Lecture 3

Convergence of sequences in metric spaces

In the previous section, we introduced the idea of a **metric space** (X, d): A set X along with a **metric** d that assigns distances between any two elements $x, y \in X$. It now remains to discuss the important idea of **convergence** in such metric spaces. This will be easy, since you have already seen the idea of convergence of sequences of real numbers.

Recall that a sequence $\{x_n\}$ of **real numbers** is said to converge to a limit $x \in \mathbb{R}$ if the following holds true:

Given any $\epsilon > 0$, there exists an integer $N_\epsilon > 0$ such that

$$|x_n - x| < \epsilon \qquad \text{for all} \;\; n > N_\epsilon. \tag{1}$$

The ϵ-subscripted N, i.e., N_ϵ, indicates that the value of N will generally be dependent upon the value of ϵ. (Most probably, the smaller we make ϵ, thus "squeezing" the tail of the sequence closer to x, the larger the required value of N.)

But look again at the LHS of the inequality in (1) – not surprisingly, it is the **distance** between x_n and the limit x, i.e., $|x_n - x| = d(x_n, x)$, where the metric d on the set of real numbers was discussed earlier. Therefore, we can rewrite the above requirement for convergence of a sequence $\{x_n\}$ as follows,

Given any $\epsilon > 0$, there exists an integer $N_\epsilon > 0$ such that

$$d(x_n, x) < \epsilon \qquad \text{for all} \;\; n > N_\epsilon. \tag{2}$$

This requirement is equivalent to the statement,

$$\lim_{n\to\infty} d(x_n, x) = 0, \tag{3}$$

which is also written as follows

$$d(x_n, x) \to 0 \quad \text{as} \;\; n \to \infty. \tag{4}$$

The beautiful thing is that the above result carries over to metric spaces in general – regardless of what comprises these spaces, e.g., numbers, functions, sets, measures, etc.. We say that a sequence $\{x_n\}$ of elements of a metric space X, i.e., $x_n \in X$ converges to a limit $x \in X$ if

$$d(x_n, x) \to 0 \quad \text{as} \;\; n \to \infty, \tag{5}$$

acknowledging that this is a shorthand notation for the more proper ϵ-N_ϵ definition.

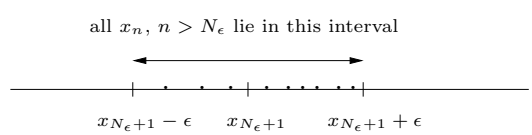
The above definition of convergence is fine if you happen to know the limit x of your sequence $\{x_n\}$. You may then be able "measure" the distances $d(x_n, x)$ and show that this sequence converges to zero. But what if you don't know the limit x? Can you still characterize a sequence $\{x_n\}$ as being convergent, or possibly convergent, if you don't actually know the limit x? The mathematician Cauchy struggled with this problem in his study of the real numbers and came up with the following definition (Cauchy 1821):

> A sequence of real numbers $\{x_n\}$ is said to be a **Cauchy sequence** if, for any $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that
>
> $$|x_n - x_m| < \epsilon \quad \text{for all} \quad n, m > N_\epsilon. \tag{6}$$

In other words, for a given $\epsilon > 0$, all the elements x_n of the infinite "tail" of the sequence, i.e., $\{x_n\,, n > N_\epsilon\}$ lie within ϵ of each another. We can go a little further than this. Given this value of $\epsilon > 0$, the inequality in (6) will be true for the particular case $m = N_\epsilon + 1$. This implies that

$$|x_n - x_{N_{\epsilon+1}}| < \epsilon \quad \text{for all} \quad n > N_\epsilon. \tag{7}$$

In other words, the distances between x_{N_ϵ} and ALL elements of the tail of the sequence, $\{x_n, n > N_\epsilon\}$, are less than ϵ as sketched in the figure below.



As we make ϵ smaller and smaller, the width 2ϵ of the interval $[x_{N_{\epsilon+1}} - \epsilon, x_{N_{\epsilon+1}} + \epsilon]$ which contains the tail in the above diagram will get smaller and smaller. Of course, as ϵ changes, i.e., gets smaller, the point x_{N_ϵ} could change but it will still have to lie inside one of the previous subintervals of width 2ϵ. If this sequence has a limit (it looks like it will have to have one, but there are still a couple of technicalities that have to be worked out) it will have to lie inside these 2ϵ-width intervals.

Cauchy then proved the following remarkable result:

> (Cauchy 1821) A sequence of real numbers $\{x_n\}$ is convergent (to a limit $x \in \mathbb{R}$) if and only if it is a Cauchy sequence.

The "if and only if", i.e., equivalence, of Cauchy sequences and convergence to a limit, is true because of the **completeness** of the real number system. We'll return to this idea in a moment.

The idea of a Cauchy sequence is easily brought into a general metric space setting:

> Let (X, d) be a metric space. Then a sequence of elements $\{x_n\}$ is said to be a **Cauchy sequence** in (X, d) if, for any $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that
>
> $$d(x_n, x_m) < \epsilon \quad \text{for all} \quad n, m > N_\epsilon. \tag{8}$$

We have simply replaced the distance function $|x-y|$ on the real number line with the metric $d(x, y)$ in our metric space. This definition is now applicable to real numbers, ordered n-tuples in $\mathbb{R}^n$, functions, etc..

In general, however, Cauchy's "if and only if" result for sequences of real numbers may not hold. We do have the following result, however, which is the $B \Rightarrow A$ part of Cauchy's $A \Leftrightarrow B$ result for real numbers:

Theorem: Let $\{x_n\}$ be a convergent sequence in a metric space (X, d). Then $\{x_n\}$ is a Cauchy sequence.

The proof of this theorem is so straightforward that it is worthy of mention here. (It is a direct analogue of the proof for real numbers.)

Proof: Since $\{x_n\}$ is convergent, there exists an $x \in X$ (the limit of the sequence) such that the following holds: For an $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that $d(x_n, x) < \epsilon/2$ for all $n > N_\epsilon$. (Note that we're using $\epsilon/2$ on the RHS, for reasons that will be clear below – it's quite OK to do this.) Then, for any $n, m > N_\epsilon$, it follows, from the triangle inequality, that

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x) + d(x, x_m) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned} \tag{9}$$

Just to repeat the final conclusion: $d(x_n, x_m) < \epsilon$ for all $n, m > N_\epsilon$. Therefore, by definition, the sequence $\{x_n\}$ is Cauchy, proving the theorem.

And what about the converse, i.e., the $A \Rightarrow B$ part of Cauchy's Theorem? Given a metric space X, are all Cauchy sequences convergent? In other words, if we pick any Cauchy sequence $\{x_n\} \subset X$, does it converge to a limit x in the set X? As much as it might seem that the answer should be "Yes", the answer is actually "No"! The existence of a limit depends upon a particular property of the metric space X in which we are working. Here is an example.

Example: Let $X = \mathbf{Q}$, the set of rational numbers, with metric $d(x, y) = |x - y|$. Consider the following sequence,

$$x_1 = 3, \quad x_2 = \frac{31}{10}, \quad x_3 = \frac{314}{100}, \quad x_4 = \frac{3141}{1000}, \cdots. \tag{10}$$

The element x_n is obtained by truncating the decimal expansion of π to $n - 1$ decimal digits. It can be shown (left as an exercise – it's not that difficult) that this sequence is Cauchy. And it darn well looks like it is convergent, i.e., converging to π. But this limit point π is NOT in the metric space $\mathbf{Q}$ since it is an irrational number. Therefore the sequence $\{x_n\}$ is NOT convergent.

Of course, if you extended the metric space to be the set of real numbers $\mathbb{R}$, then the sequence $\{x_n\}$ is convergent – it converges to an element of $\mathbb{R}$, the non-rational number π. As you may already know, the real numbers are said to be the **completion** of the rational numbers.

For this reason, we have the following definition:

Definition: A metric space (X, d) is said to be **complete** if every Cauchy sequence $\{x_n\}$ converges (to an element $x \in X$).

In other words, completeness does the trick. If a metric space X is complete then Cauchy's result for the real numbers applies to X.

We can now come up with some easy examples:

1. The metric space $X = \mathbb{R}$, the set of real numbers, is complete.

2. The metric space $X = \mathbf{Q}$, the set of rational numbers, is not complete.

3. The closed interval $X = [a, b]$ of real numbers, is complete.

4. The interval $X = (a, b]$ is not complete. (It's possible to construct a sequence which converges to a, which is not an element of X.) Likewise, the interval (a, b) is not complete.

In a short time, we shall arrive at some interesting conclusions regarding metric spaces of **functions**. First, however, we examine the consequence of convergence with respect to a couple of important metrics.

Convergence of sequences of continuous functions with respect to the d_∞ and d_2 metrics

Convergence in d_∞ metric

Here we examine the consequence of convergence of sequences of continuous functions with respect to the d_∞ metric. Recall that for two functions $f, g \in C[a, b]$, the distance between them according to this metric is given by

$$d_\infty(f, g) = \max_{a \le x \le b} |f(x) - g(x)|. \tag{11}$$

The distance between f and g is the maximum difference of their values on $[a, b]$. An important consequence of this is the following: If

$$d_\infty(f, g) < \epsilon \ \text{ implies that } \ |f(x) - g(x)| < \epsilon \ \ \forall\, x, y \in [a, b]. \tag{12}$$

This implies that for all $x \in [a, b]$, the difference between $f(x)$ and $g(x)$ is less than ϵ. This is a very "tight" closeness between the graphs of $f(x)$ and $g(x)$.

Now suppose that we have a sequence of functions $f_n \in C[a, b]$ that converges to a limit function $f \in C[a, b]$ with respect to the d_∞ metric. This means that

$$d_\infty(f_n, f) \to 0 \quad \text{as} \ \ n \to \infty. \tag{13}$$

Using the ϵ definition: For any $\epsilon > 0$, there exists an $N_\epsilon > 0$ such that

$$d_\infty(f_n, f) < \epsilon \quad \text{for all} \ \ n > N_\epsilon. \tag{14}$$

But from (12), this implies that for all $x \in [a, b]$,

$$|f_n(x) - f(x)| < \epsilon \quad \text{for all} \ \ n > N_\epsilon. \tag{15}$$

First of all, let us consider a **fixed** value of x in $[a, b]$. As we push ϵ toward 0, the values $f_n(x)$ for $n > N_\epsilon$ are forced to be closer and closer toward $f(x)$. In other words, the real numbers $f_n(x)$

converge to $f(x)$, i.e.,

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{or simply} \quad f_n(x) \to f(x). \tag{16}$$

This is true for any $x \in [a, b]$. This type of convergence is known as **pointwise convergence**.

But there is another important aspect of this result: the convergence is **uniform**: **For any** $x \in [a, b]$, the values of $f_n(x)$ lie within ϵ of $f(x)$.

Another way to see this is to note that Eq. (15) implies the following (from the property of absolute values):

$$f(x) - \epsilon < f_n(x) < f(x) + \epsilon \quad \text{for all} \quad n > N_\epsilon. \tag{17}$$

In other words, the graphs of the $f_n(x)$ functions for $n > N_\epsilon$ **lie inside an ϵ-ribbon** of the graph of $f(x)$. Letting ϵ go to zero reduces the width of this ribbon. The fact that for a given $\epsilon > 0$, the **same** N_ϵ will work for **all** values of x in $[a, b]$ implies uniform convergence. This is made possible by the fact that the interval $[a, b]$ is **closed** and **bounded**.

Convergence in d_2 metric

Recall that for two functions $f, g \in C[a, b]$, the distance between them according to the d_2 metric is given by

$$d_2(f, g) = \left[\int_a^b [f(x) - g(x)]^2 \, dx\right]^{1/2}. \tag{18}$$

Note that

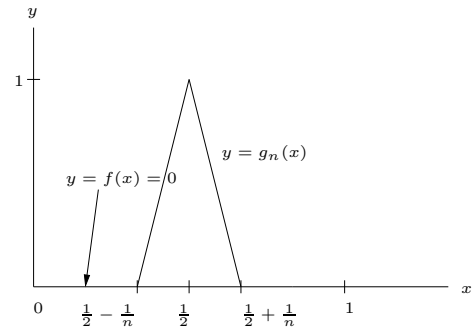
$$d_2(f, g) < \epsilon \quad \text{implies that} \quad \left[\int_a^b [f(x) - g(x)]^2 \, dx\right]^{1/2} < \epsilon. \tag{19}$$

This metric involves an integration over the functions f and g and not a comparison of their values over the interval. Unlike the case for the d_∞ metric, Eq. (19) does **not** imply that the maximum difference between $f(x)$ and $g(x)$ values must be less that ϵ. It is possible that the values of $f(x)$ and $g(x)$ can differ significantly, but over a sufficiently small interval.

To illustrate, consider the following example. Let $f(x) = 0$ and $g_n(x)$, $n = 2, 3, \cdots$, be a sequence of functions be defined over $[0, 1]$ as follows: For $n \geq 2$, define

$$g_n(x) = \begin{cases} 0, & 0 \leq x \leq \frac{1}{2} - \frac{1}{n}, \\ 1 + n\left(x - \frac{1}{2}\right), & \frac{1}{2} - \frac{1}{n} \leq x \leq \frac{1}{2}, \\ 1 - n\left(x - \frac{1}{2}\right), & \frac{1}{2} \leq x \leq \frac{1}{2} + \frac{1}{n}, \\ 0, & \frac{1}{2} + \frac{1}{n} \leq x \leq 1. \end{cases} \tag{20}$$

The graph of this seemingly complicated formula is a "hat" centered at $x = 1/2$, as sketched below.



We now compute the d_2 distance between f and g_n. After a little algebra, we find that

$$\begin{aligned} d_2(f, g_n) &= \left[\int_0^1 [f(x) - g_n(x)]^2 \, dx\right]^{1/2} \\ &= \left[\int_{\frac{1}{2}-\frac{1}{n}}^{\frac{1}{2}+\frac{1}{n}} [g_n(x)]^2 \, dx\right]^{1/2} \\ &= \sqrt{\frac{2}{3n}}. \end{aligned} \tag{21}$$

We can make n as large as we please, thereby making the d_2 distance between f and g_n as small as desired. In other words, $d_2(f, g_n) \to 0$ as $n \to \infty$.

But note that for all $n \geq 2$, $f(1/2) - g(1/2) = 1$, which implies that $d_\infty(f, g_n) = 1$ for all $n \geq 0$. In summary,

1. $d_2(f, g_n) \to 0$ as $n \to \infty$, implying that the (continuous) functions $g_n(x)$ are converging to the (continuous) function $f(x) = 0$ in the d_2 metric. In other words, the set of continuous functions $\{g_n\}$ is convergent in the metric space $(C[0, 1], d_\infty)$ metric, the space of continuous functions on [0,1] with d_∞ metric.

2. Since $d_\infty(f, g_n) = 1$ for all $n \geq 0$, the functions $g_n(x)$ remain at a constant distance of 1 away from the function $f(x) = 0$ for all $n \geq 1$. There is no way that $d_\infty(f, g_n) \to 0$ as $n \to \infty$. In fact, $g_\infty(f, g_n) \to 1$ as $n \to \infty$. Therefore, the (continuous) functions g_n **cannot converge** to the (continuous) function $f(x) = 0$ in the d_∞ metric. In other words, the set of continuous functions $\{g_n\}$ in the metric space $(C[0, 1], d_2)$, the space of continuous functions on [0,1] with d_2 metric.

Comparing these two results, we see that the d_2 metric is, in some way, a more "forgiving" metric. It allows a portion of the functions $g_n(x)$ to deviate significantly from $f(x)$ over small intervals which decrease in length as $n \to \infty$ and gives, in a sense, a "stamp of approval," i.e., it indicates that the

sequence $\{g_n\}$ converges. But there is a complication with the limit of this sequence as we now discuss.

Let us make an observation regarding the g_n functions. If we focus our attention on a fixed value of $x \in [a, b]$, note that the behaviour of the **real numbers** $g_n(x)$ as $n \to \infty$ may be summarized as follows:

1. If $x = 1/2$, then $g_n(1/2) = 1$ for all $n \geq 2$, implying that

$$\lim_{n\to\infty} g_n(1/2) = 1. \tag{22}$$

2. If $x \neq 1/2$, then for some $N > 0$, $g_n(x) = 0$ for all $n > N$. (The hat just gets thinner and thinner.). This implies that

$$\lim_{n\to\infty} g_n(x) = 0, \qquad x \neq 1/2. \tag{23}$$

To summarize, from a pointwise perspective, the sequence of functions $g_n(x)$, in the limit $n \to \infty$, approach the function

$$g(x) = \begin{cases} 0, & x \neq 1/2, \\ 1, & x = 1/2. \end{cases} \tag{24}$$

For this reason, $g(x)$ is known as the **pointwise limit function** of the sequence of functions $\{g_n(x)\}$.

Clearly, $g(x)$ is not a continuous function on [0,1] since it is discontinuous at $x = 1/2$, i.e.,

$$\lim_{x\to(\frac{1}{2})^-} g_n(x) = \lim_{x\to(\frac{1}{2})^+} g_n(x) = 0 \quad \text{but} \quad g\left(\frac{1}{2}\right) = 1\,. \tag{25}$$

Note that $g(x)$ differs from the zero function $f(x) = 0$ only at $x = 1/2$. From the perspective of the d_2 metric, however, these functions are identical, i.e.,

$$d_2(f, g) = \left[\int_0^1 [f(x) - g(x)]^2\, dx\right]^{1/2} = 0. \tag{26}$$

The difference of the functions at the single point $x = 1/2$ does not contribute to the above integral. Another way to look at it: The integral does not "register" the difference of the functions at the single point $x = 1/2$. More on this later.

The question that you may now ask is, "So, which metric is better?" The answer is, "It depends on what you want." If you are working on a problem that demands functions to be close to each other over an interval, e.g., uniform convergence, then you use the d_∞ metric. On the other hand, if the

difference between functions is more aptly described in terms of integrals, then the d_2 metric might be more useful.

In signal and image processing, the space of continuous functions is too restrictive. This should be quite evident in the case of images which, by their very nature, are not continuous functions – every edge of an image, e.g., boundary of an object, represents a curve of discontinuity of the function $u(x, y)$ representing the image.

The fact that you can have convergence of functions in the d_2 metric while still having function values differing by significant amounts will help us to understand the phenomenon of "ringing", i.e., the Gibbs phenomenon, that occurs when approximating functions with Fourier series.

Finally, we comment that the above phenomenon observed for the d_2 metric will also be observed for the other d_p metrics, $p \geq 1$, that are formulated in terms of integrals.

Appendix: Completeness/incompleteness of the function space $C[a, b]$

The material in this section was not covered in the lecture. It is included here for the purpose of "completeness" for those interested. It will not be covered on any examination.

We now return to the idea of completeness, as applied to the space of continuous functions $C[a, b]$ on a closed interval. We shall simply state the following result. Its proof is probably discussed in an advanced course on analysis (and certainly covered in AMATH 731, "Applied Functional Analysis").

Theorem: The metric space $(C[a, b], d_\infty)$ is complete.

In other words, if the metric d_∞ is employed, a Cauchy sequence of continuous functions $\{f_n\} \subset C[a, b]$ converges to a limit function f which is continuous, i.e., $f \in C[a, b]$. (Remember that the sequence $\{f_n\}$ is Cauchy **as measured by the metric** d_∞.)

On the other hand:

Theorem: The metric space $(C[a, b], d_2)$ is not complete.

In other words, sequence of of continuous functions $\{f_n\} \subset C[a, b]$ which is Cauchy with respect to the d_2 metric may converge to a function f that is NOT continuous, i.e., NOT an element of $C[a, b]$. An example was presented earlier – the "hat function" centered at $x = 1/2$. Here is another example.

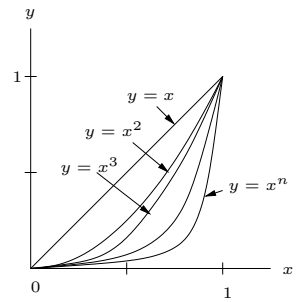
Example: Let $[a, b] = [0, 1]$ and consider the sequence of functions $f_n(x) = x^n$, $n = 1, 2, \cdots$, sketched schematically below.

Clearly, $f_n \in C[a, b]$. Let's examine the behaviour of these functions as $n \to \infty$.

1. For any x such that $0 \leq x < 1$, $x^n \to 0$ as $n \to \infty$.

2. For $x = 1$, clearly, $x^n = 1$.

We'll simply state the fact that the sequence $\{f_n\}$ is Cauchy in d_2 metric (and leave it to the reader as an optional exercise). Instead, we'll focus on the limit f of this sequence which, in this case, can



Some graphs of x^n on $[0, 1]$.

be found rather easily.

Pointwise, it appears that the functions $f_n(x)$ are converging to the function $f(x)$ given by

$$f(x) = \begin{cases} 0, & 0 \le x < 1 \\ 1, & x = 1. \end{cases} \tag{27}$$

Clearly, $f(x)$ is not a continuous function – it is discontinuous at $x = 1$. Let us now examine the distances $d_2(f_n, f)$:

$$\begin{aligned} d_2(f_n, f) &= \left[\int_0^1 [f_n(x) - f(x)]^2 \, dx\right]^{1/2} \\ &= \left[\int_0^1 [x^n - 0]^2 \, dx\right]^{1/2} \\ &= \frac{1}{\sqrt{2n+1}}. \end{aligned} \tag{28}$$

(The single point $x = 0$ at which $f(x) = 1$ can be ignored in the integration, since its contribution to the integral is zero: $f_n(1) - f(1) = 0$. Even if $f_n(1) \neq f(1)$, the contribution to the integral would be zero.) Clearly, $d_2(f_n, f) \to 0$ as $n \to \infty$, i.e., the sequence of functions $\{f_n\}$ is converging to f in d_2 metric.

In summary, we have constructed a sequence of continuous functions f_n that converges, in d_2 metric, to a discontinuous function, i.e., a function that does not belong to $C[0, 1]$. Therefore, the metric space $(C[a, b], d_2)$ is not complete.

The moral of the story

The moral of the story is that if we wish to talk about convergence in the d_2 metric – which we shall have to do in our applications to signal and image processing – then we'll have to extend the class of

functions considered. The space $C[a,b]$ of continuous functions is insufficient. We must be prepared to allow discontinuous functions.

The "completion" of the space $C[a,b]$ with respect to the d_2 metric is the space denoted as "$L^2[a,b]$. It is the metric space of functions for which the following distance is finite,

$$d_2(f,g) = \left[\int_a^b [f(x) - g(x)]^2 \, dx\right]^{1/2} < \infty. \tag{29}$$

More on this later.

End of Appendix

Lecture 4

Normed linear spaces

Metric spaces – sets with distance functions – are very nice, but they are not sufficient for the applications we wish to study. We would like to be able to add and subtract signals and images, even take linear combinations of them. For this reason, we would like to work with spaces that have a vector space structure. In the literature, vector spaces are also called "linear spaces."

Definition: Let X be a real (or complex) vector space. A real-valued function $\|x\|$ defined on X is a norm on X if the following properties are satisfied:

1. (positivity) $\|x\| \geq 0$ for all $x \in X$.
2. (strict positivity) $\|x\| = 0$ if and only if $x = 0$.
3. (triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$.
4. (homogeneity) $\|\alpha x\| = |\alpha|\|x\|$ for any scalar α ($\alpha \in \mathbb{R}$ or $\alpha \in \mathbb{C}$) for any $x \in X$.

$\|x\|$ is the "length" of $x \in X$.

The pair $X, \|\cdot\|)$ is called a **normed linear space**.

The norm $\|\cdot\|$ defines a metric on X,

$$d(x, y) = \|x - y\|, \quad x, y \in X. \tag{30}$$

It can be verified that this metric satisfies all of the required properties of a metric. In particular, the triangle inequality for norms given above guarantees that the above metric satisfies the triangle inequality for metrics: If we replace x with $x - z$ and y with $-y + z$ in Property 3 above, we obtain

$$\|x - y\| \leq \|x - z\| + \|y - z\|, \tag{31}$$

or

$$d(x, y) \leq d(x, z) + d(z, y). \tag{32}$$

As a result, **normed linear spaces are metric spaces**.

If we now consider a normed linear space X as a metric space with metric $d(x, y) = \|x - y\|$, then we may ask whether or not it is complete, i.e., whether all Cauchy sequences in X with respect to this metric converge to an element in X. If so, then we say that the normed linear space X is **complete**.

Definition: A complete normed space is called a **Banach space**.

This is in honour of Stefan Banach (1892-1945), a distinguished Polish mathematician, considered to be the founder of modern functional analysis.

Examples:

1. The space $X = \mathbb{R}^n$. Of course, you are very familiar with the Euclidean magnitude of a vector, i.e., an n-tuple, $x = (x_1, x_2, \cdots, x_n)$ to characterize its length as follows,

$$\|x\|_2 = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}. \tag{33}$$

   This is a particular example of a family of norms that can be assigned to elements of $\mathbb{R}^n$, the p-norms,

$$\|x\|_p = [|x_1|^p + |x_2|^p + \cdots + |x_n|^p]^{1/p}, \quad p \geq 1. \tag{34}$$

   As we saw for the p-metrics, the limit $p \to \infty$ is a special case:

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|. \tag{35}$$

   By virtue of the completeness property of the real numbers, the normed linear spaces $(\mathbb{R}^n, \|\cdot\|_p)$ are complete for all $p \geq 1$ as well as the case $p = \infty$. They are all Banach spaces.

2. $X = C[a, b]$, the space of continuous functions on the interval $[a, b]$. Here, we use the "infinity norm:" For an $f \in C[a, b]$,

$$\|f\|_\infty = \max_{a \leq x \leq b} |f(x)|. \tag{36}$$

   It generates the d_∞ metric studied earlier: For $f, g \in C[a, b]$,

$$d_\infty(f, g) = \|f - g\|_\infty = \max_{a \leq x \leq b} |f(x) - g(x)|. \tag{37}$$

3. The l^p sequence spaces: For a $p \geq 1$,

$$l^p = \{x = (x_1, x_2, \cdots) \mid \sum_{i]1} |x_i|^p < \infty\} \tag{38}$$

We define the p-norm as follows: For an $x = (x_1, x_2, \cdots)$,

$$\|x\|_p = \left[\sum_{i=1}^{\infty} |x_i|^p\right]^{1/p}. \tag{39}$$

As an example, consider the sequence,

$$x = \left(1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{n}, \cdots\right), \tag{40}$$

$x \in l^p$ for $p \geq 2$ but $x \notin l^1$. This follows from the fact that the sum $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges for $p = 2, 3, 4, \cdots$ but not for $p = 1$.

4. The space of real-valued, p-integrable functions on $[a, b]$, denoted as $L^p[a, b]$ and defined as follows:

$$L^p[a, b] = \{f : [a, b] \to \mathbb{R} \mid \int_a^b |f(x)|^p \, dx < \infty\}, \quad p = 1, 2, \cdots. \tag{41}$$

The associated p-norms are given by

$$\|f\|_p = \left[\int_a^b |f(x)|^p \, dx\right]^{1/p}. \tag{42}$$

These norms define the d_p metrics introduced in a previous lecture: For $f, g \in L^p[a, b]$,

$$d_p(f, g) = \|f - g\|_p = \left[\int_a^b |f(x) - g(x)|^p \, dx\right]^{1/p}. \tag{43}$$

The most commonly used normed linear space in this family is the case $p = 2$, namely, the space $L^2[a, b]$ of **square integrable** functions with norm,

$$\|f\|_2 = \left[\int_a^b |f(x)|^2 \, dx\right]^{1/2}. \tag{44}$$

(We say that a function $f \in L^2[a, b]$ is "square integrable" since the integration of its squared magnitude, $|f(x)|^2$, over the interval $[a, b]$ is finite.) The metric associated with this norm is the usual "L^2 metric",

$$d_2(f, g) = \|f - g\|_2 = \left[\int_a^b |f(x) - g(x)|^2 \, dx\right]^{1/2}. \tag{45}$$

As you may recall – and as will certainly be discussed very shortly – this is the function space that is relevant to Fourier series.

Note that for any $p \geq 1$, the space $L^p[a,b]$ includes the space of continuous functions $C[a,b]$. This follows from the fact that continuous functions on a closed interval are bounded: For each $f \in C[a,b]$, there exists an $M \geq 0$ such that $|f(x)| \leq M$ for all $x \in [a,b]$. It then follows that

$$\begin{aligned} \int_a^b |f(x)|^p \, dx &\leq \int_a^b M^p \, dx \\ &= M^p \int_a^b dx \\ &= M^p(b-a) \\ &< \infty. \end{aligned} \tag{46}$$

Therefore $f \in L^p$.

But the L^p spaces include a great deal more: discontinuous functions, even unbounded functions. (The latter won't be needed in our applications.) And the L^p spaces are not identical. For example, consider the example $[a,b] = [0,1]$. The function $f(x) = x^{-1/2}$ is an element of $L^1[0,1]$ since

$$\int_0^1 |f(x)| \, dx = \int_0^1 x^{-1/2} \, dx = 2 < \infty. \tag{47}$$

But it is **not** an element of $L^2[0,1]$ since

$$\int_0^1 |f(x)|^2 \, dx = \int_0^1 x^{-1} \, dx \quad \text{diverges.} \tag{48}$$

Therefore, there are functions which are in $L^1[0,1]$ but not in $L^2[0,1]$, suggesting that $L^2[a,b] \subset L^1[a,b]$. This result can be generalized to $L^p[a,b] \subset L^q[a,b]$ when $p > q$. But this is beyond the scope of the course.

Note: The L^p spaces may easily be extended to include **complex-valued** functions $f[a,b] \to \mathbb{C}$. The definition in (41) holds, with "$\mathbb{R}$" simply replaced by "$\mathbb{C}$".

5. The space of real/complex-valued p-integrable functions on the real line $\mathbb{R}$, i.e.

$$L^p(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \text{ (or } \mathbb{C}) \mid \int_{-\infty}^{\infty} |f(x)|^p \, dx < \infty\}, \quad p = 1, 2, \cdots, \tag{49}$$

with associated p-norms,

$$\|f\|_p = \left[\int_{-\infty}^{\infty} |f(x)|^p \, dx \right]^{1/p}. \tag{50}$$

It will be necessary to consider functions defined on the entire real line in this course. From your previous encounters with improper integrals, you will not be surprised to know that functions

in $L^p(\mathbb{R})$ must satisfy a rather stringent condition – that

$$f(x) \to 0 \quad \text{as} \quad |x| \to \infty. \tag{51}$$

It's actually a little more complicated than this – the **rate** of decay will depend on p.

"Best approximation" in normed linear spaces

We now come to an extremely important concept of this course. You are, of course, very familiar with finite-dimensional normed linear spaces, such as $X = \mathbb{R}^n$. In such cases, which we shall denote as "$\dim(X) = n$", suppose that we have a set of n linearly independent elements $u_i \in X$, $1 \le i \le n$, i.e.,

$$c_1 u_1 + c_2 u_2 + \cdots + c_n u_n = 0 \quad \text{if and only if} \quad c_1 = c_2 = \cdots = c_n = 0. \tag{52}$$

As you know, the $\{u_i\}$ form a **basis** for X: Given an element $v \in X$, there exists a set of coefficients $c_i \in \mathbb{R}$, $1 \le i \le n$, such that

$$v = c_1 u_1 + c_2 u_2 + \cdots + c_n u_n. \tag{53}$$

This course is concerned with the approximation of functions which, as stated in a previous lecture, are elements of infinite-dimensional spaces, i.e., $\dim(X) = \infty$. In such cases, we have to be a little more careful. Suppose that we have a set of linearly independent elements $u_i \in X$, $1 \le i \le n$. Define S_n to be the span of the u_i, i.e.,

$$\begin{aligned} S_n &= \text{span}\,\{u_1, u_2, \cdots, u_n\} \\ &= \{x \in X \mid x = c_1 u_1 + c_2 u_2 + \cdots + c_n u_n \quad \text{for some} \quad (c_1, c_2, \cdots, c_n) \in \mathbb{R}^n\}. \end{aligned} \tag{54}$$

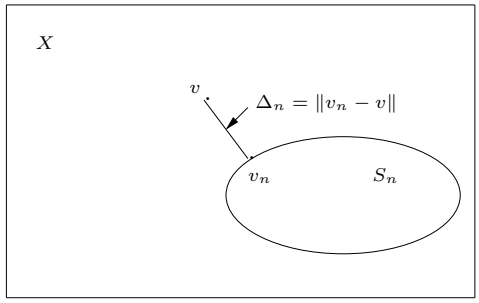
S_n is an n-dimensional subspace of the infinite-dimensional space X.

Now let v be an arbitrary element of X, depicted schematically below. The **best approximation** in S_n to $v \in X$ is given by the element $v_n \in S_n$ that is closest to v in terms of the distance function defined by the norm on X, i.e.,

$$d(v, x) = \|v - x\|. \tag{55}$$

We shall denote this minimum distance as

$$\Delta_n = \min_{x \in S_n} \|x - v\| = \|v_n - v\|. \tag{56}$$



Best (finite-dimensional) approximation $v_n \in S_n$ to $v \in X$.

Once again, v_n is the element in S_n that lies closest to v, **as measured by the metric that is defined by the norm** $\|\cdot\|$ in X.

The above statement regarding v_n, which involves the preceding equation (56), may be written mathematically and more compactly by means of the so-called "arg" notation. We write,

$$v_n = \arg \min_{x \in S_n} \|x - v\| \,. \tag{57}$$

This statement may be read as follows: "v_n is the 'argument' or 'element' in S_n at which the minimization over S_n is achieved". In other words, "v_n is the element in S_n that minimizes the distance to v, i.e., $\|x - v\|$.

(**Note:** At this point we should make a parenthetical remark that, technically speaking, after the phrase "given by the element $v_n \in S_n$" should be added "(provided it exists)". As well, the "min" in the equation should be replaced by "inf". There are pathological situations where a minimum value for the error is approached in some limit by a sequence of approximations, with the actual best approximation not existing. But in most applications, including those considered here, this will not be the case, i.e., Eq. (56) will hold. And in the case of Hilbert spaces, i.e., complete inner product spaces – to be discussed next – such a "minimizer" always exists, and it is unique.)

Since the approximation v_n lies in S_n, it will have an expansion in the basis set u_i, i.e.,

$$v_n = c_1 u_1 + c_2 u_2 + \cdots + c_n u_n. \tag{58}$$

We then write that

$$v \approx v_n = c_1 u_1 + c_2 u_2 + \cdots + c_n u_n, \tag{59}$$

often adding the phrase "in the norm $\|\cdot\|$ or associated metric on X" (e.g., "in L^2 norm," "in d_2 metric" or even "in L^2 metric"). The error of this approximation or simply "approximation error" is

$$\Delta_n = \|v - v_n\| \,. \tag{60}$$

As we increase n, i.e., the number of basis elements u_i employed, the approximation cannot get worse, i.e., the error Δ_n cannot increase. (If we employ basis elements u_k that are not used, i.e., their associated coefficients are zero, then the approximation error will remain the same.) Hopefully, the Δ_n will decrease as n increases. And, indeed, we would hope that $\Delta_n \to 0$ as $n \to \infty$. Of course, if, for some n, $v \in S_n$, then $\Delta_n = 0$, i.e., $v = v_n$, admitting the expansion in (59).

Note: In the discussion above, there has been no mention of **orthogonality**. This is because normed linear spaces, in general, are not necessarily inner product spaces that are equipped with an inner product, from which comes the property of orthogonality. The special case of best approximation in inner product spaces will be discussed in the next section of this course.

We now consider some examples of best approximations in normed linear spaces.

1. **Example No. 1:** Let $X = C[a,b]$, the normed linear space of continuous functions on $[a,b]$. The functions $u_k(x) = x^{k-1}$, $k = 1, 2, \cdots$, i.e., $\{1, x, x^2, \cdots\}$, form a linearly independent set, hence a basis in $C[0,1]$. In this case, given a function $f \in C[0,1]$, the best approximation in S_n would be the $n-1$-degree polynomial approximation to f having the form,

$$f(x) \cong v_n = c_0 + c_1 x + \cdots c_{n-1} x^{n-1}. \tag{61}$$

If we use the $\|\cdot\|_\infty$ norm on this space, i.e.,

$$\|f\|_\infty = \max_{x\in[a,b]} |f(x)| \,, \tag{62}$$

the best approximation $v_n \in S_n$ is obtained by solving the following optimization problem,

$$\min_{c_0,\cdots,c_{n-1}} \|f - v_n\| = \min_{c_0,\cdots,c_{n-1}} \max_{x\in[a,b]} |f(x) - c_1 - c_1 x - \cdots - c_{n-1} x^{n-1}| \,. \tag{63}$$

From a practical point of view, this is a very complicated optimization problem.

(a) The special case $n = 1$. Here, we use only one function from the basis set, i.e., $u_1(x) = 1$. The best approximation $v_1 \in S_1$ is the best constant approximation to $f(x)$ in the d_∞

metric, i.e., $f(x) \simeq c$, where c is obtained by minimizing the distance function,

$$\Delta_\infty(c) = d_\infty(f,c) = \|f - c\|_\infty = \max_{x\in[a,b]} |f(x) - c| \,. \tag{64}$$

If the formula for $f(x)$ exists, we may be able to remove the absolute value by considering the various intervals over which the integrand, $f(x)-c$, is positive and negative. In Problem Set No. 1, you are asked to find the best constant approximation to $f(x) = x^2$ on [0,1] using the above distance function.

2. **Example No. 2:** Let $X = L^1[a,b]$, the normed linear space of functions on $[a,b]$ satisfying the condition that

$$\|f\|_1 = \int_a^b |f(x)|\, dx < \infty \,. \tag{65}$$

Recall that the metric associated with this norm is given by

$$d_1(f,g) = \|f - g\|_1 = \int_a^b |f(x) - g(x)|\, dx \quad f,g \in L^1[a,b] \,. \tag{66}$$

Recall that the space of continuous functions $C[a,b]$ is a subset of this space. The functions $u_k(x) = x^{k-1}$, $k = 1, 2, \cdots$, also form a basis in this space. Given a function $f \in L^1[0,1]$, its best approximation in S_n is given by the solution to the following optimization problem,

$$\min_{c_0,\cdots,c_{n-1}} \|f - v_n\|_1 = \min_{c_0,\cdots,c_{n-1}} \int_a^b |f(x) - c_0 - c_1 x - \cdots - c_{n-1}x^{n-1}|\, dx \,. \tag{67}$$

This problem actually turns out to be a bit more tractable – it can be solved by the method of **linear programming** since the optimization problem in linear in the unknowns c_i.

(a) The special case $n = 1$. Here, we use only one function from the basis set, i.e., $u_1(x) = 1$. The best approximation $v_1 \in S_1$ is the best constant approximation to $f(x)$ in the L^1 metric, i.e., $f(x) \simeq c$, where we minimize the distance function,

$$\Delta_1(c) = d_1(f,c) = \|f - c\|_1 = \int_a^b |f(x) - c|\, dx \,, \tag{68}$$

where we have replaced c_0 with c. Because the integrand contains an absolute value we cannot use differentiation methods to find the c-value which minimizes $\Delta_1(c)$. If the formula for $f(x)$ exists and is not too complicated, it may be able to evaluate the integral by evaluating it on separate intervals over which $f(x) - c$ is positive and negative. In Problem Set No. 1, you are asked to find the best constant approximation to $f(x) = x^2$ on [0,1] using the above distance function.

3. **Example No. 3:** Let $X = L^2[a,b]$, the normed linear space of functions on $[a,b]$ satisfying the condition

$$\|f\|_2^2 = \int_a^b [f(x)]^2\,dx < \infty\,. \tag{69}$$

Recall that the metric associated with this norm is given by

$$d_2(f,g) = \|f-g\|_2 = \left[\int_a^b [f(x)-g(x)]^2\,dx\right]^{1/2}, \quad f,g \in L^1[a,b]\,. \tag{70}$$

Once again, the space of continuous functions $C[a,b]$ is a subset of this space. The functions $u_k(x) = x^{k-1}$, $k = 1,2,\cdots$, also form a basis in this space. Given a function $f \in L^2[0,1]$, its best approximation in S_n is given by the solution to the following optimization problem,

$$\min_{c_0,\cdots,c_{n-1}} \|f - v_n\|_2^2 = \min_{c_0,\cdots,c_{n-1}} \int_a^b [f(x) - c_0 - c_1 x - \cdots - c_{n-1}x^{n-1}]^2\,dx\,. \tag{71}$$

Note that we have chosen to minimize the **squared L^2 distance** since it avoids square roots. (Minimizing the square of a positive function $h(x)$ is the same as minimizing $h(x)$.) This problem can, in principle, be solved analytically by finding the stationary points of the squared distance function,

$$\Delta_2^2(c_0,c_1,\cdots,c_{n-1}) = \int_a^b [f(x) - c_0 - c_1 x - \cdots - c_{n-1}x^{n-1}]^2\,dx\,. \tag{72}$$

This is one of the reasons that the metric associated with the space L^2 is used in most signal and image processing applications. We'll illustrate with two special cases.

(a) The special case $n = 1$. Here again, we use only one function from the basis set, i.e., $u_1(x) = 1$. The best approximation $v_1 \in S_1$ is once again the best constant approximation to $f(x)$, but this time with respect to the L^2 metric, i.e., $f(x) \simeq c$, where we minimize the squared L^2 distance function,

$$h(c) = \Delta_1^2(c) = [d_2(f,c)]^2 = \|f-c\|_2^2 = \int_a^b [f(x)-c]^2\,dx\,. \tag{73}$$

There are two ways to solve this minimization problem.

- **Method No. 1:** Expand the integrand and integrate to produce an expression for $\Delta_2^2(c)$ in terms of c:

$$\begin{aligned} h(c) = \Delta_2^2(c) &= \int_a^b [f(x)]^2\,dx - 2c\int_a^b f(x)\,dx + c^2\int_a^b dx \\ &= \int_a^b [f(x)]^2\,dx - 2c\int_a^b f(x)\,dx + c^2(b-a)\,. \end{aligned} \tag{74}$$

Find critical points:

$$h'(c) = -2\int_a^b f(x)\,dx + 2c(b-a)\,. \tag{75}$$

Setting $h'(c) = 0$ yields the result,

$$c = \frac{1}{b-a}\int_a^b f(x)\,dx\,. \tag{76}$$

Note that $h''(c) = 2(b-a) > 0$, so the critical point is a global minimum.

In summary, the best constant approximation to a function $f(x)$ using the L^2 metric on $[a,b]$ is given by

$$f(x) \simeq \frac{1}{b-a}\int_a^b f(x)\,dx\,. \tag{77}$$

Note that this constant is the **mean** or **average value** of $f(x)$ over $[a,b]$, often denoted as $\bar{f}_{[a,b]}$. This is a very well known result in signal and image processing.

- **Method No. 2:** Use Leibniz' Rule to differentiate the integral in (73),

$$h'(c) = 2\int_a^b [f(x) - c]\,dx = 0\,, \tag{78}$$

which yields the same result for c.

In Problem Set No. 1, you are asked to provide the best piecewise-constant approximation to the function $f(x) = x^2$ on [0,1] using the above method.

(b) The special case $n = 2$. We now use two functions rom the basis set, i.e., $u_1(x) = 1$ and $u_2(x) = x$ to produce the best approximation, $v_2(x)$, to $f(x)$ having the form,

$$f(x) \simeq c_0 + c_1 x\,. \tag{79}$$

We must now minimize the following squared L^2 distance function,

$$h(c_0, c_1) = \Delta_2^2(c_0, c_1) = \|f - c_0 - c_1 x\|_2^2 = \int_a^b [f(x) - c_0 - c_1 x]^2\,dx\,, \tag{80}$$

which is now a function of two variables, c_0 and c_1. Critical points (c_0, c_1) must satisfy the following stationarity conditions,

$$\begin{aligned} \frac{\partial h}{\partial c_0} &= -2\int_a^b [f(x) - c_0 - c_1 x]\,dx = 0\,, \\ \frac{\partial h}{\partial c_1} &= -2\int_a^b [f(x) - c_0 - c_1 x]\,x\,dx = 0\,,. \end{aligned} \tag{81}$$

These conditions yield the following set of inhomogeneous linear equations in c and d,

$$\begin{aligned} \left[\int_a^b dx\right] c_0 + \left[\int_a^b x\,dx\right] c_1 &= \int_a^b f(x)\,dx \\ \left[\int_a^b x\,dx\right] c_0 + \left[\int_a^b x^2\,dx\right] c_1 &= \int_a^b x f(x)\,dx\,. \end{aligned} \tag{82}$$

The integrals on the LHS are easily evaluated in terms of the endpoints a and b. Assuming that the integrals on the RHS involving $f(x)$ can be evaluated, the system is easily solved using Cramer's Rule.

This system of equations may be viewed as a continuous version of the classical method of least squares best approximation of data points $(x_i, f(x_i))$ by the straight line $y = c_0 + c_1 x$.

4. **Example No. 4:** We return to the approximations that are yielded by partial sums of the Fourier series of a function $f(x)$ defined on the interval $[-\pi, \pi]$, i.e.,

$$f(x) \simeq S_N(x) = a_0 + \sum_{n=1}^{N} [a_n \cos nx + b_n \sin nx]\,, \tag{83}$$

where the coefficients a_n and b_n are given in Eq. (2) in Lecture 2. The relevant normed linear space is $L^2[-\pi, \pi]$ which is also an inner product space. The approximation $S_N(x)$ is an element of a $2N+1$-dimensional space that is spanned by the basis functions,

$$1,\, \cos x,\, \sin x, \cdots,\, \cos Nx,\, \sin Nx\,, \tag{84}$$

Here, we simply state that $S_N(x)$ is the best approximation to $f(x)$ in this space and that the coefficients a_n and b_n are obtained from the inner product defined on this space. We shall justify these statements in the next section, where we discuss approximations in an inner product, or "Hilbert", space.

Once again, we mention that in the discussion preceding the above examples, nothing was said about the "orthogonality" of the basis set u_n. That is because nothing could be said! We have been working with **normed linear spaces**, which are not necessarily **inner product spaces**. Only in inner product spaces, the subject of the next section, can we have the property of orthogonality.

Inner product spaces

Of course, you are familiar with the idea of inner product spaces – at least finite-dimensional ones. Let X be an abstract vector space with an **inner product**, denoted as "$\langle\ ,\rangle$", a mapping from $X \times X$ to $\mathbb{R}$ (or perhaps $\mathbb{C}$). The inner product satisfies the following conditions,

1. $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle, \quad \forall x, y, z \in X$,
2. $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle, \quad \forall x, y \in X, \quad \alpha \in \mathbb{R}$,
3. $\langle x, y\rangle = \langle y, x\rangle, \quad \forall x, y \in X$,
4. $\langle x, x\rangle \geq 0, \ \ \forall x \in X, \quad \langle x, x\rangle = 0$ if and only if $x = 0$.

We then say that $(X, \langle\ ,\rangle)$ is an **inner product space**.

In the case that the field of scalars is $\mathbb{C}$, then Property 3 above becomes

3. $\langle x, y\rangle = \overline{\langle y, x\rangle}, \quad \forall x, y \in X$,

where the bar indicates complex conjugation. Note that this implies that, from Property 2,

$$\langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle, \quad \forall x, y \in X, \quad \alpha \in \mathbb{C}.$$

Note: For anyone who has taken courses in Mathematical Physics, the above properties may be slightly different than what you have seen, as regards the complex conjugations. In Physics, the usual convention is to complex conjugate the first entry, i.e. $\langle \alpha x, y\rangle = \overline{\alpha}\langle x, y\rangle$.

The inner product defines a norm as follows,

$$\langle x, x\rangle = \|x\|^2 \quad \text{or} \quad \|x\| = \sqrt{\langle x, x\rangle}. \tag{85}$$

(**Note:** An inner product always generates a norm. But not the other way around, i.e., a norm is not always expressible in terms of an inner product. You may have seen this in earlier courses – the norm has to satisfy the so-called "parallelogram law.")

And where there is a norm, there is a metric: The norm defined by the inner product $\langle\ ,\rangle$ will define the following metric,

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y\rangle}, \quad \forall x, y \in X. \tag{86}$$

And where there is a metric, we can discuss convergence of sequences, etc..

Lecture 5

Inner product spaces (cont'd)

A complete inner product space is called a **Hilbert space**, in honour of the celebrated mathematician David Hilbert (1862-1943).

Finally, the inner product satisfies the following relation, called the "Cauchy-Schwarz inequality,"

$$|\langle x, y\rangle| \leq \|x\|\|y\|, \qquad \forall x, y \in X. \tag{87}$$

You probably saw this relation in your studies of finite-dimensional inner product spaces, e.g., $\mathbb{R}^n$. It holds in abstract spaces as well.

Examples:

1. $X = \mathbb{R}^n$. Here, for $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$,

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n. \tag{88}$$

   The norm induced by the inner product is the familiar Euclidean norm, i.e.

$$\|x\| = \|x\|_2 = \left[\sum_{i=1}^{n} x_i^2\right]^{1/2}. \tag{89}$$

   And associated with this norm is the Euclidean metric, i.e.,

$$d(x, y) = \|x - y\| = \left[\sum_{i=1}^{n} [x_i - y_i]^2\right]^{1/2}. \tag{90}$$

   You'll note that the inner product generates bf only the $p = 2$ norm or metric. $\mathbb{R}^n$ is also complete, i.e., it is a Hilbert space.

2. $X = \mathbb{C}^n$ – a minor modification of the real vector space case. Here, for $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$,

$$\langle x, y\rangle = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}. \tag{91}$$

   The norm induced by the inner product will be

$$\|x\| = \|x\|_2 = \left[\sum_{i=1}^{n} |x_i|^2\right]^{1/2}, \tag{92}$$

and the associated metric is

$$d(x,y) = \|x-y\| = \left[\sum_{i=1}^{n} |x_i - y_i|^2\right]^{1/2}. \tag{93}$$

$\mathbb{C}^n$ is complete, therefore a Hilbert space.

3. The sequence space l^2 introduced earlier: Here, $x = (x_1, x_2, \cdots) \in l^2$ implies that $\sum_{i=1}^{\infty} x_i^2 < \infty$. For $x, y \in l^2$, the inner product is defined as

$$\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i. \tag{94}$$

Note that l^2 is the only l^p space for which an inner product exists. It is a Hilbert space.

4. $X = C[a,b]$, the space of continuous functions on $[a,b]$ is NOT an inner product space.

5. The space of (real- or complex-valued) square-integrable functions $L^2[a,b]$ introduced earlier. Here,

$$\|f\|^2 = \langle f, f\rangle = \int_a^b |f(x)|^2\, dx < \infty. \tag{95}$$

The inner product in this space is given by

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\, dx, \tag{96}$$

where we have allowed for complex-valued functions.

As in the case of sequence spaces, L^2 is the only L^p space for which an inner product exists. It is also a Hilbert space.

6. The space of (real- or complex-valued) square-integrable functions $L^2(\mathbb{R})$ on the real line, also introduced earlier. Here,

$$\|f\|^2 = \langle f, f\rangle = \int_{-\infty}^{\infty} |f(x)|^2\, dx < \infty. \tag{97}$$

The inner product in this space is given by

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\, dx, \tag{98}$$

Once again, L^2 is the only L^p space for which an inner product exists. It is also a Hilbert space, and will the primary space in which we will be working for the remainder of this course.

Orthogonality in inner product spaces

An important property of inner product spaces is "orthogonality." Let X be an inner product space. If $\langle x, y\rangle = 0$ for two elements $x, y \in X$, then x and y are said to be orthogonal (to each other). Mathematically, this relation is written as "$x \perp y$," just as is done for vectors in $\mathbb{R}^n$.

We're now going to need a few ideas and definitions for the discussion that is coming up.

- Recall that a **subspace** Y of a vector space X is a nonempty subset $Y \subset X$ such that for all $y_1, y_2 \in Y$, and all scalars c_1, c_2, the element $c_1 y_1 + c_2 y_2 \in Y$, i.e., Y is itself a vector space. This implies that Y must contain the zero element, i.e., $y = 0$.

- Moreover the subspace Y is **convex**: For every $x, y \in Y$, the "segment" joining $x + y$, i.e., the set of all convex combinations,

$$z = \alpha x + (1 - \alpha)y, \qquad 0 \leq \alpha \leq 1, \tag{99}$$

  is contained in Y.

- A vector space X is said to be the **direct sum** of two subspaces, Y and Z, written as follows,

$$X = Y \oplus Z, \tag{100}$$

  if each $x \in Z$ has a unique representation of the form

$$x = y + z, \quad y \in Y,\ z \in Z. \tag{101}$$

  The sets Y and Z are said to be **algebraic complements** of each other.

  Note: The concept of an algebraic complement does **not** have to invoke the use of orthogonality. It was mentioned here simply to provide a slightly "larger picture" of the mathematics – in other words, "for information only."

- In $\mathbb{R}^n$ and inner product spaces X in general, it is convenient to consider spaces that are orthogonal to each other. Let $S \subset X$ be a subset of X and define

$$S^{\perp} = \{x \in X \mid x \perp S\} = \{x \in X \mid \langle x, y\rangle = 0 \ \ \forall y \in S\}. \tag{102}$$

  The set $S^{\perp}$ is said to be the **orthogonal complement** of S.

Aside: Some remarks regarding the idea of a "complement"

This section was covered very briefly in the lecture. It is included here only for supplementary purposes. You will not be examined on this material.

The concept of an algebraic complement does **not** have to invoke the use of orthogonality. With thanks to a student who once raised the question of algebraic vs. orthogonal complementarity in class, let us consider the following example.

Let X denote the (11-dimensional) vector space of polynomials of degree 10, i.e.,

$$X = \{\sum_{k=0}^{10} c_k x^k, \quad c_k \in \mathbb{R},\ 0 \leq k \leq 10\}. \tag{103}$$

Equivalently,

$$X = \text{span}\,\{1, x, x^2, \cdots, x^{10}\}. \tag{104}$$

Now define

$$Y = \text{span}\,\{1, x, x^2, x^3, x^4, x^5\}, \quad Z = \text{span}\,\{x^6, x^7, x^8, x^9, x^{10}\}. \tag{105}$$

First of all, Y and Z are subspaces of X. Furthermore, X is a direct sum of the subspaces Y and Z. However, the spaces Y and Z are not orthogonal complements of each other.

First of all, for the notion of orthogonal complmentarity, we would have to define an interval of support, e.g., $[0, 1]$, over which the inner products of the functions is defined. (And then we would have to make sure that all inner products involving these functions are defined.) Using the linearly independent functions, x^k, $- \leq k \leq 10$, one can then construct (via Gram-Schmidt orthogonalization) an orthogonal set of polynomial basis functions, $\{\phi_k(x)\}$, $0 \leq k \leq 10$, over X. It is possible that the first few members of this orthogonal set will contain the functions x^k, $0 \leq k \leq 5$, which come from the set Y. But the remaining members of the orthogonal set will contain higher powers of x, i.e., x^k, $6 \leq k \leq 10$, as well as lower powers of x, i.e., x^k¡ $0 \leq k \leq 5$. In other words, the remaining members of the orthogonal set will not be elements of the set Y – they will have nonzero components in X. See also Example 3 below.

Examples:

1. Let X be the Hilbert space $\mathbb{R}^3$ and $S \subset X$ defined as $S = \text{span}\{(1,0,0)\}$. Then $S^\perp = \text{span}\{(0,1,0),(0,0,1)\}$. In this case both S and $S^\perp$ are subspaces.

2. As before $X = \mathbb{R}^3$ but with $S = \{(c,0,0) \mid c \in [0,1]\}$. Now, S is no longer a subspace but simply a subset of X. Nevertheless $S^\perp$ is the same set as in 1. above, i.e., $S^\perp = \text{span}\{(0,1,0),(0,0,1)\}$. We have to include *all* elements of X that are orthogonal to the elements of S. That being said, we shall normally be working more along the lines of Example 1, i.e., **subspaces** and their orthogonal complements.

3. Further to the discussion of algebraic vs. orthogonal complementarity, consider the same spaces X, Y and Z as defined in Eqs. (104) and (105), but defined over the interval $[-1,1]$. The orthogonal polynomials $\phi_k(x)$ over $[-1,1]$ that may be constructed from the functions x^k, $0 \leq k \leq 10$ are the so-called **Legendre polynomials**, $P_n(x)$, listed below:

$$
\begin{array}{cc}
n & P_n(x) \\
0 & 1 \\
1 & x \\
2 & \frac{1}{2}(3x^2-1) \\
3 & \frac{1}{2}(5x^3-3x) \\
4 & \frac{1}{8}(35x^4-30x^2+3) \\
5 & \frac{1}{8}(63x^5-70x^3+15x) \\
6 & \frac{1}{16}(231x^6-315x^4+105x^2-5) \\
7 & \frac{1}{16}(429x^7-693x^5+315x^3-35x) \\
8 & \frac{1}{128}(6435x^8-12012x^6+6930x^4-1260x^2+35) \\
9 & \frac{1}{128}(12155x^9-25740x^7+18018x^5-4620x^3+315x) \\
10 & \frac{1}{256}(46189x^{10}-109395x^8+90090x^6-30030x^4+3465x^2-63)
\end{array}
\tag{106}
$$

These polynomials satisfy the following orthogonality property,

$$
\int_{-1}^{1} P_m(x)P_n(x)\,dx = \frac{2}{2n+1}\,\delta_{mn}\,, \tag{107}
$$

where δ_{mn} is the Kronecker delta,

$$\delta_{mn} = \begin{cases} 1\,, & m = n\,, \\ 0\,, & m \neq n\,. \end{cases} \tag{108}$$

From the above table, we see that the Legendre polynomials $P_n(x)$, $1 \leq n \leq 5$ belong to space Y, whereas the polynomials P_n, $5 \leq n \leq 10$, do not belong solely to Z. This suggests that the spaces Y and Z are not orthogonal complements. However, the following spaces are orthogonal complements:

$$Y' = \text{span}\{P_0, P_1, P_2, P_3, P_4, P_5\}, \qquad Z' = \text{span}\{P_6, P_7, P_8, P_9, P_{10}\}. \tag{109}$$

(Actually, Y' is identical to Y defined earlier.)

There is, however, another decomposition going on in this space, which is made possible by the fact that the interval $[-1, 1]$ is symmetric with respect to the point $x = 0$. Note that the polynomials $P_n(x)$ are either even or odd. This suggests that we should consider the following subsets of X,

$$Y'' = \{u \in X \mid u \text{ is an even function}\}, \qquad Z'' = \{u \in X \mid u \text{ is an odd function}\}. \tag{110}$$

It is a quite simple exercise to show that any function $f(x)$ defined on an interval $[-a, a]$ may be written as a sum of an even function and an odd function. This implies that any element $u \in X$ may be expressed n the form

$$u = v + w, \quad v \in Y'',\ w \in Z''. \tag{111}$$

Therefore the spaces Y' and Z' are algebraic complements. In terms of the inner product of functions over $[-1, 1]$, however, Y' and Z' are also orthogonal complements since

$$\int_{-1}^{1} f(x)g(x)\,dx = 0, \tag{112}$$

if f and g have different parity.

The discussion that follows will be centered around Hilbert spaces, i.e., complete inner product spaces. This is because we shall need the **closure** properties of these spaces, i.e., that they contain the limit points of all sequences. The following result is very important.

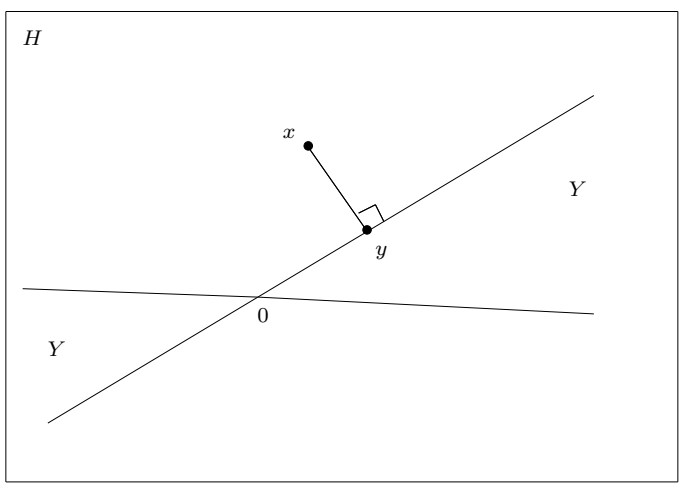
The "Projection Theorem" for Hilbert spaces

Theorem: Let H be a Hilbert space and $Y \subset H$ any closed subspace of H. (Note: This means that Y contains its limit points. In the case of finite-dimensional spaces, e.g., $\mathbb{R}^n$, a subspace is closed. But a subspace of an infinite-dimensional vector space need not be closed.) Now let $Z = Y^{\perp}$. Then for any $x \in H$, there is a unique decomposition of the form

$$x = y + z, \qquad y \in Y, \quad z \in Z = Y^{\perp}. \tag{113}$$

The point y is called the (orthogonal) projection of x on Y.

This is an extremely important result from analysis, and equally important in applications. We'll examine its implications a little later, in terms of "best approximations" in a Hilbert space. In the figure below, we provide a sketch that will hopefully illustrate the situation.



The space Y is contained between the two lines that emanate from 0. Note that Y lies on both sides of 0: If $p \in Y$, then $-p$, which lies on the "other side" of 0, is also an element of Y. The point $y \in Y$ is the orthogonal projection of x onto the set Y and may be viewed as the point of intersection betwenn a line which extends from x to the set Y in such a way that it is "perpendicular" to Y. The examples that we consider should clarify this idea.

From Eq. (113), we can define a mapping $P_Y : H \to Y$, the **projection** of H onto Y so that

$$P_Y : x \to y. \tag{114}$$

Note that

$$P_Y : H \to Y, \quad P_Y : Y \to Y, \quad P_Y : Y^\perp \to \{0\}. \tag{115}$$

Furthermore, P_Y is an **idempotent** operator, i.e.,

$$P_Y^2 = P_Y. \tag{116}$$

This follows from the fact that $P_Y(x) = y$ and $P_Y(y) = y$, implying that $P_Y(P_Y(x)) = P_Y(x)$.

Finally, note that

$$\|x\|^2 = \|y\|^2 + \|z\|^2. \tag{117}$$

This follows from the fact that the norm is defined by means of the inner product:

$$\begin{aligned} \|x\|^2 &= \langle x, x\rangle \\ &= \langle y+z, y+z\rangle \\ &= \langle y, y\rangle + \langle y, z\rangle + \langle z, y\rangle + \langle z, z\rangle \\ &= \|y\|^2 + \|z\|^2, \end{aligned} \tag{118}$$

where the final line results from the fact that $\langle y, z\rangle = \langle z, y\rangle$ since $y \in Y$ and $z \in Z = y^\perp$.

Example: Let $H = \mathbb{R}^3$, $Y = \text{span}\{(1,0,0)\}$ and $Y^\perp = \text{span}\{(0,1,0),(0,0,1)\}$. Then $x = (1,2,3)$ admits the unique expansion,

$$(1,2,3) = (1,0,0) + (0,2,3), \tag{119}$$

where $y = (1,0,0) \in Y$ and $z = (0,2,3) \in Y^\perp$. y is the unique projection of x on Y.

Orthogonal/orthonormal sets of a Hilbert space

Let H denote a Hilbert space. A set $\{u_1, u_2, \cdots u_n\} \subset H$ is said to form an **orthogonal set** in H if

$$\langle u_i, u_j\rangle = 0 \quad \text{for} \quad i \neq j. \tag{120}$$

If, in addition,

$$\langle u_i, u_i\rangle = \|u_i\|^2 = 1, \quad 1 \leq i \leq n, \tag{121}$$

then the $\{u_i\}$ are said to form an **orthonormal set** in H.

You will not be surprised by the following result, since you have most probably seen it in earlier courses in linear algebra.

Theorem: An orthogonal set $\{u_1, u_2, \cdots\}$ not containing the element $\{0\}$ is linearly independent.

Proof: Assume that there are scalars $c_1, c_2, \cdots, c_n$ such that

$$c_1 u_1 + c_2 u_2 \cdots c_n u_n = 0. \tag{122}$$

For each $k = 1, 2, \cdots n$, form the inner product of both sides of the above equation with u_k, i.e.,

$$\langle u_k, c_1 u_1 + c_2 u_2 \cdots + c_n u_n \rangle = \langle u_k, 0 \rangle = 0. \tag{123}$$

By the orthogonality of the u_i, the LHS of the above equation reduces to $c_k \|u_k\|^2$, implying that

$$c_k \|u_k\|^2 = 0, \quad k = 1, 2, \cdots, n. \tag{124}$$

By assumption, however, $u_k \neq 0$, implying that $\|u_k\|^2 \neq 0$. This implies that all scalars c_k are zero, which means that the set $\{u_1, u_2, \cdots, u_n\}$ is linearly independent.

Note: As you have also seen in courses in linear algebra, given a linearly independent set $\{v_1, v_2, \cdots, v_n\}$, we can always construct an orthonormal set $\{e_1, e_2, \cdots, e_n\}$ via the **Gram-Schmidt** orthogonalization procedure. Moreover,

$$\text{span}\{v_1, v_2, \cdots, v_n\} = \text{span}\{e_1, e_2, \cdots, e_n\}. \tag{125}$$

More on this later.

We have now arrived at the most important result of this section.

Best approximation in Hilbert spaces

Recall the idea of the **best approximation** in normed linear spaces, discussed a couple of lectures ago. Let X be an infinite-dimensional normed linear space. Furthermore, let $u_i \in X$, $1 \leq i \leq n$, be a set of n linearly independent elements of X and define the n-dimensional subspace,

$$S_n = \text{span}\{u_1, u_2, \cdots, u_n\}. \tag{126}$$

Then let x be an arbitrary element of X. We wish to find the best approximation to x in the subspace S_n. It will be given by the element $y_n \in S_n$ that lies closest to x, i.e.,

$$y_n = \arg\min_{v \in S_n} \|x - v\|. \tag{127}$$

(The variables used above may be different from those used in the earlier lecture.)

We are going to use the same idea of best approximation, but in a Hilbert space setting, where we have the additional property that an inner product exists in our space. This, of course, opens the door to the idea of **orthogonality**, which will play an important role.

The best approximation in Hilbert spaces may be phrased as follows:

Theorem: Let $\{e_1, e_2, \cdots, e_n\}$ be an orthonormal set in a Hilbert space H. Define $Y = \text{span}\{e_i\}_{i=1}^n$. Y is a subspace of H. Then for any $x \in H$, the best approximation of x in Y is given by the **unique element**

$$y = P_Y(x) = \sum_{k=1}^{n} c_k e_k \qquad \text{(projection of } x \text{ onto } Y\text{)}, \tag{128}$$

where

$$c_k = \langle x, e_k \rangle, \quad k = 1, 2, \cdots, n. \tag{129}$$

The c_k are called the **Fourier coefficients** of x w.r.t. the set $\{e_k\}$.
Furthermore,

$$\sum_{k=1}^{n} |c_k|^2 \leq \|x\|^2. \tag{130}$$

Proof: Any element $v \in Y$ may be written in the form

$$v = \sum_{k=1}^{n} c_k e_k. \tag{131}$$

The best approximation to x in Y is the point $y \in Y$ that minimizes the distance $\|x - v\|$, $v \in Y$, i.e.,

$$y = \arg \min_{v \in Y} \|x - v\|. \tag{132}$$

In other words, we must find scalars $c_1, c_2, \cdots, c_n$ such that the distance,

$$f(c_1, c_2, \cdots, c_n) = \left\| x - \sum_{k=1}^{n} c_k e_k \right\|, \tag{133}$$

is minimized. Here, $f : \mathbb{R}^n$ (or $\mathbb{C}^n$) $\to \mathbb{R}$. It is easier to consider the non-negative squared distance function,

$$g(c_1, c_2, \cdots, c_n) = \left\| x - \sum_{k=1}^{n} c_k e_k \right\|^2 \geq 0\,. \tag{134}$$

Minimizing g is equivalent to minimizing f.

But we now make use of the fact that we are working in an inner product space - the Hilbert space H. The squared distance function may then be expressed as follows,

$$g(c_1, c_2, \cdots, c_n) = \left\langle x - \sum_{k=1}^{n} c_k e_k,\ x - \sum_{l=1}^{n} c_l e_l \right\rangle \geq 0. \tag{135}$$

Here, we consider the real-scalar-valued case, i.e., $c_i \in \mathbb{R}$, which is somewhat simpler than the complex-valued case. (The latter which, of course, includes the former, is proved in an Appendix at the end of this lecture.) Using the linearity properties of the inner product, we can first right-hand side into four components as follows,

$$\begin{aligned} &\left\langle x - \sum_{k=1}^{n} c_k e_k,\ x - \sum_{l=1}^{n} c_l e_l \right\rangle \\ &\quad = \langle x, x \rangle - \left\langle x, \sum_{l=1}^{m} c_l e_l \right\rangle - \left\langle \sum_{k=1}^{n} c_k e_k, x \right\rangle - \left\langle \sum_{k=1}^{n} c_k e_k, \sum_{l=1}^{n} c_l e_l \right\rangle . \end{aligned} \tag{136}$$

The first term on the RHS is simply $\langle x, x \rangle = \|x\|^2$. The second term, using once again the linearity properties of the inner product, may be expressed as follows,

$$\left\langle x, \sum_{l=1}^{n} c_l e_l \right\rangle = \sum_{l=1}^{n} \langle x, c_l e_l \rangle = \sum_{l=1}^{n} c_l \langle x, e_l \rangle\,. \tag{137}$$

Likewise, the third term becomes

$$\left\langle \sum_{k=1}^{n} c_k e_k, x \right\rangle = \sum_{k=1}^{n} c_k \langle e_k, x \rangle = \sum_{k=1}^{n} c_k \langle x, e_k \rangle\,, \tag{138}$$

which we see to be equal to the second term.

The final inner product on the RHS of Eq. (136) becomes,

$$\begin{aligned}\left\langle \sum_{k=1}^{n} c_k e_k, \sum_{l=1}^{n} c_l e_l \right\rangle &= \sum_{k=1}^{n}\sum_{l=1}^{n} c_k c_l \langle e_k, e_l\rangle \\ &= \sum_{k=1}^{n} c_k^2 \,, \end{aligned} \tag{139}$$

where the final line is a result of the orthonormality of the e_n.

From all of these calculations, Eq. (136) becomes,

$$g(c_1, c_2, \cdots, c_n) = \|x\|^2 - 2\sum_{k=1}^{n} c_k \langle x, e_k\rangle + \sum_{k=1}^{n} c_k^2. \tag{140}$$

Recall that we would like to minimize this function of n variables. We first impose the necessary stationarity conditions for a minimum,

$$\frac{\partial g}{\partial c_p} = -2\langle x, e_p\rangle + 2c_p = 0, \quad p = 1, 2, \cdots, n. \tag{141}$$

Therefore,

$$c_p = \langle x, e_p\rangle, \quad p = 1, 2, \cdots, n, \tag{142}$$

which identifies a unique point $\mathbf{c} \in \mathbb{R}^n$, to which corresponds a unique element $y \in Y$. In order to check that this point corresponds to a minimum, we examine the second partial derivatives,

$$\frac{\partial^2 g}{\partial c_q \partial c_p} = 2\delta_{qp}. \tag{143}$$

In other words, the *Hessian* matrix is diagonal and positive definite for all $\mathbf{c} \neq \mathbf{0}$. Therefore the point correponds to a **global minimum**.

Note: There is another way to establish that the point $\mathbf{c}$ is a global minimum: by "completing the square" in Eq. (140). This is actually done in the complex case – see the Appendix.

Finally, substitution of these (real or complex) values of c_k into the squared distance function in Eq. (135) yields the result

$$\begin{aligned} g(c_1, c_2, \cdots c_n) &= \|x\|^2 - 2\sum_{k=1}^{n} |c_k|^2 + \sum_{k=1}^{n} |c_k|^2 \\ &= \|x\|^2 - \sum_{k=1}^{n} |c_k|^2 \\ &\geq 0, \end{aligned} \tag{144}$$

which then implies Eq. (130). The proof of the Theorem is complete.

Some additional comments regarding the best approximation:

1. The above result implies that the element $x \in H$ may be expressed uniquely as

$$x = y + z, \qquad y \in Y = \text{span}\{e_i\}_{i=1}^n, \quad z \in Z = Y^\perp. \tag{145}$$

   To see this, define

$$z = x - y = x - \sum_{k=1}^n c_k e_k, \tag{146}$$

   where the c_k are given by (129). For $l = 1, 2, \cdots, n$, take the inner product of e_l with both sides of this equation to give

$$\begin{aligned} \langle z, e_l \rangle &= \langle x, e_l \rangle - \sum_{k=1}^n c_k \langle e_k, e_l \rangle \\ &= \langle x, e_l \rangle - c_l \\ &= 0. \end{aligned} \tag{147}$$

   Therefore, $z \perp e_l$, $l = 1, 2, \cdots, n$, implying that $z \in Y^\perp$.

2. The term z in (145) may be viewed as the **residual** in the approximation $x \approx y$. Its norm then defines the error of approximation,

$$\|x - y\| = \|z\|. \tag{148}$$

3. As the dimension n of the orthonormal set $\{e_1, e_2, \cdots, e_n\}$ is increased, we expect to obtain better approximations to the element x – unless, of course, x is an element of one of these finite-dimensional spaces, in which case we arrive at zero approximation error, and no further improvement is possible. Let us designate $Y_n = \text{span}\{e_1, e_2, \cdots, e_n\}$, and $y_n = P_{Y_n}(x)$ the best approximation to x in Y_n. Then the error of this approximation is given by

$$\|z_n\| = \|x - y_n\|. \tag{149}$$

   We expect that $\|z_{n+1}\| \leq \|z_n\|$.

4. Note also that the inequality

$$\sum_{k=1}^n |c_k|^2 \leq \|x\|^2 \tag{150}$$

holds for all appropriate values of $n > 0$. (If H is finite-dimensional, i.e., $\dim(H) = N$, then $n = 1, 2, \cdots, N$.) In other words, the partials sums on the left are bounded from above. This inequality, known as **Bessel's inequality**, will have important consequences.

Appendix: Best approximation in the complex-scalar case

We now consider the squared distance function $g(\mathbf{a})$ in Eq. (135) in the case that H is a complex Hilbert space, i.e., $\mathbf{a} \in \mathbb{C}^n$.

$$\begin{aligned}
g(c_1, c_2, \cdots, c_n) &= \left\| x - \sum_{k=1}^{n} c_k e_k \right\|^2 \\
&= \langle x - \sum_{k=1}^{n} c_k e_k, x - \sum_{l=1}^{n} c_l e_l \rangle \\
&= \|x\|^2 - \langle \sum_{k=1}^{n} c_k e_k, x \rangle - \langle x, \sum_{l=1}^{n} c_l e_l \rangle + \sum_{k=1}^{n} \sum_{l=1}^{n} \langle c_k e_k, c_l e_l \rangle \\
&= \|x\|^2 - \sum_{k=1}^{n} \left[\overline{c}_k \langle x, e_k \rangle + c_k \overline{\langle x, e_k \rangle} + |c_k|^2 \right] \\
&= \|x\|^2 + \sum_{k=1}^{n} \left[|\langle x, e_k \rangle|^2 - \overline{c}_k \langle x, e_k \rangle - c_k \overline{\langle x, e_k \rangle} + |c_k|^2 \right] - \sum_{k=1}^{n} |\langle x, e_k \rangle|^2 \\
&= \|x\|^2 + \sum_{k=1}^{n} |\langle x, e_k \rangle - c_k|^2 - \sum_{k=1}^{n} |\langle x, e_k \rangle|^2. \qquad (151)
\end{aligned}$$

The first and last terms are fixed. The middle term is a sum of nonnegative numbers. The minimum value is achieved when all of these terms are zero. Consequently, $f(c_1, \cdots, c_n)$ is a minimum if and only if $c_k = \langle x, e_k \rangle$ for $k = 1, 2, \cdots, n$. As in the real case, we have

$$\|x\|^2 - \sum_{k=1}^{n} |\langle x, e_k \rangle|^2 \ \geq \ 0. \qquad (152)$$