Reproducing Kernel Hilbert Spaces (RKHS)

The fundamental problem we are looking at in this chapter of the notes is how to fit a function to a series of point evaluations. We are given data

$$(\boldsymbol{t}_1, y_1), \ldots, (\boldsymbol{t}_M, y_m), \quad \text{with } \ \boldsymbol{t}_m \in \mathbb{R}^D \text{ and } y_m \in \mathbb{R},$$

and want to find $f : \mathbb{R}^D \to \mathbb{R}$ such that

$$f(\boldsymbol{t}_m) \approx y_m, \quad m = 1, \ldots, M. \tag{1}$$

If $\boldsymbol{f}$ is in a Hilbert space $\mathcal{S}$, we can search for a least-squares solution by solving

$$\underset{\boldsymbol{f} \in \mathcal{S}}{\text{minimize}} \sum_{m=1}^{M} |y_m - f(\boldsymbol{t}_m)|^2 + \delta \|\boldsymbol{f}\|_S^2$$

We have seen that introducing a basis allows us to find the least-squares solution by solving a linear algebra problem. But we also saw in the last set of notes that if we could write $f(\boldsymbol{t}_m)$ as an inner product,

$$f(\boldsymbol{t}_m) = \langle \boldsymbol{f}, \boldsymbol{k} \rangle,$$

for some $\boldsymbol{k} \in \mathcal{S}$ (that would depend on $\boldsymbol{t}_m$), then we can also solve the (possibly infinite dimensional) least squares problem using linear algebra. Working in a *reproducing kernel Hilbert space* (RKHS) allows us to do just that.

Technically, a RKHS is a Hilbert space of functions where the sampling operation at every point (i.e. the operator that takes a function in $\mathcal{S}$ and returns the value of that function at a single point) is a continuous linear functional. We will start slowly, though, by really getting a handle on what the phrase "continuous linear functional" means.

Linear functionals on Hilbert spaces

We start with some definitions that will allow our discussion to be very precise.

A **functional** takes an element of a Hilbert space $\mathcal{S}$ and returns a scalar. We assume throughout these notes that the scalar field is $\mathbb{R}$; the discussion changes very little for complex scalars, you just have to put complex conjugates in the right places. We will use script letters, like this

$$\mathscr{F} : \mathcal{S} \rightarrow \mathbb{R}$$

to denote functionals — this will hopefully make it clear that it operator on a Hilbert space rather than one of the functions that is in the Hilbert space.

Examples:

1. $\mathcal{S} = \mathbb{R}^N$, and

$$\mathscr{F}(\boldsymbol{x}) = x_1 + x_2 + \cdots + x_N,$$

   (sum all the entries in a vector).

2. $\mathcal{S} = \mathbb{R}^3$,

$$\mathscr{F}(\boldsymbol{x}) = 3x_1^2 - x_2 x_3 + x_3.$$

3. $\mathcal{S} = L_2(\mathbb{R})$. Fix $w(t) \in L_2(\mathbb{R})$, and take

$$\mathscr{F}_w(\boldsymbol{f}) = \int_{-\infty}^{\infty} w(t) f(t)\; dt.$$

4. Any norm (not just the induced norm) on any Hilbert space $\mathcal{S}$.

5. Sampling (discrete). Take $\mathcal{S} = \mathbb{R}^N$ and
$$\mathscr{F}_i(\boldsymbol{x}) = x_i.$$
(Returns the ith entry of a vector.)
6. Sampling (continuum). Take $\mathcal{S} = L_2([0,1])$ and
$$\mathscr{F}_\tau(\boldsymbol{f}) = f(\tau).$$
(Returns the function $f(t)$ evaluated at $t = \tau$.)
7. $\mathcal{S} = L_2(\mathbb{R})$. Fix $\lambda > 0$ and set
$$\mathscr{F}_\lambda(\boldsymbol{f}) = \begin{cases} 1, & \sup_{t\in\mathbb{R}} |f(t)| > \lambda, \\ 0, & \text{otherwise} \end{cases}.$$
(Test if a function ever exceeds λ in magnitude.)

Other examples abound.

A functional $\mathscr{F}(\cdot)$ is **continuous** if for every $\epsilon > 0$ there exists a $\delta > 0$ such that
$$\|\boldsymbol{x} - \boldsymbol{y}\|_S \le \delta \quad \Rightarrow \quad |\mathscr{F}(\boldsymbol{x}) - \mathscr{F}(\boldsymbol{y})| \le \epsilon.$$
Qualitatively, this means that as $\boldsymbol{x}$ and $\boldsymbol{y}$ get closer together, $\mathscr{F}(\boldsymbol{x})$ and $\mathscr{F}(\boldsymbol{y})$ must also get closer together.

Examples:

1. All norms are continuous functionals. You proved for homework that
$$|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \ \le \ \|\boldsymbol{x} - \boldsymbol{y}\|.$$
Thus $\|\boldsymbol{x} - \boldsymbol{y}\| \le \epsilon \Rightarrow |\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \le \epsilon$.

2. Let $\mathcal{S}$ be an arbitrary Hilbert space, and let $\boldsymbol{c} \in \mathcal{S}$ be one of its members. Then

$$\mathscr{F}_{\boldsymbol{c}}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{c} \rangle.$$

is continuous. This is a direct result of Cauchy-Schwarz:

$$\begin{aligned} |\mathscr{F}_{\boldsymbol{c}}(\boldsymbol{x}) - \mathscr{F}_{\boldsymbol{c}}(\boldsymbol{y})| &= |\langle \boldsymbol{x}, \boldsymbol{c} \rangle - \langle \boldsymbol{y}, \boldsymbol{c} \rangle| \\ &= |\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{c} \rangle| \\ &\leq \|\boldsymbol{x} - \boldsymbol{y}\| \, \|\boldsymbol{c}\|. \end{aligned}$$

Thus $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \epsilon / \|\boldsymbol{c}\| \;\Rightarrow\; |\mathscr{F}_{\boldsymbol{c}}(\boldsymbol{x}) - \mathscr{F}_{\boldsymbol{c}}(\boldsymbol{y})| \leq \epsilon$.

A functional $\mathscr{F}$ is **linear** if it obeys

$$\mathscr{F}(a\boldsymbol{x} + b\boldsymbol{y}) = a\mathscr{F}(\boldsymbol{x}) + b\mathscr{F}(\boldsymbol{y})$$

for all scalars a, b and $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$.

Examples:

1. Let $\mathcal{S}$ be an arbitrary Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and fix $\boldsymbol{c} \in \mathcal{S}$. Then

$$\mathscr{F}_{\boldsymbol{c}}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{c} \rangle$$

is linear. This follows directly from the properties of inner product:

$$\mathscr{F}_{\boldsymbol{c}}(a\boldsymbol{x} + b\boldsymbol{y}) = \langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{c} \rangle = a\langle \boldsymbol{x}, \boldsymbol{c} \rangle + b\langle \boldsymbol{y}, \boldsymbol{c} \rangle.$$

Combined with the result above, this means that $\mathscr{F}_{\boldsymbol{c}}$ is a "continuous linear functional".

2. Let $\mathcal{S} = L_2([0,1])$ be the space of square-integrable functions on the interval $[0,1]$. Fix $\tau \in [0,1]$. Then the point-evaluation functional

$$\mathscr{F}_\tau(\boldsymbol{f}) = f(\tau),$$

is linear. This follows immediately from how we have defined multiplying functions by a scalar and adding functions together (these are both done pointwise).

We have seen in the examples above that on *every* Hilbert space $\mathcal{S}$ that

$$\mathscr{F}_{\boldsymbol{c}}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{c} \rangle$$

is a **continuous linear functional**. There are two natural questions at this point:

1. Can every continuous linear functional be represented as $\langle \cdot, \boldsymbol{c} \rangle$ for some $\boldsymbol{c} \in \mathcal{S}$? We will see below that this answer is "yes"; this is precisely the content of the *Riesz representation theorem.*

2. Is every linear functional continuous? We will see below that in finite dimensions, the answer is "yes", but in infinite dimensions, the answer is (perhaps surprisingly) "no".

Let's look a little more closely at Question 2 above. We start by showing that in $\mathbb{R}^N$ (and then in any finite dimensional Hilbert space) all linear functionals can be represented as an inner product with a fixed vector. By our results above, this also means that all linear functionals on a finite dimensional Hilbert space are continuous.

Proposition Let $\mathscr{F}$ be a linear functional on an N-dimensional

Hilbert space $\mathcal{S}$. Then there exists a $\boldsymbol{c} \in \mathcal{S}$ such that

$$\mathscr{F}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{c} \rangle.$$

Proof. This is easy. Let's start with $\mathcal{S} = \mathbb{R}^N$, where the inner product can be written

$$\langle \boldsymbol{x}, \boldsymbol{c} \rangle = \boldsymbol{c}^{\mathrm{T}} \boldsymbol{x} = \sum_{n=1}^{N} x_n c_n.$$

Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N$ be the standard basis for $\mathbb{R}^N$ (so $e_j[n] = 1$ for $n = j$ and zero for $n \neq j$). For any $\boldsymbol{x} \in \mathbb{R}^N$, we have the trivial expansion

$$\boldsymbol{x} = \sum_{n=1}^{N} x_n \boldsymbol{e}_n,$$

where the x_n are the entries of $\boldsymbol{x}$. Then

$$\begin{aligned}
\mathscr{F}(\boldsymbol{x}) &= \mathscr{F}\left(\sum_{n=1}^{N} x_n \boldsymbol{e}_n \right) \\
&= \sum_{n=1}^{N} x_n \mathscr{F}(\boldsymbol{e}_n), \quad \text{(by linearity of } \mathscr{F}\text{)} \\
&= \langle \boldsymbol{x}, \boldsymbol{c} \rangle,
\end{aligned}$$

where

$$\boldsymbol{c} = \begin{bmatrix} \mathscr{F}(\boldsymbol{e}_1) \\ \mathscr{F}(\boldsymbol{e}_2) \\ \vdots \\ \mathscr{F}(\boldsymbol{e}_N) \end{bmatrix}.$$

This result also extends easily to all finite dimensional Hilbert spaces. If $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N$ is an orthobasis for an N-dimensional Hilbert space, then

$$
\begin{aligned}
\mathscr{F}(\boldsymbol{x}) &= \sum_{n=1}^{N} \langle \boldsymbol{x}, \boldsymbol{\psi}_n \rangle \mathscr{F}(\boldsymbol{\psi}_n) \\
&= \left\langle \boldsymbol{x}, \sum_{n=1}^{N} \mathscr{F}(\boldsymbol{\psi}_n) \boldsymbol{\psi}_n \right\rangle \\
&= \langle \boldsymbol{x}, \boldsymbol{c} \rangle,
\end{aligned}
$$

where

$$
\boldsymbol{c} = \sum_{n=1}^{N} \mathscr{F}(\boldsymbol{\psi}_n) \boldsymbol{\psi}_n.
$$

Since this sum has a finite number of terms, there is no question about whether or not it converges.

■

So now we know that all linear functionals on a finite-dimensional Hilbert space are continuous and can be represented by a fixed vector in that same Hilbert space (the $\boldsymbol{c}$ above). We also know that in infinite dimensions, some linear functionals, those that can be written $\langle \boldsymbol{x}, \boldsymbol{c} \rangle$ for some $\boldsymbol{c} \in \mathcal{H}$, are continuous. We now look at a simple (but extremely important) example that shows that in an infinite dimensional Hilbert space is continuous, linear functionals need not be continuous (and hence cannot be represented using an inner product).

Let $\mathcal{S} = L_2([0,1])$, and let $\mathscr{F}_\tau$ be the **sampling** or **point evaluation** (sampling) at τ operator:

$$
\mathscr{F}_\tau(\boldsymbol{f}) = f(\tau).
$$

As we discussed above, $\mathscr{F}_\tau$ takes a function $f(t)$, and simply returns the value of that function at $t = \tau$:
Sketch:

Sampling is obviously linear. But on L_2, it is not continuous — there are square-integrable functions on $[0,1]$ that are arbitrarily close to one another, but have radically different samples at a single point. Take, for example

$$f_1(t) = 1, \ 0 \leq t \leq 1, \qquad f_2(t) = \begin{cases} 1, & 0 \leq t \leq 1/2 - \delta^2/2, \\ 0, & 1/2 - \delta^2/2 < t \leq 1/2 + \delta^2/2, \\ 1, & 1/2 + \delta^2/2 < t \leq 1. \end{cases}$$

Sketch:

The distance between $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ is

$$\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_2 = \sqrt{\int_0^1 |f_1(t) - f_2(t)|^2 \, dt} = \delta,$$

so they can be made arbitrarily close to one another by choosing

$\delta > 0$ small enough. But it is always that case that samples at $t = 1/2$ disagree by the same amount:

$$\mathscr{F}_{1/2}(\boldsymbol{f}_1) = 1, \quad \text{and} \quad \mathscr{F}_{1/2}(\boldsymbol{f}_2) = 0.$$

The samples are not getting closer together as δ gets smaller, so $\mathscr{F}_{1/2}$ is discontinuous.

This inconvenient fact, that sampling on L_2 is discontinuous, is the cause of much consternation in signal processing, control systems, and quantum mechanics. Scientists and engineers desperately want to write a sample as an inner product against some function $c(t)$

$$f(\tau) \stackrel{?}{=} \int f(t)c(t)\, dt = \langle \boldsymbol{f}, \boldsymbol{c} \rangle, \quad \boldsymbol{c} \in L_2.$$

But you can't, unless you change your definition of "function" to something completely inane. Typically, this is addressed by using the Dirac delta "function" $\delta(t)$ that is infinite at $t = 0$ and zero everywhere else ... and then basically decreeing that plugging in $c(t) = \delta(t - t_0)$ into the integral expression above yields the sample $f(t_0)$. To say that this is very ill-defined and mathematically problematic would be a dramatic understatement, and whatever this $\delta(t)$ function is, it is certainly not square-integrable.

So, not all linear functionals on infinite dimensional Hilbert spaces are continuous. But all continuous linear functionals can be represented by an inner product with a fixed vector. This is the content of our next result.

Riesz representation theorem. Let $\mathscr{F}$ be a continuous linear functional on $\mathcal{S}$. Then there exists a $\boldsymbol{c} \in \mathcal{S}$ such that

$$\mathscr{F}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{c} \rangle \quad \text{for all} \quad \boldsymbol{x} \in \mathcal{S}.$$

The proof of this fact is in the Technical Details section below. The proof in constructive in that it tells us how to create $\boldsymbol{c}$ for a given $\mathscr{F}$: if $\{\boldsymbol{\psi}_n\}_{n=1}^{\infty}$ is an orthobasis for $\mathcal{S}$, we will have

$$\boldsymbol{c} = \sum_{n=1}^{\infty} \mathscr{F}(\boldsymbol{\psi}_n)\boldsymbol{\psi}_n. \tag{2}$$

The main part of the proof is making sure that the infinite sum above actually makes sense ... that is, it converges in $\mathcal{S}$. As we have seen before, this is the same as showing that $\sum_n |\mathscr{F}(\boldsymbol{\psi}_n)|^2 < \infty$.

So, just to make this absolutely clear, in $L_2([a,b])$ with the standard inner product, all continous linear functionals (though not all linear functions in general) can be written as

$$\mathscr{F}(\boldsymbol{f}) = \int_a^b f(t)c(t)dt, \quad \text{for some } \boldsymbol{c} \in L_2([a,b]).$$

Reproducing kernel Hilbert spaces

Let's return to the particular operation of sampling or point evaluation. We have already seen that we cannot write the point evaluation operator as an inner product on L_2. But there are plenty of (infinite dimensional) Hilbert spaces where point evaluation is continuous. These are called reproducing kernel Hilbert spaces (RKHS). We will explain the "reproducing kernel" part of this name as we proceed below.

Definition. An RKHS is a Hilbert space $\mathcal{K}$ of real-valued functions with domain $\mathbb{R}^D$ in which the sampling operator $\mathscr{S}_\tau : \mathcal{K} \to \mathbb{R}$ is continuous. That is, for each $\boldsymbol{\tau} \in \mathbb{R}^D$ there exists a $\boldsymbol{k}_\tau \in \mathcal{K}$ such that

$$\langle \boldsymbol{f}, \boldsymbol{k}_\tau \rangle = f(\boldsymbol{\tau}), \quad \text{for all} \quad \boldsymbol{f} \in \mathcal{K}.$$

Of course, the $\boldsymbol{k}_\tau$ is itself a function on $\mathbb{R}^D$, and it will in general be a different function for each different $\boldsymbol{\tau}$. We will find it convenient to collect the *ensemble* of functions for all $\boldsymbol{\tau}$ into a single entity called the **kernel**:

$$k(\cdot,\cdot) : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}, \quad k(\boldsymbol{t}, \boldsymbol{\tau}) = k_\tau(\boldsymbol{t}).$$

If our space comes equipped with the standard integral inner product, combining the expressions above gives us a formula for "reproducing" a member of the "Hilbert space" using the "kernel":

$$f(\boldsymbol{t}) = \int_{\boldsymbol{s} \in \mathbb{R}^D} k(\boldsymbol{s}, \boldsymbol{t}) f(\boldsymbol{s}) \, d\boldsymbol{s}.$$

We can quickly construct the kernel in terms of an orthobasis for the RKHS. Let $\{\psi_n\}_{n=1}^\infty$ be an orthobasis for $\mathcal{K}$. Then applying (2) above, we take

$$k(\boldsymbol{t}, \boldsymbol{\tau}) = \sum_{n=1}^\infty \psi_n(\boldsymbol{\tau}) \, \psi_n(\boldsymbol{t}).$$

Of course, there are an infinite number of choices for an orthobasis, but they all yield the same kernel (the individual terms in the sum above may vary, but their sum is the same for every orthobasis).

The Riesz representation theorem and (2) also give us a way to test if a Hilbert space is a RKHS. If $\{\boldsymbol{\psi}_n\}_{n=1}^\infty$ is an orthobasis for a Hilbert space $\mathcal{S}$, then

$$\mathcal{S} \text{ is an RKHS} \quad \Leftrightarrow \quad \sum_{n=1}^{\infty} |\psi_n(\boldsymbol{\tau})|^2 < \infty, \;\; \forall \boldsymbol{\tau} \in \mathbb{R}^D.$$

Kernels and nonorthogonal bases. Sometimes it is more convenient to decompose members of a Hilbert space using a non-orthogonal (Riesz) basis. This is certainly the case with our spline spaces (for which we have a very nice basis, consisting of shifts of a B-spline, that does not happen to be orthogonal). If $\{\boldsymbol{\psi}_n\}_{n=1}^{\infty}$ is a Riesz basis for $\mathcal{S}$, then we can write

$$\boldsymbol{x} = \sum_{n=1}^{\infty} \alpha_n \boldsymbol{\psi}_n, \quad \alpha_n = \langle \boldsymbol{x}, \widetilde{\boldsymbol{\psi}}_n \rangle,$$

where $\{\widetilde{\boldsymbol{\psi}}_n\}$ is the dual basis (see notes on non-orthogonal basees from earlier in the course). We can now say that $\mathcal{S}$ is an RKHS with kernel

$$k(\boldsymbol{t}, \boldsymbol{\tau}) = \sum_{n=1}^{\infty} \psi_n(\boldsymbol{\tau}) \widetilde{\psi}_n(\boldsymbol{t})$$

if and only if

$$\sum_{n=1}^{\infty} |\psi_n(\boldsymbol{\tau})|^2 < \infty, \quad \forall \boldsymbol{\tau} \in \mathbb{R}^D.$$

Examples of RKHS

Remember, RKHS are always spaces of functions.

1. Any finite dimensional Hilbert space is an RKHS. We know this because all linear functionals on a finite dimensional Hilbert

space (of which point evaluation is one) are continuous. Also, if $\mathcal{K}$ is finite dimensional with a basis $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N$, we know that

$$\sum_{n=1}^{N} |\psi_n(\boldsymbol{\tau})|^2 < \infty$$

since there are a finite number of terms in the sum.

2. The space $B_\Omega(\mathbb{R})$ of bandlimited signals with the standard (integral) inner product. We have already seen that a sample of $\boldsymbol{x} \in B_\Omega(\mathbb{R})$ can be written as

$$f(\tau) = \int_{-\infty}^{\infty} f(t) k_\tau(t)\, dt, \quad \text{where} \quad k_\tau(t) = \frac{\sin(\pi(t-\tau)/T)}{\pi(t-\tau)},$$

for any $\tau \in \mathbb{R}$.

3. The space $\mathcal{S}_{L,\mathbb{Z}}(\mathbb{R})$ of Lth-order polynomial splines on the real line. We have seen that a (Riesz) basis for this space is given by

$$\psi_n(t) = b_L(t-n), \quad n \in \mathbb{Z}.$$

Even though there are an infinite number of basis functions, we know

$$\sum_{n=-\infty}^{\infty} |\psi_n(\tau)|^2 < \infty,$$

since for every τ, there are only a finite number of non-zero terms in the sum above (since $b_L(t)$ is non-zero only on a finite interval).

One key thing that we will realize is that it is often far more natural to characterize a RKHS by the kernel $k(\boldsymbol{t}, \boldsymbol{\tau})$ rather than a direct description of the elements in the space, or a basis for the space.

Technical Details: Proof of Riesz Representation Theorem

Let $\{\boldsymbol{\psi}_n\}_{n=1}^{\infty}$ be an orthobasis for $\mathcal{S}$. Let $\boldsymbol{\beta}$ be the sequence of numbers created by applying the linear operator to each of the members of the orthobasis:

$$\beta_n = \mathscr{F}(\boldsymbol{\psi}_n)$$

For a fixed vector $\boldsymbol{x} \in \mathcal{H}$, let

$$\boldsymbol{x} = \sum_{n=1}^{\infty} \alpha_n \boldsymbol{\psi}_n, \quad \text{where} \quad \alpha_n = \langle \boldsymbol{x}, \boldsymbol{\psi}_n \rangle,$$

be its expansion in the orthobasis. Define $\boldsymbol{x}_N$ as the N-term approximation of the above

$$\boldsymbol{x}_N = \sum_{n=1}^{N} \alpha_n \boldsymbol{\psi}_n.$$

It is clear that

$$\mathscr{F}(\boldsymbol{x}_N) = \sum_{n=1}^{N} \alpha_n \mathscr{F}(\boldsymbol{\psi}_n) = \sum_{n=1}^{N} \alpha_n \beta_n.$$

We know that $\boldsymbol{x}_N \to \boldsymbol{x}$ as $N \to \infty$. Since $\mathscr{F}$ is continuous, this also means that

$$\mathscr{F}(\boldsymbol{x}_N) \to \mathscr{F}(\boldsymbol{x}), \quad \text{as } N \to \infty.$$

Thus we can safely write

$$\mathscr{F}(\boldsymbol{x}) = \sum_{n=1}^{\infty} \alpha_n \beta_n. \tag{3}$$

It is a fact that all continuous linear functionals are also bounded; that is, there exists a constant C such that

$$|\mathscr{F}(\boldsymbol{x})| \leq C\,\|\boldsymbol{x}\|. \tag{4}$$

(It is important to note that C is independent of $\boldsymbol{x}$.) We will prove this fact below. Then (3) becomes

$$\left|\sum_{n=1}^{\infty} \alpha_n \beta_n\right| \leq C \left(\sum_{n=1}^{\infty} \alpha_n^2\right)^{1/2},$$

where we have used the Parseval theorem to replace $\|\boldsymbol{x}\|$ with $(\sum_n \alpha_n^2)^{1/2}$. Since (3) holds for all $\boldsymbol{x}$ and hence all $\{\alpha_n\}$, it holds for the particular choice of

$$\alpha_n = \begin{cases} \beta_n, & n = 1, \ldots, N, \\ 0, & \text{otherwise.} \end{cases}$$

Thus we have

$$\left|\sum_{n=1}^{N} \beta_n^2\right| \leq C \left(\sum_{n=1}^{N} \beta_n^2\right)^{1/2},$$

or in other words

$$\left(\sum_{n=1}^{N} \beta_n^2\right)^{1/2} \leq C.$$

Since C is a constant, the left-hand side is uniformly bounded for all N, and thus converges to something finite as $N \to \infty$,

$$\left(\sum_{n=1}^{\infty} \beta_n^2\right)^{1/2} \leq C.$$

The take-away is that the sequence $\{\beta_n\}$ is square-summable. Thus by the Parseval theorem

$$\mathscr{F}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{c} \rangle, \quad \text{where } \boldsymbol{c} = \sum_{n=1}^{\infty} \beta_n \boldsymbol{\psi}_n = \sum_{n=1}^{\infty} \mathscr{F}(\boldsymbol{\psi}_n)\boldsymbol{\psi}_n$$

Note that as a corollary, we now know that $\mathscr{F}$ is a continuous linear functional on $\mathcal{S}$ if and only if

$$\sum_{n=1}^{\infty} |\mathscr{F}(\boldsymbol{\psi}_n)|^2 \ < \ \infty$$

for all orthobases $\{\boldsymbol{\psi}_n\}$ for $\mathcal{S}$.

Continuous linear functionals are bounded

We now prove (4). Since $\mathscr{F}$ is continuous at $\mathbf{0}$, for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\|\boldsymbol{v} - \mathbf{0}\| \leq \delta \quad \Rightarrow \quad |\mathscr{F}(\boldsymbol{v}) - \mathscr{F}(\mathbf{0})| \leq \epsilon.$$

Taking $\epsilon = 1$ and realizing that $\mathscr{F}(\mathbf{0}) = 0$, this means that there exists a $\delta > 0$ such that

$$|\mathscr{F}(\boldsymbol{v})| \leq 1 \ \text{ for all } \boldsymbol{v} \text{ such that } \|\boldsymbol{v}\| \leq \delta.$$

Now let $\boldsymbol{x}$ be an arbitrary vector in $\mathcal{S}$. Then

$$\begin{aligned} |\mathscr{F}(\boldsymbol{x})| &= \frac{\|\boldsymbol{x}\|}{\delta} \left| \mathscr{F}\left(\delta \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right) \right| \\ &\leq \frac{\|\boldsymbol{x}\|}{\delta}. \end{aligned}$$

So we can take $C = 1/\delta$ in (4).