}{p_1}} \right\|_{\ell^{p_1}(I_0)} .$$

Equation (5.12) leads to the following application of Hölder's inequality, using the conjugate exponents p_1/q_2 and $(p_1 : q_2)/q_2$.

$$\begin{aligned} \|f\|_Q^{q_2} &= \sum_{i \in I_0} \left(\|h_i\|_{\ell^{p_2}_{u_2}(J(i))} \, u_1(i)^{\frac{1}{p_1}} \right)^{q_2} \left(\frac{\|h_i\|_{\ell^{q_2}_{v_2}(J(i))} \, v_1(i)^{\frac{1}{q_1}}}{\|h_i\|_{\ell^{p_2}_{u_2}(J(i))} \, u_1(i)^{\frac{1}{p_1}}} \right)^{q_2} \\ &\le \left\| \|h_i\|_{\ell^{p_2}_{u_2}(J(i))} \, u_1(i)^{\frac{1}{p_1}} \right\|_{\ell^{p_1}(I_0)}^{q_2} \left(\sum_{i \in I_0} \left(\frac{\|h_i\|_{\ell^{q_2}_{v_2}(J(i))} \, v_1(i)^{\frac{1}{q_1}}}{\|h_i\|_{\ell^{p_2}_{u_2}(J(i))} \, u_1(i)^{\frac{1}{p_1}}} \right)^{p_1 : q_2} \right)^{\frac{q_2}{p_1 : q_2}} . \end{aligned}$$

Therefore

$$\|f\|_Q \le \left\| \frac{\|h_i\|_{\ell^{q_2}_{v_2}(J(i))}}{\|h_i\|_{\ell^{p_2}_{u_2}(J(i))}} C_1(i) \right\|_{\ell^{p_1 : q_2}(I_0)} \|f\|_{\sigma(P)} .$$

Different choices of the function f can provide any disjoint collection $\{J(i) : i \in I_0\}$ of subsets of I_0 and functions h_i, each supported on $J(i)$. These functions h_i can be chosen so as to bring the ratio $\|h_i\|_{\ell^{q_2}_{v_2}(J(i))} / \|h_i\|_{\ell^{p_2}_{u_2}(J(i))}$ as near as desired to the local best constant $C(I_0)$. Therefore, the supremum $C(I_0, J_0)$ of $\|f\|_Q / \|f\|_{\sigma(P)}$ over $f \in L^+(I \times J)$ supported on $I_0 \times J_0$ (and not almost everywhere zero) can be expressed as follows:

$$C(I_0, J_0) = \sup_f \frac{\|f\|_Q}{\|f\|_{\sigma(P)}} = \sup_{\{J(i)\}} \|C_1(i) C_2(J(i))\|_{\ell^{p_1 : q_2}(I_0)} ,$$

where the latter supremum is over all disjoint collections $\{J(i) : i \in I_0\}$ of subsets of J_0. Because $q_2 < p_2$, Corollary 2.6.9 gives the formula

$$C_2(J(i)) = \left\| v_2(j)^{\frac{1}{q_2}} u_2(j)^{-\frac{1}{p_2}} \right\|_{\ell^{p_2 : q_2}(J(i))} = \|C_2(j)\|_{\ell^{p_2 : q_2}(J(i))} .$$

Therefore

$$\begin{aligned} C(I_0, J_0) &= \sup_{\{J(i)\}} \left\| C_1(i) \, \|C_2(j)\|_{\ell^{p_2 : q_2}(J(i))} \right\|_{\ell^{p_1 : q_2}(I_0)} \\ &= \sup_{\{J(i)\}} \left(\sum_{i \in I_0} \left(\sum_{j \in J(i)} C_2(j)^{p_2 : q_2} \right)^{\frac{p_1 : q_2}{p_2 : q_2}} C_1(i)^{p_1 : q_2} \right)^{\frac{1}{p_1 : q_2}} . \end{aligned} \tag{5.13}$$

By Corollary 5.2.2, the supremum of $C(I_0, J_0)$ over finite subsets I_0 and J_0 is C itself. As J_0 comes closer to J and the collection $\{I(j) : j \in J\}$ comes closer to partitioning I, more terms are introduced and the expression in (5.13) grows. Therefore, the supremum

$$\begin{aligned} C^{p_1 : q_2} &= \sup_{I_0, J_0} C(I_0, J_0)^{p_1 : q_2} = \sup_{\{J(i)\}} \sum_{i \in I} C_1(i)^{p_1 : q_2} \left(\sum_{j \in J(i)} C_2(j)^{p_2 : q_2} \right)^{\frac{p_1 : q_2}{p_2 : q_2}} \\ &= GPP_{\frac{p_1 : q_2}{p_2 : q_2}} \left(C_1(i)^{p_1 : q_2}, C_2(j)^{p_2 : q_2} \right) , \end{aligned}$$

where $0 < (p_1 : q_2) / (p_2 : q_2) < 1$. Finally, Lemma 5.3.1 gives the best constants for singletons.

□

Chapter 6

Two-variable summary

The following summary clarifies the cases in the two-variable inclusion problem, and how the results fit together to solve it. Assume that X_1 and X_2 are measurable spaces and, for each $k = 1, 2$, there are non-zero σ-finite measures μ_k and ν_k on X_k. Bounds and, where possible, values are provided for the best constant C in one of the following inequalities.

$$\left(\int_{X_2}\left(\int_{X_1}|f|^{q_1}\,d\nu_1\right)^{q_2/q_1}d\nu_2\right)^{1/q_2} \le C\left(\int_{X_2}\left(\int_{X_1}|f|^{p_1}\,d\mu_1\right)^{p_2/p_1}d\mu_2\right)^{1/p_2} \tag{6.1}$$

$$\left(\int_{X_2}\left(\int_{X_1}|f|^{q_1}\,d\nu_1\right)^{q_2/q_1}d\nu_2\right)^{1/q_2} \le C\left(\int_{X_1}\left(\int_{X_2}|f|^{p_1}\,d\mu_2\right)^{p_1/p_2}d\mu_1\right)^{1/p_1} \tag{6.2}$$

The corresponding mixed-norm inclusion holds if $C < \infty$ and fails if the least constant $C = \infty$. For each $k = 1, 2$, let $C_k \in [0, \infty]$ denote the least constant such that, for any $f_k \in L^+(X_k)$, $\|f_k\|_{L^{q_k}_{\nu_k}(X_k)} \le C_k \|f_k\|_{L^{p_k}_{\mu_k}(X_k)}$.

Inequality (6.1) is the "unpermuted case" solved in Section 3.4, with a simple solution given by Theorem 3.4.1.

$$C = C_1C_2,$$

so that inclusion holds if and only if each $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$, or equivalently $C_k < \infty$. (Each $C_k > 0$ because no measure μ_k or ν_k is zero.)

Inequality (6.2) is the "permuted case" introduced in Chapter 3 and treated in various sub-cases. The following questions step through solving an instance of the problem.

1. Do both one-variable inclusions $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$ hold, for $k = 1, 2$?
   - By Proposition 3.3.3, $C \ge C_1C_2$.
   - Therefore, unless both inclusions hold, inclusion is impossible ($C = \infty$).
2. Is this the Minkowski case, where $\min(p_1, q_1) \le \max(p_2, q_2)$?
   - If so, by Theorem 3.6.1, $C = C_1C_2$.

From this point, we must have both one-variable inclusions, for which absolute continuity $\nu_k \ll \mu_k$ is necessary. Assume that each $\frac{d\nu_k}{d\mu_k} > 0$, μ_k-almost everywhere, if necessary by reducing to functions supported where this is so. (Proposition 4.2.1) With this assumption, note that each μ_k has exactly the same atoms and the same null sets as ν_k, so we can speak of "atoms of X_k". Similarly, we can discuss whether or not X_k or any measurable subset is atomless, as well as whether or not it is purely atomic.

Now decompose $X_1 = E_0 \dot\cup (\bigcup_{i\in I} E_i)$ and $X_2 = F_0 \dot\cup (\bigcup_{j\in J} F_j)$, where E_0 and F_0 are atomless and each E_i or F_j is an atom.

3. Is either X_1 or X_2 purely atomic? (That is, is either E_0 or F_0 a null set?)

   - If not, $C = \infty$. Theorem 4.2.16 shows that the Minkowski sufficient condition (Theorem 3.6.1) is in this case also necessary.

What remains is the non-Minkowski case, $\max(p_2, q_2) < \min(p_1, q_1)$, with one or both of X_1 and X_2 purely atomic.

4. Is X_1 purely atomic, but X_2 not?

   - If so, inclusion holds if and only if the sequence $M_i := \nu_1(E_i)^{1/q_1}\mu_1(E_i)^{-1/p_1}$ is $(p_1 : p_2)$-summable. (Recall $p_1 : p_2 = (p_2^{-1} - p_1^{-1})^{-1}$.) In this case,

$$C_2(F_0)\,\|M_i\|_{p_1\,:\,p_2} \le C \le C_2\,\|M_i\|_{p_1\,:\,p_2}\,.$$

5. Is X_2 purely atomic, but X_1 not?

   - If so, inclusion holds if and only if the sequence $N_j := \nu_2(F_j)^{1/q_2}\mu_2(F_j)^{-1/p_2}$ is $(q_1 : q_2)$-summable. In this case,

$$C_1(E_0)\,\left\|N_j\right\|_{q_1\,:\,q_2} \le C \le C_1\,\left\|N_j\right\|_{q_1\,:\,q_2}\,.$$

Otherwise, both measures are purely atomic. The non-Minkowski $\max(p_2, q_2) < \min(p_1, q_1)$ divides into four more specific cases, with different solutions. The following are true in each of these cases.

- From factorable functions (Proposition 3.3.3): $C_1C_2 \le C$.

- From "diagonal" functions (Theorem 4.2.11 applied to functions of the same form as Theorem 5.3.4): $\|M^*N^*\|_{p_1\,:\,q_2} \le C$.

- From the use of Hölder's inequality and Tonelli's theorem (Propositions 4.2.19, 4.2.20, 5.1.1, and 5.1.2):

$$C \le C_1\,\|N\|_{q_1\,:\,q_2}\,, \qquad C \le C_1\,\|N\|_{p_1\,:\,q_2}\,,$$
$$C \le C_2\,\|M\|_{p_1\,:\,p_2}\,, \qquad \text{and } C \le C_2\,\|M\|_{p_1\,:\,q_2}\,.$$

6. Is $p_2 \leq q_2 < p_1 \leq q_1$? If so,
   - Theorem 5.3.4 gives the best constant with "diagonal" functions, $C = \|M^*N^*\|_{p_1 : q_2}$.
7. Is $p_2 \leq q_2 < q_1 < p_1$? If so, by Theorem 5.5.1,

$$C = \left(GPP_{\frac{p_1 : q_2}{p_1 : q_1}}\left(C_2(j)^{p_1 : q_2}, C_1(i)^{p_1 : q_1}\right)\right)^{\frac{1}{p_1 : q_2}}.$$

8. Is $q_2 < p_2 < p_1 \leq q_1$? If so, by Theorem 5.5.2,

$$C = \left(GPP_{\frac{p_1 : q_2}{p_2 : q_2}}\left(C_1(i)^{p_1 : q_2}, C_2(j)^{p_2 : q_2}\right)\right)^{\frac{1}{p_1 : q_2}}.$$

(Refer to Section 5.4 for a description of the generalized partition problem (GPP).)

9. Otherwise ($q_2 < p_2 < q_1 < p_1$, X_1 and X_2 purely atomic), the problem is not fully solved. It is not clear that there is any tractable condition to be found which is both necessary and sufficient, and it is feared that it may be at least as tricky as the computationally difficult of the GPP cases.

   There are, at least, the separate necessary conditions and sufficient conditions given above, plus necessary conditions derived by using either particular functions or special classes of functions, like block-factorable functions.

(Note that not all purely atomic spaces can produce such difficult problems. For one simple example, when dealing only with unweighted ℓ^p, a special case of Theorem 7.6.2 shows that the Minkowski condition is both necessary and sufficient.)

Chapter 7

Multiple-variable case

7.1 Notation

Let $(X_1, \Sigma_1), \ldots, (X_n, \Sigma_n)$ be measurable spaces, each Σ_k a σ-algebra on X_k. Denote the product by $X = X_1 \times \cdots \times X_n$, with its σ-algebra Σ. Let $L^+(X_k)$ represent the space of nonnegative measurable functions on (X_k, Σ_k), and let $L^+(X)$ represent the space of nonnegative measurable functions on (X, Σ).

For any $k \in \{1. \ldots, n\}$, any $p_k \in (0, \infty]$, and any σ-finite measure μ_k on X_k, let $\rho_{\mu_k}^{p_k}$ be a map taking $L^+(X)$ to itself, defined by

$$\rho_{\mu_k}^{p_k}(f)(x_1, \ldots, x_n) = \begin{cases} \left(\int_{X_k} f(x_1, \ldots, x_n)^{p_k} d\mu_k\right)^{1/p_k} & \text{if } p_k < \infty \\ \operatorname{ess\,sup}_{x_k \in (X_k, \mu_k)} f(x_1, \ldots, x_n) & \text{if } p_k = \infty \end{cases}$$

The idea of this map is that it takes an L^{p_k} norm (a norm, at least, for $p_k \geq 1$) in the variable x_k, resulting in a function which is constant in x_k and depends only on the remaining variables.

By composing such maps, we can express the L^P mixed norms described by Benedek and Panzone in [4]. Given an n-tuple of exponents $P = (p_1, \ldots, p_n)$, each $p_k \in (0, \infty]$ and σ-finite measures $\mu_1, \ldots, \mu_n$, each μ_k on X_k, define for each $k \in \{1, \ldots, n\}$ a map $\rho_k := \rho_{\mu_k}^{p_k}$ as specified above. The composition is denoted by

$$\rho = \rho_n \circ \cdots \circ \rho_1.$$

Observe that, for any $f \in L^+(X)$, $\rho(f)$ is a constant function. When each $p_k \geq 1$, its sole value is that of the mixed norm $\|f\|_P$ in the notation used by Benedek and Panzone, computed with respect to the measures $\mu_1, \ldots, \mu_n$. (The same procedure is used with at least one exponent less than 1, but the result is not, strictly speaking, a norm.) Making the natural identification of numbers with constant functions, ρ is that mixed norm. We then define the mixed norm space

$$L^\rho = \{f \in L^+(X) \mid \rho(f) < \infty\}, \tag{7.1}$$

modulo the identification of any pair of functions which agree almost everywhere. Of course, the full mixed-norm space L^ρ will include all measurable real-valued or complex-valued functions f such that the modulus $|f| \in L^\rho$ as defined above, but since all of the work here depends only on the modulus, with no addition of functions, only those functions in $L^+(X)$ need be considered.

The composition need not be $\rho_n \circ \cdots \circ \rho_1$ in numeric order, however. By introducing a permutation to the composition, we can obtain what Fournier [16] called "permuted mixed norms." As in the two-variable case, it turns out that the question of embeddings has a trivial answer unless the mixed norms involved are differently permuted.

Given exponents $p_1, \ldots, p_n \in (0, \infty]$ and measures $\mu_1, \ldots, \mu_n$, and additionally a permutation $\sigma \in S_n$, where as before $\rho_k = \rho_{\mu_k}^{p_k}$, form the composition

$$\rho = \rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}.$$

Such (permuted) mixed norms are those for which we consider the inclusion problem.

7.2 Mixed-norm Hölder and Minkowski inequalities

Benedek and Panzone [4] provide a mixed-norm version of Hölder's inequality, in several formulations. The following is a consequence of their Theorem 2 in Section 2. Unlike Minkowski's integral inequality, which involves mixed norms with different permutations, the two mixed norms which appear in Hölder's inequality, ρ_P and $\rho_{P'}$ below, have their variables in the same order.

Theorem 7.2.1. *Let $(X_1, \mu_1), \ldots, (X_n, \mu_n)$ be σ-finite measure spaces with product (X, μ), $P = (p_1, \ldots, p_n)$ with each $p_k \in [1, \infty]$, and $P' = \left(p_1', \ldots, p_n'\right)$, meaning that each $\frac{1}{p_k} + \frac{1}{p_k'} = 1$. For each $k \in \{1, \ldots, n\}$, define $\rho_k = \rho_{\mu_k}^{p_k}$, $\rho_P = \rho_{p_n} \circ \cdots \circ \rho_{p_1}$, $\rho_k' = \rho_{\mu_k}^{p_k'}$, and $\rho_{P'} = \rho_n' \circ \cdots \circ \rho_1'$. Then, for any $f, g \in L^+(X)$,*

$$\int_X f(x)g(x)d\mu(x) \leq \rho_P(f)\rho_{P'}(g). \tag{7.2}$$

Furthermore, for any fixed f, $\rho_P(f)$ is the least constant C such that, for any $g \in L^+(X)$,

$$\int_X f(x)g(x)d\mu(x) \leq C\rho_{P'}(g). \tag{7.3}$$

Proof. Theorem 2 in Section 2 of Benedek and Panzone says that, for any measurable function f on X,

$$\rho_P(f) = \sup_{g \in U_{P'}} \left| \int_X f(x)g(x)d\mu(x) \right|,$$

where $U_{P'}$ denotes the unit sphere in $L^{P'}$, i.e. the set of functions g for which $\rho_{P'}(g) = 1$. If $g = 0$ μ-a.e., then (7.2) holds trivially, as does (7.3) with $C = 0$. As long as $\rho_{P'}(g) < \infty$, let $\tilde{g}(x) = (\rho_{P'}(g))^{-1}\, g(x)$ and observe that $\rho_{P'}(\tilde{g}) = 1$, so $\tilde{g} \in U_{P'}$. Therefore

$$\begin{aligned}
&\int_X f(x)\tilde{g}(x)d\mu(x) \leq \rho_P(f), \text{ so} \\
&\int_X f(x)g(x)d\mu(x) = \rho_{P'}(g) \int_X f(x)\tilde{g}(x)d\mu(x) \\
&\qquad\qquad\qquad\qquad \leq \rho_{P'}(g)\rho_P(f).
\end{aligned}$$

For $1 \le k \le n$, because μ_k is σ-finite there is a sequence $E_{k,1} \subset E_{k,2} \subset \cdots$ with each $\mu_k E_{k,m} < \infty$ and $\bigcup_m E_{k,m} = X_k$. For each $m \ge 1$, define

$$g_m(x_1, \dots, x_n) = \min(m, g(x_1, \dots, x_n)) \prod_{k=1}^{n} \chi_{E_{k,m}}(x_k).$$

Since $g_m \le m$ and each set $E_{k,m}$ has finite measure, $\rho_{P'}(g_m) < \infty$. The preceding argument then shows that $\int_X f(x) g_m(x) d\mu(x) \le \rho_P(f)\rho_{P'}(g_m)$. Of course, (g_m) is an increasing sequence of sets converging pointwise to g, so by monotone convergence

$$\int_X f(x)g(x) \le \rho_P(f)\rho_{P'}(g).$$

(Although the monotone convergence theorem is technically only for integrals, dealing with $p'_k < \infty$, it is also valid to conclude that any $\rho_{\mu_k}^{p'_k}(g_m) \to \rho_{\mu_k}^{p'_k}(g)$ when $p'_k = \infty$.)

To see that $\rho_P(f)$ is the least value of C for the second inequality, note that the second inequality applies to any $g \in U_{P'}$. For such functions g, it states that $\int_X f(x)g(x)d\mu(x) \le C$, which when combined with Benedek and Panzone's Theorem 2 means that $\rho_P(f) \le C$. □

Hölder's inequality has another formulation involving more than two mixed norms, analogous to the version of Hölder's inequality which uses more than two L^p norms. Previously stated in [4], this result is presented with the notation used here for convenience.

Theorem 7.2.2. *Given any $m \ge 1$, for $i \in \{1, \dots, m\}$ let $P_i = (p_i(1), \dots, p_i(n))$ be an n-tuple of exponents from $[1, \infty]$. Suppose that $\sum_{i=1}^{m} (P_i)^{-1} = 1$ interpreted coordinatewise, i.e. for each $j \in \{1, \dots, n\}$, $\sum_{i=1}^{m} p_i(j)^{-1} = 1$. Then, for any $f_1, \dots, f_m \in L^+(X)$,*

$$\int_X f_1 \cdots f_m d\mu_1 \cdots d\mu_n \le \prod_{i=1}^{m} \rho_{P_i}(f_i),$$

where each ρ_{P_i} is the mixed norm defined by

$$\rho_{P_i}(f) = \left(\int_{X_n} \cdots \int_{X_2} \left(\int_{X_1} f^{p_i(1)} d\mu_1 \right)^{p_i(2)/p_i(1)} d\mu_2 \cdots d\mu_n \right)^{1/p_i(n)}$$

for any $f \in L^+(X)$, replacing by an essential supremum if any exponent is ∞.

Proof. Simply apply the traditional m-function version of Hölder's inequality, variable by variable. □

Minkowski's integral inequality, as stated in Corollary 3.5.2, can be generalized to situations involving more than two variables. Such a mixed-norm Minkowski's integral inequality was described by Fournier in [16]; although the Theorem 2.2 there explicitly describes only the greatest and least permutations, those tend to be particularly useful cases, and there is no reason to believe that Fournier would have been unaware of the ideas in this version. This mixed-norm Minkowski's integral inequality is presented not as a novel result, but to provide a statement compatible with the notation used here, and so as to include some examples of how the mixed-norm versions of Hölder's and Minkowski's inequalities can be useful.

Proposition 7.2.3. *Let $(X_1, \mu_1), \ldots, (X_n, \mu_n)$ be σ-finite measure spaces, with product space X. Suppose that $i, j \in \{1, \ldots, n\}$, $i \neq j$, and $1 \leq p_i \leq p_j \leq \infty$. For any $f \in L^+(X)$,*

$$\rho_{\mu_j}^{p_j} \circ \rho_{\mu_i}^{p_i}(f) \leq \rho_{\mu_i}^{p_i} \circ \rho_{\mu_j}^{p_j}(f).$$

Proof. Take any $f \in L^+(X)$. For fixed values of $x_1, \ldots, x_n$, excluding x_i and x_j, define $g(x_i, x_j) = f(x_1, \ldots, x_n)$, so that $g \in L^+(X_i \times X_j)$. Because $1 \leq p_i \leq p_j \leq \infty$, Corollary 3.5.2 provides that

$$\left\| \|g\|_{L_{\mu_i}^{p_i}} \right\|_{L_{\mu_j}^{p_j}} \leq \left\| \|g\|_{L_{\mu_j}^{p_j}} \right\|_{L_{\mu_i}^{p_i}}. \tag{7.4}$$

The above one-variable norms in x_i and x_j produce functions of the remaining variables; because (7.4) holds for every value of these variables, one of these functions is greater than the other. The quantities can also be viewed as functions of $(x_1, \ldots, x_n)$ which happen to be constant in x_i and x_j. Either way, (7.4) means that $\rho_{\mu_j}^{p_j} \circ \rho_{\mu_i}^{p_i}(f) \leq \rho_{\mu_i}^{p_i} \circ \rho_{\mu_j}^{p_j}(f)$. □

By repeatedly applying adjacent permutations, we can achieve the following generalization of Minkowski's integral inequality to mixed norms. (Since the number s of one-variable L^p computations involved could be less than the number n of spaces, this result also applies to compositions which are not mixed norms, for their values are functions of $n - s$ variables rather than constants.)

Theorem 7.2.4 (Mixed-norm Minkowski integral inequality)**.** *Let $(X_1, \mu_1), \ldots, (X_n, \mu_n)$ be σ-finite measure spaces. For each $k \in \{1, \ldots, n\}$, let $\rho_k = \rho_{\mu_k}^{p_k}$, for some exponent $0 < p_k \leq \infty$. Let S be a subset of $\{1, \ldots, n\}$, enumerated by $k_1, \ldots, k_s$ and $l_1, \ldots, l_s$, where $s = |S|$. Define a permutation of S by $\tau(a) = b$ if and only if $k_a = l_b$. If $p_{k_i} < p_{k_j}$ whenever $i < j$ and $\tau(i) > \tau(j)$, then*

$$\rho_{k_s} \circ \cdots \circ \rho_{k_1} \leq \rho_{l_s} \circ \cdots \circ \rho_{l_1}.$$

Proof. The proof is by induction on s. Both $s = 0$ (with the identity as a trivial composition) and $s = 1$ are trivial, providing base cases. Recall the hypothesis that $i < j$ and $\tau(i) > \tau(j)$ imply $p_{k_i} < p_{k_j}$. Whenever $i < \tau^{-1}(1)$, of course $\tau(i) > 1$, so this hypotheses yields $p_{k_i} < p_{k_{\tau^{-1}(1)}} = p_{l_1}$. For such i, Proposition 7.2.3 then shows that $\rho_{l_1} \circ \rho_{k_i} \leq \rho_{k_i} \circ \rho_{l_1}$. Repeatedly applying these gives

$$\begin{aligned}
\rho_{k_{\tau^{-1}(1)}} \circ \cdots \circ \rho_{k_1} &= \rho_{l_1} \circ \rho_{k_{\tau^{-1}(1)-1}} \circ \cdots \circ \rho_{k_1} \\
&\leq \rho_{k_{\tau^{-1}(1)-1}} \circ \rho_{l_1} \circ \rho_{k_{\tau^{-1}(1)-2}} \circ \cdots \circ \rho_{k_1} \\
&\vdots \\
&\leq \rho_{k_{\tau^{-1}(1)-1}} \circ \cdots \circ \rho_{k_1} \circ \rho_1.
\end{aligned}$$

Because furthermore every ρ_k is monotone, in the sense that $f \leq g$ implies $\rho_k(f) \leq \rho_k(g)$, precompose $\rho_{k_s} \circ \cdots \circ \rho_{k_{\tau^{-1}(1)+1}}$ to obtain

$$\rho_{k_s} \circ \cdots \circ \rho_{k_1} \leq \rho_{k_s} \circ \cdots \circ \widehat{\rho_{k_{\tau^{-1}(1)}}} \circ \cdots \circ \rho_{k_1} \circ \rho_{l_1}$$

where $\widehat{\rho_{k_{\tau^{-1}(1)}}}$ indicates that this term is omitted, appearing instead at the beginning of the composition as ρ_{l_1}. The inductive hypothesis, applied to $S \setminus \left\{k_{\tau^{-1}(1)}\right\} = S \setminus \{l_1\}$, provides

$$\rho_{k_s} \circ \cdots \circ \widehat{\rho_{k_{\tau^{-1}(1)}}} \circ \cdots \circ \rho_{k_1} \leq \rho_{l_s} \circ \cdots \circ \rho_{l_2},$$

which completes the inductive step with

$$\rho_{k_s} \circ \cdots \circ \rho_{k_1} \leq \rho_{l_s} \circ \cdots \circ \rho_{l_1}.$$

□

Although the hypothesis looks a bit complicated at first glance, the idea is that Minkowski's integral inequality applies to any permutation which is formed from adjacent transpositions, each of which moves norms with larger exponents earlier in the composition; the mixed norm resulting from such a permutation always gives larger values than the original mixed norm. On the other hand, if each adjacent transposition moves norms with larger exponents later in the composition, then the resulting norm always gives smaller values than the original mixed norm.

For a theorem with a simpler statement, we have the following useful corollary.

Corollary 7.2.5. *Let $(X_1, \mu_1), \ldots, (X_n, \mu_n)$ be σ-finite measure spaces. For each $k \in \{1, \ldots, n\}$, let $\rho_k = \rho_{\mu_k}^{p_k}$ for some $0 < p_k \leq \infty$. If $\sigma \in S_n$ is a permutation such that $p_i < p_j$ whenever both $1 \leq i < j \leq n$ and $\sigma^{-1}(i) > \sigma^{-1}(j)$, then*

$$\rho_n \circ \cdots \circ \rho_1 \leq \rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}.$$

Proof. Simply apply Theorem 7.2.4 with $k_i = i$ and $l_i = \sigma(i)$. Then $\tau = \sigma^{-1}$, since for each $i \in \{1, \ldots, n\}$, $l_{\sigma^{-1}(i)} = \sigma(\sigma^{-1}(i)) = i = k_i$. □

The last result provides a partial order on those mixed norms produced by permuting some particular starting norm. (In terms of group actions, this partial order is on an orbit of the action of the symmetric group S_n on mixed norms in n variables.) Among these, there are always a maximum and a minimum based on the order of the exponents. This is given by the following special case, which describes the same idea which Fournier proved in [16] as Theorem 2.2.

Corollary 7.2.6. *For each $k \in \{1, \ldots, n\}$, let (X_k, μ_k) be a σ-finite measure space, let $0 < p_k \leq \infty$, and let $\rho_k = \rho_{\mu_k}^{p_k}$. Suppose that $\sigma, \tau \in S_n$ are permutations such that*

$$p_{\sigma(1)} \leq p_{\sigma(2)} \leq \cdots \leq p_{\sigma(n)},$$
$$p_{\tau(1)} \geq p_{\tau(2)} \geq \cdots \geq p_{\tau(n)}.$$

Then

$$\rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)} \leq \rho_n \circ \cdots \circ \rho_1 \leq \rho_{\tau(n)} \circ \cdots \circ \rho_{\tau(1)}.$$

That is, of all permutations of the mixed norm $\rho_n \circ \cdots \circ \rho_1$, one which puts the exponents in decreasing order gives the greatest values, and thus the smallest mixed-norm space. At the other extreme, a permutation which puts the exponents in increasing order gives the least values and the largest space. (When some exponents are equal, such extreme permutations may not be unique, but by Tonelli's theorem the resulting mixed-norm spaces will be the same.)

7.3 Problem statement

Let $(X_1, \Sigma_1), \ldots, (X_n, \Sigma_n)$ be measurable spaces with corresponding non-zero, σ-finite measures μ_1 to μ_n and ν_1 to ν_n, and let $X = X_1 \times \cdots \times X_n$. Also, let $P = (p_1, \ldots, p_n)$ and $Q = (q_1, \ldots, q_n)$ be vectors of exponents, each in $(0, \infty]$, and both σ and η be permutations of $\{1, \ldots, n\}$. As noted above, for each $k \in \{1, \ldots, n\}$ let $\rho_k = \rho_{\mu_k}^{p_k}$ and form the mixed norm $\rho = \rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}$. Similarly, for each k let $\xi_k = \rho_{\nu_k}^{q_k}$, then define $\xi = \xi_{\eta(n)} \circ \cdots \circ \xi_{\eta(1)}$.

We aim to find necessary and sufficient conditions for $L^\rho \subset L^\xi$ or, equivalently, for there to exist $C < \infty$ (and, ideally, to determine the least C) such that, for every $f \in L^+(X)$,

$$\xi(f) \leq C\rho(f), \tag{7.5}$$

which would be written in terms of single-variable norms as

$$\xi_{\eta(n)} \circ \cdots \circ \xi_{\eta(1)}(f) \leq C\rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}(f).$$

When $n = 1$, the permutations are irrelevant and this is solved by Theorem 2.1.8. The case $n = 2$ is the subject of the two-variable chapters, with the unpermuted case covering $\sigma = \eta$ and the permuted case covering $\sigma \neq \eta$. For example, when $n = 2$ and $\sigma = \eta$ is the identity, the inequality (7.5) expands to $\xi_2 \circ \xi_1(f) \leq C\rho_2 \circ \rho_1(f)$, i.e. $\rho_{\nu_2}^{q_2} \circ \rho_{\nu_1}^{q_1}(f) \leq C\rho_{\mu_2}^{p_2} \circ \rho_{\mu_1}^{p_1}(f)$, which represents, in the notation used in Chapter 3,

$$\Big\| \|f\|_{L_{\nu_1}^{q_1}(X_1)} \Big\|_{L_{\nu_2}^{q_2}(X_2)} \leq C \Big\| \|f\|_{L_{\mu_1}^{p_1}(X_1)} \Big\|_{L_{\mu_2}^{p_2}(X_2)}$$

The following results address the problem of determining when $L^\rho \subset L^\xi$ for more general n, finding conditions for inclusion to exist between permuted mixed norm spaces. By reordering the variables according to η, we can reduce to one permutation. Suppose that we replace each X_k by $X_{\eta(k)}$, reorder the corresponding exponents p_k and q_k as well as the measures μ_k and ν_k, and replace σ by $\tilde{\sigma} = \sigma \circ \eta^{-1}$. With this relabeling of coordinates, inequality (7.5) is equivalent to

$$\xi_n \circ \cdots \xi_1(f) \leq C\rho_{\tilde{\sigma}(n)} \circ \cdots \rho_{\tilde{\sigma}(1)}(f) \tag{7.6}$$

So, we can always let η be the identity when solving this problem, which is exactly what was done in Chapter 3. Henceforth, η is therefore not used.

Why is $L^\rho \subset L^\xi$ equivalent to the existence of $C < \infty$? A first step is to sweep away the annoying technicality that the presence of an exponent less than 1 means that we do not, strictly speaking, have norms. This can be achieved through the following lemma, which uses a trick to adjust all exponents involved in the problem simultaneously.

Lemma 7.3.1. *Fix an arbitrary real number $t > 0$ and, for each $k \in \{1, \ldots, n\}$, define $\tilde{\rho}_k$ the same way as ρ_k, but replacing p_k by tp_k, i.e.*

$$\tilde{\rho}_k(f)(x_1, \ldots, x_n) = \begin{cases} \left(\int_{X_k} f(x_1, \ldots, x_n)^{tp_k} d\mu_k \right)^{1/tp_k} & \text{if } p_k < \infty \\ \operatorname{ess\,sup}_{x_k \in (X_k, \mu_k)} f(x_1, \ldots, x_n) & \text{if } p_k = \infty \end{cases}.$$

and similarly define $\tilde{\xi}_k$ which differs from ξ_k only in its use of tq_k in place of q_k. Let $C \in [0, \infty]$ denote the least constant such that

$$\xi(f) \leq C\rho(f) \tag{7.7}$$

holds for all $f \in L^+(X)$, and $D \in [0,\infty]$ denote the least constant such that

$$\tilde{\xi}(h) \leq D\tilde{\rho}(h) \tag{7.8}$$

holds for all $h \in L^+(X)$. Then $C = D^t$.

Furthermore, $L^\rho \subset L^\xi$ if and only if $L^{\tilde{\rho}} \subset L^{\tilde{\xi}}$.

Proof. Given any $h \in L^+(X)$, let $f = h^t$. Observe that

$$\begin{aligned}\tilde{\xi}(h) &= \tilde{\xi}_n \circ \cdots \circ \tilde{\xi}_1(h) \\ &= \left(\int_{X_n} \cdots \left(\int_{X_1} f(x_1,\ldots,x_n)^{tp_1} d\mu_1(x_1)\right)^{tp_2/tp_1} \cdots d\mu_n(x_n)\right)^{1/tp_n} = \xi(f)^{1/t}\end{aligned}$$

and

$$\begin{aligned}\rho(f) &= \rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}(f) \\ &= \left(\int_{X_{\sigma(n)}} \cdots \left(\int_{X_{\sigma(1)}} f(x_1,\ldots,x_n)^{tp_1}\, d\mu_{\sigma(1)}(x_{\sigma(1)})\right)^{p_{\sigma(2)}/p_{\sigma(1)}} \cdots d\mu_{\sigma(n)}(x_{\sigma(n)})\right)^{1/p_{\sigma(n)}} = \tilde{\rho}(h)^t,\end{aligned}$$

with appropriate use of the essential supremum if any exponent is ∞. Therefore, with C the least constant such that (7.7) holds,

$$\tilde{\xi}(h) = \xi(f)^{1/t} \leq (C\rho(f))^{1/t} = C^{1/t}\tilde{\rho}(h).$$

Because D is the least constant for (7.8), $D \leq C^{1/t}$, so $D^t \leq C$.

To obtain the reverse inequality, for any $f \in L^+(X)$, let $h = f^{1/t}$. Of course, this implies that $f = h^t$, as above, so the same equalities apply, and

$$\xi(f) = \tilde{\xi}(h)^t \leq (D\tilde{\rho}(h))^t = D^t\rho(f).$$

Because C is the least constant for (7.7), $C \leq D^t$.

Finally, to see equivalence of the inclusion problems, suppose that $L^\rho \subset L^\xi$. For any $h \in L^+(X)$, let $f = h^t$ and note that if $h \in L^{\tilde{\rho}}$, then $\rho(f) = \tilde{\rho}(h)^t < \infty$, so $\tilde{\xi}(h) = \xi(f)^{1/t} < \infty$; consequently, $L^{\tilde{\rho}} \subset L^{\tilde{\xi}}$. The converse is proven similarly. □

With this generalization of 2.2.4 to multiple variables, it is possible to convert the original inclusion problem to one involving Banach function spaces. The following generalization of Proposition 2.2.6 takes advantage of this to show that the mixed-norm inclusion problem can be formulated as one of finding a best constant.

Proposition 7.3.2. *There is a constant $C < \infty$ such that $\xi \leq C\rho$ if and only if $L^\rho \subset L^\xi$.*

Proof. If any exponent p_k or q_k, for $k \in \{1,\ldots,n\}$, is strictly less than 1, let

$$t = \max(p_1^{-1},\ldots,p_n^{-1},q_1^{-1},\ldots,q_n^{-1})$$

and use this t in Lemma 7.3.1 to convert the inclusion problem of proving or refuting $L^\rho \subset L^\xi$ to the equivalent problem concerning $L^{\tilde{\rho}} \subset L^{\tilde{\xi}}$; at the same time, go from finding the least constant in $\xi \leq C\rho$ to finding the least constant in $\tilde{\xi} \leq D\tilde{\rho}$. (If either C or D exists, the other does and they are related by $C = D^t$, by Lemma 7.3.1. Consequently, if one of them fails to exist, so does the other.) From now on, use ρ and ξ to refer to the values $\tilde{\rho}$ and $\tilde{\xi}$, and observe that no exponent involved has a value less than 1.

This means that the spaces L^ρ and L^ξ are normed spaces; furthermore, as observed by Benedek and Panzone in [4], they are Banach function spaces, defined in such references as [5]. One direction is clear; the existence of $C < \infty$ immediately gives a bounded (equivalently, continuous) inclusion map $L^\rho \hookrightarrow L^\xi$.

Conversely, suppose that $L^\rho \subset L^\xi$. Because these are Banach function spaces there is a constant $C < \infty$ such that $\xi \leq C\rho$, by Theorem 1.8 in [5]. That is, the inclusion map must be bounded. □

Now that the inclusion problem $L^\rho \subset L^\xi$ is known to amount to determining whether the best constant C is finite, the following definition summarizes the notation for the multi-variable inclusion problem.

Definition 7.3.3. Take measurable spaces $(X_1, \Sigma_1), \ldots, (X_n, \Sigma_n)$, where for $k = 1, \ldots, n$ each X_k has measures μ_k and ν_k on the σ-algebra Σ_k. Given exponents $p_1, \ldots, p_n, q_1, \ldots, q_n \in (0, \infty]$, define, for any $f \in L^+(X)$,

$$\rho_k(f)(x_1, \ldots, x_n) = \begin{cases} \left(\int_{X_k} f(x_1, \ldots, x_n)^{p_k} d\mu_k\right)^{1/p_k} & \text{if } p_k < \infty \\ \operatorname{ess\,sup}_{x_k \in (X_k, \mu_k)} f(x_1, \ldots, x_n) & \text{if } p_k = \infty \end{cases},$$

$$\xi_k(f)(x_1, \ldots, x_n) = \begin{cases} \left(\int_{X_k} f(x_1, \ldots, x_n)^{q_k} d\nu_k\right)^{1/q_k} & \text{if } q_k < \infty \\ \operatorname{ess\,sup}_{x_k \in (X_k, \nu_k)} f(x_1, \ldots, x_n) & \text{if } q_k = \infty \end{cases}.$$

For any permutation σ of $\{1, \ldots, n\}$, define mixed norms

$$\rho = \rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}, \qquad \xi = \xi_n \circ \cdots \circ \xi_1,$$

where the constant value of $\rho(f)$ is understood as the value of the mixed norm, and similarly for ξ. The mixed-norm spaces L^ρ and L^ξ are then defined by

$$L^\rho = \{f \in L^+(X) \mid \rho(f) < \infty\}, \qquad L^\xi = \{f \in L^+(X) \mid \xi(f) < \infty\}.$$

Let C denote the least constant in $[0, \infty]$ such that $\xi \leq C\rho$, i.e. for all $f \in L^+(X)$,

$$\xi(f) \leq C\rho(f).$$

7.4 All subinclusions are necessary

The permuted mixed-norm inclusion $L^\rho \subset L^\xi$ has, as necessary conditions, inclusions involving subsets of the variables $x_1, \ldots, x_n$, corresponding to subsets $S \subset \{1, \ldots, n\}$. Important special cases include the one-variable inclusions $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X)$ obtained by considering singleton S and, two-variable subinclusions, for which the results in the two-variable problem will be helpful. A first step is to establish absolute continuity of measures, again.

Lemma 7.4.1. *If $L^\rho \subset L^\xi$, then every $\nu_k \ll \mu_k$.*

Proof. Were any particular $\nu_j \not\ll \mu_j$, then there would be some measurable $S_j \subset X_j$ with $\mu_j S_j = 0$, yet $\nu_j S_j > 0$. For each $k \neq j$, let $S_k = X_k$; because ν_k is non-zero, $\nu_k S_k > 0$. Let $f = \chi_{\prod_{k=1}^n S_k} = \prod_{k=1}^n \chi_{S_k}$ and then compute

$$\xi(f) = \prod_{k=1}^n \left\|\chi_{S_k}\right\|_{L^{q_k}_{\nu_k}(X_k)} > 0$$

since each factor is positive; if $q_k < \infty$, then $\left\|\chi_{S_k}\right\|_{L^{q_k}_{\nu_k}(X_k)} = (\nu_k S_k)^{1/q_k} > 0$, while the essential supremum of χ_{S_k} is 1 since $\nu_k S_k > 0$. On the other hand, since $\mu_j S_j = 0$, $\chi_{S_j} = 0$ μ_j-a.e., so

$$\rho(f) = \prod_{k=1}^n \left\|\chi_{S_k}\right\|_{L^{p_k}_{\mu_k}} = 0$$

Consequently, there can be no constant C such that $\xi(f) \le C\rho(f)$. □

Once more, since $\nu \ll \mu$ is a basic necessary condition, it is assumed.

Assumption For the rest of the n-variable problem, assume that $\nu_k \ll \mu_k$ for $k \in \{1, \ldots, n\}$.

For this theorem, recall that the permutation σ is included in the definition of the composition $\rho = \rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}$, while $\xi = \xi_n \circ \cdots \xi_1$ is defined in a standard order.

Theorem 7.4.2. *Suppose that $L^\rho \subset L^\xi$, i.e. there is some $C < \infty$ such that $\xi \le C\rho$. For any $S \subset \{1, \ldots, n\}$, let $s = |S|$. Write $S = \{k_1, \ldots, k_s\} = \{l_1, \ldots, l_s\}$ where $k_1 < \cdots < k_s$ and $\sigma^{-1}(l_1) < \cdots < \sigma^{-1}(l_s)$. Define a permutation τ on S by $\tau(k_i) = l_i$ for each $1 \le i \le s$. Then there is a constant $C' < \infty$ such that*

$$\xi_{k_s} \circ \cdots \circ \xi_{k_1} \le C' \rho_{l_s} \circ \cdots \circ \rho_{l_1}$$

Proof. Let $t = n - |S|$ and enumerate $\{1, \ldots, n\} \setminus S$ by $m_1, \ldots, m_t$ (in any order). For each $1 \le j \le t$, because $\nu_{m_j} \ll \mu_{m_j}$ and $\nu_{m_j} \neq 0$, there is some set E_j with positive μ_{m_j} and ν_{m_j} measure. Because these measures are σ-finite, we can ensure (taking a subset if needed) that $0 < \mu_{m_j}(E_j) < \infty$ and $0 < \nu_{m_j}(E_j) < \infty$.

For any measurable function $g(x_{k_1}, \ldots, x_{k_s})$ depending on the variables x_i with $i \in S$, define $h \in L^+(X)$ by $h = g \prod_{j=1}^t \chi_{E_j}(x_{m_j})$. Also, for $i \in S$, let $\tilde{\rho}_i$ and $\tilde{\xi}_i$ denote essentially the same one-variable Lebesgue norm computations as ρ_i and ξ_i, with the same exponents and measures, but instead of mapping $L^+(X)$ to itself, these take $L^+(\prod_{i\in S} X_i)$ to itself. Then

$$\begin{aligned}\left(\prod_{j=1}^t \left\|\chi_{E_j}\right\|_{L^{q_{m_j}}_{\nu_{m_j}}(X_{m_j})}\right) \tilde{\xi}_{k_s} \circ \cdots \circ \tilde{\xi}_{k_1}(g) &= \xi(h) \\ &\le C\rho(h) \\ &= C\left(\prod_{j=1}^t \left\|\chi_{E_j}\right\|_{L^{p_{m_j}}_{\mu_{m_j}}(X_{m_j})}\right) \tilde{\rho}_{l_s} \circ \cdots \circ \tilde{\rho}_{l_1}(g),\end{aligned}$$

so

$$\tilde{\xi}_{k_s} \circ \cdots \circ \tilde{\xi}_{k_1}(g) \le C \prod_{j=1}^{t} \left(\frac{\left\|\chi_{E_j}\right\|_{L^{p_{m_j}}_{\mu_{m_j}}(X_{m_j})}}{\left\|\chi_{E_j}\right\|_{L^{q_{m_j}}_{\nu_{m_j}}(X_{m_j})}} \right) \tilde{\rho}_{l_s} \circ \cdots \circ \tilde{\rho}_{l_1}(g).$$

Each E_j satisfies $0 < \mu_{m_j} E_j < \infty$ and $0 < \nu_{m_j} E_j < \infty$, so that $0 < \left\|\chi_{E_j}\right\|_{L^{p_{m_j}}_{\mu_{m_j}}(X_{m_j})} < \infty$ and $0 < \left\|\chi_{E_j}\right\|_{L^{q_{m_j}}_{\nu_{m_j}}(X_{m_j})} < \infty$. So, we can define $C' = C \prod_{j=1}^{t} \left(\frac{\left\|\chi_{E_j}\right\|_{L^{p_{m_j}}_{\mu_{m_j}}(X_{m_j})}}{\left\|\chi_{E_j}\right\|_{L^{q_{m_j}}_{\nu_{m_j}}(X_{m_j})}} \right)$ and note that $C' < \infty$.

It remains to show that this C' is a suitable value of the constant we seek; remember that $\xi_{k_s} \circ \cdots \circ \xi_{k_1}$ and $\rho_{l_s} \circ \cdots \circ \rho_{l_1}$ apply to functions in $L^+(X)$, i.e. of all variables $x_1, \ldots, x_n$ rather than merely $x_{k_1}, \ldots, x_{k_s}$. For any $f \in L^+(X)$ and any particular points $x_{m_1} \in X_{m_1}, \ldots, x_{m_t} \in X_{m_t}$, let $g(x_{k_1}, \ldots, x_{k_s}) = f(x_1, \ldots, x_n)$ where each x_i is one of the temporarily fixed x_{m_j} (where $i = m_j$) if $i \notin S$ or is x_{k_j} ($i = k_j$) if $i \in S$. From what we've already found,

$$\begin{aligned}
\xi_{k_s} \circ \cdots \circ \xi_{k_1}(f)(x_1, \ldots, x_n) &= \tilde{\xi}_{k_s} \circ \cdots \circ \tilde{\xi}_{k_1}(g)(x_{m_1}, \ldots, x_{m_t}) \\
&\le C' \tilde{\rho}_{l_s} \circ \cdots \circ \tilde{\rho}_{l_1}(g)(x_{m_1}, \ldots, x_{m_t}) \\
&= C' \rho_{l_s} \circ \cdots \circ \rho_{l_1}(f)(x_1, \ldots, x_n).
\end{aligned}$$

The quantities $\tilde{\xi}_{k_s} \circ \cdots \circ \tilde{\xi}_{k_1}(g)$ and $\tilde{\rho}_{l_s} \circ \cdots \circ \tilde{\rho}_{l_1}(g)$ are functions of those variables $x_{m_1}, \ldots, x_{m_t}$ included as parameters in the definition of g, so this inequality compares the (constant) functions

$$\xi_{k_s} \circ \cdots \circ \xi_{k_1}(f) \le C' \rho_{l_s} \circ \cdots \circ \rho_{l_1}(f),$$

which holds for any $f \in L^+(X)$ with the constant C'. □

Definition 7.4.3. The inequalities in the conclusion of Theorem 7.4.2, and the corresponding embeddings of s-variable mixed norm spaces, are referred to as *s-variable subinclusions* of the mixed-norm inclusion problem. It is important to note that every subinclusion with $s > 1$ inherits a permutation of the variables $l_1, \ldots, l_s$ from the permutation in the original problem.

The reason why σ^{-1} appears in the definition of $l_1, \ldots, l_s$ is that σ maps the position i on the right-hand side of inequality (7.6) to the index $\sigma(i)$ of the variable in the i^{th} place. That is, σ, maps the place to the variable, dictating which variable goes in a particular spot in the mixed norm. However, when determining whether particular variables are inverted between the left and right sides, we instead want to use σ^{-1}, which maps each index to the place where it appears. Specifically, $\sigma^{-1}(j)$ is the place where the computation of $L^{p_j}_{\mu_j}(X_j)$ ends up in the composition on the right-hand side.

Corollary 7.4.4. *If $L^\rho \subset L^\xi$, then for each $k \in \{1, \ldots, n\}$, $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$.*

Proof. Simply consider $S = \{k\}$ in Theorem 7.4.2. □

Corollary 7.4.5. *If $L^\rho \subset L^\xi$, then for each $1 \le i < j \le n$, there is a constant $C' < \infty$ such that*

$$\xi_j \circ \xi_i \le C' \rho_{\tau(j)} \circ \rho_{\tau(i)}$$

where the permutation τ on $\{i, j\}$ is the identity if $\sigma^{-1}(i) < \sigma^{-1}(j)$ or τ is the transposition swapping i and j if $\sigma^{-1}(i) > \sigma^{-1}(j)$. (That is, it places i and j in the relative order imposed on them by σ.)

Proof. Let $S = \{i, j\}$ in Theorem 7.4.2. □

Although they are simple consequences of the theorem, these corollaries turn out to be important. As following sections will show, these corollaries have converses in certain special cases. In the case of unpermuted mixed norms, all one-variable inclusions are not only necessary but also sufficient for the mixed norm inclusion. When no measure is purely atomic, having all two-variable subinclusions suffices for the complete mixed norm inclusion, so the problem reduces to the two-variable case.

7.5 Unpermuted case

When there is no permutation, i.e. σ is the identity, then the necessary condition of Corollary 7.4.4 is also sufficient.

Proposition 7.5.1. *The unpermuted mixed norm inclusion $L^\rho \subset L^\xi$ holds if and only if every $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$. Furthermore, in this case the norm of the inclusion operator $L^\rho \hookrightarrow L^\xi$ is the product of the norms of the inclusions $L^{p_k}_{\mu_k}(X_k) \hookrightarrow L^{q_k}_{\nu_k}(X_k)$.*

In other words, if for each $k \in \{1, \ldots, n\}$, $C_k \in [0, \infty]$ is the least constant such that $\xi_k \le C_k \rho_k$, then the least constant in $\xi \le C\rho$ is $C = C_1 \cdots C_n$.

Proof. Corollary 7.4.4 has already established that every single-variable inclusion is necessary. For the converse, take any $g \in L^+(X)$ and note that, for $k \in \{1, \ldots, n\}$, for every fixed $(x_1, \ldots, \hat{x}_k, \ldots, x_n)$ (where $\hat{x}_k$ denotes that x_k is omitted), $g(x_1, \ldots, x_n)$ is a measurable function of x_k, with its $L^{p_k}_{\mu_k}(X_k)$ norm given by the value of $\rho_k(g)(x_1, \ldots, x_n)$, which does not depend on x_k. The same is true of the $L^{q_k}_{\nu_k}(X_k)$ norm, so that the inclusion $L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$ implies that $\xi_k(g) \le C_k \rho_k(g)$.

For any $f \in L^+(X)$, successively applying this result to each variable in turn provides

$$\begin{aligned}
\xi(f) &= \xi_n \circ \cdots \circ \xi_1(f) \\
&\le C_1 \xi_n \circ \cdots \circ \xi_2 \circ \rho_1(f) \\
&\quad \vdots \\
&\le C_1 \cdots C_n \rho_n \circ \cdots \circ \rho_1(f) \\
&= C_1 \cdots C_n \rho(f).
\end{aligned}$$

In the particular case that $f \in L^\rho$, this shows that $f \in L^\xi$ as well. To see that $C_1 \cdots C_n$ is actually the norm of the inclusion, we need only consider a factorable function $f = f_1 \cdots f_n$, where each $f_k \in L^{p_k}_{\mu_k}(X_k) \subset L^{q_k}_{\nu_k}(X_k)$. In this case,

$$\rho(f) = \prod_{k=1}^{n} \|f_k\|_{L^{p_k}_{\mu_k}(X_k)} = \prod_{k=1}^{n} \rho_k(f)$$

$$\xi(f) = \prod_{k=1}^{n} \|f_k\|_{L^{q_k}_{\nu_k}(X_k)} = \prod_{k=1}^{n} \xi_k(f)$$

By choosing $f_1, \ldots, f_n$ such that the ratio $\xi_k(f)/\rho_k(f)$ is arbitrarily close to C_k, which is possible since that is the norm of inclusion, the ratio $\xi(f)/\rho(f)$ can be brought as close as desired to $C_1 \cdots C_n$. □

7.6 Minkowski criterion with no purely atomic measure

Now, suppose that none of the measures μ_k or ν_k is purely atomic. As noted in Section 4.2.1, we can restrict our attention to functions supported where every $\frac{d\nu_k}{d\mu_k} > 0$, since non-zero values off that set only increase $\rho(f)$, not $\xi(f)$, and thus take us further from $C = \sup_{f\in L^+(X)} \frac{\xi(f)}{\rho(f)}$. This can be effected by restricting our measures, after which we have $\frac{d\nu_k}{d\mu_k} > 0$ μ_k-a.e. and, consequently, either both μ_k and μ_k are purely atomic or neither of them is.

In this case, it is not only necessary that every two-variable subinclusion hold to have $L^\rho \subset L^\xi$, but also sufficient. Furthermore, as we learned from studying the two-variable case, with no purely atomic measure, this occurs when the exponents satisfy a condition based on Minkowski's inequality.

Theorem 7.6.1. *Suppose that, for each $k \in \{1, \ldots, n\}$, neither μ_k nor ν_k is purely atomic and that there is a constant $C_k < \infty$ such that $\xi_k \le C_k\rho_k$. Then the following are equivalent.*

1. *For any $1 \le i < j \le n$ there exists some constant $C' < \infty$ such that*

$$\xi_j \circ \xi_i \le C'\rho_{\tau(j)} \circ \rho_{\tau(i)}$$

   where τ is the identity if $\sigma^{-1}(i) < \sigma^{-1}(j)$ or τ swaps i and j if $\sigma^{-1}(i) > \sigma^{-1}(j)$.

2. *For any $1 \le i < j \le n$, if $\sigma^{-1}(i) > \sigma^{-1}(j)$ then $q_i \le p_j$.*

3. *There is a finite $C < \infty$ such that*

$$\xi \le C\rho.$$

Furthermore, if these are true, then the least possible values of C and $C_1, \ldots, C_n$ satisfy $C = C_1 \cdots C_n$.

Proof. Suppose the first condition holds and we have every two-variable inclusion. In every case where $\sigma^{-1}(i) > \sigma^{-1}(j)$ while $i < j$, the two-variable inclusion is permuted, so Theorem 4.2.16 provides that $\min(p_i, q_i) \le \max(p_j, q_j)$. Furthermore, Theorem 2.1.8 provides that $q_i \le p_i$ and $q_j \le p_j$, so that the Minkowski criterion on exponents is $q_i \le p_j$. This proves the second condition.

Assuming the second condition, we have the Minkowski sufficient condition described in Section 3.6, so every permuted (non-identity τ) inclusion in the first condition holds. Of course, when τ is the identity, unpermuted inclusion follows from single-variable inclusions, by Theorem 3.4.1. Therefore we have every two-variable inclusion required for the first condition.

Corollary 7.4.5 says that the third condition implies the first, so all that remains is to establish, with the assumed single-variable inclusions and atomless measures, that the second condition implies the third. The aim is to prove that all subinclusions (each corresponding to some $S \subset \{1, \ldots, n\}$) hold, by induction on the size of the subset. While this is a superficially stronger result, Theorem 7.4.2 tells us that it is actually equivalent to the overall inclusion.

Claim: For any $S \subset \{1, \ldots, n\}$, let $s = |S|$ and write $S = \{k_1, \ldots k_s\} = \{l_1, \ldots, l_s\}$, where $k_1 < \cdots < k_s$ and $\sigma^{-1}(l_1) < \cdots < \sigma^{-1}(l_s)$. Then the least constant $C \in [0, \infty]$ such that

$$\xi_{k_s} \circ \cdots \circ \xi_{k_1} \le C\rho_{l_s} \circ \cdots \circ \rho_{l_1} \tag{7.9}$$

is $C = C_{k_1} \cdots C_{k_s}$.

The proof is by induction on s. The base case $s = 1$ is in the hypothesis. Assume that all subinclusions for subsets S with $|S| < s$ hold, each with the specified constant, the product of all single-variable constants. (The case of factorable functions $f_{k_1} \cdots f_{k_s}$, each $f_{k_i} \in L^+(X_{k_i})$, gives $C \geq C_{k_1} \cdots C_{k_s}$, so it remains only to establish (7.9) with $C = C_{k_1} \ldots C_{k_s}$.)

Define a permutation τ on $\{1, \ldots, s\}$ by, for each $i, j \in \{1, \ldots, s\}$, $\tau(i) = j$ if and only if $k_i = l_j$. Let $i_0 = \tau^{-1}(1)$, so $k_{i_0} = l_1$. The single-variable inclusion $\xi_{l_1} \leq C_{l_1}\rho_{l_1}$ implies that

$$\xi_{k_{i_0}} \circ \cdots \circ \xi_{k_1} \leq C_{l_1}\rho_{l_1} \circ \xi_{k_{i_0-1}} \circ \cdots \circ \xi_{k_1}. \tag{7.10}$$

For any $i < i_0$, $\tau(i) > 1 = \tau(i_0)$, so l_i occurs after l_{i_0} in the l ordering of S, i.e. $\sigma^{-1}(i) > \sigma^{-1}(i_0)$. By the second condition, $q_{k_i} \leq p_{k_{i_0}} = p_{l_1}$. Repeatedly apply Minkowski's integral inequality (Proposition 7.2.3) to all $i < i_0$.

$$\begin{aligned} \rho_{l_1} \circ \xi_{k_{i_0-1}} \circ \cdots \circ \xi_{k_1} &\leq \xi_{k_{i_0-1}} \circ \rho_{l_1} \circ \xi_{k_{i_0-2}} \circ \cdots \circ \xi_{k_1} \\ &\vdots \\ &\leq \xi_{k_{i_0-1}} \circ \cdots \circ \xi_{k_1} \circ \rho_{l_1} \end{aligned}$$

Composition with $\xi_{k_s} \circ \cdots \circ \xi_{k_{i_0+1}}$ gives

$$\xi_{k_s} \circ \cdots \circ \xi_{k_{i_0+1}} \circ \rho_{l_1} \circ \xi_{k_{i_0-1}} \circ \cdots \circ \xi_{k_1} \leq \xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1} \circ \rho_{l_1},$$

where $\widehat{\xi_{k_{i_0}}}$ indicates that this term is missing from the composition. With (7.10), this means that

$$\xi_{k_s} \circ \cdots \circ \xi_{k_1} \leq C_{l_1}\xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1} \circ \rho_{l_1}. \tag{7.11}$$

Apply the inductive hypothesis to $S \setminus \{l_1\} = S \setminus \{k_{i_0}\}$ to find that

$$\xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1} \leq C_{l_s} \cdots C_{l_2}\rho_{l_s} \circ \cdots \circ \rho_{l_2}$$

which, together with (7.11), gives

$$\begin{aligned} \xi_{k_s} \circ \cdots \circ \xi_{k_1} &\leq C_{l_1}\xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1} \circ \rho_{l_1} \\ &\leq C_{l_1} \cdots C_{l_s}\rho_{l_s} \circ \cdots \circ \rho_{l_1}, \end{aligned} \tag{7.12}$$

as desired. □

The case of nonatomic measures is not the only time when the multivariable problem reduces to two-variable subproblems. There is a similar result involving counting measure.

Theorem 7.6.2. *Suppose that, for each $k \in \{1, \ldots, n\}$, both of μ_k and ν_k are (unweighted) counting measure and that there is a constant $C_k < \infty$ such that $\xi_k \leq C_k\rho_k$. (When this happens, the least constant is $C_k = 1$.) Then the following are equivalent.*

1. *For any $1 \leq i < j \leq n$ there exists some constant $C' < \infty$ such that*

$$\xi_j \circ \xi_i \leq C'\rho_{\tau(j)} \circ \rho_{\tau(i)},$$

where τ is the identity if $\sigma^{-1}(i) < \sigma^{-1}(j)$ or τ swaps i and j if $\sigma^{-1}(i) > \sigma^{-1}(j)$. (When this is true, the least such constant is $C' = 1$.)

2. *For any* $1 \le i < j \le n$, *if* $\sigma^{-1}(i) > \sigma^{-1}(j)$ *then* $p_i \le q_j$.

3. *There is a finite* $C < \infty$ *such that*
$$\xi \le C\rho.$$
(When this is true, the least constant is $C = 1$.*)*

Proof. The proof proceeds much as in Theorem 7.6.1. One difference to note is that each single-variable inclusion $\ell^{p_k}(X_k) \subset \ell^{q_k}(X_k)$ holds if and only if $p_k \le q_k$, and in this case $C_k = 1$, by Corollary 2.6.10.

Assume the first condition. Whenever $i < j$ and $\sigma^{-1}(i) > \sigma^{-1}(j)$, the permuted two-variable inclusion requires $p_i \le q_j$ by Proposition 5.3.5. This proves the second condition.

Assume the second condition. The hypothesis provides one-variable inclusions and, consequently, $p_i \le q_i$ and $p_j \le q_j$. Theorem 3.4.1 shows that the one-variable inclusions suffice in the first condition's unpermuted case, with $C' = C_i C_j = 1$. The permuted case, with non-identity τ, has $\sigma^{-1}(i) > \sigma^{-1}(j)$ and therefore $p_i \le q_j$. Proposition 5.3.5 then gives permuted two-variable inclusion with $C' = 1$, finishing the proof of the first condition.

Subinclusions are necessary by Corollary 7.4.5, so the third condition implies the first. It remains only to prove that the second condition implies the third. As in Theorem 7.6.1, the following claim is established by induction on s, assuming the second condition.

Claim: For any $S \subset \{1, \dots, n\}$, let $s = |S|$ and write $S = \{k_1, \dots k_s\} = \{l_1, \dots, l_s\}$, where $k_1 < \cdots < k_s$ and $\sigma^{-1}(l_1) < \cdots < \sigma^{-1}(l_s)$. Then the least constant $C \in [0, \infty]$ such that

$$\xi_{k_s} \circ \cdots \circ \xi_{k_1} \le C\rho_{l_s} \circ \cdots \circ \rho_{l_1} \tag{7.13}$$

is $C = 1$.

The hypothesis includes the base case $s = 1$. Assume that all subinclusions for subsets S with $|S| < s$ hold with constant 1. (Factorable functions give $C \ge C_{k_1} \cdots C_{k_s} = 1$.)

Define a permutation τ on $\{1, \dots, s\}$ by, for each $i, j \in \{1, \dots, s\}$, $\tau(i) = j$ if and only if $k_i = l_j$. Let $i_0 = \tau^{-1}(s)$, so $k_{i_0} = l_s$. The single-variable inclusion $\xi_{l_s} \le \rho_{l_s}$ implies that

$$\xi_{k_s} \circ \cdots \circ \xi_{k_{i_0}} \le \xi_{k_s} \circ \dots \xi_{k_{i_0+1}} \circ \rho_{l_s}. \tag{7.14}$$

For any $j > i_0$, $\tau(j) < n = \tau(i_0)$, so l_j occurs before l_{i_0} in the l ordering of S, i.e. $\sigma^{-1}(j) < \sigma^{-1}(i_0)$. By the second condition, $p_{l_s} = p_{k_{i_0}} \le q_{k_j}$. Repeatedly apply Minkowski's integral inequality (Proposition 7.2.3) to all $j > i_0$.

$$\begin{aligned}\xi_{k_s} \circ \dots \xi_{k_{i_0+1}} \circ \rho_{l_s} &\le \xi_{k_s} \circ \dots \xi_{k_{i_0+2}} \circ \rho_{l_s} \circ \xi_{k_{i_0+1}} \\ &\vdots \\ &\le \rho_{l_s} \circ \xi_{k_s} \circ \cdots \circ \xi_{k_{i_0+1}}\end{aligned}$$

Composition with $\xi_{k_{i_0-1}} \circ \cdots \circ \xi_1$ gives

$$\xi_{k_s} \circ \cdots \circ \xi_{k_{i_0+1}} \circ \rho_{l_s} \circ \xi_{k_{i_0-1}} \circ \cdots \circ \xi_{k_1} \le \rho_{l_s} \circ \xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1},$$

where $\widehat{\xi_{k_{i_0}}}$ indicates that this term is missing from the composition. With (7.14), this means that

$$\xi_{k_s} \circ \cdots \circ \xi_{k_1} \le \rho_{l_s} \circ \xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1}. \tag{7.15}$$

Apply the inductive hypothesis to $S \setminus \{l_s\} = S \setminus \{k_{i_0}\}$ to find that

$$\xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1} \leq \rho_{l_{s-1}} \circ \cdots \circ \rho_{l_1}$$

which, together with (7.15), gives

$$\begin{aligned} \xi_{k_s} \circ \cdots \circ \xi_{k_1} &\leq \rho_{l_s} \circ \xi_{k_s} \circ \cdots \circ \widehat{\xi_{k_{i_0}}} \circ \cdots \circ \xi_{k_1} \\ &\leq \rho_{l_s} \circ \cdots \circ \rho_{l_1}. \end{aligned} \tag{7.16}$$

□

In each of these cases, the multivariable problem has been reduced to two-variable subproblems. Where these two-variable subproblems are permuted, inclusion in either case requires the Minkowski criterion. Additionally, each one-variable subproblem requires a particular order of exponents. For measures which are not purely atomic, one-variable inclusions demand $q_k \leq p_k$ so that the Hölder criterion for inclusion can apply; with this, the Minkowski case of $\min(p_i, q_i) \leq \max(p_j, q_j)$ is reduced to $q_i \leq p_j$.. For unweighted ℓ^p, $p_k \leq q_k$ is both necessary and sufficient, and it reduces the Minkowski case to $p_i \leq q_j$. Without these special conditions, it is not clear that just having the Minkowski condition for each permuted two-variable subinclusion suffices to establish the full inclusion. However, this has been proven to suffice in all cases with three or four variables, and computation has furthermore ruled out a counterexample in five or six variables. Therefore, the following conjecture is suggested.

Conjecture 7.6.3. *For each $k \in \{1, \ldots, n\}$, let $C_k \in [0, \infty]$ denote the least constant so that $\xi_k \leq C_k \rho_k$. Assume that every permuted two-variable subinclusion satisfies the Minkowski criterion. That is, assume that for any $1 \leq i < j \leq n$, if $\sigma^{-1}(i) > \sigma^{-1}(j)$, then $\min(p_i, q_i) \leq \max(p_j, q_j)$. Then the least constant C such that*

$$\xi \leq C\rho$$

is $C = C_1 \cdots C_n$.

7.7 Multiple-variable summary

Let $X_1, \ldots, X_n$ be measurable spaces such that each X_k has non-zero σ-finite measures μ_k and ν_k. For each $k \in \{1, \ldots, n\}$, let ρ_k denote the one-variable $L^{p_k}_{\mu_k}(X_k)$ norm, while ξ_k denotes the one-variable $L^{q_k}_{\nu_k}(X_k)$ norm. Let σ be a permutation of $\{1, \ldots, n\}$ and let C denote the least constant such that

$$\xi_n \circ \cdots \circ \xi_1 \leq C\rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}. \tag{7.17}$$

There is an inclusion between the corresponding mixed-norm spaces if and only if $C < \infty$.

Factorable functions yield the lower bound $C \geq C_1 \cdots C_n$.

1. (Theorem 7.4.2) Does any subinclusion fail? If so, then $C = \infty$.

   Given a subset $S \subset \{1, \ldots, n\}$ with cardinality $s = |S|$, the corresponding s-variable subinclusion is obtained by removing from the compositions in

$$\xi_n \circ \cdots \circ \xi_1 \leq C\rho_{\sigma(n)} \circ \cdots \circ \rho_{\sigma(1)}$$

any one-variable norms in variables x_k with $k \notin S$. The order in which one-variable norms are computed in the subinclusion is inherited from their order in the original inclusion problem. Although the best constant may differ among the subinclusions, if $\xi \leq C\rho$ with $C < \infty$, there must be a finite best constant for the subinclusion corresponding to any S. There are two special cases worth noting.

- (Corollary 7.4.4) Every one-variable subinclusion $\xi_k \leq C_k\rho_k$ must hold (and let C_k denote the best constant there) if mixed-norm inclusion is to be possible, by considering singleton S.
- (Corollary 7.4.5) Every two-variable subinclusion must hold if mixed-norm inclusion is to be possible, by considering S with two elements.

The standard results on embeddings among Lebesgue spaces solve the one-variable subproblems, while the two-variable methods developed here solve many two-variable cases fully, and provide insight even into the more difficult cases. If any necessary condition can show that a subinclusion in one or two variables fails, then the full mixed-norm problem certainly fails, as well.

One unfortunate side to this is that proving a mixed-norm inclusion establishes every subinclusion, so it must be at least as difficult a problem as each of its subproblems. Since certain difficult two-variable cases have been found, the most difficult problems in three or more variables must be at least as troublesome. Therefore, a neat solution to all cases is not expected. Certain cases, however, are solved.

2. (Proposition 7.5.1) Is σ the identity? If so, then if all one-variable inclusions hold (each $C_k < \infty$), so does the mixed-norm inclusion, with $C = C_1 \cdots C_n$.

3. (Theorem 7.6.1) Is no measure μ_k or ν_k purely atomic? In this case, the mixed-norm inclusion holds if and only if every two-variable subinclusion does.

   (Unpermuted two-variable subproblems, of course, reduce to their one-variable pieces, while in this case permuted two-variable subinclusions hold if and only if both the one-variable inclusions and the Minkowski criterion hold.)

4. (Theorem 7.6.2) Is every measure space a countably infinite set with counting measure? In this case as well, the problem reduces to its two-variable subproblems. Permuted two-variable inclusion holds if and only if one-variable inclusions and the Minkowski criterion hold.

5. (Conjecture 7.6.3) It is conjectured that, if every two-variable subinclusion holds, and every permuted two-variable subinclusion satisfies the Minkowski criterion, then the full mixed-norm inclusion holds.

Bibliography

[1] Robert Algervik and Viktor I. Kolyada, *On Fournier-Gagliardo mixed norm spaces*, Ann. Acad. Sci. Fenn. Math. **36** (2011), no. 2, 493–508. MR 2865509 (2012i:46034)

[2] S. Barza, L.-E. Persson, and J. Soria, *Multidimensional rearrangement and Lorentz spaces*, Acta Math. Hungar. **104** (2004), no. 3, 203–224. MR 2075921 (2005e:46043)

[3] Sorina Barza, Anna Kamińska, Lars-Erik Persson, and Javier Soria, *Mixed norm and multidimensional Lorentz spaces*, Positivity **10** (2006), no. 3, 539–554. MR 2258957 (2007f:46024)

[4] A. Benedek and R. Panzone, *The spaces L^P with mixed norm*, Duke Math. J. **26** (1961), 301–324.

[5] Colin Bennett and Robert Sharpley, *Interpolation of operators*, Pure and Applied Mathematics, vol. 129, Academic Press, Inc., Boston, MA, 1988. MR 928802 (89e:46001)

[6] Ron Blei, *Analysis in integer and fractional dimensions*, Cambridge University Press, Cambridge, United Kingdom, 2001.

[7] Ron C. Blei, *Fractional Cartesian products of sets*, Ann. Inst. Fourier (Grenoble) **29** (1979), no. 2, v, 79–105. MR 539694 (81h:43008)

[8] Ron C. Blei and John J. F. Fournier, *Mixed-norm conditions and Lorentz norms*, Commutative harmonic analysis (Canton, NY, 1987), Contemp. Math., vol. 91, Amer. Math. Soc., Providence, RI, 1989, pp. 57–78. MR 1002588 (90h:46048)

[9] A. P. Blozinski, *Multivariate rearrangements and Banach function spaces with mixed norms*, Trans. Amer. Math. Soc. **263** (1981), no. 1, 149–167. MR 590417 (81k:46023)

[10] A. Boccuto, A. V. Bukhalov, and A. R. Sambucini, *Some inequalities in classical spaces with mixed norms*, Positivity **6** (2002), 393.

[11] V. I. Bogachev, *Measure theory*, Springer-Verlag, Berlin, Heidelberg, 2007.

[12] Nadia Clavero and Javier Soria, *Mixed norm spaces and rearrangement invariant estimates*, J. Math. Anal. Appl. **419** (2014), no. 2, 878–903. MR 3225412

[13] Andreas Defant, Leonhard Frerick, Joaquim Ortega-Cerdà, Myriam Ounaïes, and Kristian Seip, *The Bohnenblust-Hille inequality for homogeneous polynomials is hypercontractive*, Ann. of Math. (2) **174** (2011), no. 1, 485–497. MR 2811605 (2012e:32005)

[14] Andreas Defant, Dumitru Popa, and Ursula Schwarting, *Coordinatewise multiple summing operators in Banach spaces*, J. Funct. Anal. **259** (2010), no. 1, 220–242. MR 2610385 (2011d:47046)

[15] Gerald B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication. MR 1681462 (2000c:00001)

[16] John J. F. Fournier, *Mixed norms and rearrangements: Sobolev's inequality and Littlewood's inequality*, Ann. Mat. Pura Appl. (4) **148** (1987), 51–76. MR 932758 (89e:46037)

[17] Emilio Gagliardo, *Proprietà di alcune classi di funzioni in più variabili*, Ricerche Mat. **7** (1958), 102–137. MR 0102740 (21 #1526)

[18] W. Grey and G. Sinnamon, *The inclusion problem for mixed-norm spaces*, Trans. Amer. Math. Soc. (to appear).

[19] G. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, United Kingdom, 1953.

[20] V. I. Kolyada, *Iterated rearrangements and Gagliardo-Sobolev type inequalities*, J. Math. Anal. Appl. **387** (2012), no. 1, 335–348. MR 2845754

[21] Matthieu Kowalski, *Sparse regression using mixed norms*, Applied and Computational Harmonic Analysis **27** (2009), no. 3, 303–324.

[22] J.E. Littlewood, *On bounded bilinear forms in an infinite number of variables*, Quarterly J. Math. **1** (1930), 164–174.

[23] W. A. J. Luxemburg, *On the measurability of a function which occurs in a paper by A.C. Zaanen*, Proc. Netherl. Acad. Sci. (A) **61** (1958), 259–265.

[24] A. G. Miamee, *The inclusion* $L^p(\mu) \subseteq L^q(\nu)$, Amer. Math. Monthly **98** (1991), no. 4, 342–345. MR 1103190 (92d:46078)

[25] Mario Milman, *Notes on interpolation of mixed norm spaces and applications*, Quart. J. Math. Oxford Ser. (2) **42** (1991), no. 167, 325–334. MR 1120993 (92f:46042)

[26] L. Nirenberg, *On elliptic partial differential equations*, Ann. Scuola Norm. Sup. Pisa (3) **13** (1959), 115–162. MR 0109940 (22 #823)

[27] Luboš Pick, Alois Kufner, Oldřich John, and Svatopluk Fučík, *Function spaces. Vol. 1*, second ed., De Gruyter Series in Nonlinear Analysis and Applications, vol. 14, Walter de Gruyter & Co., Berlin, 2013. MR 3024912

[28] Dumitru Popa and Gord Sinnamon, *Blei's inequality and coordinatewise multiple summing operators*, Publ. Mat. **57** (2013), no. 2, 455–475. MR 3114778

[29] B. Subramanian, *On the inclusion* $L^p(\mu) \subset L^q(\mu)$, Amer. Math. Monthly **85** (1978), no. 6, 479–481. MR 0482134 (58 #2221)

[30] Yubao Sun, Qingshan Liu, Jinhui Tang, and Dacheng Tao, *Learning discriminative dictionary for group sparse representation*, Image Processing, IEEE Transactions on **23** (2014), no. 9, 3816–3828.

[31] Alfonso Villani, *Another note on the inclusion $L^p(\mu) \subset L^q(\mu)$*, Amer. Math. Monthly **92** (1985), no. 7, 485–487. MR 801221 (86k:46044)

[32] Shengzheng Wang, Jing Peng, and Wei Liu, *An ℓ^2/ℓ^1 regularization framework for diverse learning tasks*, Signal Processing **109** (2015), no. 0, 206 – 211.

[33] P. Zhao, G. Rocha, and B. Yu, *The composite absolute penalties family for grouped and hierarchical variable selection*, Annals of Statistics **37** (2009), no. 6 A, 3468–3497.

Appendix A

Applications of mixed-norm Hölder and Minkowski inequalities

Although they are not necessary to address the central inclusion problem, the mixed-norm versions of the Hölder and Minkowski integral inequalities were included in Chapter 7. In part, this is because Minkowski's integral inequality does describe a special class of inclusions, specifically those among permuted copies of some particular mixed norm. For example, if

$$\rho = \rho_3 \circ \rho_2 \circ \rho_1$$

is a mixed norm for functions on $X_1 \times X_2 \times X_3$, where each ρ_k is a one-variable norm over X_k, then Minkowski's integral inequality describes inclusions among the various mixed norms $\rho_{\sigma(3)} \circ \rho_{\sigma(2)} \circ \rho_{\sigma(1)}$, where σ is any permutation of $\{1, 2, 3\}$.

However, another reason is because the mixed-norm Hölder and Minkowski can be used together effectively for certain proofs, as mentioned in the introduction. Several examples are included here as simple demonstrations, hopefully to encourage readers to take advantage of these tools when appropriate.

The first example involves Littlewood's 4/3 inequality (from [22], described and generalized in numerous ways since) which directly concerns bilinearity, but has a proof which relies on a certain inequality involving mixed norms. That inequality is stated and proven below.

Proposition A.0.1. *For any doubly-indexed collection $\left(a_{i,j}\right)_{i,j\in\mathbb{N}}$ of nonnegative numbers,*

$$\left(\sum_{i,j} a_{i,j}^{4/3}\right)^{3/4} \leq \left(\sum_j \left(\sum_i a_{i,j}^2\right)^{1/2}\right)^{1/2} \left(\sum_i \left(\sum_j a_{i,j}^2\right)^{1/2}\right)^{1/2} .$$

Proof. For $P = (p_1, p_2)$, define a mixed norm

$$\rho_P(c_{i,j}) = \left(\sum_j \left(\sum_i c_{i,j}^{p_1}\right)^{p_2/p_1}\right)^{1/p_2},$$

with appropriate replacements by the essential supremum if either or both of p_1 and p_2 is ∞.

As described in Theorem 7.2.1, $\left(3, \frac{3}{2}\right)' = \left(\frac{3}{2}, 3\right)$. Let $\alpha_{i,j} = \beta_{i,j} = a_{i,j}^{2/3}$ and, by that very theorem,

$$\sum_{i,j} a_{i,j}^{2/3} = \sum_{i,j} \alpha_{i,j}\beta_{i,j} \leq \rho_{(3,\frac{3}{2})}(\alpha_{i,j})\rho_{(\frac{3}{2},3)}(\beta_{i,j})$$
$$= \left(\sum_j \left(\sum_i \alpha_{i,j}^3\right)^{1/2}\right)^{2/3} \left(\sum_j \left(\sum_i \beta_{i,j}^{3/2}\right)^2\right)^{1/3}.$$

Notice that the $(\frac{3}{2}, 3)$ mixed norm has its smaller exponent $\frac{3}{2}$ on the inside of the nested sums and its larger exponent 3 on the outside. Even the everyday two-variable Minkowski establishes that swapping the order of those one-variable norms would give greater values. Therefore

$$\sum_{i,j} a_{i,j}^{2/3} \leq \left(\sum_j \left(\sum_i \alpha_{i,j}^3\right)^{1/2}\right)^{2/3} \left(\sum_i \left(\sum_j \beta_{i,j}^3\right)^{1/2}\right)^{2/3}$$
$$= \left(\sum_j \left(\sum_i a_{i,j}^2\right)^{1/2}\right)^{2/3} \left(\sum_i \left(\sum_j a_{i,j}^2\right)^{1/2}\right)^{2/3},$$

after which taking the $3/4$ power on each side gives the desired result. □

Notice that this demonstration used no property specific to sums, and in fact would work perfectly well for integrals over any σ-finite measure spaces instead of series. The next demonstration takes advantage of this by deriving the result for integrals first, with the series version as a corollary.

Proposition A.0.2. *Let $(X_1, \mu_1), \ldots, (X_3, \mu_3)$ be σ-finite measure spaces. Let $X = X_1 \times X_2 \times X_3$ and let μ denote the product measure on X. Then, for any $f \in L^+(X)$,*

$$\left(\int_X f^{6/5} d\mu\right)^{5/6} \leq \left(\int_{X_2\times X_3} \left(\int_{X_1} f(x_1, x_2, x_3)^2 d\mu_1(x_1)\right)^{1/2} d\mu_2(x_2)d\mu_3(x_3)\right)^{1/3}$$
$$\left(\int_{X_1\times X_3} \left(\int_{X_2} f(x_1, x_2, x_3)^2 d\mu_2(x_2)\right)^{1/2} d\mu_1(x_1)d\mu_3(x_3)\right)^{1/3}$$
$$\left(\int_{X_1\times X_2} \left(\int_{X_3} f(x_1, x_2, x_3)^2 d\mu_3(x_3)\right)^{1/2} d\mu_1(x_1)d\mu_2(x_2)\right)^{1/3}.$$

Proof. Define three ordered triples by

$$P_1 = \left(5, \frac{5}{2}, \frac{5}{2}\right) \qquad P_2 = \left(\frac{5}{2}, 5, \frac{5}{2}\right) \qquad P_3 = \left(\frac{5}{2}, \frac{5}{2}, 5\right).$$

Denoting the j^{th} coordinate of P_i by $p_i(j)$, notice that $\sum_i P_i^{-1} = 1$, in the sense that for each particular j, $\sum_i p_i(j)^{-1} = 1$. This permits the use of a three-function mixed-norm Hölder's inequality, as given in Theorem 7.2.2. Letting $g_1 = g_2 = g_3 = f^{2/5}$,

$$\int_X g_1 g_2 g_3 d\mu \leq \rho_{P_1}(g_1)\rho_{P_2}(g_2)\rho_{P_3}(g_3), \tag{A.0}$$

where each mixed norm ρ_{P_i} is defined by, for $f \in L^+(X)$,

$$\rho_{P_i} = \rho_{\mu_3}^{p_i(3)} \circ \rho_{\mu_2}^{p_i(2)} \circ \rho_{\mu_1}^{p_i(1)},$$

e.g. $\rho_{P_1} = \rho_{\mu_3}^{5/2} \circ \rho_{\mu_2}^{5/2} \circ \rho_{\mu_1}^{5}$. Two of these mixed norms have their largest exponent, 5, somewhere other than the inside, coming in the middle or last in the composition of one-variable norms. Let ξ_2 denote the version of ρ_{P_2} thus permuted and ξ_3 denote the version of ρ_{P_3} thus permuted, so that

$$\xi_2 = \rho_{\mu_3}^{5/2} \circ \rho_{\mu_1}^{5/2} \circ \rho_{\mu_2}^{5} \quad \text{and} \quad \xi_3 = \rho_{\mu_2}^{5/2} \circ \rho_{\mu_1}^{5/2} \circ \rho_{\mu_3}^{5}.$$

The mixed-norm Minkowski inequality in Theorem 7.2.4 (or, perhaps more conveniently here, Corollary 7.2.6) shows that $\rho_{P_2} \le \xi_2$ and $\rho_{P_3} \le \xi_3$. Therefore, continuing from (A.0),

$$\int_X g_1 g_2 g_3 d\mu \le \rho_{P_1}(g_1)\xi_2(g_2)\xi_3(g_3).$$

When each side is expanded, this turns out to be the desired inequality. □

As an immediate corollary, here is Lemma 2 on page 430 of [6], a mixed-norm lemma leading up to a trilinear result.

Corollary A.0.3. *Let $a_{i,j,k} \ge 0$ be triply indexed over $\mathbb{N}$. Then*

$$\left(\sum_{i,j,k} a_{i,j,k}^{6/5}\right)^{5/6} \le \left(\sum_{j,k}\left(\sum_i a_{i,j,k}^2\right)^{1/2}\right)^{1/3} \left(\sum_{i,k}\left(\sum_j a_{i,j,k}^2\right)^{1/2}\right)^{1/3} \left(\sum_{i,j}\left(\sum_k a_{i,j,k}^2\right)^{1/2}\right)^{1/3}.$$

A more substantial application is the following theorem, which is Theorem 2.1 in [28]. Just over three pages of argument can be replaced by the following application of mixed-norm techniques. First, relevant definitions from the paper.

Definition A.0.4. Let (M_j, μ_j) be σ-finite measure spaces for $j = 1, 2, \dots, n$, and introduce the product measure spaces (M^n, μ^n) and (M_j^n, μ_j^n) by

$$M^n = \prod_{k=1}^n M_k, \qquad \mu^n = \prod_{k=1}^n \mu_k, \qquad M_j^n = \prod_{\substack{k=1\\k\ne j}}^n M_k, \qquad \mu_j^n = \prod_{\substack{k=1\\k\ne j}}^n \mu_k.$$

Note that $M_n^n = M^{n-1}$.

Theorem A.0.5. *If $n \ge 2$ and positive indices $q_1, \dots, q_n$ satisfy $\sum_{j=1}^n \frac{1}{q_j} \le 1$ then for any non-negative μ^n-measurable functions $f_1, f_2, \dots, f_n$,*

$$\int_{M^n} f_1 f_2 \cdots f_n d\mu^n \le \prod_{j=1}^n \left(\int_{M_j}\left(\int_{M_j^n} f_j^{q_j} d\mu_j^n\right)^{p_j/q_j} d\mu_j\right)^{1/p_j} \quad \textit{and}$$

$$\int_{M^n} f_1 f_2 \cdots f_n d\mu^n \le \prod_{j=1}^n \left(\int_{M_j^n}\left(\int_{M_j} f_j^{q_j} d\mu_j\right)^{s_j/q_j} d\mu_j^n\right)^{1/s_j}.$$

Here $\frac{1}{p_j} = \frac{1}{q_j} + 1 - \sum_{k=1}^n \frac{1}{q_k}$ and $\frac{1}{s_j} = \frac{1}{q_j} + \frac{1}{n-1}\left(1 - \sum_{k=1}^n \frac{1}{q_k}\right)$.

Proof. To prove the first inequality, define, for each $j, k \in \{1, \dots, n\}$,

$$\rho_{j,k}(f) = \begin{cases} \left(\int_{M_k} f^{p_j} d\mu_k\right)^{1/p_j} & \text{if } j = k, \\ \left(\int_{M_k} f^{q_j} d\mu_k\right)^{1/q_j} & \text{if } j \neq k. \end{cases}$$

For each $j \in \{1, \dots, n\}$, let $\rho_j = \rho_{j,n} \circ \cdots \rho_{j,1}$, which can also be specified as ρ_{P_j} where each k^{th} entry of the n-tuple P_j is $P_{j,k} = q_j$, except for the j^{th}, $P_{j,j} = p_j$.

For each $k \in \{1, \dots, n\}$, let $P_j(k)$ denote the k^{th} entry in P_j and observe that

$$\sum_{j=1}^{n} P_j(k)^{-1} = p_k^{-1} + \sum_{j \neq k} q_j^{-1} = q_k^{-1} + 1 - \sum_{j=1}^{n} q_j^{-1} + \sum_{j \neq k} q_j^{-1} = 1,$$

so that $\sum_{j=1}^{n} P_j^{-1} = 1$, in the coordinatewise sense used in Theorem 7.2.2, a mixed-norm Hölder's inequality. This theorem then implies that

$$\int_{M^n} f_1 f_2 \cdots f_n d\mu^n \leq \prod_{j=1}^{n} \rho_{P_j}(f_j).$$

By hypothesis, each $p_j \leq q_j$, so $\rho_{j,j}$ has the least exponent, p_j, compared to the q_j exponents on the other $\rho_{j,k}$. Therefore Minkowski's integral inequality, in its Corollary 7.2.6 form, implies that, if we sort the composition ρ_j into

$$\tilde{\rho}_j = \rho_{j,n} \circ \cdots \rho_{j,j+1} \circ \rho_{j,j-1} \circ \cdots \circ \rho_{j,1} \circ \rho_{j,j},$$

then $\rho_j \leq \tilde{\rho}_j$. Therefore

$$\int_{M^n} f_1 f_2 \cdots f_n d\mu^n \leq \prod_{j=1}^{n} \tilde{\rho}_j(f_j),$$

which expands into the desired first inequality in the conclusion. For the second inequality, let, for $j, k \in \{1, \dots, n\}$,

$$\xi_{j,k}(f) = \begin{cases} \left(\int_{M_k} f^{q_j} d\mu_k\right)^{1/q_j} & \text{if } j = k, \\ \left(\int_{M_k} f^{s_j} d\mu_k\right)^{1/s_j} & \text{if } j \neq k. \end{cases}$$

For each $j \in \{1, \dots, n\}$, let $\xi_j = \xi_{j,n} \circ \cdots \circ \xi_{j,1}$, a mixed norm defined by an n-tuple S_j, where each entry $S_j(k) = s_j$, except for $S_j(j) = q_j$. Observe that $\sum_{j=1}^{n} S_j^{-1} = 1$, since for each $k \in \{1, \dots, n\}$,

$$\begin{aligned} \sum_{j=1}^{n} S_j(k)^{-1} &= q_k^{-1} + \sum_{j \neq k} s_j^{-1} \\ &= q_k^{-1} + \sum_{j \neq k} \left[q_j^{-1} + \frac{1}{n-1} \left(1 - \sum_{m=1}^{n} q_m^{-1} \right) \right] \\ &= \sum_{j=1}^{n} q_j^{-1} + \sum_{j \neq k} \frac{1}{n-1} - \sum_{j \neq k} \frac{1}{n-1} \sum_{m=1}^{n} q_m^{-1} \\ &= \sum_{j=1}^{n} q_j^{-1} + 1 - \sum_{m=1}^{n} q_m^{-1} = 1, \end{aligned}$$

so by Hölder's inequality (Theorem 7.2.2) and Minkowski's integral inequality (Corollary 7.2.6),

$$\int_{M^n} f_1 f_2 \cdots f_n d\mu^n \le \prod_{j=1}^{n} \xi_j(f_j) \le \prod_{j=1}^{n} \tilde{\xi}_j(f_j),$$

where, for each $j \in \{1, \ldots, n\}$,

$$\tilde{\xi}_j = \xi_{j,j} \circ \xi_{j,n} \circ \cdots \circ \xi_{j,j+1} \circ \xi_{j,j-1} \circ \cdots \circ \xi_{j,1}.$$

Minkowski's integral inequality implies that each $\xi_j \le \tilde{\xi}_j$ because each $s_j \le q_j$, so $\xi_{j,j}$ has a greater exponent, q_j, than the other $\xi_{j,k}$ with their s_j exponents. □

For another example, take Lemma 1 from [13], itself a rather restricted special case of Lemma 5.3 in Blei's [7]. Again, there are preliminary definitions.

Definition A.0.6. For two positive integers m and n, both assumed to be larger than 1, define

$$M(m,n) = \{i = (i_1, \ldots, i_m) : i_1, \ldots, i_m \in \{1, \ldots, n\}\}.$$

Definition A.0.7. Given an index i in $M(m,n)$, we set $i^k = (i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_m)$, which is then an index in $M(m-1, n)$.

Lemma A.0.8. *For all families $(c_i)_{i \in M(m,n)}$ of complex numbers, we have*

$$\left(\sum_{i \in M(m,n)} |c_i|^{\frac{2m}{m+1}} \right)^{\frac{m+1}{2m}} \le \prod_{1 \le k \le m} \left[\sum_{i_k=1}^{n} \left(\sum_{i^k \in M(m-1,n)} |c_i|^2 \right)^{\frac{1}{2}} \right]^{\frac{1}{m}}.$$

Note that each such family (c_i) is naturally identified with a complex-valued function $f(i_1, \ldots, i_m)$ on $\{1, \ldots, n\}^m$ defined by $f(i_1, \ldots, i_m) = c_i$. The above result can be substantially generalized, and is a special case of the following proposition.

Proposition A.0.9. *Let $(X_1, \mu_1), \ldots, (X_m, \mu_m)$ be σ-finite measure spaces with product space (X, μ). For any $k \in \{1, \ldots, n\}$, let*

$$X^k = X_1 \times \cdots \times \widehat{X_k} \times \cdots \times X_m,$$

with its product measure μ^k, where the notation $\widehat{X_k}$ indicates that this factor is omitted. For any μ-measurable function $f : X \to \mathbb{C}$,

$$\left(\int_X |f|^{\frac{2m}{m+1}} d\mu \right)^{\frac{m+1}{2m}} \le \prod_{1 \le k \le m} \left[\int_{X_k} \left(\int_{X^k} |f|^2 d\mu^k \right)^{\frac{1}{2}} d\mu_k \right]^{\frac{1}{m}}.$$

Proof. For $k \in \{1, \ldots, m\}$, let $g_k = |f|^{\frac{2}{m+1}}$. Let $P_k = (p_{k,1}, \ldots, p_{k,n})$, where each

$$p_{k,i} = \begin{cases} m+1 & \text{if } i \ne k \\ \frac{m+1}{2} & \text{if } i = k \end{cases}.$$

For each $i \in \{1, \ldots, m\}$, $\sum_{k=1}^{m} p_{k,i}^{-1} = (m-1) \cdot \frac{1}{m+1} + \frac{2}{m+1} = 1$, so by Hölder's inequality (Theorem 7.2.2)

$$\int_X |f|^{\frac{2m}{m+1}} = \int_X \prod_{k=1}^{m} g_k \leq \prod_{k=1}^{m} \rho_{P_k}(g_k).$$

Let $\tilde{\rho}_k$ denote the mixed norm which results from permuting ρ_{P_k} so that its exponents are in decreasing order; the least exponent $\frac{m+1}{2}$ comes last in $\tilde{\rho}_k$. By Minkowski's integral inequality (Corollary 7.2.6), each $\rho_{P_k} \leq \tilde{\rho}_k$. Therefore

$$\int_X |f|^{\frac{2m}{m+1}} \leq \prod_{k=1}^{m} \tilde{\rho}_k \left(|f|^{\frac{2}{m+1}} \right) = \prod_{1 \leq k \leq m} \left[\int_{X_k} \left(\int_{X^k} |f|^2 d\mu^k \right)^{\frac{1}{2}} d\mu_k \right]^{\frac{2}{m+1}}.$$

Finally, take the $\frac{m+1}{2m}$ power of each side. □

Furthermore, Lemma 5.3 from [7] can be readily obtained with these methods, even though it is a somewhat more complicated result. Once more, the preliminary notation. The wording does not match Blei's original, but should be similar, and correctly reflect the meaning.

Definition A.0.10. Consider integers $J > K > 0$. Let $N = \binom{J}{K}$ and let $S_1, \ldots, S_N$ enumerate the size-K subsets of $\{1, \ldots, J\}$, and for $1 \leq \alpha \leq N$, let $\sim S_\alpha$ denote the complement $\{1, \ldots, J\} \setminus S_\alpha$ of S_α.

Lemma A.0.11. *For any J-fold indexed sequence $\left(a_{j_1,\ldots,j_J}\right)$, i.e. function on the product $(\mathbb{Z}^+)^J$,*

$$\left(\sum_{j_1,\ldots,j_J} \left|a_{j_1,\ldots,j_J}\right|^{\frac{2J}{K+J}} \right)^{\frac{K+J}{2J}} \leq \prod_{\alpha=1}^{N} \left[\sum_{S_\alpha} \left(\sum_{\sim S_\alpha} \left|a_{j_1,\ldots,j_J}\right|^2 \right)^{1/2} \right]^{1/N}.$$

As always with these methods, the result does not depend on exactly which σ-finite measure spaces are considered, much less the fact that the integration here is summation, so a generalization will be proven.

Proposition A.0.12. *For any σ-finite measure spaces $(X_1, \mu_1), \ldots, (X_J, \mu_J)$ and any measurable function $f(x_1, \ldots, x_J)$ on $X_1 \times \cdots \times X_J$,*

$$\left(\int_{\{1,\ldots,J\}} |f|^{\frac{2J}{K+J}} \right)^{\frac{K+J}{2J}} \leq \prod_{\alpha=1}^{N} \left[\int_{S_\alpha} \left(\int_{\sim S_\alpha} |f|^2 \right)^{1/2} \right]^{1/N},$$

where the notation $\int_S$, for $S \subset \{1, \ldots, J\}$, denotes integration over the product space $\prod_{k \in S} X_k$.

Proof. For $1 \leq \alpha \leq N$, let $g_\alpha(x_1, \ldots, x_n) = |f|^{\frac{2J}{N(K+J)}}$ and define a J-tuple P_α by

$$P_\alpha(i) = \begin{cases} \frac{N(K+J)}{J} & \text{if } i \notin S_\alpha \\ \frac{N(K+J)}{2J} & \text{if } i \in S_\alpha \end{cases}$$

for $i \in \{1, \ldots, J\}$. For each i, the number of sets S_α which include a particular i is $\binom{J-1}{K-1}$, since any set of size K known to include i is determined by the other $K-1$ elements chosen out of the $J-1$ in $\{1,\ldots,J\}\setminus\{i\}$. The number of sets S_α which exclude that i is $\binom{J-1}{K}$, because all K elements of each such set are chosen from the $J-1$ other values than i. Therefore

$$\begin{aligned}\sum_{\alpha=1}^{N} P_\alpha(i)^{-1} &= \binom{J-1}{K-1}\frac{2J}{N(K+J)} + \binom{J-1}{K}\frac{J}{N(K+J)}\\ &= \frac{J}{K+J}\left(2\frac{\binom{J-1}{K-1}}{\binom{J}{K}} + \frac{\binom{J-1}{K}}{\binom{J}{K}}\right)\\ &= \frac{J}{K+J}\left(2\frac{(J-1)!}{(K-1)!(J-K)!}\frac{K!(J-K)!}{J!} + \frac{(J-1)!}{K!(J-K-1)!}\frac{K!(J-K)!}{J!}\right)\\ &= \frac{J}{K+J}\left(\frac{2K}{J} + \frac{J-K}{J}\right)\\ &= \frac{J}{K+J}\cdot\frac{K+J}{J} = 1.\end{aligned}$$

Since $\sum_{\alpha=1}^{N} P_\alpha^{-1} = 1$ coordinatewise, by Hölder's inequaliity (Theorem 7.2.2),

$$\int_{\{1,\ldots,J\}} |f|^{\frac{2J}{K+J}} = \int_{\{1,\ldots,J\}} \prod_{\alpha=1}^{N} g_\alpha \le \prod_{\alpha=1}^{N} \rho_{P_\alpha}(g_\alpha).$$

Let $\tilde\rho_{P_\alpha}$ denote the mixed norm obtained by permuting ρ_{P_α} so that its greatest exponents $\frac{N(K+J)}{J}$ occur first, in coordinates from S_α, and its least exponents $\frac{N(K+J)}{2J}$ occur last, in coordinates from $\sim S_\alpha$. By Minkowski's integral inequality (Corollary 7.2.6), each $\rho_{P_\alpha} \le \tilde\rho_{P_\alpha}$, so

$$\int_{\{1,\ldots,J\}} |f|^{\frac{2J}{K+J}} \le \prod_{\alpha=1}^{N} \tilde\rho_{P_\alpha}(g_\alpha) = \prod_{\alpha=1}^{N}\left[\int_{S_\alpha}\left(\int_{S_\alpha} |f|^2\right)^{1/2}\right]^{\frac{2J}{N(K+J)}}.$$

Take $\frac{K+J}{2J}$ powers on each side to obtain the desired inequality. □

Appendix B

Index of Notation

Notation	Page	Description
$\lVert\cdot\rVert_{L^p_\mu(X)}$	6	L^p norm over (X,μ)
$L^p_\mu(X)$	7	L^p space over (X,μ)
$\lVert\cdot\rVert_p$	7	L^p norm on an inferred measure space
$\lVert\cdot\rVert_{\ell^p_w(I)}, \lVert\cdot\rVert_{\ell^p(I)}, \lVert\cdot\rVert_{\ell^p}$	7	L^p norms w.r.t (possibly weighted) counting measure
$L^+(X)$	7	Nonnegative measurable functions on X
$\ll$	8	Absolute continuity of measures
$\frac{d\nu}{d\mu}$	8	Radon-Nikodym derivative
$[f]_\mu$	9	Equivalence class of f for μ-a.e. agreement
p'	12	Conjugate exponent to p for Hölder's inequality
$\sim$	18	Equivalence of atoms by null symmetric difference
$L^P, \lVert\cdot\rVert_P$	23	(Unpermuted) mixed-norm space L^P and its norm
$L^{\sigma(P)}, \lVert\cdot\rVert_{\sigma(P)}$	24	Permuted mixed-norm space $L^{\sigma(P)}$ and its norm
C_1, C_2	27	Norms of inclusions $L^{p_k}_{\mu_k}(X_1) \hookrightarrow L^{q_1}_{\nu_1}(X_1)$, $L^{p_2}_{\mu_2}(X_2) \hookrightarrow L^{q_2}_{\nu_2}(X_2)$
α, β	33	Exponents dependent on p_1, q_1, p_2, and q_2
$C_1(A_i), C_2(B_j)$	49	Best one-variable constants on subsets $A_i \subset X_1$, $B_j \subset X_2$
$p : q$	49	Relative conjugate of p with respect to q
$C_{BF}(A, B)$	51	Best constant for block factorable functions on $(A_i \times B_i)$
E_0, F_0	56	Atomless parts of X_1 and X_2
$(E_i)_{i\in I}, (F_j)_{j\in J}$	56	Atoms in X_1 and X_2
u_1, v_1, u_2, v_2	57	Weights from measures of atoms on X_1 and X_2
$C(I_0, J_0)$	67	Best constant for functions supported on $I_0 \times J_0$
$C_1(i), C_2(j)$	74	Best one-variable constants for functions on atoms
$C(i, j)$	74	Best constant for functions supported on $\{i\} \times \{j\}$
K^*	75	Decreasing rearrangement of a sequence
$\rho^p_{\mu_k}$	87	One-variable $L^p_{\mu_k}(X_k)$ norm for functions on $X_1 \times \cdots \times X_n$
ρ, ξ	94	Mixed norms on multiple variables
L^ρ, L^ξ, C	94	Mixed norm spaces on multiple variables, best constant
$C_1, \ldots, C_n$	97	Norms of $L^{p_k}_{\mu_k}(X_k) \hookrightarrow L^{q_k}_{\nu_k}(X_k)$, best constants in $\xi_k \le C_k\rho_k$

Curriculum Vitae

Name:	Wayne Grey
Post-Secondary Education and Degrees:	University of British Columbia Vancouver, B.C., Canada 1998 - 2004 B.Sc.
	University of Washington Seattle, Washington, USA 2005 - 2008 M.Sc.
	University of Western Ontario London, Ontario, Canada 2010 - 2015 Ph.D.
Honours and Awards:	Queen Elizabeth II Graduate Scholarship 2012 - 2013
Related Work Experience:	Teaching Assistant University of Washington 2005 - 2008
	Teaching Assistant University of Western Ontario 2010 - 2014

Publications:

W. Grey and G. Sinnamon, *The inclusion problem for mixed-norm spaces,* Trans. Amer. Math. Soc. (to appear).