The second term belongs to the finite-dimensional space spanned by \( \{ e_\beta \}_{\beta \in F} \), so by the isometry of that space with a \( \mathbb{K}^n \) (or by direct computation), the norm-squared of that term is the sum of the squares of the components: we thus have

\[
\left\| x - \sum_{\alpha \in F} \lambda_\alpha e_\alpha \right\|^2 = \left\| x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \right\|^2 + \sum_{\beta \in F} |\langle x, e_\beta \rangle - \lambda_\beta|^2 e_\beta.
\]

From the sum-of-squares form of the second term here, it is immediate that the minimum is attained if and only if \( \lambda_\beta = \langle x, e_\beta \rangle \) for each \( \beta \in F \).

A similar computation gives

**Lemma** [Bessel?]: If \( \{ e_\alpha \}_{\alpha \in A} \) is an orthonormal set in \( H \), then for any \( x \in H \) and finite subset \( F \subseteq A \),

\[
\left\| x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \right\|^2 = \left\| x \right\|^2 - \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2.
\]

**Proof.**

\[
\left\| x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \right\|^2 = \left\langle x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha, x \right\rangle - \left\langle \sum_{\beta \in F} \langle x, e_\beta \rangle e_\beta, x \right\rangle + \left\langle \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha, \sum_{\beta \in F} \langle x, e_\beta \rangle e_\beta \right\rangle \\
= \left\langle x, x \right\rangle - \sum_{\beta \in F} |\langle x, e_\beta \rangle|^2 - \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 + \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 \\
= \left\langle x, x \right\rangle - \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2.
\]

**Corollary** [Bessel’s Inequality, First Version]: If \( \{ e_\alpha \}_{\alpha \in A} \) is an orthonormal set in \( H \), then for any \( x \in H \) and finite subset \( F \subseteq A \),

\[
\sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2
\]

with equality if and only if \( x \) belongs to the linear space spanned by \( \{ e_\alpha \}_{\alpha \in F} \).

**Corollary**: If \( \{ e_\alpha \}_{\alpha \in A} \) is an orthonormal set in \( H \), then the net \( \left\{ \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 \right\}_{F \subseteq A} \) of finite sums of the absolute squares of the Fourier coefficients of \( x \)—an “unordered series of nonnegative terms”—converges (in \( \mathbb{R}^+ \)) to its supremum, which is \( \leq \|x\|^2 \).

For obvious reasons, one denotes the limit of such a net of finite sums by \( \sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2 \). It is easy to see that such a convergent “unordered series” can have at most countably many nonzero terms; applying the usual argument is left to the reader.

**Corollary** [Bessel’s Inequality, Second Version]: If \( \{ e_\alpha \}_{\alpha \in A} \) is an orthonormal set in \( H \), then

\[
\sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2
\]

with equality if and only if \( x \) belongs to the norm closure of the linear space spanned by \( \{ e_\alpha \}_{\alpha \in A} \).

**Proof.** The statement makes sense because we now know the series converges, and the inequality follows from the first version of Bessel’s inequality. If equality holds, then given any \( \epsilon > 0 \) we can find finite \( F \subseteq A \) for which \( \|x - \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 \) is less than \( \epsilon^2 \); but that says \( \|x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \| < \epsilon^2 \), so \( x \) can be approximated arbitrarily closely by finite linear combinations of the \( \{ e_\alpha \}_{\alpha \in A} \). On the other hand, if \( x \) can be approximated arbitrarily closely by finite linear combinations of the \( \{ e_\alpha \}_{\alpha \in A} \), then—since the best approximation to \( x \) of the form \( \sum_{\alpha \in F} \lambda_\alpha e_\alpha \) is \( \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \)—given any \( \epsilon > 0 \) one can find \( F \subseteq A \) that makes \( \|x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \| < \epsilon \), which means that one can make \( 0 \leq \langle x, x \rangle - \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 < \epsilon \), and that identifies the sum of the absolute squares of the Fourier coefficients.
This version of Bessel’s inequality shows that for any orthonormal set \( \{e_\alpha\}_{\alpha \in A} \subseteq H \) the Fourier coefficients (relative to that set) of any \( x \in H \) belong to \( \ell^2(A) \). We already know that if \( \#A = n < \aleph_0 \), then the linear space spanned by \( \{e_\alpha\}_{\alpha \in A} \) is isometrically isomorphic—i.e., isomorphic with the inner product preserved—to \( K^n = \ell^2(n) \). We can extend this to arbitrary \( A \), but we need to know a little about unordered series in normed spaces.

An unordered series in \( (X, \| \cdot \|_X) \) is an indicated unordered sum \( \sum_{\alpha \in A} x_\alpha \), where the “terms” \( x_\alpha \) belong to \( X \). It converges if the (filter based on the) net \( \left\{ \sum_{\alpha \in F} x_\alpha \right\} \) converges. The limit (if it exists) is called the sum of the series, logically enough, and denoted by \( \sum_{\alpha \in A} x_\alpha \). It is easy to verify that convergent series can be multiplied term-by-term by scalars, and that convergent series indexed by the same set can be added term-by-term; the reader can easily supply the details of a verification, which requires nothing more than the continuity of the operations in \( X \). If “\( X \)” is a Hilbert space with the inner-product norm, then the continuity of the inner product also gives

\[
\left\langle \sum_{\alpha \in A} x_\alpha, \sum_{\beta \in B} y_\beta \right\rangle = \sum_{(\alpha, \beta) \in A \times B} \langle x_\alpha, y_\beta \rangle.
\]

In the most interesting cases the space \( (X, \| \cdot \|_X) \) is a Banach space, and the Cauchy condition for the convergence of unordered series takes an interesting and useful form:

**Proposition**: Let \( \sum_{\alpha \in A} x_\alpha \) be an unordered series in a Banach space. Then the series converges if and only if: for every \( \epsilon > 0 \) there exists a finite set \( F \subseteq A \) (depending, in general on \( \epsilon \)) such that if \( G \subseteq A \) is a finite set disjoint from \( F \), then \( \| \sum_{\alpha \in G} x_\alpha \| \leq \epsilon \).

**Proof.** It is necessary and sufficient for convergence that the net of finite subsums of the series be a Cauchy net. If that net is Cauchy, then given \( \epsilon > 0 \) there exists finite \( F \subseteq A \) such that if \( F_1, F_2 \subseteq A \) are two finite subsets of \( A \) containing \( F \), then \( \| \sum_{\alpha \in F_1} x_\alpha - \sum_{\alpha \in F_2} x_\alpha \| \leq \epsilon \). If \( G \subseteq A \) is a finite subset disjoint from \( F \), then taking \( F_2 = G \cup F \) and \( F_2 = F \) in that inequality yields

\[
\| \sum_{\alpha \in G} x_\alpha \| = \| \sum_{\alpha \in G \cup F} x_\alpha - \sum_{\alpha \in F} x_\alpha \| \leq \epsilon
\]

so the condition is necessary for convergence. On the other hand, if the condition holds then let \( \epsilon > 0 \) be given and take a finite set \( F \subseteq A \) with the property that if \( G \subseteq A \) is a finite set disjoint from \( F \), then \( \| \sum_{\alpha \in F} x_\alpha \| \leq \epsilon/2 \). Given finite \( F_1, F_2 \subseteq A \) both of which contain \( F \), put \( G_1 = F_1 \setminus F \) and \( G_2 = F_2 \setminus F \). Because the terms with indices belonging to \( F \) cancel in the difference of the sums, we then have

\[
\| \sum_{\alpha \in F_1} x_\alpha - \sum_{\alpha \in F_2} x_\alpha \| = \| \sum_{\alpha \in G_1} x_\alpha - \sum_{\alpha \in G_2} x_\alpha \| \leq \| \sum_{\alpha \in G_1} x_\alpha \| + \| \sum_{\alpha \in G_2} x_\alpha \| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\]

and so the net of finite subsums is Cauchy.

**Corollary**: A convergent unordered series in a normed space can have only countably many nonzero terms.

Indeed, for each \( n \in \mathbb{N} \) one can find finite \( F_n \subseteq A \) for which \( G \cap F_n = \emptyset \Rightarrow \| \sum_{\alpha \in G} x_\alpha \| \leq 1/n \), and thus in particular \( \alpha \in A \setminus F_n \Rightarrow \| x_\alpha \| \leq 1/n \). It follows that any term whose index does not belong to \( \bigcup_{n=1}^\infty F_n \) must be zero.

The reader can easily verify that a series of scalars (\( \mathbb{K} = \mathbb{R} \) or \( \mathbb{C} \)) is unordered convergent if and only if it is absolutely convergent; it follows that if \( \dim X < \infty \) then the unordered convergent series are exactly
those for which $\sum_{\alpha \in A} \|x_{\alpha}\| < \infty$. In $C[0,1], \| \cdot \|_{\infty}$, on the other hand, a series $\sum_{\alpha \in A} f_{\alpha}$ is unordered convergent if and only if the series $\sum_{\alpha \in A} |f_{\alpha}|$ is unordered convergent,\(^{29}\) but it is possible for that to happen even though $\sum_{\alpha \in A} \|f_{\alpha}\|_{\infty} = \infty$. Conversely, however, it is easy to see that for a series $\sum_{\alpha \in A} x_{\alpha}$ in any Banach space $X$, the convergence of the series $\sum_{\alpha \in A} \|x_{\alpha}\|$ of positive scalars—necessarily converging to the supremum of its finite subsums, so its convergence can be characterized as $\sum_{\alpha \in A} \|x_{\alpha}\| < \infty$—implies the condition given in the proposition above, and thus implies the convergence of the given vector series in $X$.

In a Hilbert space, the condition for convergence of (unordered) series with orthogonal terms “scalarizes” in a very neat way:

**Proposition:** Let $\sum_{\alpha \in A} x_{\alpha}$ be an unordered series in a Hilbert space $(H, \langle \cdot, \cdot \rangle)$, and suppose that its terms are pairwise orthogonal, i.e., that $\alpha \neq \beta \Rightarrow \langle x_{\alpha}, x_{\beta} \rangle = 0$. Then the series converges if and only if the (unordered nonnegative) scalar series $\sum_{\alpha \in A} \|x_{\alpha}\|^2 < \infty$.

**Proof.** For any finite $G \subseteq A$ we have the Pythagorean

$$\left\| \sum_{\alpha \in G} x_{\alpha} \right\|^2 = \left( \sum_{\alpha \in G} \langle x_{\alpha}, x_{\beta} \rangle \right) = \sum_{\alpha \in G} \langle x_{\alpha}, x_{\alpha} \rangle = \sum_{\alpha \in G} \|x_{\alpha}\|^2 .$$

It follows that the scalar series in $\mathbb{R}^+$ satisfies the necessary and sufficient condition for Cauchy-ness established in the proposition above if and only if the vector series in $H$ satisfies it.

Putting this together with the Bessel inequality, we have

**Theorem:** If $\{e_\alpha\}_{\alpha \in A}$ is an orthonormal set in a Hilbert space $H$, then for each $x \in H$ the unordered series

$$\sum_{\alpha \in A} \langle x, e_\alpha \rangle e_\alpha$$

converges (in the norm metric of $H$) to the point nearest to $x$ in the norm closure of the linear space spanned by $\{e_\alpha\}_{\alpha \in A}$. In particular, the following three conditions are equivalent:

1. The vector series $\sum_{\alpha \in A} \langle x, e_\alpha \rangle e_\alpha$ converges to $x$;
2. $x$ belongs to the norm closure of the linear space spanned by $\{e_\alpha\}_{\alpha \in A}$;
3. $\sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2 = \|x\|^2$.

**Proof.** The series obviously has pairwise orthogonal terms, so its convergence is guaranteed by the proposition just proved—joined with Bessel’s inequality, which gives a bound on the (scalar) series of squares of the lengths of the terms:

$$\sum_{\alpha \in A} \|\langle x, e_\alpha \rangle e_\alpha\|^2 = \sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2 .$$

The sum of the series evidently belongs to the norm closure of the linear space spanned by $\{e_\alpha\}_{\alpha \in A}$, and since

$$\left\langle \sum_{\alpha \in A} \langle x, e_\alpha \rangle e_\alpha, e_\beta \right\rangle = \langle x, e_\beta \rangle - \langle x, e_\beta \rangle = 0$$

holds for each $\beta \in A$, the difference between $x$ and the sum of the series is orthogonal to all the $\{e_\beta\}_{\beta \in A}$ and thus to the closure of the subspace that they span. It follows that the sum of the series is the projection of $x$ on that subspace. As to the equivalence of the three conditions listed: (1) $\Rightarrow$ (2) is clear. (2) $\Rightarrow$ (3) is the sufficient condition for equality in Bessel’s inequality (second version, p. 45 above). Finally, (3) and the relation

$$\left\| x - \sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha \right\|^2 = \|x\|^2 - \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2$$

of the Lemma on p. 45 above show that when (3) holds the limit of the finite sums $\sum_{\alpha \in F} \langle x, e_\alpha \rangle e_\alpha$ is precisely the vector $x \in H$ with which one started.

\(^{29}\) Checking this assertion is elementary but not entirely trivial; it makes a nice exercise for the reader.
The equivalence among (1), (2) and (3) above is sometimes called the Parseval relation, and (3) is called the Parseval identity or Parseval equation for the norm-closed subspace generated by the \( \{e_{\alpha}\}_{\alpha \in A} \)—although that usage is frequently restricted to the case in which that subspace is all of the Hilbert space \( H \).

It is now appropriate to look at the relation between the norm-closed subspace generated by an orthonormal \( \{e_{\alpha}\}_{\alpha \in A} \) and the orthonormal set itself—but from the viewpoint of the subspace rather than from that of the orthonormal set.

**Definition:** Let \((H, \langle \cdot, \cdot \rangle)\) be a Hilbert space and \( M \subseteq H \) a norm-closed subspace. An orthonormal set \( \{e_{\alpha}\}_{\alpha \in A} \subseteq M \) is called an orthonormal basis (or base) of \( M \) if \( M \) is the norm closure of the linear subspace generated by \( \{e_{\alpha}\}_{\alpha \in A} \).

One skips the "of \( M \)" in the case \( M = H \) and simply refers to an orthonormal basis.

**Proposition:** Let \((H, \langle \cdot, \cdot \rangle)\) be a Hilbert space and \( M \subseteq H \) a norm-closed subspace. The following conditions on an orthonormal set \( \{e_{\alpha}\}_{\alpha \in A} \subseteq M \) are logically equivalent:

1. \( \{e_{\alpha}\}_{\alpha \in A} \) is an orthonormal basis of \( M \);
2. If \( y \in M \) is orthogonal to all the \( \{e_{\alpha}\}_{\alpha \in A} \), then \( y = 0 \);
3. \( \{e_{\alpha}\}_{\alpha \in A} \) is a maximal orthonormal subset of \( M \), i.e., one not properly contained in a larger orthonormal set;
4. For every \( x \in M \), the vector series \( \sum_{\alpha \in A} \langle x, e_{\alpha} \rangle e_{\alpha} \) converges to \( x \), i.e.,
   \[
   x = \sum_{\alpha \in A} \langle x, e_{\alpha} \rangle e_{\alpha} ;
   \]
5. For every \( x \in M \), \( \sum_{\alpha \in A} |\langle x, e_{\alpha} \rangle|^2 = \|x\|^2 \).

**Proof.** The equivalence of (1), (4) and (5) follows from the theorem just proved. (2) \( \Leftrightarrow \) (3) is clear: (2) prevents one from finding a new \( e_{\beta} \) that is orthogonal to all the \( e_{\alpha} \)’s one already has, while if (3) holds then any vector \( y \in M \) orthogonal to all the \( \{e_{\alpha}\}_{\alpha \in A} \) must be zero, because otherwise one could set \( e = y/\|y\| \) and then \( \{e_{\alpha}\}_{\alpha \in A} \cup \{e\} \subseteq M \) would be orthonormal, contrary to maximality. (4) \( \Rightarrow \) (2): if all the coefficients in the series are zero, as they will be if \( x \perp \{e_{\alpha}\}_{\alpha \in A} \), then \( x = 0 \). (2) \( \Rightarrow \) (4): Let \( x \in M \) be given; the nearest point to \( x \) in the norm closure of the linear span of \( \{e_{\alpha}\}_{\alpha \in A} \) is known to be \( \sum_{\alpha \in A} \langle x, e_{\alpha} \rangle e_{\alpha} \). Put
   \[
   y = x - \sum_{\alpha \in A} \langle x, e_{\alpha} \rangle e_{\alpha} .
   \]
   For each \( \beta \in A \) we have \( \langle y, e_{\beta} \rangle = \langle x, e_{\beta} \rangle - \langle x, e_{\beta} \rangle = 0 \), so by (2) we have \( y = 0 \) and therefore \( x = \sum_{\alpha \in A} \langle x, e_{\alpha} \rangle e_{\alpha} \), which is (4).

**Corollary:** Every norm-closed subspace \( M \neq \{0\} \) of a Hilbert space possesses orthonormal bases; indeed, every orthonormal set in such an \( M \) can be enlarged to an orthonormal basis. In particular, every orthonormal basis of a subspace \( M \subseteq H \) can be enlarged to an orthonormal basis of \( H \).\(^{30}\)

**Proof.** Let \( \{e_{\alpha}\}_{\alpha \in B} \subseteq M \) be an orthonormal set. The family \( \mathcal{E} \) of orthonormal subsets of \( M \) that contain \( \{e_{\alpha}\}_{\alpha \in B} \) is partially ordered under set-inclusion (in \( 2^M \)) and it is obvious that the union of a chain of orthonormal subsets of \( M \) is again an orthonormal subset of \( M \). Zorn’s lemma produces maximal orthonormal subsets \( \{e_{\alpha}\}_{\alpha \in A} \subseteq M \), and the proposition just proved implies that these are orthonormal bases.

It is surprisingly complicated to give a proof of the following proposition for uncountable cardinals, so we content ourselves with the countable cases. For the uncountable cases see, e.g., N. Dunford and J. T. Schwartz, *Linear Operators*, Vol. I, IV.4.14, pp. 253–255.

\(^{30}\) In these as in purely algebraic considerations about dimension and bases, one can avoid dealing with very-low-dimensional exceptions by agreeing to say that the zero subspace has the empty set as its basis, and that the empty set is (vacuously) orthonormal.
Lemma: A Hilbert space $H$ is separable (i.e., possesses a countable subset dense in its norm metric topology) if and only if every orthonormal basis of $H$ is either finite or countably infinite, and for this it suffices that $H$ possess one countable orthonormal basis.

Proof. Let $\{e_\alpha\}_{\alpha \in A} \subseteq M$ be an orthonormal basis of $H$. We have already observed that all but countably many of the terms in a series $x = \sum_{\alpha \in A} \langle x, e_\alpha \rangle e_\alpha$ must have zero coefficients. Thus if $\{x_n\}_{n=1}^\infty$ is dense in $H$ and for each $n \in \mathbb{N}$ we put $A_n = \{\alpha \in A : \langle x_n, e_\alpha \rangle \neq 0\}$, then $\bigcup_{n=1}^\infty A_n \subseteq A$ is countable. If there were some $\beta \in A \setminus \bigcup_{n=1}^\infty A_n$ it would have the property that $\langle x_n, e_\beta \rangle = 0$ for all $x_n$ belonging to a dense subset of $H$; approximating $e_\beta$ arbitrarily closely by $x_n$'s, we see that $\|e_\beta\|^2 = 0$ which is impossible for an element of an orthonormal set. Thus $A = \bigcup_{n=1}^\infty A_n$ and it is countable. Conversely, if $\{e_n\}_{n=1}^\infty \subseteq H$ is a countable orthonormal basis of $H$, then the set of all finite linear combinations $\sum_{j=1}^k \lambda_n e_n_j$ (coefficients in $\mathbb{K}$) of its elements is dense in $H$, and since each coefficient $\lambda_n$ can be approximated arbitrarily closely in $\mathbb{K}$ by elements of $\mathbb{Q}$ (if $\mathbb{K} = \mathbb{Q}$) or $\mathbb{Q}[i]$ (if $\mathbb{K} = \mathbb{C}$), the existence of the countable orthonormal basis forces $H$ to be separable.

Proposition: Let $H$ be a separable Hilbert space. Then if $M \subseteq H$ is a closed linear subspace the cardinality of all orthonormal bases of $M$ is the same, and is either a finite cardinal or $\aleph_0$ (the cardinality of $\mathbb{N}$); hereinafter this cardinal is called the dimension of $M$ and written $\dim M$. If $M \subseteq N$ are two closed linear subspaces, then $\dim M \leq \dim N$. In general, all orthonormal bases of a given Hilbert space $H$, separable or not, have the same cardinality.

Proof. If $M$ is finite-dimensional (and thus automatically closed, as we observed on p. 44 above) then the invariance of the basis cardinality is a well-known theorem of linear algebra. If $M$ is not finite-dimensional, then it is separable (as a subspace of a separable metric space) and the lemma shows that the cardinality of any orthonormal base must be $\aleph_0$. The fact that dimension grows with subspaces follows from the fact that if $M \subseteq N$ are subspaces, then any orthonormal basis of $M$ can be enlarged to an orthonormal basis of $N$. For the general case of the invariance of cardinality of orthonormal bases, see Dunford & Schwartz, loc. cit.

It is natural to define an isomorphism of Hilbert spaces, also called an isometric isomorphism, an orthogonal transformation (if $\mathbb{K} = \mathbb{R}$) or a unitary transformation (if $\mathbb{K} = \mathbb{C}$) to be an algebraic isomorphism $U : H_1 \to H_2$ of one Hilbert space onto another, such that $\langle Ux, Uy \rangle_2 = \langle x, y \rangle_1$ holds for all $x, y \in H_1$. The considerations for finite linear spans on p. 44 above can be viewed as having shown us that all Hilbert spaces of dimension $n < \aleph_0$ are isometrically isomorphic to $\mathbb{K}^n$ with its usual inner product; we want to extend this to give us “standard models” of all Hilbert spaces.

Proposition: Let $H$ be a Hilbert space with orthonormal basis $\{e_\alpha\}_{\alpha \in A} \subseteq M$. Then the mapping

$$\ell^2(A) \to H \quad (\text{coefficients for } \ell^2 \text{ from the scalar field of } H)$$

$$(\ldots, \lambda_\alpha, \ldots) \mapsto \sum_{\alpha \in A} \lambda_\alpha e_\alpha$$

is an isometric isomorphism (where $\ell^2(A)$ is given the inner product that goes with $L^2(A, 2^A, \#)$, namely $\langle (\ldots, \lambda_\alpha, \ldots), (\ldots, \mu_\alpha, \ldots) \rangle = \sum_\alpha \lambda_\alpha \overline{\mu_\alpha}$).

Proof. Not much is left to prove. If $(\ldots, \lambda_\alpha, \ldots) \in \ell^2(A)$, then the formal unordered series $\sum_{\alpha \in A} \lambda_\alpha e_\alpha$ has orthogonal terms, and $\sum_{\alpha \in A} |\lambda_\alpha|^2 < \infty$ (the definition of belonging to $\ell^2(A)$) is exactly the necessary and sufficient condition for this series to converge. Thus the mapping is well-defined, and it is trivial to verify—extending what one knows in the finite-sum case by continuity—that the mapping is linear and preserves the inner product. The mapping is onto in view of the theorem on p. 47 above.

31 The original axiomatization of Hilbert spaces insisted that they be separable and thus finessed cardinality complications.
Of course the first application one wants to consider in this setting is

**Proposition:** For the Hilbert space \( L^2(\mathbb{R}, dm) \) equipped with the inner product

\[
\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)g(\theta)\,d\theta,
\]

the orthonormal set \( \{e^{in\theta}\}_{n \in \mathbb{Z}} \) is an orthonormal basis. Thus the unitary mappings

\[
L^2 \rightarrow \ell^2 \quad \text{(Fourier transform)}
\]

\[
f \mapsto \hat{f} = \left( n \mapsto \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta \right)
\]

\[
\ell^2 \rightarrow L^2 \quad \text{(inverse Fourier transform)}
\]

\[
(\ldots, \lambda_k, \ldots) \mapsto \sum_{k \in \mathbb{Z}} \lambda_k e^{ik\theta}
\]

(these mappings are each other’s inverses) are isomorphisms of Hilbert spaces between the space of functions \( L^2(\mathbb{R}, dm) \) and the “space of square-summable Fourier coefficients” \( \ell^2(\mathbb{Z}) \).

**Proof.** One needs only to show that the orthonormal set \( \{e^{in\theta}\}_{n \in \mathbb{Z}} \) is an orthonormal basis, since the correspondence between inner products is easily verified. It suffices to show that the linear span of \( \{e^{in\theta}\}_{n \in \mathbb{Z}} \) is dense in \( L^2 \). Because the “identity injection” of \( (C[-\pi, \pi], \| \cdot \|_\infty) \) into \( (L^2, \| \cdot \|_2) \) is linear and norm-continuous, to prove density it suffices to observe two things: (1) The Stone-Weierstraß theorem implies that the linear span of \( \{e^{in\theta}\}_{n \in \mathbb{Z}} \), which consists exactly of the trigonometric polynomials, is uniformly dense in \( L^2 \). (2) The continuous functions on \( [-\pi, \pi] \) which take the same value at both endpoints are dense in \( L^2 \), because the characteristic functions of any measurable set \( E \subseteq [-\pi, \pi] \) can be \( L^2 \)-norm approximated by the characteristic function of a measurable set \( F \) whose closure \( \overline{F} \) is contained in \( (-\pi, \pi) \), and any such function can be approximated in the \( L^2 \)-norm metric by a continuous function \( 0 \leq f \leq 1 \) whose support is also contained in \( (-\pi, \pi) \). (The reader should check these assertions; they follow easily from the countable additivity of \( m_1 \).) Since any such function can be uniformly approximated—and thus \( L^2 \)-norm approximated—by a trigonometric polynomial, it follows that any simple function can be thus \( L^2 \)-norm approximated, and that suffices to prove the density in \( L^2 \) of the trigonometric polynomials.

What you just read, of course, was a completely “soft” proof that the Fourier series of an \( L^2 \)-function converges to the function in the \( L^2 \) norm. The only bit of “hard” analysis in the proof amounted to checking that the Taylor series of \( \sqrt{1+x} \) has coefficients that are \( o(1/n^\alpha) \) for \( 1 < \alpha < 3/2 \) and that its series thus converges absolutely and uniformly for \( |x| \leq 1 \). The second-semester course will offer much more precise information on the \( L^p \)-convergence of Fourier series.

There are ways to get orthonormal bases of separable Hilbert spaces that don’t involve infinite choice processes. One favorite is to take a sequence of vectors \( \{x_n\}_{n=1}^\infty \subseteq H \) whose (finitistic) linear span is dense in \( H \) and orthogonalize them by the Gram-Schmidt orthogonalization process (for the details, see any good first-course-in-linear-algebra textbook, or see Kenneth M. Hoffman & Ray A. Kunze, *Linear Algebra*, 2nd ed., Prentice-Hall (1971), p. 280 and p. 287(33)). For measures on intervals of \( \mathbb{R} \), a favorite choice of the sequence of vectors is the monomials \( \{x^n\}_{n=0}^\infty \subseteq H \); this choice has certain peculiar properties that lead to closed-form formulas in a number of classical cases. For two examples: if the measure is \( dm_1(x) = \frac{dm(x)}{\sqrt{1-x^2}} \) on the interval \([-1, 1]\), then Gram-Schmidt orthogonalization leads to the Chebyshev polynomials \( T_n(x) \) which are characterized by the fact that \( T_n(\cos \theta) = \cos n\theta \) for all \( n \in \mathbb{N} \). (Easy verification: make the substitution

---

32 Of course, this is the example that started the whole thing.

33 In making the alternative between this and a good (etc.), I am not necessarily implying that this is not a good book, just that it’s not a first-course-(etc.) book. *Honi soit qui mal y pense.*
It is evident that the mapping is real-linear, complex-conjugate-linear, and sends $H$ into $H^*$; indeed, the Schwarz inequality shows that the $H^*$-norm of $(x \mapsto \langle x, y \rangle)$ is $\leq \|y\|$, and since equality is attained in the Schwarz inequality for $x = y$, the mapping is an isometry (and therefore 1-1). To see that it is onto, let $h^* \in H^*$ be given. If $h^* = 0$ there is nothing to prove, so assume $h^* \neq 0$; then its null space $M = \{x : h^*(x) = 0\}$ is a proper closed (because $h^*$ is continuous) subspace of $H$ of codimension 1. We can write $H = M \oplus M^\perp$ where $M^\perp \neq \{0\}$; indeed, by elementary linear algebra, $\dim M^\perp = 1$. Let $e \in M^\perp$ be a unit vector (evidently $\{e\}$ is an orthogonal basis of $M^\perp$) and let $y = h^*(e)e$ (the complex conjugation is vacuous if $\mathbb{K} = \mathbb{R}$). Then for any $x \in H$ orthogonal decomposition gives us

$$x = (x - \langle x, e \rangle e) + \langle x, e \rangle e \quad \text{(evidently } x - \langle x, e \rangle e \in [M^\perp] = M)$$

$$h^*(x) = h^*(x - \langle x, e \rangle e) + \langle x, e \rangle h^*(e)$$

$$= 0 + \langle x, h^*(e)e \rangle = \langle x, y \rangle$$

so $h^* = (x \mapsto \langle x, y \rangle)$.

**Corollary** [Hahn-Banach Theorem for Hilbert spaces]: Let $M$ be a subspace of a Hilbert space $H$ and let $\Psi \in M^*$. Then $\Psi$ can be extended to an element of $H^*$ without change of norm.

**Proof.** Since $\Psi \in M^*$ is continuous, it can be extended to $M$ in a unique way by simply taking limits.35 Thus there is no loss of generality in assuming that $M$ is closed; so $M$ is a Hilbert space in the relativized inner product of $H$, and thus—by the theorem—there is a (unique) element $y \in M$ for which $\Psi(x) \equiv \langle x, y \rangle$ for $x \in M$. But the r. h. s. of that identity defines a linear functional $x \mapsto \langle x, y \rangle$ which obviously extends $\Psi$, and since the norms of $\Psi$ and $y$ were the same and the norm of $x \mapsto \langle x, y \rangle$ on $H$ is the same as $\|y\|$, we have extended $\Psi$ from $M$ to $H$ without increase of norm.

{Remark: In the Hilbert-space situation, the norm-preserving extension of $\Psi$ to an element of $H^*$ is unique: if $x \mapsto \langle x, z \rangle$ is an(other) extension of $\Psi$, then $z - y \in M^\perp$ and so the resolution of $z$ into orthogonal components along $M$ and $M^\perp$ is

$$z = y + (z - y) \in M \oplus M^\perp$$

$$\|z\|^2 = \|y\|^2 + \|z - y\|^2$$

where the norm relation follows from the Pythagorean theorem. But clearly if $z \neq y$ then $\|z\| > \|y\| = \|\Psi\|$, so any choice of extension other than $x \mapsto \langle x, y \rangle$ will increase the norm of the extension. Uniqueness of norm-preserving extension does not hold for Banach spaces in general: any element of $\ell^\infty(2)$ of the form $(1, \mu)$ with $|\mu| \leq 1$ will have $\|\|(1, \mu)\|\| = 1$ and will extend the linear functional $\Psi$ defined on the subspace of $\ell^1(2)$ spanned by the first standard basis vector $(1, 0)^T$ by $\Psi : (\lambda, 0) \mapsto \lambda$. The fact that the $\ell^\infty$ unit ball has “flats” is responsible for this situation. Similar and worse things happen in $L^1([0, 1], \text{Lebesgue}, m_1).$}

---

34 A modest and disarming title for a book that might well have been called “an exceedingly lovely introduction to some basic classical mathematics with which every analyst who doesn’t want to look like a complete idiot should be acquainted.”

35 The argument involved here is the one showing that a uniformly continuous function from a dense subspace of a metric space $(X,d)$ to a complete metric space $(Y,\rho)$ has a unique continuous extension to all of $X$, obtained by taking limits along convergent sequences.