Inner Product Spaces

So far in this chapter, we have studied sets with various levels of structure. We motivated that discussion by a desire to quantify how different any two elements of a set are from each other, leading us to define the concept of a divergence. Stronger assumptions give us a metric. When the underlying set is a vector space, we added yet more requirements to define a norm. Finally, in this section we will introduce another quantity that some vector spaces have for relating their elements: an *inner product*. Rather than telling us how different its arguments are, however, an inner product is more related to the extent to which its arguments are "pointing in the same direction." However, as we'll see, an inner product can always be used to define a norm, making such vector spaces more structured than normed spaces.

Inner product spaces come up in essentially every branch of mathematics, and the extensive theory from functional analysis provides us with a powerful toolbox for working with them. This section will give an overview of the basics of that theory, while focusing on three main ways that it connects to our study of probability. First, one of the L_p spaces ($p = 2$) is an inner product space, so measurable functions in L_2 inherit the rich inner product theory that we will learn about. Second, recall that because L_p spaces assign a norm to probability densities, they can be thought of as providing a norm for the space of probability measures. In this way, L_2 also provides an inner product structure on the space of probability measures (IS THIS REALLY what i wanted to say here?). Finally, we will consider the special case of Bochner expectation in which the range of the random vector is an inner product space.

REVISE THIS INTRO BECAUSE THE original plan was split into three separate sections.

1 Definitions and basics

1.1 Inner product space

An *inner product*(SIDENOTE: Intuitively, you should think of the concept of inner product as a generalization of dot product.) on a real vector space V is a mapping from $V \times V$ to $\mathbb{R}$ such that for any $x, y, z \in V$ and $\alpha, \beta \in \mathbb{R}$, the following following three conditions are satisfied:

- **Symmetry:** $\langle x, y\rangle = \langle y, x\rangle$ (SIDENOTE: If you want to worry about complex vector spaces, you need to use a slightly more general definition of inner product as a mapping to $\mathbb{C}$. The "symmetry" property is replaced by "conjugate symmetry": $\langle x, y\rangle$ equals the complex conjugate of $\langle y, x\rangle$.)
- **Linearity:** $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$ (SIDENOTE: To be more precise, this condition should be called "linearity in the first argument." However, real inner products are symmetric, so linearity in the first argument implies linearity in the second argument as well. For complex inner products, one can use the conjugate symmetry property to see how a linear expression in the second argument works.)
- **Positive definiteness:** $\langle x, x\rangle \geq 0$ with equality iff x is the zero vector

An inner product space is a vector space along with an inner product on its elements. (SIDENOTE: A more general treatment of inner product says that it is a mapping to the vector space's scalar field. In this book, we will limit ourselves to real vector spaces and thus *real inner products*.)

1.2 Example

As a prototypical example, the vector space $\mathbb{R}^n$ along with the familiar *dot product* constitute an inner product space, with $\langle x, y\rangle := x'y$. This is known as *Euclidean n-space*. (SIDENOTE: Notice that this includes the real numbers ($n = 1$); the product operation constitutes an inner product on $\mathbb{R}$.)

In fact, any positive definite $n \times n$ [real symmetric] matrix M provides us with an inner product on $\mathbb{R}^n$. Simply define $\langle x, y\rangle := x'My$. Notice that the ordinary dot product is the case in which M is the identity.

1. Show that $\langle x, y\rangle := x'My$ is indeed an inner product.

This example demonstrates that a single vector space can have multiple inner products defined on it.

1.3 Inner product norm

Any inner product can be used to define a norm by

$$\|x\| := \sqrt{\langle x, x\rangle} \tag{1}$$

Of course, we need to verify that this quantity does indeed meet the norm conditions.

2. Consider $\|x\|$ as defined in (1); recall the definition of a norm from (REF SECTION). First, explain why $\|x\|$ will be a finite real number. Then, we need to confirm three additional conditions to establish that it is truly a norm. Show that it has absolute homogeneity and that it separates points. (To finish proving that this is a norm, you will show that it satisfies the triangle inequality in Exercise (NUMBER).)

Every inner product space can also be considered a normed space by using its inner product norm. Whenever we use the norm notation in the context of an inner product space, we will be referring to the inner product norm unless otherwise specified.

Although every inner product can be used to define a norm, it should be noted that not every norm comes from some corresponding inner product. One common way to prove that a particular norm doesn't correspond to an inner product is to identify a particular pair of vectors doesn't satisfy the *parallelogram equality*.

3. Show that any inner product norm has the parallelogram equality:

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

Why do you think it's called the parallelogram equality?

When a norm *does* correspond to an inner product, that inner product can be recovered in terms of the norm by the *polarization identity*.

4. Prove the *polarization identity* (for real inner product spaces):

$$\langle x, y\rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2)$$

1.4 Cauchy-Schwarz inequality

The following theorem gives the Cauchy-Schwarz inequality, an essential tool that we will use a number of times in the course of this book.

Theorem 1.1 (Cauchy-Schwarz inequality)**.** *Let x and y be two vectors in an inner product space. Then*

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

You will prove this in Exercise (NUMBER).

1.5 Orthogonality

Using an inner product, we can define *angles* between vectors by the relationship

$$\cos\theta = \frac{\langle x, y\rangle}{\|x\|\|y\|} \tag{2}$$

To make sure this θ is well-defined, we need to know that the right-hand side is in $[-1, 1]$. The Cauchy-Schwarz inequality tells us that the absolute value of the numerator is bounded by the denominator, so indeed this definition works. Thinking in terms of angles can be more intuitive in many cases. (SIDENOTE: Given any x and y, infinitely many values of θ will satisfy (2). We will prefer the smallest positive solution, which will lie in $[0, \pi]$.)

When $\langle x, y\rangle = 0$ (and consequently $\theta = \pi/2$), the vectors x and y are said to be *orthogonal* (denoted $x \perp y$), a concept that generalizes the familiar notion of "perpendicular" vectors in Euclidean space. If x is orthogonal to every y in a set S, then we say that x is orthogonal to S ($x \perp S$). If every pair of vectors in S is orthogonal, then we call S orthogonal as well. If in addition to being orthogonal, all the vectors in S have unit norm, then we call S orthonormal.

5. Show that an orthonormal set is linearly independent.
6. Show that the inner product of any vector with the zero vector is zero.
7. Prove the Cauchy-Schwarz inequality (Theorem 1.1). [Hint: start by expanding $\|x - y\|^2$ for unit vectors x and y.] When is equality achieved?
8. Use the Cauchy-Schwarz inequality to compare a vector's L_1 norm to its L_2 norm.
9. Prove that the triangle inequality holds for the norm induced by an inner product.
10. Show that the Pythagorean theorem is true for any Hilbert space.
11. Prove that every inner product is continuous by showing that if $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.
12. Assume $\langle\cdot,\cdot\rangle$ is a symmetric, linear mapping from $V \times V$ to $\mathbb{R}$ that is also positive *semi*-definite. Show how to construct a true inner product on equivalence classes of V.

1.6 Orthonormal basis

With an inner product structure, we can now introduce a special type of "basis." An *orthonormal basis* is an orthonormal set whose linear span is dense in the inner product space. (SIDENOTE: Recall that the linear span of a set of vectors is the set of all finite linear combinations of those vectors.) Note that in general this is not a Hamel basis, because a given vector in the inner product space can't necessarily be represented as a finite linear combination from the orthonormal basis. But it can be approximated arbitrarily well by a finite linear combination from the orthonormal basis. When an orthonormal basis is countable, it is a Schauder basis.

Let H be an inner product space with an orthonormal basis B. The set of inner products of a vector x with the basis vectors $\{\langle x, b\rangle : b \in B\}$ are called the Fourier coefficients of x with respect to B. A remarkable fact is that *any vector only has countably many non-zero Fourier coefficients* (Kreyszig Lemma 3.5-3); let B_x be the countable set indexing the basis vectors that produce non-zero Fourier coefficients for x. It can be shown that x is equal to the countable summation

$$x = \sum_{i \in B_x} \langle x, b_i\rangle b_i \tag{3}$$

Recall from (section on Schauder basis) what this means. The the norm between x and the summation goes to zero as you add terms.

In fact, knowing that $\{b_i : i \in B_x\}$ is a Schauder basis, it is easy to see that the inner products must be the coefficients $\{x_i : i \in B_x\}$, as demonstrated below. (SIDENOTE: This is why the inner products are called the Fourier coefficients.)

$$\begin{aligned} \langle x, b_j\rangle &= \left\langle \sum_i x_i b_i, b_j \right\rangle \\ &= \sum_i x_i \langle b_i, b_j\rangle \qquad \text{recall } \langle b_i, b_j\rangle = 1 \text{ when } i = j\text{, zero otherwise} \\ &= x_j \end{aligned}$$

Understanding that $x_j = \langle x, b_j\rangle$, we express (3) more simply as

$$x = \sum_i x_i b_i$$

This is useful if you want to find the representation of x with respect to $\{b_i : i \in B_x\}$. You can find each coefficient separately, without making any reference to the others.

We can also derive a more concrete representation of inner products in H, which is $\langle x, y\rangle = \sum_i x_i y_i$, where in the sum, i has to range over $B_{xy} := B_x \cup B_y$ which is countable. That is, the inner product is equal to sum of the products of the vectors' non-zero Fourier coefficients.

$$\begin{aligned} \langle x, y\rangle &= \left\langle \sum_i x_i b_i, \sum_j y_j b_j \right\rangle \\ &= \sum_{i,j} x_i y_j \langle b_i, b_j\rangle \qquad \text{valid because inner product is continuous} \\ &= \sum_i x_i y_i \underbrace{\langle b_i, b_i\rangle}_{1} + \sum_{i \neq j} x_i y_j \underbrace{\langle b_i, b_j\rangle}_{0} \\ &= \sum_i x_i y_i \end{aligned} \tag{4}$$

This fact is called *Parseval's identity*. In the finite dimensional case, we call (4) the dot-product. Note that any choice of orthonormal basis is valid; in some cases, one particular choice is more convenient than others. (CITE Pollard and Kreyszig)

Parseval's identity also shows us that the squared norm of a vector is just the sum of its squared non-zero Fourier coefficients.

$$\begin{aligned} \|x\|^2 &:= \langle x, x\rangle \\ &= \sum_i x_i^2 \end{aligned} \tag{5}$$

For a countable orthonormal set S in general (that isn't necessarily an orthonormal basis), we have *Bessel's inequality*:

$$\|x\|^2 \geq \sum_{i \in S} x_i^2$$

where the x_i are the inner products of x with the vectors in S.

COVER Gram-Schmidt here. OR if I've already covered it in Chapter 0, then point out that it extends to general Hilbert spaces.

INCLUDE SOME EXAMPLES from Kreyszig section 3.4 and 3.5 (including Trigonometric series).

I SHOULD use different notations for inner product spaces (H) and Hilbert spaces ($\mathcal{H}$). And likewise for normed spaces (B) and Banach spaces ($\mathcal{B}$).

2 Hilbert spaces

Again, completeness turns out to be a crucial property for some aspects of the theory of inner product spaces. When we studied normed spaces, we gave complete normed spaces a special name: Banach spaces. Likewise, complete inner product spaces (SIDENOTE: To be clear, this completeness must be with respect to the inner product norm.) have their own special name: *Hilbert spaces*.

Recall from (CITE from the normed spaces section), that all finite-dimensional normed spaces are complete. Because inner product spaces are normed spaces, all finite-dimensional inner product spaces are complete. Likewise (CITE section or theorem) tells us that a subspace of a Hilbert space is complete iff it is closed.

Any inner product space can be completed (CITE Kreyszig page 139). That is, given an inner product space G, there exists a Hilbert space H for which G is dense in H. This *completion* is unique up to isomorphism.

In the context of inner product spaces, an isomorphism is a bijective linear operator T that preserves inner products: $\langle Tx, Ty\rangle = \langle x, y\rangle$. Because the inner products determine the norms, any inner product space isomorphism is also a normed space isomorphism. They are identical as inner product spaces.

Any [non-trivial] Hilbert space H has an orthonormal basis. In fact, every orthonormal basis in H has the same cardinality, which is called the *Hilbert dimension* of H. (Kreyszig page 168) (SIDENOTE: The trivial Hilbert space $\{0\}$ is defined to have Hilbert dimension zero.) A remarkable fact about Hilbert spaces is that, in a sense, there "aren't very many" of them: *all Hilbert spaces of the same cardinality are isomorphic to each other.* (SIDENOTE: This includes infinite cardinalities, i.e. there is one countably infinite Hilbert space, one Hilbert space with the cardinality of the reals, etc.)

Recall that for Banach spaces, existence of a Schauder basis implies separability, but not vice versa. For Hilbert spaces, however, these two properties are equivalent: a Hilbert space has a countable orthonormal basis iff it is separable. (Kreyszig 3.6-4)

(WHY is HAVING A countable orthonormal basis is equivalent to having a Schauder basis for Hilbert spaces? Make this an exercise if it's easy.)

Let's return to our prototypical inner product space example. REVISIT Euclidean space $\mathbb{R}^n$ to point out that it is complete (i.e. a Hilbert space) - easy to prove? - if so, do it or make it an exercise. And because it is a finite-dimensional normed space, it is also separable.

2.1 Orthogonal projection

Recall (CITE section) that we defined the divergence between an element x and a [non-empty] set S by

$$D(x, S) := \inf_{y \in S} D(x, y)$$

It is often useful to know whether or not there is some $y \in S$ that actually achieves this infimum divergence, and if so, whether or not it is the only one. (SIDENOTE: One might, for instance, need to select an element from S to use as an approximation for x.) This question is called "existence and uniqueness of the minimizer." It is easy to imagine situations in which existence and uniqueness fail. For instance, if S is a sphere centered at x, then every point in S is the minimizer. On the other hand, if S is the complement of a closed ball centered at x in Euclidean space, then no element in S achieves $D(x, S)$. When there is a unique minimizer, we will call it the *projection* of x onto S. (SIDENOTE: POINT OUT that "affine projections" are a thing too, so be careful with terminology.)

(WILL the above discussion already have happened in a previous section on divergences or examples of divergence spaces?)

For inner product spaces, the divergence in question is the inner product norm of the difference between vectors. The following result is central.

Theorem 2.1 (Orthogonal projection)**.** *Given any x in an inner product space and a [non-empty] complete convex subset C, there exists a vector $x_C \in C$ that is the unique projection of x onto C. (Kreyszig Thm 3.3-1) (SIDENOTE: This result is not true for normed spaces in general.) It is characterized by $\langle x - x_C, y - x_C\rangle \leq 0$ for all $y \in C$. In particular, if C is a subspace (and thus a Hilbert space), then $x - x_C$ is orthogonal to C. (Kreyszig Lemma 3.3-2 and Rychlik Theorem 1) (SIDENOTE: In the case that C is a subspace, the projection x_C is often called the* orthogonal projection *of x onto C.)*

(HOW WOULD ONE SHOW the part of this theorem about characterizing a projection? Maybe proof by contradiction. If there is a $y \in C$ such that the inner product is positive, then PROBABLY that y has a smaller distance from x than x_C does. WORK THIS OUT.)

To prove the last part of Theorem (2.1), one should first show that there exists an $x_C \in C$ such that $x - x_C$ is orthogonal to C. Then by the Pythagorean identity, for any $y \in C$

$$\|x - y\|^2 = \|x - x_C\|^2 + \|x_C - y\|^2$$

The first term doesn't depend on our choice of y; the best we can do is to choose y to minimize the second term. So the choice of $y \in C$ minimizing $\|x - y\|^2$ is $y = x_C$. (SIDENOTE: Because squaring is a monotonically increasing transformation on the non-negative reals, minimizing a squared norm is equivalent to minimizing a norm.)

When C is a complete subspace, the fact that $x - x_C$ is orthogonal to S is often used to find x_C. (IN THE CONVEX [not-necessarily-subspace] case, $x - x_C$ is orthogonal to a supporting hyperplane of S, right? Along with the Pythagorean inequality? This could be an exercise if it's easy enough.)

13. Let x and y be vectors in an inner product space H, and let C be the span of $\{x\}$. Is there a projection of y onto C? If so, find it.

(ANOTHER EXERCISE: R^n projection matrix - derive $M(M'M)^{-1}M'$. AND some of the useful facts about projection matrices - e.g. projection iff symmetric and idempotent. OR will some of this be covered in the Algebra section from Chapter 0? I COULD just state in Chapter 0 that much of it will be put off until it can be covered more generally, even if I *have to* go ahead and state some of the results.) In light of Exercise (NUMBER), we see that the Fourier coefficients of a vector are its projections onto the spans of the orthonormal basis vectors.

For any set C, its *orthogonal complement*, denoted $C^\perp$ is the set of all vectors that are orthogonal to C. It is easy to see that $C^\perp$ is closed under linear combination and is thus a subspace. It can also be shown that $C^\perp$ is a closed set; so if the context is a Hilbert space, then $C^\perp$ is complete as well. In that case, we know that the orthogonal projection $x_{C^\perp}$ exists. In fact,

Theorem 2.2 (Hilbert space)**.** *If H is a Hilbert space and C is a complete subspace, then any $x \in H$ is the sum of its projections onto C and $C^\perp$:*

$$x = x_C + x_{C^\perp}$$

This tells us that $H = C \oplus C^\perp$. (Kreyszig Thm 3.3-4) (SIDENOTE: POINT to DIRECT SUMS coverage in Chapter 0 - Kreyszig page 146.)

ORTHOGONAL projection operator (Kreyszig pages 147-148) - exercise: show bounded and linear. NOTATION: $\mathrm{proj}_C x$

2.2 Representations of inner products

Let's start this section with an exercise to warm up.

14. Consider an inner product $\langle\cdot, y\rangle$ with one of the arguments fixed to be some particular vector y. Show that this is a bounded linear functional.

So any fact that we know for bounded linear functionals also holds for inner products with one fixed argument.

15. Show that if $\langle x, y_1\rangle = \langle x, y_2\rangle$ for all x, then y_1 and y_2 must be the same vector.

This tells us that each vector produces a different bounded linear functional.

Amazingly, it turns out that for Hilbert spaces, the converse of this is also true. Every bounded linear functional is an inner product with one fixed argument.

Theorem 2.3 (Riesz-Fréchet representation theorem)**.** *Let f be a bounded linear functional on a Hilbert space H. Then*

$$f(x) \equiv \langle x, y\rangle$$

for some $y \in H$. This y is uniquely determined by f, and $\|y\|$ is equal to the operator norm of f.

(CITE Kreyszig Theorem 3.8-1 and point to him for the proof.) This result provides a useful characterization of the dual space of a Hilbert space and shows that any Hilbert space is isomorphic to its own dual space.

16. In the context of Theorem 2.3, show that $\|f\| = \|y\|$, and prove that y is unique.

A similar result for bilinear forms is also worth knowing. (MAKE sure bilinear form and quadratic form are defined in the Algebra section of Chapter 0. Refer to that section here.) A bilinear form f whose arguments belong to normed spaces (SIDENOTE: Recall that the arguments of a bilinear form don't have to belong to the same vector space.) is called a *bounded bilinear form* if $|f(x, y)| \leq c\|x\|\|y\|$ for some $c \in \mathbb{R}$; the smallest such c is the bilinear form's norm. (SIDENOTE: One can verify that this satisfies the defining properties of a norm, but we will omit the details.) Similarly to linear functionals, this norm has a more straight-forward representation:

$$\|f\| = \sup_{\|x\|, \|y\| = 1} |f(x, y)|$$

We can see that any inner product is a bounded bilinear form. The bilinearity comes from the linearity and symmetry defining an inner product. The boundedness (with operator norm 1) is a direct observation of the Cauchy-Schwarz inequality. We can generalize this idea in a way that allows the bilinear form arguments to come from different spaces. Let H be an inner product space, B be a normed space, and T be a bounded linear operator from B to H. Then $\langle T\cdot, \cdot\rangle$ is a bounded bilinear form with an argument from B and an argument from H. The bilinearity comes from the bilinearity of inner products along with the linearity of T. The boundedness comes from Cauchy-Schwarz and the boundedness of T: for $x \in B$ and $y \in H$,

$$\begin{aligned} |\langle Tx, y\rangle| &\leq \|Tx\|\|y\| \\ &\leq \|T\|\|x\|\|y\| \end{aligned}$$

Thus every bounded linear operator used within an inner product gives us a bounded bilinear form.

Conversely, for Hilbert spaces, every bounded bilinear form corresponds to using some bounded linear operator within the inner product.

Theorem 2.4 (Riesz representation theorem)**.** *Let H_1 and H_2 be Hilbert spaces and f be a bounded bilinear form mapping from $H_1 \times H_2$ to $\mathbb{R}$. Then*

$$f(x, y) \equiv \langle Tx, y\rangle$$

where T is a bounded linear operator from H_1 to H_2. T is uniquely determined by f and their operator norms are equal.

For a proof of this result, see Kreyszig Theorem 3.8-4 (CITE). (SIDENOTE: By symmetry, the roles of H_1 and H_2 can be interchanged.)

(SIDENOTE: There is some disagreement about the names of these two theorems, along with other related "Riesz representation theorems.")

2.3 Reproducing kernel Hilbert spaces

Let $\mathcal{H}$ be a Hilbert space of real-valued functions on a set $\mathcal{X}$. (SIDENOTE: An element h of any finite-dimensional or countable-dimensional Hilbert space can be thought of as a function mapping from $\mathcal{X} = \{1, 2, \ldots\}$ to the coordinates of h with respect to some basis.) Then a bivariate function K mapping $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a *reproducing kernel* (SIDENOTE: Unfortunately, the word "kernel" has quite a number of different uses in mathematics. Don't strain yourself too much trying to make sense of it!) *of* $\mathcal{H}$ if it satisfies the following two properties:

- For every $x \in \mathcal{X}$, the function $k_x := K(\cdot, x)$ is in $\mathcal{H}$.
- For every $x \in \mathcal{X}$ and $h \in \mathcal{H}$, point evaluations of h are equivalent to taking inner product with the the corresponding kernel functions: $h(x) = \langle h, k_x\rangle$.

If a Hilbert space has a reproducing kernel, it is called a *reproducing kernel Hilbert space* (RKHS). It can be shown that the reproducing kernel of an RKHS is unique.

Let's see another important characterization of RKHSs.

Theorem 2.5. *$\mathcal{H}$ is an RKHS iff for every $x \in \mathcal{X}$, the point evaluation functional L_x is a bounded functional on $\mathcal{H}$.*

Proof. Assume $\mathcal{H}$ has a reproducing kernel K. Then we can show boundedness of the point evaluation functional by using the Cauchy-Schwarz inequality.

$$\begin{aligned} |L_x(h)| &= |h(x)| \\ &= |\langle h, k_x\rangle| \\ &\leq \|h\|\|k_x\| \end{aligned}$$

Recall that for any x, the function k_x is in $\mathcal{H}$ and so must have finite inner product norm $\|k_x\|$. So for any x, the operator norm of L_x is $\|k_x\|$.

Conversely, assume L_x is bounded (SIDENOTE: Recall also that point evaluation is a linear functional - CITE section.) for every $x \in \mathcal{X}$. By Riesz-Fréchet (Theorem 2.3), there exists some $g_x \in \mathcal{H}$ such that $L_x(h) = \langle h, g_x\rangle$. Because $L_x(h) = h(x)$, and because such a g_x exists for any $x \in \mathcal{X}$, this $g_{(\cdot)}$ satisfies the definition of a reproducing kernel for $\mathcal{H}$. □

Recall that all linear functionals on finite-dimensional normed spaces are bounded. So we can see that any finite dimensional Hilbert space must be an RKHS (with "point evaluation" corresponding to finding a coordinate with respect to some basis (SIDENOTE: See the discussion in CITE SECTION - from "examples of divergence and metric spaces")).

In fact, we can guarantee that an even larger set of Hilbert spaces are RKHSs. Any separable Hilbert space has a countable basis; define $\pi_1, \pi_2, \ldots$ to be the coordinate functionals (each mapping $\mathcal{H} \to \mathbb{R}$) with respect to some particular countable basis. Coordinate functionals are bounded linear functionals (see exercise NUMBER - from Bochner expectation or covariance section), and they are also the point evaluation functionals when $h \in \mathcal{H}$ are considered as functions from $\{1, 2, \ldots\}$ to $\mathbb{R}$. So any separable Hilbert space is an RKHS —- EMBEDDED IN SOMETHING???

NEED TO CLARIFY the RKHS is a subspace in a separable Hilbert space (which is of course also a RKHS).

BUT THIS SITUATION is still sometimes true in uncountably infinite dimensional cases - THAT's the point of RKHS - the spaces in some ways continue to behave like Euclidean spaces.

To better understand RKHSs, let's start by thinking about them in finite dimensions. Let $V \subseteq \mathbb{R}^n$ be a vector subspace with an inner product defined on it. In finite dimensions, point evaluation means finding a coordinate, so the kernel vectors must be defined by

$$e_i'v = \langle v, k_i\rangle$$

We know from our discussion above that this is satisfied by a unique $k_i \in V$. We can represent the kernel K as the $n \times n$ matrix of these column vectors $(k_1, \ldots, k_n)$, and $K(i, j)$ means $e_i'Ke_j$. Interestingly, in the finite-dimensional case, there are two other definitions of the reproducing kernel of $(V, \langle\cdot,\cdot\rangle)$ that are equivalent to the original definition.

- The column space of K equals V, and $\langle k_j, k_i\rangle = K_{ij}$.
- $K = u_1u_1' + \ldots + u_ru_r'$ for any orthonormal basis $u_1, \ldots, u_r$ for V.

(POINT READER to the paper that verifies these.) Notice that the second definition makes it clear that K is symmetric. The third definition tells us that K must be psd, because any $a'Ka$ is a sum of squares.

17. Consider the Mahalanobis inner product $\langle u, v\rangle := u'Mv$ for some positive definite $M \in \mathbb{R}^{n \times n}$. Find the kernel K for this inner product space $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$.

In fact, given any n, every Hilbert space $\mathcal{H}$ living in $\mathbb{R}^n$ has a unique kernel $K_{\mathcal{H}} \in \mathbb{R}^{n \times n}$. Likewise, every psd matrix in $\mathbb{R}^{n \times n}$ is the kernel of a unique Hilbert space in $\mathbb{R}^n$. Thus, there is a one-to-one correspondence between psd matrices and Hilbert spaces of the form $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$. So any question regarding one of the types of objects can be transformed into a question regarding the other type, and sometimes that transformation makes the question more manageable.

In infinite-dimensions - does this mean every psd function on $\mathcal{X}$ is the kernel of some Hilbert space on $\mathbb{R}^{\mathcal{X}}$? (NEED to define psd function in a general sense - presumably a symmetric bivariate function on the vectors $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is non-negative when its arguments are equal and that takes $(0, 0)$ to zero.)

Kernel as covariance matrix - extends to countably infinite matrix for stochastic processes.

emphasize isomorphism between kernels and RKHSs. SOMETIMES transforming from one context to the other is useful.

THEREFORE EVERY inner product on R^n is a Mahalonobis inner product! IN FACT, all RKHSs take a generalization of the Mahalanobis inner product form, with K^{-1} taking over the role of Q.

embedding - creating a higher dimensional space with specified distances between points.

3 Examples

REVISIT EXAMPLES FROM BANACH SPACE section (credit Kreyszig for these examples)

- prove that they are inner product spaces? - include $L_2(\mu)$. (Point out that Holder's inequality implies the Cauchy-Schwarz inequality in this case. But of course, we already have the Cauchy-Schwarz inequality because it's an inner product space. So Holder's inequality could perhaps be thought of as a generalization of Cauchy-Schwarz from $L_2(\mu)$ to $L_p(\mu)$.) - VARIANCE (and a generalization) - also define standard deviation. - any example from previous sections that is an inner product space should be shown again. - give an example of a normed space that is not an inner product space and say how it can be proven (parallelogram equality).

- point out completeness and separability or lacks thereof

18. Draw a Venn diagram to show the relationships among the following: inner product spaces, normed spaces, Hilbert spaces, metric spaces, topological spaces, sets, Banach spaces, complete metric spaces, divergence spaces, and complete divergence spaces.