TWENTY ANSWERS FOR THE FINAL QUESTIONS

1. (a) Let \( (X, \| \cdot \|) \) and \( (Y, \| \cdot \|) \) be normed spaces. Show that the following conditions on a family \( \mathcal{T} \subseteq \mathcal{L}(X, Y) \) are logically equivalent: (1) \( \{ \| T \| : T \in \mathcal{T} \} \) is bounded in \( \mathbb{R}^+ \); (2) \( \{ T : T \in \mathcal{T} \} \) is equicontinuous at some point of \( X \); (3) \( \{ T : T \in \mathcal{T} \} \) is equicontinuous at \( 0 \in X \); (4) \( \{ T : T \in \mathcal{T} \} \) is equicontinuous at every point of \( X \). (This order may not be the most efficient one to prove round-robin implications.)

(2) \( \Rightarrow \) (3): Suppose \( \mathcal{T} \) is equicontinuous at \( x_0 \in X \). If \( \| x - x_0 \| < \delta \Rightarrow \| T(x) - T(x_0) \| < \varepsilon \) for every \( T \in \mathcal{J} \), then \( \| \Delta x \| < \delta \Rightarrow \| T(\Delta x) \| = \| T(x_0) + T(\Delta x) - T(x_0) \| = \| T(x_0 + \Delta x) - T(x_0) \| < \varepsilon \), showing that \( \mathcal{T} \) is equicontinuous at \( 0 \in X \) (with the same “modulus of equicontinuity”). (3) \( \Rightarrow \) (1): Take \( \varepsilon = 1 \) in the definition of equicontinuity at zero; then for any corresponding \( \delta > 0 \) we see that \( \| x \| < \delta \Rightarrow \| dx \| < \delta \Rightarrow \| T(dx) \| < 1 \Rightarrow \| T(x) \| < \frac{1}{\delta} \), so the norms of the \( T \in \mathcal{T} \) are uniformly bounded by \( 1/\delta \). (1) \( \Rightarrow \) (4): if \( \| T \| \leq K \) for \( T \in \mathcal{T} \), then \( \| x - x_0 \| < \varepsilon/K \Rightarrow \| T(x) - T(x_0) \| = \| T(x - x_0) \| < K \cdot \varepsilon/K = \varepsilon \) holds for every \( x \) and \( x_0 \) in \( X \). (4) \( \Rightarrow \) (2) has nothing to prove.

(b) Let \( (X, \| \cdot \|) \) be a Banach space and \( \{ \Psi_k \}_{k=1}^{\infty} \) a pointwise convergent sequence of linear functionals (elements of \( X^* \)), i.e., such that \( \lim_{k \to \infty} \Psi_k(x) \) exists for each \( x \in X \). Show that there is a single bound for all the \( \{ \| \Psi_k \| \}_{k=1}^{\infty} \), and that the linear mapping \( X \to \mathbb{K} \) defined by the pointwise limit is continuous on \( X \). {Use (a) and the Baire continuity theorem.}

It is routine to check that the mapping defined by the pointwise limit is linear. The Baire continuity theorem says that a sequence of continuous functions on a complete metric space that converges pointwise is equicontinuous on a dense \( G_\delta \); in particular, it is equicontinuous somewhere (and the pointwise limit \( \Psi(x) = \lim_{k \to \infty} \Psi_k(x) \) is continuous there, with the same modulus of continuity). Applying (a), we also see that \( \Psi \) is continuous everywhere and that there is a single bound for \( \{ \| \Psi \| \} \cup \{ \| \Psi_k \| \}_{k=1}^{\infty} \).

A very restricted and simple continuity theorem, called the Uniform Boundedness Principle, suffices to establish (b). Suppose \( \mathcal{T} \subseteq \mathcal{L}(X, Y) \) is a set of linear operators that is pointwise bounded, i.e., for each \( x \in X \) the set \( \{ T(x) : T \in \mathcal{T} \} \) is a norm-bounded set in \( Y \) (equivalently, the set \( \{ \| T(x) \| : T \in \mathcal{T} \} \) is a set of real numbers bounded by some \( K_x \geq 0 \)—which may depend on the choice of \( x \in X \)). Then \( \mathcal{T} \) has the properties (1)–(4) above. To see this, for each \( n \in \mathbb{N} \) let \( S_n = \{ x \in X : \sup \{ \| T(x) \| : T \in \mathcal{T} \} \leq n \} \). Each \( S_n \) is a norm-closed set in \( X \) because the \( T \)'s are continuous, and the pointwise boundedness of \( \mathcal{T} \) implies that \( X = \bigcup_{n=1}^{\infty} S_n \). \( X \) is complete in the norm metric, so the Baire category theorem implies that at least one \( S_n \) has interior points. If \( x_0 \) is an interior point of \( S_n \) and the ball \( \{ x \in X : \| x - x_0 \| \leq \delta \} \subseteq S_n \), then—because \( S_n \) contains with each pair of its points \( z \) and \( w \) the point \( \frac{1}{2}(z - w) \)—the ball \( \{ x \in X : \| x \| \leq \delta \} \subseteq S_n \).

{Each point \( \Delta x \) with \( \| \Delta x \| \leq \delta \) can be written in the form \( \Delta x = \frac{1}{2}[(x_0 + \Delta x) - (x_0 - \Delta x)] \), and both expressions in parentheses belong to the ball \( \{ x \in X : \| x - x_0 \| \leq \delta \} \subseteq S_n \}. So \( \| \Delta x \| \leq \delta \Rightarrow \| T(\Delta x) \| \leq n/\delta \). So a pointwise bounded \( \mathcal{T} \) is uniformly norm-bounded and thus equicontinuous. Since a sequence \( \{ \Psi_n \}_{n=1}^{\infty} \subseteq X^* \) for which \( \{ \Psi_n(x) \}_{n=1}^{\infty} \) is convergent for each \( x \in X \) is obviously pointwise bounded (convergent sequences in \( \mathbb{K} \) are bounded), such a sequence is (uniformly) bounded in norm, and from its equicontinuity the continuity of the pointwise limit functional defined by \( \Psi(x) = \lim_{n \to \infty} \Psi_n(x) \) follows immediately.\(^{(1)}\)

2. Use 1. to show: if \( (X, \mathcal{M}, \mu) \) is a (nonnegative) \( \sigma \)-finite measure space and \( f \) is an \( \mathcal{M} \)-measurable scalar-valued function for which \( \int_X f h d\mu \) (exists and) is finite for every \( h \in L^q \), where \( 1 \leq q \leq \infty \), then \( f \in L^p \).

The case \( q = \infty \) is easy, because we may take \( h = \begin{cases} 1/|f| & \text{if } f \neq 0 \\ 0 & \text{if } f = 0 \end{cases} \). Then
\[
\int f h d\mu = \int |f| d\mu < \infty
\]

\(^{(1)}\) It goes without saying that the functional analysts have generalized this argument out of all proportion: interested parties might want to look at the definition of *tonnelé* or *barrelled* locally convex spaces.
and thus \( f \in L^1(\mu) \). To handle the case \( 1 \leq q < \infty \), let \( \{ E_n \}_{n=1}^{\infty} \) be an increasing sequence of \( \mathcal{M} \)-measurable sets whose union is \( X \), each with \( \mu(E_n) < \infty \), and define

\[
f_n(x) = \begin{cases} f(x) \cdot \chi_{E_n}(x) & \text{where } |f(x)| \leq n \\ 0 & \text{elsewhere} \end{cases}
\]

Each \( f_n \) is bounded and is identically zero off \( E_n \), so it certainly belongs to \( L^p(\mu) \) and thus \( h \mapsto \int f_n h \, d\mu \) is a well-defined element of the dual space of \( L^q(\mu) \). For each \( h \in L^q \) a routine use of the dominated convergence theorem (with \( |f| \cdot |h| \) as the dominating function) shows that \( \int f_n h \, d\mu = \lim_{n \to \infty} \int f_n h \, d\mu \) and thus, by (a), the linear mapping \( h \mapsto \int f h \, d\mu \) is an element of \( L^q(\mu)^* \). Since we know what \( L^q(\mu)^* \) is, we see that there is a function \( g \in L^p(\mu) \) such that \( \int gh \, d\mu = \int f h \, d\mu \) holds for all \( h \in L^q(\mu) \). It is routine to verify that \( f \) and \( g \) cannot differ on a set of positive measure (but notice that the \( \sigma \)-finiteness of \((X, \mathcal{M}, \mu)\) plays a rôle here, because one needs to know that every set of positive \( \mu \)-measure is the union of a sequence of sets, each of which has finite positive \( \mu \)-measure), and so \( f \in L^p(\mu) \) as advertised. \{ There exist spaces \((X, \mathcal{M}, \mu)\) some of whose sets have infinite measure but intersect each set of finite measure in a null set. The characteristic function of such a set will integrate against an \( L^p \)-function to give zero, because an \( L^p \)-function must equal zero off a countable union of sets of finite measure; however, such a characteristic function does not belong to \( L^q \). Hence the need to employ \( \sigma \)-finiteness does not reflect a lack of technique. }\n
It is possible to avoid the use of functional-analytic tools in solving this problem; one mimics the proof of the converse Hölder inequality. Suppose the measurable function \( f \) has the property that \( \int f h \, d\mu \) exists for every \( h \in L^q \); then (as multiplying and dividing by \( \text{sgn}(f) \) shows) so does \( |f| \). Consider first the case \( 1 \leq q < \infty \), and suppose it were true that \( \int |f|^q \, d\mu = +\infty \). Let \( f_n(x) = \min\{|f|(x), n\} \); then

\[
\int_E |f_n|^p \, d\mu < +\infty \quad \text{whenever } E \in \mathcal{M} \text{ is a set of finite } \mu \text{-measure. If } \{ E_n \}_{n=1}^{\infty} \subseteq \mathcal{M} \text{ is an increasing sequence of sets of finite } \mu \text{-measure whose union is } X, \text{ then since the } p \text{-th power integral of } f \text{ is finite the sequence of integrals } \int_{E_n} |f_n|^p \, d\mu \text{ must increase to } +\infty \text{ by the monotone convergence theorem. By passing to a subsequence if necessary, we may assume } \|f_n \cdot \chi_{E_n}\|_p \geq 4^n. \text{ As in the proof of the converse Hölder inequality, for each } f_n \text{ there is an } h_n \in L^q \text{ of norm } 1 \text{ such that } \int f_n \chi_{E_n} h_n \, d\mu = \|f_n \cdot \chi_{E_n}\|_p; \text{ in fact, using } p = (p - 1)q \text{ one may take } h_n = \frac{f_n^{p-1}}{\|f_n\|_p^p} \chi_{E_n} \text{ for } p > 1, \text{ while for } p = 1 \text{ one just takes } \chi_{E_n}. \text{ If we now form } h = \sum_{n=1}^{\infty} \frac{1}{2^n} h_n, \text{ we see that } \|h\|_q \leq 1, \text{ by the monotone convergence theorem (or by the completeness of } L^q). \text{ However, the monotone convergence theorem also gives us}
\]

\[
\int f h \, d\mu = \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \frac{1}{2^n} \int f_n h_n \, d\mu \geq \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \frac{1}{2^n} \cdot 4^n = +\infty
\]

and so we have exhibited a function \( h \in L^q \) such that \( \int f h \, d\mu = +\infty \), contrary to assumption. The case \( p = \infty \) is similar: again we may restrict our attention to nonnegative functions, and if \( f \) is a \( \mathcal{M} \)-measurable function on \( X \) that is not essentially bounded, then for every \( n \in \mathbb{N} \) there exists (by the \( \sigma \)-finiteness of \((X, \mathcal{M}, \mu)\)) a set \( E_n \in \mathcal{M} \) of nonzero finite \( \mu \)-measure such that \( f(x) \geq 4^n \) on \( E_n \). If we form

\[
h = \sum_{n=1}^{\infty} \frac{1}{2^n} \mu(E_n) \chi_{E_n}, \text{ we see that it is an } L^1 \text{ function of norm } 1 \text{ for which the monotone convergence theorem gives us, contrary to hypothesis on } f, \text{ the integral}
\]

\[
\int f h \, d\mu = \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \frac{1}{2^n} \mu(E_n) \int_{E_n} f \, d\mu \geq \lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \frac{1}{2^n} \mu(E_n) 4^n \mu(E_n) = \lim_{m \to \infty} \sum_{n=1}^{m} 2^n = +\infty.
\]
3. Use 2. to show: if \( f \in L^1(\mathbb{R}) \) and \( g \in L^p(\mathbb{R}) \), where \( 1 \leq p \leq \infty \), then the formula \((f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x - t) \, dt\) defines a function at almost all \( x \in \mathbb{R} \), the function belongs to \( L^p(\mathbb{R}) \), and \( \|f * g\|_p \leq \|f\|_1 \|g\|_p \). {Of course Lebesgue measure and Lebesgue-measurable sets are intended. Fubini-Tonelli are essential here.}

{The following argument does not employ 2 above.} First, we need to show that \( f(t)g(x - t) \) is a (jointly) measurable function of \((x, t) \in \mathbb{R}^2\). Let \( G_1(x, t) = g(x) \). Then the measurability of \( g \) as a function on \( \mathbb{R} \) implies the measurability of \( G_1 \) as a function on \( \mathbb{R}^2 \), because for each \( a \in \mathbb{R} \) the set

\[
\{(x, t) : G_1(x, t) > a\} = \{(x, t) : g(x) > a, \ t \in \mathbb{R}\}
\]

is measurable (as an infinite cylinder with measurable base \( \{x : g(x) > a\} \)). Now consider the nonsingular linear transformation \( x = u - v, \ y = u + v \). The function \( G_2(u, v) = G_1(u - v, u + v) \) is measurable on \( \mathbb{R}^2 \), and thus so is \( g(u - v) = G_2(u, v) \). It follows that \( f(t)g(x - t) \) is the product of two measurable functions on \( \mathbb{R}^2 \) and thus is measurable. To show that the convolution \((f * g)(x)\) defines a function at almost all \( x \in \mathbb{R} \) it will suffice to show that \( f(t)g(x - t) \) is an integrable function of \((x, t) \).

**Case 1: \( p = 1 \).**

Since \( f(t)g(x - t) \) is measurable, we can use Tonelli’s theorem to give

\[
\int \int |f(t)| |g(x - t)| \, dx \, dt = \int |f(t)| \left[ \int |g(x - t)| \, dx \right] \, dt \\
= \int |f(t)| \int |g(x)| \, dx \\
= \|f\|_1 \|g\|_1 < \infty.
\]

Thus \( F(x, t) = f(x)g(x - t) \in L^1(\mathbb{R}^2) \), and so, by Fubini’s theorem, \( F(x, t) \) is integrable in \( t \) for almost every \( x \). Since \( |(f * g)| \leq |f| * |g| \) (as can be seen, e.g., by writing \( f \) and \( g \) as differences of nonnegative functions), we have

\[
\|f * g\|_1 = \int |f * g| \, dx \leq \int (|f| * |g|) \, dx = \|f\|_1 \|g\|_1 < \infty.
\]

**Case 2: \( 1 < p < \infty \).**

Because \( g \in L^p \), we have \( |g|^p \in L^1 \). Using case 1, we see that for almost every \( x \in \mathbb{R} \) the function \( |f(t)|^{1/p} |g(x - t)| \in L^p(\mathbb{R}) \) as a function of \( t \). If \( q \) is the conjugate exponent of \( p \) (so \( \frac{1}{p} + \frac{1}{q} = 1 \)) we have \( |f|^{1/q} \in L^q \), so for those \( x \in \mathbb{R} \) for which the function \( |f(t)|^{1/p} |g(x - t)| \in L^p(\mathbb{R}) \) as a function of \( t \), the Hölder inequality gives the inequality

\[
|f(t)| |g(x - t)| = |f(t)|^{1/p} |g(x - t)| \left| f(t) \right|^{1/q} \in L^1
\]

(as a function of \( t \)), valid for those \( x \in \mathbb{R} \) for which \( t \mapsto |f(t)|^{1/p} |g(x - t)| \in L^p(\mathbb{R}) \). Integrating this Hölder inequality gives

\[
|(f * g)(x)| \leq (|f| * |g|)(x) = \left[ \int |f(t)| |g(x - t)|^p \, dt \right]^{1/p} \left[ \int |f(t)| \, dt \right]^{1/q}
\]

whose \( p \)-th power is then

\[
|(f * g)(x)|^p \leq (|f| * |g|)^p(x) = \int |f(t)| |g(x - t)|^p \, dt \left[ \int |f(t)| \, dt \right]^{p/q}.
\]

Applying case 1 to the first integral on the r. h. side (which is clearly a convolution integral) we get

\[
\int \int |f(t)| |g(x - t)|^p \, dt \, dx \leq \|f\|_1 \|g|^p \|_1 = \|f\|_1 \|g\|_p^p.
\]
It follows that
\[ \| (f \ast g) \|_p^p = \int |(f \ast g)(x)|^p \, dx \leq \int |f| \cdot |g| \cdot |h|(x) \, dx \]
\[ \leq \| f \|_1 \| g \|_p \| h \|_1^{p/q} \]
and so finally (taking p-th roots)
\[ \| (f \ast g) \|_p \leq \| f \|_1^{1/p+1/q} \| g \|_p = \| f \|_1 \| g \|_p \]
as advertised.

**Case 3:** \( p = \infty \).

In this case there is almost nothing to prove, since for each \( x \in \mathbb{R} \) the function \( t \mapsto g(x-t) \) has the same essential bound as \( g \) and
\[ |(f \ast g)(x)| \leq |(f \ast g)|_1(x) = \int |f(t)||g(x-t)||h(x)| \, dt \leq \| f \|_1 \| g \|_{\infty} \cdot \]
This holds for all \( x \in \mathbb{R} \), giving \( \| f \ast g \|_{\infty} \leq \| f \|_1 \| g \|_{\infty} \).\(^{(2)} \)

An approach to this problem using 2. might proceed as follows. Let \( f \in L^1 \) be given and consider, for \( g \in L^p \) and \( h \in L^q \) (where \( p \) and \( q \) are conjugate exponents, as usual), the function \( (x,t) \mapsto f(x)g(x-t)h(t) \) on \( \mathbb{R}^2 \). This function is measurable, as considerations in the preceding solution showed. Applying Tonelli’s theorem to the function \( (x,t) \mapsto |f(t)||g(x-t)||h(x)| \) gives
\[ \int \int |f(t)||g(x-t)||h(x)| \, d(t \times x) = \int |f(t)| \left( \int |g(x-t)||h(x)| \, dx \right) \, dt \leq \int |f(t)||g||_p \| h \|_q \, dt = \| f \|_1 \| g \|_p \| h \|_q \]
so the iterated integral \( \int \left( \int |f(t)||g(x-t)||h(x)| \, dx \right) \, dt \) must also exist and be bounded by the same bound. In particular, \( \int |f(t)||g(x-t)| \, dt < \infty \) must hold for almost all \( x \in \mathbb{R} \) (we could have picked \( h \) to be nonzero at every \( x \in \mathbb{R} \)) and (therefore) \( x \mapsto \int f(x)g(x-t) \, dt \) is almost-everywhere defined, measurable, and satisfies \( \left| \int f(t)g(x-t) \, dt \right| \leq \int |f(t)||g(x-t)| \, dt \). Fubini now tells us that the iterated integral that we get by taking the absolute-value signs off satisfies
\[ \left| \int \left( \int f(t)g(x-t) \, dt \right) h(x) \, dx \right| \leq \| f \|_1 \| g \|_p \| h \|_q \]
and so the function \( x \mapsto \int f(x)g(x-t) \, dt \) satisfies the hypotheses of 2 and thus belongs to \( L^p(\mathbb{R}) \). Because the value of the linear functional on \( L^q(\mathbb{R}) \) defined by \( h \mapsto \int \left( \int f(t)g(x-t) \, dt \right) h(x) \, dx \) is bounded by \( \| f \|_1 \| g \|_p \| h \|_q \), the \( L^p(\mathbb{R}) \)-function \( f \ast g = (x \mapsto \int f(x)g(x-t) \, dt) \) has \( L^p \) norm at most \( \| f \|_1 \| g \|_p \) by the converse of the Hölder inequality (or the fact that \( (L^q)^* \) is identified with \( L^p \) in norm as well as algebraically); thus we have all we desired.

\(^{(2)}\) Using the fact that the continuous functions of compact support are dense in \( L^1 \), it is easy to show that the functions in \( L^1 \ast L^\infty \) not only are bounded on \( \mathbb{R} \) but also are uniformly continuous on \( \mathbb{R} \).
Let \((X, \mathcal{M})\) be a measurable space, \(\mu \geq 0\) with \(\mu(X) = 1\) a measure on \((X, \mathcal{M})\), and \(\mathcal{M}_0 \subseteq \mathcal{M}\) a sub-\(\sigma\)-algebra of \(\mathcal{M}\). Show that for each \(f \in L^1(X, \mathcal{M}, d\mu)\) there is a unique function (class) \(f_0 \in L^1(X, \mathcal{M}_0, d\mu)\) for which the relation \(\int_X f g d\mu = \int_X f_0 g d\mu\) holds for every \(\mathcal{M}_0\)-measurable \(g\) for which the integrals are finite. This function is called \((a \ \text{version of the}) \ \text{conditional expectation} \ \text{of} \ f \ \text{with respect to} \ \mathcal{M}_0 \ \text{and written} \ E[f\mid\mathcal{M}_0]. \ \{\text{Use Radon-Nikodým on the measure} \ f \cdot \mu, \ \text{considered as acting on} \ (X, \mathcal{M}_0).\} \)

To construct \(E[f\mid\mathcal{M}_0]\) one simply applies the Radon-Nikodým theorem, since the set function

\[
E \mapsto \int_E f \ d\mu
\]

is well defined on \((X, \mathcal{M}_0)\) and inherits total finiteness, countable additivity and absolute continuity from the set function with the same defining formula that is defined on all of \(\mathcal{M}\). Radon-Nikodým gives an \(\mathcal{M}_0\)-integrable function \(f_0\), uniquely defined up to an \(\mathcal{M}_0\)-measurable null function, for which \(\int_E f \ d\mu = \int_E f_0 \ d\mu\), or equivalently \(\int_X f 1_E \ d\mu = \int_X f_0 1_E \ d\mu\), holds for every \(E \in \mathcal{M}_0\). The uniqueness of \(f_0\) implies that if we write \(f = f^+ - f^-\) and form \((f^+)_0\) and \((f^-)_0\) (which we may assume to be finite-valued everywhere, without loss of generality), then \(f_0 = (f^+)_0 - (f^-)_0\) must hold.\(^{(3)}\) If \(g\) is \(\mathcal{M}_0\)-measurable and \(\int_X f g d\mu\) exists then so does \(\int_X |f| |g| d\mu\), so it suffices to consider the case in which both \(f\) and \(g\) are nonnegative—look at \(f^\pm\) and \(g^\pm\) separately. (Because the positive and negative parts of a given function \(h\) are the \textit{smallest} nonnegative functions \(h_1\) and \(h_2\) for which \(h = h_1 - h_2\), we must have \((f^+)_0 \geq (f_0)^+\) and \((f^-)_0 \geq (f_0)^-\), so if the integrals \(\int_X |g|(f^\pm)_0 \ d\mu\) exist [and are finite], the integrals \(\int_X |g|(f_0)^\pm \ d\mu\) must also exist [and be, respectively, no larger].) There exist increasing sequences of \(\mathcal{M}_0\)-measurable simple functions converging pointwise everywhere to \(g\). For a typical \(\mathcal{M}_0\)-measurable simple function one has

\[
\int_X f \cdot \left[ \sum_{j=1}^n a_j 1_{E_j} \right] \ d\mu = \sum_{j=1}^n a_j \int_{E_j} f \ d\mu = \sum_{j=1}^n a_j \int_{E_j} f_0 \ d\mu = \int_X \left[ \sum_{j=1}^n a_j 1_{E_j} \right] f_0 \ d\mu
\]

and for nonnegative \(f\) this relation is preserved under taking monotone limits, by the monotone convergence theorem; so \(\int_X g f \ d\mu = \int_X g f_0 \ d\mu\) holds for nonnegative \(g\) and \(f\) and thus holds in general.

5. In the situation of 4. above, show that \(f \mapsto E[f\mid\mathcal{M}_0]\) has the following properties:

(a) \(E[\cdot\mid\mathcal{M}_0]\) is linear;

(b) \(f \geq 0 \Rightarrow E[f\mid\mathcal{M}_0] \geq 0\);

(c) \(E[g\cdot f\mid\mathcal{M}_0] = g \cdot E[f\mid\mathcal{M}_0]\) for \(g \in L^\infty(X, \mathcal{M}_0, d\mu)\), and in all cases if both \(f, g \geq 0\), \(f\) is \(\mathcal{M}\)-measurable and \(g\) is \(\mathcal{M}_0\)-measurable;

(d) For \(1 \leq p \leq \infty\), \(f \mapsto E[f\mid\mathcal{M}_0]\) sends \(L^p(X, \mathcal{M}, d\mu) \rightarrow L^p(X, \mathcal{M}_0, d\mu)\) with \(\|E[\cdot\mid\mathcal{M}_0]\| = 1\);

(e) For \(p = 2\), \(L^2(X, \mathcal{M}_0, d\mu)\) is a norm-closed subspace of \(L^2(X, \mathcal{M}, d\mu)\), and \(f \mapsto E[f\mid\mathcal{M}_0]\) is the orthogonal projection onto it.

\(^{(3)}\) Note that these functions may not be the positive and negative parts of \(f_0\); the reader can easily find examples—on \([0, 1]\) equipped with Lebesgue measure—in which \(\mathcal{M}_0\) has four elements and \(f_0\) is positive although \(f\) is not.
which \( g \) is a \( \mathcal{M}_0 \)-measurable simple function, because for \( E \in \mathcal{M}_0 \) we have for any \( \mathcal{M}_0 \)-measurable simple function \( g = \sum_{j=1}^{n} a_j \chi_{E_j} \)

\[
\int_{E} \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] f \, d\mu = \sum_{j=1}^{n} a_j \int_{E \cap E_j} f \, d\mu = \sum_{j=1}^{n} a_j \int_{E \cap E_j} f_0 \, d\mu = \int_{E} \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] f_0 \, d\mu ,
\]

(\$)

showing that \( g \cdot f_0 \) is a \( \mathcal{M}_0 \)-measurable) Radon-Nikodým derivative of \( E \mapsto \int_{E} g \, f \, d\mu \). In general, \( g \in L^\infty(X, \mathcal{M}_0, d\mu) \) is a uniform limit on \( X \) of \( \mathcal{M}_0 \)-measurable simple functions and it does not require the dominated or monotone convergence theorems to see that the equality (\$) above is preserved under taking uniform limits of the simple functions: the estimates follow easily because \( |f| \in L^1(X, \mathcal{M}, d\mu) \). If \( 0 \leq f \in L^1(X, \mathcal{M}, d\mu) \) and \( g \geq 0 \) is \( \mathcal{M}_0 \)-measurable, then one takes an increasing sequence of nonnegative \( \mathcal{M}_0 \)-measurable simple functions converging pointwise everywhere to \( g \) and applies the monotone convergence theorem to both sides of (\$), obtaining the desired equality and noting that one side remains finite if and only if the other side does.

One can begin consideration of (d) by observing that because \( f \mapsto E[f|\mathcal{M}_0] \) preserves positivity, the relation \(-|f| \leq f \leq |f|\) implies \(-E[|f||\mathcal{M}_0] \leq E[f|\mathcal{M}_0] \leq E[|f||\mathcal{M}_0]\), and thus \( E[|f|\mathcal{M}_0] \leq E[|f|\mathcal{M}_0] \); therefore, if we can show that \( f \mapsto E[f|\mathcal{M}_0] \) decreases \( L^p \) norm on nonnegative functions, then we will know that it decreases \( L^p \) norm on signed functions. Now if \( h \) is a nonnegative \( \mathcal{M}_0 \)-measurable function, then its \( L^p(X, \mathcal{M}_0, d\mu) \)-norm can be computed (even if it is \( +\infty \)) as the supremum of all integrals \( \int_X h \cdot g \, d\mu \) in which \( g \) is a nonnegative \( \mathcal{M}_0 \)-measurable simple function of \( L^q \) norm at most 1 (this follows easily from the converse of the Hölder inequality, approximating the \( L^q \) function by simple functions). Applying this fact to \( E[f|\mathcal{M}_0] \) where \( 0 \leq f \in L^p(X, \mathcal{M}, d\mu) \), we see that

\[
\|E[f|\mathcal{M}_0]\|_p = \sup \left\{ \int_X E[f|\mathcal{M}_0] \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] \, d\mu \right\} = \sup \left\{ \int_X f \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] \, d\mu \right\}
\]

where the indicated supremum is taken over the set of all nonnegative \( \mathcal{M}_0 \)-measurable simple functions of \( L^q \) norm \( \leq 1 \). But that set is contained in the set of all nonnegative \( \mathcal{M} \)-measurable simple functions of \( L^q \) norm \( \leq 1 \), so the Hölder inequality gives

\[
\sup \left\{ \int_X f \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] \, d\mu \right\} \leq \|f\|_p \cdot 1
\]

and thus \( \|E[f|\mathcal{M}_0]\|_p \leq \|f\|_p \) as desired. So the linear mapping \( f \mapsto E[f|\mathcal{M}_0] \) decreases \( L^p \) norm, if anything, and its operator norm is thus \( \leq 1 \). But its operator norm is also \( \geq 1 \), because it leaves fixed any \( f \in L^p(X, \mathcal{M}, d\mu) \) that is \( \mathcal{M}_0 \)-measurable—in particular, it leaves fixed the identically-1 function, whose \( L^p \) norm is 1 for all \( 1 \leq p \leq \infty \). Thus the operator norm of \( f \mapsto E[f|\mathcal{M}_0] \) as an element of \( L^p(X, \mathcal{M}, d\mu) \) is exactly 1. Finally, to prove (e) that the linear operator \( f \mapsto E[f|\mathcal{M}_0] \) is the orthogonal projection of \( L^2(X, \mathcal{M}, d\mu) \) onto \( L^2(X, \mathcal{M}_0, d\mu) \), it suffices to prove two things: that it maps \( L^2(X, \mathcal{M}, d\mu) \) onto \( L^2(X, \mathcal{M}_0, d\mu) \) leaving \( L^2(X, \mathcal{M}_0, d\mu) \) fixed, and that for every \( L^2(X, \mathcal{M}, d\mu) \), the vector \( f - E[f|\mathcal{M}_0] \) is orthogonal to \( L^2(X, \mathcal{M}_0, d\mu) \). We have already checked that it maps \( L^2(X, \mathcal{M}, d\mu) \) into \( L^2(X, \mathcal{M}_0, d\mu) \), and that it leaves \( L^2(X, \mathcal{M}_0, d\mu) \) fixed is obvious. To check the orthogonality requirement it suffices to check orthogonality to a dense subspace of \( L^2(X, \mathcal{M}_0, d\mu) \), and of course we use the subspace consisting of \( \mathcal{M}_0 \)-measurable simple functions. For these we have

\[
\int_X \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] f \, d\mu = \int_X \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] f_0 \, d\mu , \quad \text{or}
\]

\[
\int_X \left[ \sum_{j=1}^{n} a_j \chi_{E_j} \right] \left[ f - E[f|\mathcal{M}_0] \right] d\mu = 0 ,
\]

which is the desired orthogonality relation.
6. Let \( \{f_j(x)\}_{j=1}^\infty \) and \( \{g_k(y)\}_{k=1}^\infty \) be orthonormal bases for the Hilbert spaces \( L^2(\mathbb{R}^n, \text{Lebesgue}, m_n) \) and \( L^2(\mathbb{R}^m, \text{Lebesgue}, m_m) \) respectively. Show that the doubly-indexed set \( \{f_j(x) \cdot g_k(y)\}_{j,k=1}^\infty \) is an orthonormal basis for the Hilbert space \( L^2(\mathbb{R}^{n+m}, \text{Lebesgue}, m_{n+m}) \).

It is straightforward to check that these functions are measurable on \( \mathbb{R}^{n+m} \). The Tonelli theorem gives

\[
\int_{\mathbb{R}^{n+m}} |f_j(x)g_k(y)|^2 \, d(x \times y) = \left[ \int_{\mathbb{R}^n} |f_j(x)|^2 \, dx \right] \cdot \left[ \int_{\mathbb{R}^m} |g_k(y)|^2 \, dy \right] < \infty
\]

and so the functions belong to \( L^2(\mathbb{R}^{n+m}, \text{Lebesgue}, m_{n+m}) \). Fubini then gives

\[
\int_{\mathbb{R}^{n+m}} f_j(x)g_k(y) \overline{f_p(y)g_q(y)} \, d(x \times y) = \left[ \int_{\mathbb{R}^n} f_j(x) \overline{f_p(x)} \, dx \right] \cdot \left[ \int_{\mathbb{R}^m} g_k(y) \overline{g_q(y)} \, dy \right] = \delta_{jp} \delta_{kq}
\]

and so the set \( \{f_j(x) \cdot g_k(y)\}_{j,k=1}^\infty \) is orthonormal. To complete the proof we need to show that the linear space spanned by these functions is dense in \( L^2(\mathbb{R}^{n+m}, \text{Lebesgue}, m_{n+m}) \).

Now the linear space spanned by functions of the form \( \chi_E(x) \cdot \chi_F(y) \), where \( E \subseteq \mathbb{R}^n \) and \( F \subseteq \mathbb{R}^m \) are measurable sets of finite measure, is dense in \( L^2(\mathbb{R}^{n+m}, \text{Lebesgue}, m_{n+m}) \): indeed, it would suffice to take \( E \) and \( F \) to be partly-open intervals. The reason is that the linear space of simple functions is dense in \( L^2(\mathbb{R}^{n+m}, \text{Lebesgue}, m_{n+m}) \), and every set \( G \subseteq \mathbb{R}^{n+m} \) which is measurable of finite measure can be covered by a countable union of disjoint partly-open intervals whose measure is only slightly larger than that of \( G \); the characteristic function of a suitably-chosen finite sub-union of those intervals will then approximate \( \chi_G(x,y) \) except for a slight error, and it is routine to check that the “slight” errors just mentioned can be chosen to make the \( L^2 \)-norm distance between \( \chi_G(x,y) \) and the sum of the characteristic functions of those partly-open intervals be as small as one pleases. It follows that to show that the linear space spanned by functions of the form \( f_j(x)g_k(y) \) is dense it will suffice to show that every function of the form \( \chi_E(x) \cdot \chi_F(y) \) (with \( E, F \) measurable of finite measure, or partly-open intervals if you wish) belongs to the closed linear space spanned by the \( f_j(x)g_k(y) \)'s. That this is true follows from the respective orthonormal sets being bases in their respective spaces, so that the respective Parseval equalities hold: for

\[
\| \chi_E \|_2^2 = \sum_{j=1}^\infty |\langle \chi_E, f_j \rangle|^2 \text{ and } \| \chi_F \|_2^2 = \sum_{k=1}^\infty |\langle \chi_F, g_k \rangle|^2 \text{ then imply}
\]

\[
\| \chi_E \cdot \chi_F \|_2^2 = \int_{\mathbb{R}^{n+m}} |\chi_E(x)\chi_F(y)|^2 \, d(x \times y) = \left[ \int_{\mathbb{R}^n} |\chi_E(x)|^2 \, dx \right] \cdot \left[ \int_{\mathbb{R}^m} |\chi_F(y)|^2 \, dy \right]
\]

\[
\| \chi_E \cdot \chi_F \|_2^2 = \sum_{j,k=1}^\infty |\langle \chi_E, f_j \rangle|^2 \sum_{k=1}^\infty |\langle \chi_F, g_k \rangle|^2 = \sum_{j,k=1}^\infty |\langle \chi_E, f_j \rangle\langle \chi_F, g_k \rangle|^2
\]

\[
= \sum_{j,k=1}^\infty \left[ \int_{\mathbb{R}^n} \chi_E(x)f_j(x) \, dx \right] \cdot \left[ \int_{\mathbb{R}^m} \chi_F(y)g_k(y) \, dy \right]^2
\]

and since this is the Parseval equality for \( \chi_E \cdot \chi_F \) with respect to the orthonormal set \( \{f_j(x) \cdot g_k(y)\}_{j,k=1}^\infty \), the Parseval theorem says that \( \chi_E \cdot \chi_F \) belongs to that set’s closed linear span. This suffices to establish the desired result.

Another approach is possible but some care is required. Let \( h \in L^2(\mathbb{R}^{n+m}, \text{Lebesgue}, m_{n+m}) \) be given. Then \( |h(x,y)|^2 \) is integrable \( d(x \times y) \), and therefore \( x \mapsto |h(x,y)|^2 \) is integrable for \( y \notin A \), where \( A \subseteq \mathbb{R}^m \) is a null set. It is also true that \( x \mapsto h(x,y) \) is measurable for \( y \) out of a certain null set; one can prove this by approximation with simple functions, or one can get this from the Fubini theorem by observing that if \( E \subseteq \mathbb{R}^n \) and \( F \subseteq \mathbb{R}^m \) are sets of finite measure then \( (x,y) \mapsto h(x,y)\chi_E(x)\chi_F(y) \) is measurable and integrable.
on $\mathbb{R}^{n+m}$, so $x \mapsto h(x,y)\chi_E(x)\chi_F(y)$ is measurable for $y$ out of a certain null set; letting $E$ run through an increasing sequence $\{E_t\}_{t=1}^\infty$ of sets of finite measure whose union is $\mathbb{R}^n$ and staying off the union of the exceptional null sets of $y$'s, one sees that the pointwise limit $x \mapsto h(x,y)\chi_F(y)$ is measurable for almost all $y$; then letting $F$ run through an increasing sequence $\{F_t\}_{t=1}^\infty$ of sets of finite measure whose union is $\mathbb{R}^m$ one sees that the pointwise limit $x \mapsto h(x,y)$ is measurable for $y$ out of a certain null set $B \subseteq \mathbb{R}^m$. With $N = A \cup B$ we see that $(x \mapsto h(x,y)) \in L^2(\mathbb{R}^n, \text{Lebesgue, } m_n)$ for $y \notin N$. For each $j \in \mathbb{N}$ and $y \notin N$ the integral
$$\int h(x,y)f_j(x)\ dm_n(x)$$
is then well defined, and in fact
$$\int |h(x,y)||f_j(x)|\ dm_n(x) \leq \sqrt{\int |h(x,y)|^2\ dm_n(x)}$$
by the Schwarz inequality. The r. h. side of this inequality is square-integrable $dm_m(y)$ by the Schwartz inequality. The a. e. defined function $y \mapsto \int h(x,y)f_j(x)\ dm_n(x)$, which can be shown to be measurable by the same kind of argument employed above, is a square-integrable function of $y \in \mathbb{R}^m$. Now if $h$ were orthogonal to all the functions $(f_j(x) \cdot g_k(y))_{j,k=1}^\infty$, then we would have for each $j \in \mathbb{N}$ the (orthogonality relations)
$$\int \left[ \int h(x,y)f_j(x)\ dm_n(x) \right] g_k(y)\ dm_m(y) = \int h(x,y)f_j(x)g_k(y)\ dm_n+m(x \times y) = 0$$for all $k \in \mathbb{N}$; these say that the square-integrable functions $y \mapsto \int h(x,y)f_j(x)\ dm_n(x)$, all of which are defined for $y \notin N$, are orthogonal to all the $g_k$'s. The $j$-th of these functions therefore takes the value zero for all $y \notin N \cup N_j \subseteq \mathbb{R}^m$, where $m_j(N_j) = 0$. If $y \notin N \cup \bigcup_{j=1}^\infty N_j$, which is again a null set, we then have
$$\int h(x,y)f_j(x)\ dm_n(x) = 0$$
for all $j \in \mathbb{N}$. Since $x \mapsto h(x,y)$ is a square-integrable function of $x$ for $y \notin N \cup \bigcup_{j=1}^\infty N_j$, orthogonality to the orthonormal basis $(f_j)_{j=1}^\infty$ implies that
$$\int |h(x,y)|^2\ dm_n(x) = 0$$
for $y \notin N \cup \bigcup_{j=1}^\infty N_j$, and now by Tonelli—the integral $dm_m(y)$ ignoring the null set $N \cup \bigcup_{j=1}^\infty N_j$—we have
$$\|h\|^2_2 = \int |h(x,y)|^2\ dm_n(x) = 0$$
for $y \notin N \cup \bigcup_{j=1}^\infty N_j$. So if $h$ is orthogonal to all the $(f_j(x) \cdot g_k(y))_{j,k=1}^\infty$, then $h$ is the zero element of $L^2(\mathbb{R}^{n+m}, \text{Lebesgue, } m_{n+m})$. This shows that $(f_j(x) \cdot g_k(y))_{j,k=1}^\infty$ is an orthogonal basis of $L^2(\mathbb{R}^{n+m}, \text{Lebesgue, } m_{n+m})$, as desired.

7. Let $M \subseteq L^2([0,1], \text{Lebesgue, } m_1)$ be a (n $L^2$-norm-)closed linear subspace consisting entirely of continuous functions (i.e., classes each of which has a [necessarily unique] continuous representative). It can be shown (using the closed graph theorem) that there must exist a constant $K \geq 0$ for which $\|f\|_\infty \leq K\|f\|_2$ for all $f \in M$. Assuming this,

(a) Show that for each $x \in [0,1]$ there exists a function $g_x \in M$ such that $f(x) = \int_0^1 f(t)g_x(t)\ dt$ holds for all $f \in M$.

Note that $M$ is a closed subspace of a Hilbert space and therefore it is a Hilbert space under the relativized inner product. The linear transformation from $M$ to the scalars that sends $f$ to $f(x)$ is a bounded functional on $M$, because $|f(x)| \leq \|f\|_\infty \leq K\|f\|_2$. So by the Riesz representation theorem we find that there exists a $g_x \in M$ such that $f(x) = \langle f, g_x \rangle$ for all $f \in M$. Of course $\langle f, g_x \rangle = \int_0^1 f\overline{g_x}\ dm_1$.

(b) Show that the dimension of $M$ is at most $K^2$. Let $(f_j)_{j \in J}$ be an orthonormal set in $M$, show that
$$\sum_{j \in F} |f_j(x)|^2 \leq K^2$$
for any finite subset $F \subseteq J$, and consider the implications. The corresponding result for $p \neq 2$ is considerably more difficult.}

For this part, note that if $(f_j)_{j \in J}$ is an orthonormal base for $M$, then we have $g_x = \sum_j \langle g_x, f_j \rangle f_j$ and also $\|g_x\|^2_2 = \sum_{j \in J} |\langle g_x, f_j \rangle|^2$. But from part (a) we know that $|\langle g_x, f_j \rangle| = |\langle f, g_x \rangle| = |f_j(x)|$ and so for every finite subset $F$ of the index set $J$ we have
$$\sum_{j \in F} |f_j(x)|^2 \leq \sum_{j \in J} |f_j(x)|^2 = \|g_x\|^2_2.$$
On the other hand, $\|g_r\|_2$ is equal to the operator norm of the linear functional $f \mapsto f(x)$ and thus is less than or equal to $K$. So we have $\sum_{j \in F} |f_j(x)|^2 \leq K^2$ for all $x \in [0, 1]$. By integrating this inequality over $[0, 1]$, we get $\sum_{j \in F} \int_0^1 |f_j(x)|^2 \, dm_1 \leq K^2$. But $\{f_j\}_{j \in J}$ is an orthonormal set. $\int_0^1 |f_j(x)|^2 \, dm_1 = 1$ holds for all $j \in J$, and thus the left-hand side of that inequality is the number of elements of $F$. This shows that no finite subset of the index set can have more than $K^2$ elements, and thus the left-hand side of that inequality is the number of elements of $F$. This shows that no finite subset of the index set can have more than $K^2$ elements, and thus the left-hand side of that inequality is finite, and thus that the set $\{f_j\}_{j \in J}$ is finite, with at most $K^2$ elements. Since $\{f_j\}_{j \in J}$ is an orthonormal base for $M$, it follows that $M$ is finite-dimensional and $\dim(M) \leq K^2$.

8. Show that if $f \in \mathcal{L}^p \cap \mathcal{L}^\infty$ for some $p < \infty$ (any measure space will do), then $f \in \mathcal{L}^r$ for all $p \leq r \leq \infty$ and $\lim_{r \to \infty} \|f\|_r = \|f\|_\infty$.

For $0 < p < \infty$ we have

$$\|f\|_r = |f|^p \cdot |f|^{r-p} \leq |f|^p \cdot \|f\|_\infty^{r-p} \quad \text{and integration gives}$$

$$\|f\|_r \leq \|f\|_p \cdot \|f\|_\infty^{r-p}$$

$$\|f\|_r \leq \|f\|_p^{r/p} \cdot \|f\|_\infty^{1/(p-r)} \quad \text{(*)}$$

{NB: This is one of the “endpoint cases” of the fact that the $L^p$ norm is a logarithmically convex function of $f$: if $p < r < \infty$ then $0 = \frac{1}{r} < \frac{1}{p} < 1$ for some $0 \leq \lambda \leq 1$, so in convex-combination terms we have $\frac{1}{r} = (1-\lambda) \frac{1}{p} + \lambda \frac{1}{\infty}$, giving $\frac{p}{r} = (1-\lambda)$, $1 - \frac{p}{r} = \lambda$, and thus}

$$\|f\|_r \leq \|f\|_p^{1-\lambda} \cdot \|f\|_\infty^\lambda$$

or

$$\log \|f\|_r \leq (1-\lambda) \log \|f\|_p + \lambda \log \|f\|_\infty$$

an inequality displaying convexity. The same relation holds when both exponents are finite: if $f \in \mathcal{L}^{p_1} \cap \mathcal{L}^{p_2}$ and $\frac{1}{r} = (1-\lambda) \frac{1}{p_1} + \lambda \frac{1}{p_2}$ for some $0 \leq \lambda \leq 1$, then $1 = (1-\lambda) \frac{r}{p_1} + \lambda \frac{r}{p_2}$ (#). We may consequently write

$$|f|^r = |f|^{(1-\lambda)r} |f|^{\lambda r} = \left[|f|^{p_1}\right]^{(1-\lambda)r/p_1} \cdot \left[|f|^{p_2}\right]^{\lambda r/p_2}$$

with the first factor on the r.h.s. in $\mathcal{L}^{p_1/(1-\lambda)r}$ and the second factor in $\mathcal{L}^{p_2/\lambda r}$. The relation (#) says that these are conjugate exponents, so the Hölder inequality implies that $|f|^r$ is integrable and

$$\int |f|^r \, d\mu \leq \left[\int |f|^{p_1} \, d\mu\right]^{(1-\lambda)r/p_1} \cdot \left[\int |f|^{p_2} \, d\mu\right]^{\lambda r/p_2}$$

$$\left[\int |f|^r \, d\mu\right]^{1/r} \leq \left\{ \left[\int |f|^{p_1} \, d\mu\right]^{1/p_1} \right\}^{1-\lambda} \cdot \left\{ \left[\int |f|^{p_2} \, d\mu\right]^{1/p_2} \right\}^\lambda$$

$$\|f\|_r \leq \|f\|_1 - \lambda \cdot \|f\|_2^\lambda$$

or

$$\log \|f\|_r \leq (1-\lambda) \log \|f\|_1 + \lambda \log \|f\|_2$$

the logarithmic-convexity relation in this case.) As $r \to \infty$, the inequality (*) gives

$$\limsup_{r \to \infty} \|f\|_r \leq \|f\|_\infty$$

On the other hand, for any $0 < \epsilon < 1$ we know that the set $E_\epsilon = \{x : |f(x)| > \|f\|_\infty \cdot (1-\epsilon)\}$ has positive measure, and therefore

$$\int |f|^r \, d\mu \geq \int_{E_\epsilon} |f|^r \, d\mu \geq \|f\|_\infty \cdot (1-\epsilon)^r \cdot \mu(E_\epsilon)$$

$$\|f\|_r = \sqrt[r]{\int |f|^r \, d\mu} \geq \|f\|_\infty \cdot (1-\epsilon) \cdot \sqrt[r]{\mu(E_\epsilon)}$$

$$\liminf_{r \to \infty} \|f\|_r \geq \|f\|_\infty \cdot (1-\epsilon) \cdot 1, \quad \text{and since } \epsilon \text{ is arbitrary}$$

$$\liminf_{r \to \infty} \|f\|_r \geq \|f\|_\infty$$

Those inequalities taken together show that $\lim_{r \to \infty} \|f\|_r$ exists and equals $\|f\|_\infty$. 9
9. Let $(X, \mathfrak{M}, \mu)$ be a measure space with $\mu \geq 0$ and $\mu(X) = 1$. Clearly the unit ball $\{f : f \in L^\infty, \|f\|_\infty \leq 1 \}$ is contained in $L^1(X, \mathfrak{M}, \mu)$. Show that it is norm-compact in $L^1(X, \mu, d\mu)$ if and only if $\mu$ is purely atomic. (Show that compactness is impossible if $\mu$ is purely continuous; then split $\mu$ over the decomposition $X = X_a \cup X_c$ and observe that the pieces inherit the compactness condition. The converse is quite easy.)

While it is possible to beat this to death directly, it really has to do with the conditional expectation operators of \#4 and \#5 above. To see this one needs the following

**Lemma:** Let $(X, \| \cdot \|)$ and $(Y, \| \cdot \|)$ be normed spaces, and suppose $\{T_\alpha : \alpha \in A\} \subseteq \mathcal{L}(X, Y)$ is a uniformly norm-bounded net of operators such that $\{T_\alpha(x) : \alpha \in A\}$ converges in $Y$ for each $x \in X$. Then the convergence is uniform on compact subsets of $X$.

**Proof.** Let $T$ denote the linear operator defined by $T(x) = \lim_\alpha T_\alpha(x)$; it is obviously continuous, because if $M \geq 0$ is such that $\|T_\alpha(x)\| \leq M \|x\|$ holds for all $\alpha \in A$ then also $\|T(x)\| \leq M \|x\|$. Let $K \subseteq X$ be a compact set and let $\epsilon > 0$ be given. Let $B$ denote the open unit ball of $X$ and find a finite set of points $\{x_j\}_{j=1}^n \subseteq X$ such that $K \subseteq \bigcup_{j=1}^n \left( x_j + \frac{\epsilon}{3M} B \right)$, where again $M$ is a uniform bound for the norms of the $T_\alpha$'s. Let $\gamma \in A$ be such that $\alpha \geq \gamma \Rightarrow \|T(x_j) - T_\alpha(x_j)\| \leq \epsilon/3$ for $j = 1, \ldots, n$. Given $x \in K$, let $j$ be an index, $1 \leq j \leq n$, for which $\|x - x_j\| \leq \frac{\epsilon}{3M}$. Then $\alpha \geq \gamma$ implies

$$\|T(x) - T_\alpha(x)\| \leq \|T(x) - T(x_j)\| + \|T(x_j) - T_\alpha(x_j)\| + \|T_\alpha(x_j) - T_\alpha(x)\|$$

$$\leq M \cdot \frac{\epsilon}{3M} + \frac{\epsilon}{3} + M \cdot \frac{\epsilon}{3M} = \epsilon$$

and this holds for every $x \in K$ (the choice of $j$ does not appear in the final inequality), showing that convergence is uniform on $K$.

Now let $(X, \mathfrak{M}, \mu)$ be a measure space with $\mu \geq 0$ and $\mu(X) = 1$. If $\mathfrak{F} \subseteq \mathfrak{M}$ is a finite disjoint family of non-$\mu$-null measurable sets whose union is $X$, then the family $\mathfrak{F}$ of unions of sets belonging to $\mathfrak{F}$ is a (finite) sub-$\sigma$-algebra of $\mathfrak{M}$ and hence defines a conditional expectation operator $E[\cdot | \sigma(\mathfrak{F})]$, which—for slight convenience of notation—we denote by $T_\mathfrak{F}$. One can even write a formula for this operator, namely, if $\mathfrak{F} = \{F_1, \ldots, F_n\}$ then (as follows immediately from the uniqueness of Radon-Nikodým derivatives)

$$T_\mathfrak{F}(f) = \sum_{j=1}^n \left[ \frac{1}{\mu(F_j)} \int_{F_j} f \, d\mu \right] \chi_{F_j}.$$

The $\mathfrak{F}$’s are ordered by “refinement”: if $\mathfrak{F}$ and $\mathfrak{G}$ are two such families, then $\mathfrak{G}$ is finer than $\mathfrak{F}$ if each $G \in \mathfrak{G}$ is a subset of some $F \in \mathfrak{F}$. When $\mathfrak{G}$ is finer than $\mathfrak{F}$, each element of $\mathfrak{F}$ must equal the union of the sets of $\mathfrak{G}$ that it contains (proof easy), and $T_\mathfrak{G}T_\mathfrak{F} = T_\mathfrak{G}$. The operators $\{T_\mathfrak{F}\}$ thus form a net indexed by the $\mathfrak{F}$’s ordered by refinement; on each space $L^p(X, \mathfrak{M}, \mu)$, they are operators of norm 1. For each $f \in L^p(X, \mathfrak{M}, \mu)$ we have $\lim_\mathfrak{F} T_\mathfrak{F} f = f$ in the $L^p$ norm. This is obvious for simple functions, because if $f = \sum_{j=1}^n a_j \chi_{F_j}$, then $T_\mathfrak{F} f = f$; it follows by an easy approximation-by-simple-functions argument (which the reader may provide) that $\lim_\mathfrak{F} T_\mathfrak{F} f = f$ in $L^p$ norm for all $f \in L^p$.

Now by the lemma, if the $L^\infty$ unit ball is a compact subset of $L^1$, then the $L^1$-norm limit $\lim_\mathfrak{F} T_\mathfrak{F} f = f$ is attained uniformly on it. However, this uniformity is impossible if $(X, \mathfrak{M}, \mu)$ is not purely atomic. The reason is that if the decomposition $X = X_a \cup X_c$ does not give $\mu(X_c) = 0$, then for every $\mathfrak{F} = \{F_j\}_{j=1}^n$ not

---

(4) This is actually a special case of the half of the Arzelà-Ascoli theorem that says: if a sequence (or net) of continuous functions is equicontinuous, then if it converges pointwise it converges uniformly on compacta.
all the sets of the family \( \{ F_j \cap X_c \}_{j=1}^n \) are \( \mu \)-null. Each of those non-\( \mu \)-null subsets of \( X_c \) can be written as a disjoint union \( F_j \cap X_c = F_{j,0} \cup F_{j,1} \) of two \( \mathcal{M} \)-measurable sets, each of measure \( \frac{1}{2} \mu(F_j \cap X_c) \). Consider the function defined by

\[ f = \sum_j (\chi_{F_{j,0}} - \chi_{F_{j,1}}) \]

where the sum is extended over all the \( 1 \leq j \leq n \) for which \( \mu(F_j \cap X_c) \neq 0 \). In all cases \( \| f \|_1 = \mu(X_c) \), because \( X_c = \bigcup_{j=1}^n (F_j \cap X_c) \) as a disjoint union; and because at least one of the \( \mu(F_j \cap X_c) \)'s was nonzero, \( \| f \|_\infty = 1 \). However, \( \int_{F_j} f \, d\mu = 0 \) for each \( 1 \leq j \leq n \) because the measures of \( F_{j,0} \) and \( F_{j,1} \) are equal, and therefore

\[ T_\delta(f) = \sum_{j=1}^n \left[ \frac{1}{\mu(F_j)} \int_{F_j} f \, d\mu \right] \chi_{F_j} = 0. \]

No matter how fine \( \delta \) is, then, we can exhibit a function \( f \) with \( \| f \|_\infty = 1 \) but \( \| f - T_\delta f \|_1 = \| f \|_1 = \mu(X_c) > 0 \). This contradicts the compactness of the \( L^\infty \) unit ball: by the lemma, if the \( L^\infty \) unit ball is a compact subset of \( L^1 \), then the \( L^1 \)-norm limit \( \lim_\delta T_\delta f = f \) is attained uniformly on it. So if the \( L^\infty \) unit ball is \( L^1 \)-compact, then \( \mu(X_c) = 0 \) must hold and \( (X, \mathcal{M}, \mu) \) must be purely atomic.

The converse is easy; only formalizing it is tedious. The details go as follows: if \( (X, \mathcal{M}, \mu) \) is purely atomic, let \( \{ E_j \}_{j=1}^{n \ or \ \infty} \subseteq \mathcal{M} \) be a maximal disjoint family of \( \mathcal{M} \)-measurable \( \mu \)-atoms; such a family exists by Zorn’s lemma, it is at most countably infinite, and the complement of its union is a \( \mu \)-null \( \mathcal{M} \)-measurable set (all this by a familiar argument: cf. the solution of 19 below). Essentially the same argument shows that for each \( E \in \mathcal{M} \) the set \( \bigcup \{ E_j : \mu(E \cap E_j) > 0 \} \) differs from \( E \) by a \( \mu \)-null set. Equivalently, the function

\[ \sum_{j=1}^{n \ or \ \infty} \left[ \frac{1}{\mu(E_j)} \int_{E_j} \chi_{E} \, d\mu \right] \chi_{E_j} \]

differs from \( \chi_E \) by a null function. The indicated series (assuming it is infinite—if there are only \( n \) of the \( E_j \)'s, there are no convergence questions) converges pointwise \( \mu \)-a.e. and also converges in \( L^p \) norm for \( 1 \leq p < \infty \), since

\[ \left\| \sum_{j=m}^{n} \left[ \frac{1}{\mu(E_j)} \int_{E_j} \chi_{E} \, d\mu \right] \chi_{E_j} \right\|_p = \left\| \chi_E \bigcup_{j>m} E_j \right\|_p = \sqrt{\mu(E \cap \bigcup_{j>m} E_j)} \to 0 \quad \text{as} \quad m \to \infty. \]

Because the simple functions are norm-dense in each \( L^p \), \( 1 \leq p \leq \infty \), it follows that for each \( f \in L^p \)

\[ f = \sum_{j=1}^{n \ or \ \infty} \left[ \frac{1}{\mu(E_j)} \int_{E_j} f \, d\mu \right] \chi_{E_j} \]

holds in the senses of convergence in \( L^p \) norm for \( 1 \leq p < \infty \) and also pointwise \( \mu \)-a.e. for \( 1 \leq p \leq \infty \). To each \( f \in L^p \), \( 1 \leq p \leq \infty \), there thus corresponds a sequence of scalars (indexed by \( 1 \leq j \leq n \) or by \( j \in \mathbb{N} \))

\[ f \mapsto \left( \ldots, \frac{1}{\mu(E_j)} \int_{E_j} f \, d\mu, \ldots \right), \]

and the correspondence is 1-1 between \( \mu \)-a.e. equivalence classes and sequences. It is evident that for \( f \in L^\infty(\mu) \) the corresponding sequence is bounded by \( \| f \|_\infty \). Consequently, from any sequence \( \{ f_k \}_{k=1}^{\infty} \) in the unit ball of \( L^\infty \) one can select (by the first Cantor diagonal process or the Tihonov theorem) a subsequence \( \{ f_{k_r} \}_{r=1}^{\infty} \) such that for each \( j \) the “\( j \)-th coefficient sequence” \( \left\{ \frac{1}{\mu(E_j)} \int_{E_j} f_{k_r} \, d\mu \right\}_{r=1}^{\infty} \) converges,
say to \( \beta_j \) with \( |\beta_j| \leq 1 \). But we then have \( \lim_{r \to \infty} f_{k_r} = \sum_{j=1}^{n \text{ or } \infty} \beta_j \chi_{E_j} \) in \( L^1 \) norm, for given any \( \epsilon > 0 \) we may find \( m \) for which \( \mu \left( \bigcup_{j \geq m} E_j \right) < \frac{\epsilon}{4} \), whereupon

\[
\left\| f_{k_r} - \sum_{j=1}^{n \text{ or } \infty} \beta_j \chi_{E_j} \right\|_1 \leq \sum_{j=1}^{m} \left| \frac{1}{\mu(E_j)} \int_{E_j} f_{k_r} \, d\mu - \beta_j \right| \mu(E_j) + \sum_{j=m}^{\infty} 2 \mu(E_j)
\]

and since the finite sum on the last set-off line above can be made < \( \epsilon/2 \) by taking \( r \) sufficiently large, the \( L^1 \) limit is as advertised. Since the sequence \( \{f_{k_r}\}_{r=1}^{\infty} \) in the unit ball of \( L^\infty \) was arbitrary, this shows that the unit ball of \( L^\infty \) is norm-compact as a subset of \( L^1 \), as desired.

\{NB: The treatment just given is rather “manual.” Clearly what is going on is the following: when \( (X, \mathcal{M}, \mu) \) is (totally finite and) purely atomic, the family \( \mathcal{M}_0 \) of unions of the \( \{E_j\}_{j=1}^{n \text{ or } \infty} \subseteq \mathcal{M} \) is a sub-\( \sigma \)-algebra of \( \mathcal{M} \) that is “\( \mu \)-essentially the whole thing;” every \( \mathcal{M} \)-measurable function equals an \( \mathcal{M}_0 \)-measurable function up to a \( \mu \)-null set. The mapping

\[
f \mapsto \sum_{j=1}^{n \text{ or } \infty} \left[ \frac{1}{\mu(E_j)} \int_{E_j} f \, d\mu \right] \chi_{E_j}
\]

is the conditional expectation mapping \( E[f | \mathcal{M}_0] \), but it’s also the identity mapping on the spaces \( L^p(X, \mathcal{M}, \mu) \).

The partial sums \( f \mapsto \sum_{j=1}^{M} \left[ \frac{1}{\mu(E_j)} \int_{E_j} f \, d\mu \right] \chi_{E_j} \) give bounded linear operators of finite rank that have operator norm 1 and converge pointwise to the identity operator on \( L^p \), \( 1 \leq p < \infty \) (they also converge to the identity on \( L^\infty \), but only in the senses of pointwise \( \mu \)-a.e. convergence and weak* convergence in the dual of \( L^1 \) [whatever that means]). The easily-checked fact that these finite-rank, uniformly bounded operators converge to the identity uniformly in \( L^1 \) norm on the unit ball \( B \) of \( L^\infty \) says that \( B \) has the following property as a subset of \( L^1 \): for every \( \epsilon > 0 \) one can find a compact set \( K_\epsilon \) such that every point of \( B \) is within \( \epsilon \) norm-distance of some point of \( K_\epsilon \) (in this case the \( K_\epsilon \) will just be the image of \( B \) under one of those partial sum operators, provided \( M \) is sufficiently large). Sets that have this “compact approximation” property are easily seen to be relatively compact themselves (let \( \epsilon > 0 \) be given, cover \( K_{\epsilon/2} \) with finitely many balls of radius \( \epsilon/2 \) and the balls with the same centers but radius \( \epsilon \) will cover \( B \)). So with sufficient preparation, one can make the converse proposition (purely atomic \( \Rightarrow \) \( B \) compact in \( L^1 \)) look a bit more civilized. It’s the direct proposition (\( B \) compact in \( L^1 \) \( \Rightarrow \) \( L^1(X, \mathcal{M}, \mu) \) purely atomic) that has turned out to have interesting consequences.\}

10. Show that \( \ell^\infty(\mathbb{N}) \) is not separable. Then show that it is possible to imbed copies of \( \ell^\infty(\mathbb{N}) \) in \( L^\infty(\mathbb{R}^n, \text{Lebesgue}, m_n) \)—by norm-preserving 1-1 isomorphisms of the algebra structure—and deduce that the latter space is not separable. Generalize this to abstract measure spaces to the extent that you can.

If \( S, T \subseteq \mathbb{N} \) are distinct subsets of \( \mathbb{N} \) then \( \| \chi_S - \chi_T \|_{\infty} = 1 \), and so the open balls of radius 1/3 (say) centered on the points \( \{ \chi_S : S \in 2^\mathbb{N} \} \) have (pairwise) empty intersections. It is therefore impossible to find a countable subset of \( \ell^\infty(\mathbb{N}) \) that has at least one point in common with each of these (uncountably many) balls, and so \( \ell^\infty(\mathbb{N}) \) cannot be separable. If \( (X, \mathcal{M}, \mu) \) is a measure space such that there exists a sequence
of pairwise disjoint non-$\mu$-null sets \( \{S_n\}_{n=1}^{\infty} \subseteq \mathcal{M} \), then it is routine to verify that the mapping which one can indicate formally as
\[
\ell^\infty(\mathbb{N}) \rightarrow L^\infty(X, \mathcal{M}, \mu)
\]
\[
(\alpha_1, \alpha_2, \ldots) \mapsto \sum_{n=1}^{\infty} \alpha_n \chi_{S_n}
\]
(the indicated sum has at most one nonzero term at each point \( x \in X \)) is an isometric isomorphism of \( \ell^\infty(\mathbb{N}) \) into \( L^\infty(X, \mathcal{M}, \mu) \). Since separability of metric spaces is inherited by subspaces, \( L^\infty(X, \mathcal{M}, \mu) \) cannot be separable. It is trivial to construct such a sequence of disjoint sets for Lebesgue measure on \( \mathbb{R}^n \); e.g., one may take the ball of radius 1/3 and center 0 \( \in \mathbb{R}^n \) and consider the set of its translates by integer multiples of one of the standard unit basis vectors of \( \mathbb{R}^n \).

A less-tautological sufficient condition for the non-separability of \( L^\infty(X, \mathcal{M}, \mu) \) would be that at least one of the spaces \( L^p(X, \mathcal{M}, \mu) \) be infinite-dimensional, \( 1 \leq p < \infty \). For if this condition holds, then—because the simple functions are dense in \( L^p \) for \( 1 \leq p < \infty \)—the simple functions must also form an infinite-dimensional linear space.\(^{(5)}\) It is tedious but elementary to start with a sequence of linearly independent simple functions \( (5) \) and integration is justified by the monotone convergence theorem (the fact that the terms are nonnegative makes the partial sums of the series form a monotone sequence).

11. The gamma integral is defined for \( \text{Re}[s] > 0 \) by \( \Gamma(s) = \int_0^\infty t^{s-1} e^{-t} \, dt \) and the Riemann zeta function for \( \text{Re}[s] > 1 \) by \( \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \). Give a rigorous demonstration of the relation
\[
\Gamma(s) \zeta(s) = \int_0^\infty \frac{x^{s-1}}{e^x - 1} \, dx
\]
for \( \text{Re}[s] > 1 \), using appropriate convergence theorems (assume \( s \in \mathbb{R} \) if you wish).

Assuming \( 1 < s \in \mathbb{R} \) we have
\[
\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} \, dt = \int_0^\infty n^{s-1} x^{s-1} e^{-nx} \, n \, dx \quad \text{(with } t = nx)\]
\[
\frac{1}{n^s} = \frac{1}{\Gamma(s)} \int_0^\infty x^{s-1} e^{-nx} \, dx
\]
\[
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{\Gamma(s)} \sum_{n=1}^{\infty} \int_0^\infty x^{s-1} e^{-nx} \, dx = \frac{1}{\Gamma(s)} \int_0^\infty x^{s-1} \left[ \sum_{n=1}^{\infty} e^{-nx} \right] \, dx
\]
\[
= \frac{1}{\Gamma(s)} \int_0^\infty x^{s-1} \frac{e^{-x}}{1 - e^{-x}} \, dx = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1}}{e^x - 1} \, dx
\]
where, since everything in sight is nonnegative and the integral on the last line converges by comparison with \( 1/x^{s-2} \) at the \( x \to 0^+ \) end and by comparison with \( e^{-x/2} \) at the \( x \to \infty \) end, the interchange of summation and integration is justified by the monotone convergence theorem (the fact that the terms are nonnegative makes the partial sums of the series form a monotone sequence).

There are various ways to extend this equality to the half-plane \( \text{Re}[s] > 1 \). The classical way is to justify differentiation of the series on the l. h. side and the integral on the r. h. side with respect to the complex variable \( s \) by making estimates on the difference quotients and using the Lebesgue dominated

---

\(^{(5)}\) It is fairly easy to show that every finite-dimensional normed space over \( \mathbb{R} \) or \( \mathbb{C} \) is isomorphic, with uniform continuity both ways, to \( \mathbb{R}^k \) or \( \mathbb{C}^k \) for some \( k \in \mathbb{N} \). It is therefore complete (as a metric space), and if it is a subspace of some larger normed space, it must therefore be a closed subspace of that bigger space. It follows that if the space of simple functions were both dense and finite-dimensional, it would be all of the \( L^p \)-space, which would therefore be finite-dimensional.
convergence theorem. Once one knows that the expressions on the two sides are holomorphic functions of $s$ in the half-plane, their equality for all $\text{Re } [s] > 1$ follows from the identity theorem for holomorphic functions. One way of seeing that both sides are holomorphic functions of $s$ that is particularly appealing in the context of a course in integration is the following. If $\gamma$ is a small circle in the complex plane and $f$ is (say) a continuous function defined on $\gamma$, then the function $s \mapsto \frac{1}{2\pi i} \int_{\gamma} \frac{f(\zeta)}{\zeta - s} d\zeta$ defined by the Cauchy integral formula is holomorphic in the disc of which $\gamma$ is the boundary, without any restriction on the choice of $f$ (of course one can’t expect this formula to reproduce $f$ except under the usual hypotheses of the Cauchy integral formula). The curvilinear integral can be construed as a Lebesgue integral—simply parametrize $\gamma$ in the context of a course in integration is the following. If $\gamma$ is contained in the half-plane $\text{Re } [s] > 1$, then the Fubini-Tonelli theorems and the Cauchy integral theorem allow us to write

$$\frac{1}{2\pi i} \int_{\gamma} \left[ \int_{0}^{\infty} \frac{x^{\zeta-1}}{e^x - 1} dx \right] \frac{d\zeta}{\zeta - s} = \int_{0}^{\infty} \left[ \frac{1}{2\pi i} \int_{\gamma} \frac{x^{\zeta-1}}{e^x - 1} \frac{d\zeta}{\zeta - s} \right] dx = \int_{0}^{\infty} \frac{x^{s-1}}{e^x - 1} dx$$

(of course $x^{s-1}$ means $e^{(s-1) \log x}$ where $\log x$ is the real logarithm). The l. h. expression is a holomorphic function of $s$ in the disc of which $\gamma$ is the boundary, so the integral on the r. h. side is a holomorphic function of $s$ there, and since holomorphy is a local property, the integral is holomorphic throughout the half-plane $\text{Re } [s] > 1$. (The same argument works on the series, although because the series converges uniformly in any half-plane $\text{Re } [s] \geq 1 + \delta$, nothing as heavy-duty as the Fubini theorem needs to be employed.)

12. (a) Use the addition formula for the cosine to show that the linear space spanned by the functions \{cos$n\theta$\}$^\infty_{n=0}$ is an algebra, and deduce (from what double-barrelled theorem?) that it is dense in $C[0, \pi]$ (in the uniform norm). Thus deduce that if these functions are suitably normalized, they provide an orthonormal basis of $L^2[0, \pi]$ (with the usual $L^2$ inner product and norm).

It is routine to see that for the linear span of a set of functions to form an algebra, it is sufficient to show that the product of any two elements of the spanning set again belongs to the linear span. In view of the relations

$$\cos A \cos B = \cos(A + B) - \sin A \sin B$$
$$\cos A \cos B + \sin A \sin B = \cos(A - B)$$
$$\cos A \cos B = \frac{\cos(A + B) + \cos(A - B)}{2}$$
$$\cos m\theta \cos n\theta = \frac{\cos(m+n)\theta + \cos(m-n)\theta}{2}$$

that condition is satisfied, and the linear span of the cosines is thus an algebra. (If one is using complex scalars, one notes that it is an algebra closed under conjugation.) Since $\cos \theta$ separates points of $[0, \pi]$ all by itself and $1 \equiv \cos 0$, the Stone-Weierstraß theorem implies that these “cosine polynomials” form a uniformly dense subalgebra of $C([0, \pi])$, and therefore also a dense subspace of $L^2[0, \pi]$ (continuous functions are dense in $L^2$ and uniform convergence implies $L^2$-norm convergence).

The cosines form an orthogonal set in $L^2[0, \pi]$, again by virtue of the addition formula: for $m \neq n$ we have

$$\int_{0}^{\pi} \cos m\theta \cos n\theta \, d\theta = \frac{1}{2} \int_{0}^{\pi} \left[ \cos(m+n)\theta + \cos(m-n)\theta \right] \, d\theta = \frac{1}{2} \left[ \frac{\sin(m+n)\theta}{m+n} + \frac{\sin(m-n)\theta}{m-n} \right]_{0}^{\pi} = 0$$

since $\sin k\pi = 0$ for any $k \in \mathbb{Z}$. Freshman calculus shows that $\|1\|_2^2 = \pi$ and $\|\cos n\theta\|_2^2 = \pi/2$ for $n \geq 1$, giving the normalizing constants. The normalized cosines are thus $\sqrt{\frac{1}{\pi}}, \sqrt{\frac{2}{\pi}} \cos \theta, \sqrt{\frac{2}{\pi}} \cos 2\theta, \ldots$.

(b) It is not difficult to show that for each $n \in \{0\} \cup \mathbb{N}$ there is a polynomial $T_n(x)$ of degree $n$ with the property $T_n(\cos \theta) = \cos(n\theta)$; assume this. Then use the substitution $x = \cos \theta$ and freshman calculus on Riemann integrals to show that the $\{T_n(x)\}^\infty_{n=0}$, suitably normalized, form an orthonormal basis of $L^2([-1,1], \text{Lebesgue}, \frac{dx}{\sqrt{1-x^2}})$.
Since the means are at hand to show that these polynomials \( T_n(x) \) (the famous Chebyshev polynomials) exist, we may as well do it: the addition formula for the cosine also gives

\[
\begin{align*}
\cos(n+1)\theta &= \cos n\theta \cos \theta - \sin n\theta \sin \theta \\
\cos(n-1)\theta &= \cos n\theta \cos \theta + \sin n\theta \sin \theta \\
\cos(n+1)\theta &= 2\cos n\theta \cos \theta - \cos(n-1)\theta .
\end{align*}
\]

This tells us that if we define polynomials recursively by \( T_0(x) = 1, T_1(x) = x \) and \( T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x) \) for \( n \geq 1 \), then a simple induction (unkindly left to the reader) will show that \( T_n(\cos \theta) \equiv \cos n\theta \). In the course of that induction one sees that deg \( T_n(x) \) equals all the polynomials, which makes it uniformly dense in \( C([-1,1]) \) and therefore norm-dense in \( L^2([-1,1], \text{Lebesgue}, \frac{dx}{\sqrt{1-x^2}}) \). The classical theorem on integration by substitution in Riemann integrals also tells us that with \( x = \cos \theta, \ dx = -\sin \theta \ d\theta \) we have

\[
\int_{-1}^{1} T_m(x)T_n(x) \frac{dx}{\sqrt{1-x^2}} = -\int_{0}^{\pi} \cos m\theta \cos n\theta \frac{-\sin \theta \ d\theta}{\sqrt{1-\cos^2 \theta}} = \int_{0}^{\pi} \cos m\theta \cos n\theta \ d\theta
\]

so for \( m \neq n \) we have \( T_m \perp T_n \) and for \( m = n \) we have the same normalizing constants for the \( T_n \)'s as for the \( \cos n\theta \)'s. Up to normalization we thus have an orthonormal basis for the space \( L^2([-1,1], \text{Lebesgue}, \frac{dx}{\sqrt{1-x^2}}) \).

13. Use complex exponentials (or any means at your disposal) to compute the sum

\[
D_N(t) = \sum_{n=-N}^{N} e^{int}
\]

in “closed form.” Thus compute a closed-form formula, as an integral, for the partial sum \( \sum_{n=-N}^{N} \langle f, e^{in\theta} \rangle e^{in\theta} \) of the Fourier series of a function \( f \) on \([-\pi, \pi]\) (assumed to be at least integrable, and extended to be periodic on \( \mathbb{R} \) to simplify writing the integral in question).

The sum defining \( D_N \) is, essentially, a finite geometric series:

\[
D_N(t) = \sum_{n=-N}^{N} e^{int} = e^{-iNt} \sum_{n=0}^{2N} e^{int} = \frac{e^{-i(2N+1)t}}{1 - e^{it}}
\]

\[
= e^{-iNt} \cdot \frac{e^{i(N+1/2)t} - e^{-i(N+1/2)t}}{e^{it/2} - e^{-it/2}}
\]

\[
= \sin \left( N + \frac{1}{2} \right) \frac{t}{2} .
\]

Thus

\[
\sum_{n=-N}^{N} \langle f, e^{in\theta} \rangle e^{in\theta} = \sum_{n=-N}^{N} \left[ \int f(t) e^{-i\theta \, dt} \right] e^{i\theta \, dt} = \int f(t) \left[ \sum_{n=-N}^{N} e^{in(\theta-t)} \right] dt
\]

\[
= \int f(t) \frac{\sin \left( N + \frac{1}{2} \right) (\theta-t)}{\sin \left( \frac{\theta-t}{2} \right)} \, dt ,
\]

Math 501:01 Fall 1999
where the integral is extended over any interval of length $2\pi$ (all the functions involved being periodic). Note that if one thinks of the integration as being carried out over the group $\mathbb{R}/2\pi\mathbb{Z}$—the “circle group”—the integral has the form of a convolution. The fact that the Dirichlet kernel $D_N(t)$ is not everywhere positive has given entertainment to analysts through the centuries.

14. Find the Fourier coefficients of the function $f(\theta) = \theta$ on $[-\pi, \pi]$, and use a Hilbert-space theorem to deduce from what you have found the value of $\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2}$.

Using complex exponentials and normalized Lebesgue measure on $[-\pi, \pi]$, we get the Fourier coefficients $c_n = 0$ (by symmetry) and

\[ c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta e^{-in\theta} d\theta = \frac{1}{2\pi} \left\{ \left( \frac{e^{-in\theta}}{-i} \right) \right\}^{\pi}_{-\pi} + \frac{1}{in} \int_{-\pi}^{\pi} e^{-in\theta} d\theta = \frac{\pi e^{-in\pi}}{-2\pi in} = \frac{i \cos n\pi}{n} \]

for $0 \neq n \in \mathbb{Z}$. Parseval’s equality now gives

\[ 2 \sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n \neq 0 \in \mathbb{Z}} \left| \frac{i \cos n\pi}{n} \right|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta^2 d\theta 
\]

\[ 2\zeta(2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta^2 d\theta = \frac{\pi^2}{3} \]

\[ \zeta(2) = \frac{\pi^2}{6}. \]

15. Let $\mu$ be a finite real measure on $(X, \mathcal{M})$. Show that if $\mu(X) = |\mu|(X)$, then one must have $\mu \geq 0$. Does the same implication hold for complex-valued $\mu$ (give a proof or a counterexample)?

In the real case, since the measure $|\mu| - \mu$ is a nonnegative measure and $|\mu| - \mu(X) = 0$, it is the zero measure and $\mu = |\mu|$. In the complex case one starts with $Re |\mu|(X) + i Im |\mu|(X) = |\mu|(X)$. Taking real parts of both sides gives $Re |\mu|(X) = |\mu|(X)$; but since the absolute value of $Re |\mu|$ is the smallest nonnegative measure dominating the absolute values of the values of $Re |\mu|$, one has $Re |\mu|(X) \leq |Re |\mu|||X) \leq |\mu|(X) = Re |\mu|(X)$.

This implies two things: (1) $Re |\mu|$ is a nonnegative measure, and (2) because $|\mu| - Re |\mu|$ is a nonnegative measure but $|\mu| - Re |\mu|(X) = 0$, it is the zero measure and therefore $|\mu| = Re |\mu|$. For any $E \in \mathcal{M}$ we now have

\[ Re |\mu|(E)^2 = |\mu|(E)^2 \geq |\mu|(E)^2 = |Re |\mu|(E)^2| + |Im |\mu|(E)^2| 
\]

\[ 0 \geq |Im |\mu|(E)^2| \geq 0 \]

so $Im |\mu| = 0$. Thus $\mu(X) = |\mu|(X) \implies \mu = |\mu|$ holds in the complex case also.

16. Compute \( \lim_{n \to \infty} \int_{0}^{\infty} \left( 1 + \frac{t}{n} \right)^{-n} \sin \left( \frac{t}{n} \right) dt \), justifying the calculations (by appeal to suitable convergence theorems, or otherwise).

Freshman-calculus methods suffice to attack this limit: after the substitution $t = nx$, $dt = n dx$,

\[ \int_{0}^{\infty} \left( 1 + \frac{t}{n} \right)^{-n} \sin \left( \frac{t}{n} \right) dt = n \int_{0}^{\infty} \frac{1}{(1+x)^n} \sin x dx, \]

two integrations by parts give (for $n \geq 4$)

\[ n \int_{0}^{\infty} \frac{\sin x}{(1+x)^n} dx = \frac{n}{-n+1} \left\{ \frac{\sin x}{(1+x)^{n-1}} \right\}_{0}^{\infty} - \int_{0}^{\infty} \frac{\cos x}{(1+x)^{n-1}} dx \]

\[ = \frac{n}{-n+1} \left\{ \frac{\sin x}{(1+x)^{n-1}} - \frac{1}{-n+2} \int_{0}^{\infty} \frac{\cos x}{(1+x)^{n-2}} dx \right\} \]

\[ = \frac{n}{-n+1} \left\{ \frac{1}{-n+2} - \frac{1}{-n+2} \int_{0}^{\infty} \frac{\sin x}{(1+x)^{n-2}} dx \right\} \]
and since that last remaining integral clearly remains bounded as \( n \to \infty \), this expression has limit zero. In fact, as a function of \( n \) the integral has the form of a Laplace transform: the substitution \( 1 + x = e^t \), \( dx = e^t \, dt \) rewrites it as

\[
n \int_0^\infty \frac{\sin x}{(1 + x)^n} \, dx = \int_0^\infty e^{-nt} \left[ \sin(e^t - 1) e^t \right] \, dt
\]

and since the expression in square brackets has a convergent Maclaurin series whose first few terms are given by \( \sin(e^t - 1) e^t = t + \frac{3}{2} t^2 + \frac{5}{24} t^4 + O(t^5) \), Watson’s lemma gives the asymptotic expansion

\[
n \int_0^\infty e^{-nt} \left[ \sin(e^t - 1) e^t \right] \, dt = n \left[ \frac{1}{n^2} + \frac{3}{2} \frac{2!}{n^3} + \frac{3!}{n^4} + \frac{5}{24} \frac{4!}{n^5} + O \left( \frac{1}{n^6} \right) \right]
\]

\[
\int_0^\infty \left( 1 + \frac{t}{n} \right)^n \sin \left( \frac{t}{n} \right) \, dt = \frac{1}{n} + \frac{3}{n^2} + \frac{6}{n^3} + \frac{5}{n^4} + O \left( \frac{1}{n^5} \right)
\]

for those who are picky. In the context of an course in integration, one probably ought to take the approach that, since it is clear that the pointwise limit of the integrand is zero, the Lebesgue dominated-convergence theorem could be made to play a part. For this one needs a “dominating function.” The elementary calculation

\[
\left( 1 + \frac{t}{n} \right)^n = \sum_{k=0}^{\infty} \binom{n}{k} \left( \frac{t}{n} \right)^k \geq 1 + n \frac{t}{n} + \frac{n(n-1)}{2} \frac{t^2}{n^2}
\]

implies that for \( n \geq 2 \) one has

\[
\left( 1 + \frac{t}{n} \right)^n \geq 1 + t + \frac{n-1}{n} t^2 > 1 + t + \frac{t^2}{4} = \left( 1 + \frac{t}{2} \right)^2
\]

and thus for all \( n \geq 2 \) we have \( \left( 1 + \frac{t}{n} \right)^{-n} < \left( 1 + \frac{t}{2} \right)^2 \in L^1(\mathbb{R}^+) \). So this function can be used as the dominating function in an application of the Lebesgue dominated convergence theorem, showing again that the limit of the integral as \( n \to \infty \) is zero (the integral of the pointwise limit).

17. Suppose you know how to construct Cantor sets of arbitrary measure in \([0, 1]\). Exhibit (a) a subset of \([0, 1]\) that is a dense \( G_\delta \) of measure 0; (b) a subset of \([0, 1]\) that is of first category but measure 1. Can you exhibit a closed nowhere dense subset of \([0, 1]\) that is of measure 1?

(a) This doesn’t require Cantor sets: take a dense, countable subset \( S \subset [0, 1] \) (e.g., its rational elements) and for each \( n \in \mathbb{N} \) let \( U_n \) be an open set with \( S \subseteq U_n \) and \( m_1(U_n) < 1/n \). Evidently each \( U_n \) is dense, so \( T = \bigcap_{n=1}^{\infty} U_n \) is a dense \( G_\delta \) and \( m_1(T) = 0 \). (b) doesn’t require Cantor sets either, since it is basically the same as (a) via complementation: the sets \( F_n = [0, 1] \setminus U_n \) are closed and nowhere dense (since their open complements are dense), so \([0, 1] \setminus T = \bigcup_{n=1}^{\infty} ([0, 1] \setminus U_n) \) is of first category, and as the complement of a null set it has measure 1. One can do (b) via Cantor sets if one wishes: for each \( n \in \mathbb{N} \) let \( C_n \subseteq [0, 1] \) be a Cantor set with \( m_1(C_n) \geq 1 - \frac{1}{n} \), and let \( C = \bigcup_{n=1}^{\infty} C_n \). Since each \( C_n \) is nowhere dense, \( C \) is of first category, but also \( m_1(C) = 1 \). Complementing this would also give a dense \( G_\delta \) of measure zero. To exhibit (c) a closed nowhere-dense subset of \([0, 1]\) is impossible: the complement of a closed nowhere-dense set is nonempty and open and the Lebesgue measure of a nonempty open subset of \([0, 1]\) must be positive; consequently, a nowhere-dense closed subset of \([0, 1]\) must have measure strictly < 1.

18. Assume what you know about the effect of nonsingular linear transformations of \( \mathbb{R}^n \) on \( n \)-dimensional Lebesgue measure and measurability of subsets of \( \mathbb{R}^n \). (a) Let \( N \subseteq \mathbb{R} \) be a set of Lebesgue measure zero. Show that \( N \times \mathbb{R} \subseteq \mathbb{R}^2 \) is a set of 2-dimensional Lebesgue measure zero. (b) Let \( N \subseteq \mathbb{R} \) be
a set of Lebesgue measure zero. Show that \{ (x, y) \in \mathbb{R}^2 : x - y \in N \} is a set of 2-dimensional Lebesgue measure zero.

The easiest way to do (a), since the Tonelli theorem is available, is to iteratedly integrate the characteristic function \( \chi_{N \times \mathbb{R}}(x, y) = \chi_N(x) \), obtaining

\[
m_2(N \times \mathbb{R}) = \int \int \chi_{N \times \mathbb{R}}(x, y) \, d(x \times y) = \int \left( \int \chi_N(x) \, dx \right) \, dy = \int \int_0 dy = 0.
\]

A direct covering argument also works: write \( \mathbb{R} \) as a union of sets of finite measure, e.g., the intervals \((-k, k)\) for \( k \in \mathbb{N} \), and for given \( \epsilon > 0 \) let \( U_k \supseteq N \) be an open set of measure \( < \frac{\epsilon}{k^{2k+2}} \) containing \( N \); then \( \bigcup_{k=1}^{\infty} (U_k \times (-k, k)) \supseteq N \times \mathbb{R} \) is an open set in \( \mathbb{R}^2 \) of measure \( < \epsilon \), and since \( \epsilon > 0 \) is arbitrary, \( m_2(N \times \mathbb{R}) = 0 \).

For (b) one observes that a 45° rotation and suitable dilation of \( \mathbb{R}^2 \) will send the set \{ (x, y) \in \mathbb{R}^2 : x - y \in N \} to \( N \times \mathbb{R} \), and since both those linear transformations preserve measurability and the dilation only multiplies 2-dimensional measures by its determinant, \{ (x, y) \in \mathbb{R}^2 : x - y \in N \} must also be a (measurable) set of measure zero.

19. (a) There are situations in which one encounters disjoint families \( \mathfrak{A} = \{ E_\alpha \}_{\alpha \in A} \) of measurable subsets of \( \mathbb{R}^n \) with \( 0 < m(E_\alpha) < \infty \) for each \( \alpha \in A \) that are **maximal** with respect to those properties, i.e., there is no properly larger disjoint family of measurable sets of finite positive measure that contains \( \mathfrak{A} \). Show (1) that such a family \( \mathfrak{A} \) is necessarily countable; (2) that \( \mathbb{R}^n = [\bigcup_{\alpha \in A} E_\alpha] \cup N \), where \( m(N) = 0 \). (Use what you know about Lebesgue-measurable sets.)

Let \( F \subseteq \mathbb{R}^n \) be a measurable set with \( 0 < m(F) < \infty \) and consider the sets \( E_\alpha \cap F \) for \( \alpha \in A \). Because the \( E_\alpha \)'s are disjoint, for each \( p \) in \( \mathbb{N} \) at most \( p \) of these sets can have measure \( \geq \frac{1}{p} m(F) \); therefore the set of indices \( A_F = \{ \alpha \in A : m(E_\alpha \cap F) > 0 \} \) is countable. If \( \{ F_k \}_{k=1}^{\infty} \) is an increasing sequence of measurable sets with \( 0 < m(F_k) < \infty \) and \( \mathbb{R}^n = \bigcup_{k=1}^{\infty} F_k \), then the union \( \bigcup_{k=1}^{\infty} \{ \alpha \in A : m(E_\alpha \cap F_k) > 0 \} \) of the corresponding \( A_{F_k} \)'s is countable. However, this must be all of the index set \( A \), because if \( \alpha \notin \bigcup_{k=1}^{\infty} \{ \alpha \in A : m(E_\alpha \cap F_k) > 0 \} \) then \( m(E_\alpha \cap F_k) = 0 \) for each \( k \) in \( \mathbb{N} \) and therefore \( m(E_\alpha) = \lim_{k \to \infty} m(E_\alpha \cap F_k) = 0 \), contrary to the sets of \( \mathfrak{A} \) having positive measure. This proves (1): \( \mathfrak{A} \) is countable. It is thus legitimate to form \( \bigcup_{\alpha \in A} E_\alpha \), which is a measurable set. If the measurable set \( N = \mathbb{R}^n \setminus \bigcup_{\alpha \in A} E_\alpha \) were not a null set, it would (again by the \( \sigma \)-finiteness of \( \mathbb{R}^n \) with Lebesgue measure) contain a measurable set \( E_\beta \) with \( 0 < m(E_\beta) < \infty \), and \( \mathfrak{A} \cup \{ E_\beta \} \) would then be a properly larger disjoint family of measurable sets of finite positive measure that contained \( \mathfrak{A} \), contrary to the maximality of \( \mathfrak{A} \). That proves (2).

(b) Give an example of an abstract measure space \( (X, \mathfrak{M}, \mu) \) (with \( \mu \geq 0 \)) in which the analogue of (a) above does not hold. What additional condition would you put on \( (X, \mathfrak{M}, \mu) \) to get the analogue of (a)? Show that your condition implies the desired result.

If \( (X, \mathfrak{M}, \mu) \) is \( (X, 2^X, \#) \) where \( \# \) is counting measure, then for uncountable \( X \) it is evident that the singleton subsets of \( X \) form an uncountable maximal disjoint family of sets of finite positive measure. However, the proof just given for (a) above is obviously valid for any \( \sigma \)-finite \( (X, \mathfrak{M}, \mu) \), virtually without change (just replace \( \mathbb{R}^n \) by \( X \) wherever it occurs).

20. Observe that (by a “shift and change of scale”) the set \( \{ e^{2\pi i n \theta} \}_{n \in \mathbb{Z}} \) is an orthonormal basis of \( L^2([-1/2, 1/2], \text{Lebesgue}, m_1) \) with the usual inner product \( \langle f, g \rangle = \int_0^1 f(t)g(t) \, dt \). In a similar slight variant of the usual definition of the Fourier transform, define \( \hat{f} (\lambda) = \int_{\mathbb{R}} f(t)e^{-2\pi i \lambda t} \, dt \) for \( f \in L^1(\mathbb{R}, \text{Lebesgue}, m_1) \).

(a) Show that if \( f \in L^1(\mathbb{R}) \), then the series (indexed by \( \mathbb{Z} \)) in \( L^1[-1/2, 1/2] \) whose \( n \)-th term is the restriction of \( f(t-n) \) to \([-1/2, 1/2]\)—usually written as \( \sum_{n \in \mathbb{Z}} f(t-n) \), but with the understanding that it is restricted to the interval \([-1/2, 1/2]\) is (unordered) convergent in \( L^1[-1/2, 1/2] \), and its sum has \( L^1 \)-norm on the interval at most equal to \( ||f||_1 \). Show also that if we denote this sum (only on \([-1/2, 1/2]\)) by \( \varphi(t) \),
then the Fourier coefficients of \( \varphi \) are related to the Fourier transform of \( f \) by the simple \( \langle \varphi, e^{2\pi i k \theta} \rangle = \hat{f}(k) \). [Chop \( \mathbb{R} \) up into disjoint translates of \([-1/2, 1/2)\) by elements of \( \mathbb{Z} \).]

Consider first the case in which \( f \geq 0 \). The countable additivity of \( E \mapsto \int_E f(t) \, dt \) gives the first equality, translation-invariance of Lebesgue measure gives the second equality, and the monotone convergence theorem gives the third equality in the chain

\[
\int_{\mathbb{R}} f(t) \, dt = \sum_{n \in \mathbb{Z}} \int_{[n-1/2,n+1/2)} f(t) \, dt = \sum_{n \in \mathbb{Z}} \int_{-1/2}^{1/2} f(t-n) \, dt = \int_{-1/2}^{1/2} \left[ \sum_{n \in \mathbb{Z}} f(t-n) \right] \, dt.
\]

Thus if \( f \in L^1(\mathbb{R}) \), then the series \( \sum_{n \in \mathbb{Z}} f(t-n) \) converges to a function in \( L^1([-1/2,1/2]) \). As a series with nonnegative terms, it must therefore converge pointwise a.e. on \([-1/2,1/2)\)—and, moreover, the relation

\[
\|f\|_{L^1(\mathbb{R})} = \sum_{n \in \mathbb{Z}} \|f(t-n)\|_{L^1([-1/2,1/2])} = \left\| \sum_{n \in \mathbb{Z}} f(t-n) \right\|_{L^1([-1/2,1/2])}
\]

must hold. For general \( f \in L^1(\mathbb{R}) \) the argument just given, applied to \( |f(t)e^{-2\pi i \lambda t}| \), produces the series of absolute values of the functions in the chain

\[
\int_{\mathbb{R}} f(t)e^{-2\pi i \lambda t} \, dt = \sum_{n \in \mathbb{Z}} \int_{[n-1/2,n+1/2)} f(t)e^{-2\pi i \lambda t} \, dt = \sum_{n \in \mathbb{Z}} \int_{-1/2}^{1/2} f(t-n)e^{-2\pi i \lambda (t-n)} \, dt = \int_{-1/2}^{1/2} \left[ \sum_{n \in \mathbb{Z}} f(t-n) \right] e^{-2\pi i \lambda t} \, dt \quad \text{if } \lambda \text{ is an integer}
\]

and thus shows, by the dominated convergence theorem, that this chain of equalities also holds. The series \( \sum_{n \in \mathbb{Z}} f(t-n) \) is also seen to converge absolutely pointwise a.e. in \([-1/2,1/2)\), and one also has the estimate

\[
\left\| \sum_{n \in \mathbb{Z}} f(t-n) \right\|_{L^1([-1/2,1/2])} \leq \sum_{n \in \mathbb{Z}} \|f(t-n)\|_{L^1([-1/2,1/2])} = \|f\|_{L^1(\mathbb{R})}.
\]

For \( \lambda = k \in \mathbb{Z} \) the last set-off relation also shows that the Fourier coefficients of \( \varphi(t) = \sum_{n \in \mathbb{Z}} f(t-n) \) are given by

\[
\langle \varphi, e^{2\pi i k \theta} \rangle = \int_{-1/2}^{1/2} \left[ \sum_{n \in \mathbb{Z}} f(t-n) \right] e^{-2\pi i k t} \, dt = \int_{\mathbb{R}} f(t)e^{-2\pi i k t} \, dt = \hat{f}(k) \quad \text{for } k \in \mathbb{Z}.
\]

(b) Deduce the Poisson summation formula: if both \( f \) and \( \hat{f} \) fall off sufficiently rapidly at infinity, say \( |f(t)| \leq \frac{\text{const.}}{1 + |t|^{\alpha}} \) and \( |\hat{f}(\lambda)| \leq \frac{\text{const.}}{1 + |\lambda|^{\alpha}} \) for some \( \alpha > 1 \), then

\[
\sum_{n \in \mathbb{Z}} f(t-n) = \sum_{k \in \mathbb{Z}} \hat{f}(k)e^{2\pi i k t}
\]

with the series on both sides unordered convergent uniformly and absolutely on \([-1/2,1/2)\) (and therefore, by periodicity, absolutely and uniformly on compacta in \( \mathbb{R} \)). [Note that if a Fourier series converges uniformly to a continuous function, it must converge to the same function in the \( L^2 \) norm.]

{The instructor expresses his embarrassment and apologies about this problem, which requires some form of the inversion theorem for Fourier transforms, something that was not covered in the course. The raw materials for a suitable version of this theorem are available, fortunately. We want to see that if both \( f \) and \( \hat{f} \) belong to \( L^1(\mathbb{R}) \), then \( f \) is a.e. equal to the inverse Fourier transform of \( \hat{f} \), and that the latter is a continuous function (tending to zero at infinity). The Fourier transform and the inverse Fourier transform

19

Math 501:01 Fall 1999}
differ only by the ambiguous sign in the integral \( \int_{\mathbb{R}} f(t) e^{\mp 2\pi i \lambda t} dt \). For \( f \in L^1(\mathbb{R}) \) a function given by such an expression is sequentially continuous, and thus continuous, by the Lebesgue dominated convergence theorem: if \( \lambda_n \to \lambda \) in \( \mathbb{R} \), then

\[
\lim_{n \to \infty} \int_{\mathbb{R}} f(t) e^{\mp 2\pi i \lambda_n t} dt = \int_{\mathbb{R}} f(t) e^{\mp 2\pi i \lambda t} dt
\]

(1)
since \( |f(t)| \) dominates the absolute values of all the integrands and pointwise convergence of the integrands is obvious. (The fact that Fourier transforms tend to zero at infinity is the Riemann-Lebesgue lemma.) For \( y > 0 \) and \( t \in \mathbb{R} \), consider the elementary integral

\[
\int_{\mathbb{R}} e^{-|2\pi \lambda y|} e^{2\pi i \lambda t} d\lambda = 2 \int_{0}^{\infty} e^{-2\pi \lambda y} \cos(2\pi \lambda t) d\lambda = \left[ 4\pi(e^{-2\pi \lambda y} \cdot \frac{t \sin(2\pi \lambda t) - y \cos(2\pi \lambda t)}{4\pi^2(y^2 + t^2)} \right]_{\lambda=0}^{\lambda=\infty} = \frac{y}{\pi} \frac{1}{y^2 + t^2}
\]

(2)

where the r. h. side of (2) belongs to \( L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}) \) as a function\(^{(6)}\) of \( t \). If we convolve the r. h. side of (2) with a function \( f \in L^1(\mathbb{R}) \) we get—pointwise in \( x \), because we are convolving an \( L^1 \) function with an \( L^\infty \) function—

\[
\left( f * \frac{y}{\pi} \frac{1}{y^2 + t^2} \right)(x) = \int_{\mathbb{R}} f(t) \frac{y}{\pi} \frac{1}{y^2 + (x-t)^2} dt = \int_{\mathbb{R}} f(t) \left[ \int_{\mathbb{R}} e^{\mp 2\pi \lambda y} e^{2\pi i \lambda (x-t)} d\lambda \right] dt
\]

\[
= \int_{\mathbb{R}} e^{-|2\pi \lambda y|} \left[ \int_{\mathbb{R}} e^{2\pi i \lambda t} e^{2\pi i \lambda x} d\lambda \right] dt = \int_{\mathbb{R}} e^{-|2\pi \lambda y|} \hat{f}(\lambda) e^{2\pi i \lambda x} d\lambda,
\]

(3)

the inverse Fourier transform of the \( L^1 \) function \( e^{-|2\pi \lambda y|} \hat{f}(\lambda) \). Reversing the order of integration is justified by Tonelli and Fubini, since \( f(t) \in L^1 \) as a function of \( t \), \( e^{-|2\pi \lambda y|} \in L^1 \) as a function of \( \lambda \), and the complex exponentials are bounded by 1 in absolute value. It is clear (by the Lebesgue dominated convergence theorem) that as \( y \to 0^+ \) the r. h. side of (3) tends to \( \int_{\mathbb{R}} \hat{f}(\lambda) e^{2\pi i \lambda x} d\lambda \), the value at \( x \) of the inverse Fourier transform of \( \hat{f} \); in fact, one sees easily that the convergence is uniform in \( x \). The l. h. side of (3) requires closer analysis.

The integral \( \int_{\mathbb{R}} \frac{y}{\pi} \frac{1}{y^2 + t^2} dt = 1 \) (for all \( y > 0 \)) is elementary. More generally, however, for each fixed \( \delta > 0 \) we have

\[
\int_{-\delta}^{\delta} \frac{y}{\pi} \frac{1}{y^2 + t^2} dt = \frac{2}{\pi} \arctan \left( \frac{\delta}{y} \right)
\]

(4)

which tends to 1 as \( y \to 0^+ \) irrespective of the choice of \( \delta > 0 \). The positivity of everything in sight in the relation

\[
\left\| \frac{y}{\pi} \frac{1}{y^2 + t^2} \right\|_1 = \int_{\mathbb{R}} \frac{y}{\pi} \frac{1}{y^2 + t^2} \cdot \chi(-\delta,\delta) dt + \int_{\mathbb{R}} \frac{y}{\pi} \frac{1}{y^2 + t^2} \cdot \chi_{\mathbb{R}\setminus(-\delta,\delta)} dt
\]

(5)

then tells us that the first term on the r. h. side is bounded by 1 in \( L^1 \)-norm and that the second term tends to 0 as \( y \to 0^+ \); therefore, to demonstrate that \( f * \left( \frac{y}{\pi} \frac{1}{y^2 + t^2} \right) \to f \) in \( L^1 \) norm for each \( f \in L^1(\mathbb{R}) \) it will suffice to establish that limit relation for \( f \ast \left( \frac{y}{\pi} \frac{1}{y^2 + t^2} \cdot \chi(-\delta,\delta) \right) \).

Moreover, for a continuous function \( g \) of compact support on \( \mathbb{R} \) we can write

\[^{(6)}\) The r. h. side of (2) is the Poisson kernel for the upper half of the plane \( \mathbb{R}^2 \). With \( t \) replaced by \( x \) one gets a harmonic function of \((x,y)\) in the upper half-plane. (The easiest way to see this is to observe that the r. h. side of (2) is, up to a multiplicative constant, the imaginary part of \( 1/(t+iy) \).) We shall see below that its convolution with a continuous function \( g \) on \( \mathbb{R} \) gives a harmonic function of \((x,y)\) on the upper half-plane with boundary values \( g \) on the \( x \)-axis; if \( g \) is uniformly continuous, then \( g(x) \) is attained as a boundary value uniformly in \( x \) as \( y \) approaches zero through positive values.\]
\[ \left\| f - f \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \right\|_1 \leq \]
\[ \left\| f - g \right\|_1 + \left\| g - g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \right\|_1 + \left\| g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) - f \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \right\|_1 \leq \]
\[ \left\| f - g \right\|_1 + \left\| g - g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \right\|_1 + \left\| g - f \right\|_1 \left\| \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \right\|_1 \leq \]
\[ 2 \cdot \left\| f - g \right\|_1 + \left\| g - g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \right\|_1 \]

using the norm bound for convolutions established in 3 above. Since for any \( f \in L^1(\mathbb{R}) \) one can find continuous \( g \) of compact support with \( \| f - g \|_1 \) as small as one pleases, to demonstrate that \( f \ast \frac{1}{y^2 + t^2} \to f \) in \( L^1 \) norm for each \( f \in L^1(\mathbb{R}) \) it will suffice to establish that limit relation for \( g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \) for \( g \) continuous of compact support. In fact, for \( g \) continuous of compact support the convergence is uniform on \( \mathbb{R} \), with the supports of the convolutions contained in a fixed neighborhood of the support of \( g \); that certainly implies \( L^1 \)-norm convergence. To establish this, let \( g \) be given, let \([-a, a] \subseteq \mathbb{R} \) be a fixed interval with the support of \( g \) contained in its interior, and given \( \epsilon > 0 \) let \( \delta > 0 \) be smaller than the distance from the support of \( g \) to the complement of \([-a, a] \) and such that \( |x - t| < \delta \Rightarrow |g(x) - g(t)| < \epsilon/2 \); then clearly the support of \( g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) \) is contained in \([-a, a] \). For arbitrary \( x \in \mathbb{R} \) one has the estimates

\[
\left| g(x) - g \ast \left( \frac{1}{y^2 + t^2} \cdot \chi(-\delta, \delta) \right) (x) \right| \leq \]
\[ \left| g(x) - g(x) \cdot \int_{x-\delta}^{x+\delta} \frac{y}{y^2 + (x-t)^2} \, dt \right| + \left| \int_{x-\delta}^{x+\delta} \left| g(x) - g(t) \right| \frac{y}{y^2 + (x-t)^2} \, dt \right| \leq \]
\[ \left| g(x) - g(x) \cdot \int_{x-\delta}^{x+\delta} \frac{y}{y^2 + (x-t)^2} \, dt \right| + \left| \int_{x-\delta}^{x+\delta} \left| g(x) - g(t) \right| \frac{y}{y^2 + (x-t)^2} \, dt \right| \leq \]
\[ \left| g(x) - g(x) \cdot \int_{x-\delta}^{x+\delta} \frac{y}{y^2 + (x-t)^2} \, dt \right| + \left| \frac{\epsilon}{2} \right| \]

where the second integral is \(< \frac{\epsilon}{2} \) because \( |g(x) - g(t)| < \epsilon/2 \) in the interval of integration. Since (4) tells us that the integral in the first term of the last line of (7) can be made as close to 1 as desired by taking \( y > 0 \) sufficiently small, we have established the uniform convergence.

\( \text{(c) Deduce} \sum_{n \in \mathbb{Z}} \frac{1}{(a^2 + n^2)} = \frac{\pi}{a} \coth(\pi a) \text{ for} \ a > 0 \text{ from (b) above and the famous contour integral} \)

\[ \int_{-\infty}^{\infty} \frac{1}{1 + x^2} \cos(ax) \, dx = \pi e^{-|a|} \]

beloved of generations of complex-variables students. Can you deduce the value of \( \zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} ? \)

\{NB: This relation can also be deduced from the version of the Fourier inversion theorem given above: interchanging the rôles of the direct and inverse Fourier transforms one finds that with \( f(x) = e^{-2\pi|x|} \), relation (2) says \( \hat{f}(\lambda) = \frac{1}{\pi} \frac{1}{1 + \lambda^2} \) and therefore—by Fourier inversion—\( \int_{-\infty}^{\infty} \frac{1}{1 + \lambda^2} \cos(\lambda x) \, d\lambda = \pi e^{-|x|} \) \}
Starting from \( \int_{-\infty}^{\infty} \frac{1}{1 + x^2} \cos(\alpha x) \, dx = \pi e^{-|\alpha|} \) one can replace \( x \) by \( \frac{x}{a} \) and \( \alpha \) by \( 2\pi \lambda a \) to get (assuming \( a > 0 \))
\[
\int_{-\infty}^{\infty} \frac{a}{a^2 + x^2} \cos(2\pi \lambda x) \, dx = \pi e^{-2\pi \lambda a}
\]
so the Fourier transform of \( \frac{a}{a^2 + x^2} \) is \( \pi e^{-2\pi \lambda a} \). Since both the function and its Fourier transform belong to \( L^1(\mathbb{R}) \cap C(\mathbb{R}) \), we have, **pointwise**, by Poisson summation
\[
\sum_{n \in \mathbb{Z}} \frac{a}{a^2 + (x - n)^2} = \pi \sum_{k \in \mathbb{Z}} e^{-2\pi a |k|} e^{2\pi i k x}, \quad \text{and for } x = 0
\]
\[
\sum_{n \in \mathbb{Z}} \frac{1}{a^2 + n^2} = \pi \sum_{k \in \mathbb{Z}} e^{-2\pi a |k|} = \frac{\pi}{a} \left[ 1 + 2 \sum_{k=1}^{\infty} (e^{-2\pi a})^k \right]
\]
\[
= \frac{\pi}{a} \left[ 1 + 2 \frac{e^{-2\pi a}}{1 - e^{-2\pi a}} \right] = \frac{\pi}{a} \frac{1 + e^{-2\pi a}}{1 - e^{-2\pi a}} = \frac{\pi e^a + e^{-a}}{e^a - e^{-a}}
\]
\[
= \frac{\pi}{a} \coth \left( \frac{\pi a}{a} \right).
\]

As a last shot, from the series summed above we have (since uniform convergence—established by comparing the series to the series of constants \( \sum_{n=1}^{\infty} \frac{1}{n^2} \) justifies term-by-term passage to the limit in the series as \( a \to 0 \))
\[
2 \sum_{n=1}^{\infty} \frac{1}{a^2 + n^2} = \sum_{n \neq 0 \in \mathbb{Z}} \frac{1}{(a^2 + n^2)} = \frac{\pi}{a} \coth(\pi a) - \frac{1}{a^2}
\]
\[
= \lim_{a \to 0} \left[ \frac{\pi}{a} \coth(\pi a) - \frac{1}{a^2} \right] = \lim_{a \to 0} \frac{\pi a \cosh \pi a - \sinh \pi a}{a^2 \sinh \pi a}
\]
\[
= \lim_{a \to 0} \frac{\pi a \sinh \pi a + \pi^2 a \sinh \pi a - \pi \cosh \pi a}{2a \sinh \pi a + a^2 \pi \cosh \pi a} = \lim_{a \to 0} \frac{\pi^2 \sinh \pi a}{2 \sinh \pi a + a \pi \cosh \pi a}
\]
\[
= \lim_{a \to 0} \frac{\pi^2 \cosh \pi a}{\pi^2 \cosh \pi a + \pi \cosh \pi a + a \pi^2 \sinh \pi a} = \frac{\pi^2}{3}
\]
\[
\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}
\]
following two applications of L'Hôpital's rule.