Inner Product Spaces

Every inner product space has an induced norm $\|x\| = \sqrt{\langle x, y\rangle}$. The Cauchy-Schwarz inequality relates norms and inner products.

Theorem 1, (Cauchy-Schwarz)**.** *In an inner product space,* $|\langle v, w\rangle| \leq \|v\|\|w\|$.

Proof. (see Tao [1]) In the real case, expanding $\|v-w\|^2 \geq 0$ is enough. In the complex case, we use *amplification*. Expanding $\|v-w\|^2 \geq 0$, we obtain

$$\mathfrak{Re}\,\langle v, w\rangle \leq \tfrac{1}{2}\|v\|^2 + \tfrac{1}{2}\|w\|^2$$

Observe that the phase rotation $v \mapsto e^{i\theta}v$ preserves the right but not the left-hand side,

$$\mathfrak{Re}\{e^{i\theta}\langle v, w\rangle\} \leq \tfrac{1}{2}\|v\|^2 + \tfrac{1}{2}\|w\|^2$$

Choose $\theta = -\arg\langle v, w\rangle$ to cancel out the phase of $\langle v, w\rangle$. Then,

$$|\langle v, w\rangle| \leq \tfrac{1}{2}\|v\|^2 + \tfrac{1}{2}\|w\|^2$$

Now, amplify again with the homogenization symmetry $(v, w) \mapsto (\lambda v, \frac{1}{\lambda}w)$, for some $\lambda > 0$, preserving the left-hand side but not the right,

$$|\langle v, w\rangle| \leq \tfrac{\lambda^2}{2}\|v\|^2 + \tfrac{1}{2\lambda^2}\|w\|^2$$

If $v, w \neq 0$, the minimum $\|v\|\|w\|$ is achieved when $\lambda^2 = \|w\|/\|v\|$, giving the final inequality. If either is zero, the inequality holds in the limit $\lambda \to 0$ or $\lambda \to +\infty$. ☐

The Cauchy-Schwarz inequality can be used to show that $\langle x, y\rangle$ is continuous in both arguments. It is natural to ask whether every norm comes from an inner product; in general, the answer is no, but the following exercises demonstrate that the *parallelogram law* for norms characterizes inner product spaces.

Exercise 1. (Parallelogram Law) For any inner product space,

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

Exercise 2. (Polarization Identity) Show that a norm $\|\cdot\|$ satisfying the parallelogram law is induced by the inner product

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

Definition 1. An **isometry** is a distance-preserving map. Two inner product spaces are **isomorphic** if there is an invertible isometry between them.

Independence and Span

Definition 2. Vectors $\{v_\alpha\}_{\alpha\in A} \subset V$ are **(algebraically) independent** iff every non-trivial finite linear combination is non-zero, that is, $\sum_{k=1}^n v_{\alpha_k}$ is zero only when $\alpha_1, \ldots, \alpha_n \in \mathcal{C}$ are all zero.

Definition 3. The **algebraic span** of a collection $\mathcal{V} \subset V$ of vectors is the set of finite linear combinations of vectors in $\mathcal{V}$, denoted $\mathrm{span}(\mathcal{V})$. The **closed span** or Hilbert space span is $\overline{\mathrm{span}}(\mathcal{V})$.

Orthogonality

As expected, we call two vectors $v, w \in V$ **orthogonal** if $\langle v, w \rangle = 0$, and orthonormal if both have unit length. The Pythagorean theorem $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ for orthogonal vectors follows easily from sesquilinearity. We also have the following familiar properties.

EXERCISE 3. Let V be an inner product space, and $x \in \operatorname{span}\{v_\alpha\}_{\alpha \in A}$.

I. Every orthonormal system $\{v_\alpha\}_{\alpha \in A}$ is algebraically independent.
II. (Inversion) $x = \sum_{\alpha \in A} \langle x, v_\alpha \rangle v_\alpha$ with finitely many non-zero terms.
III. (Plancherel) $\|x\|^2 = \sum_{\alpha \in A} |\langle x, v_\alpha \rangle|^2$.

EXERCISE 4. (Gram-Schmidt) Let $\mathcal{B} = \{v_1, \dots, v_n\} \subset V$ be orthonormal in an inner product space. For any $x \notin \operatorname{span} \mathcal{B}$. there exists v_{n+1} orthogonal to $\mathcal{B}$ such that $\operatorname{span}\{\mathcal{B}, v_{n+1}\} = \operatorname{span}\{\mathcal{B}, x\}$.

Hilbert Spaces

A **Hilbert space** $\mathcal{H}$ is a complete real or complex inner product space. Unless otherwise stated, we take the metric topology on $\mathcal{H}$ induced by the inner product. Naturally, a subspace of $\mathcal{H}$ is a Hilbert space if and only if it is closed. In particular, proper dense subspaces cannot be Hilbert spaces. Inner product spaces are sometimes called *pre-Hilbert* spaces, since their unique completion is a Hilbert space.

Theorem 2, (Existence of Minimizers)**.** *Let $K \subset \mathcal{H}$ be non-empty, closed, and convex. For any $x \in \mathcal{H}$, there is a unique $y \in K$ of minimum distance $\|y - x\|$ to x. Further, for any $z \in K$ different from y, $\mathfrak{Re} \langle z - y, y - x \rangle \geq 0$.*

Proof. Adapted from Tao, *Epsilon of Room, Vol. 1*, Proposition 1.4.12.

(1) Uniqueness. Observe from the parallelogram law that if y_1 and y_2 are distinct and equidistant from x, then their midpoint $\frac{1}{2}(y_1 + y_2)$ is strictly closer to x. Thus the minimizer, if it exists, is unique.
(2) Angle. If y is the distance minimizer and $z \in K$, the interpolant $(1-\theta)y + \theta z$ is at least as far from x as y for any $0 < \theta < 1$. Squaring and rearranging, we conclude $2\mathfrak{Re} \langle z - y, y - x \rangle + \theta \|z - y\|^2 \geq 0$. In the limit $\theta \to 0$ we obtain the desired inequality.
(3) Existence. Define $D \equiv \inf_{y \in K} \|x - y\| \in [0, +\infty)$. By definition of infimum, there is a sequence $y_1, y_2, \ldots \in K$ whose distances approach the infimum, $\|x - y_k\| \to D$. In fact, this sequence is Cauchy, and therefore converges to some $y \in \mathcal{H}$ with $\|x - y\| = D$. Since K is closed, $y \in K$ must be the minimizer. ∎

Orthogonal Projections

Corollary 1, (Orthogonal Projections)**.** *Let $V \preccurlyeq \mathcal{H}$ be a closed subspace. Then each $x \in \mathcal{H}$ can be expressed uniquely as $x = x_{/\!/} + x_\perp$, where $x_\perp$ is orthogonal to each $v \in V$ and $x_{/\!/} \in V$ is called the* ***orthogonal projection****.*

Proof. We show that $x_{/\!/}$ must be the distance minimizer from V to x.

(1) Uniqueness. Suppose such a decomposition exists. Since $x_{/\!/} \in V$, we have $x_{/\!/} - v \in V$ for any $v \in V$. Thus $x_\perp$ is orthogonal to $x_{/\!/} - v$, and by the Pythagorean theorem,

$$\begin{aligned}\|x - v\|^2 &= \|(x - x_{/\!/}) + (x_{/\!/} - v)\|^2 \\ &= \|x - x_{/\!/}\|^2 + \|x_{/\!/} - v\|^2 \geq \|x - x_{/\!/}\|^2\end{aligned}$$

Therefore, our only choice for $x_{/\!/}$ is the unique distance minimizer, making $x_\perp = x - x_{/\!/}$ unique as well.

(2) Existence. It remains to show that $x_\perp = x - x_{/\!/}$ satisfies the orthogonality condition when $x_{/\!/}$ is the closest point in V to x. Since V is a subspace, for any $v \in V$ and $\lambda \in \mathcal{C}$, we have $x_{/\!/} + \lambda v \in V$. Thus

$$\begin{aligned}\|x - x_{/\!/}\|^2 &\leq \|x - (x_{/\!/} + \lambda v)\|^2 \\ &= \|x - x_{/\!/}\|^2 + |\lambda|^2 \|v\|^2 - 2\mathfrak{Re}\{\lambda \langle v, x_\perp \rangle\}\end{aligned}$$

Then $|\lambda|^2 \|v\|^2 \geq 2\mathfrak{Re}\{\lambda \langle v, x_\perp \rangle\}$. Choose $\lambda \in \mathcal{C}$ to cancel the phase of $\langle v, x_\perp \rangle$ and send $\lambda \to 0$ to obtain $\langle v, x_\perp \rangle = 0$. □

Remark 1. *The* orthogonal projection operator $\pi_V : \mathcal{H} \to V$ *is a linear contraction, and in particular is a bounded, self-adjoint linear operator on $\mathcal{H}$. More to be said later.*

Orthogonal Complements

Property 1. *The **orthogonal complement** of a subspace $V \preccurlyeq \mathcal{H}$ is the set $V^\perp$ of vectors in $\mathcal{H}$ that are orthogonal to every element of V. Then,*

I. *$V^\perp$ is a closed subspace of $\mathcal{H}$ and $(\overline{V})^\perp = V^\perp$.*
II. *$(V^\perp)^\perp = \overline{V}$.*
III. *$V^\perp = \{0\}$ if and only if V is dense in $\mathcal{H}$.*
IV. *If V is closed, then $\mathcal{H} \cong V \oplus V^\perp$.*
V. *Closed $W, V \preccurlyeq \mathcal{H}$ satisfy $(V + W)^\perp = \overline{V^\perp \cap W^\perp}$.*
VI. *Closed $W, V \preccurlyeq \mathcal{H}$ satisfy $(V \cap W)^\perp = \overline{V^\perp + W^\perp}$.*

Proof. The proofs are similar to the finite-dimensional case.

Part I. $V^\perp$ is clearly a subspace and closed by continuity of inner products. The inclusion $V \subset \overline{V}$ gives $(\overline{V})^\perp \subseteq V^\perp$. Conversely, suppose $x \in V^\perp$. Since $\mathcal{H}$ is a metric space, each $v \in \overline{V}$ is the limit of some sequence $(v_n) \subset V$. By continuity again, $\langle x, v \rangle = \lim \langle x, v_n \rangle = 0$, so $x \in \overline{V}^\perp$.

Part II. Decompose $x \in (V^\perp)^\perp$ as $x = x_{/\!/} + x_\perp$, where $x_{/\!/} \in \overline{V}$ and $x_\perp \in (\overline{V})^\perp = V^\perp$. Since $x \perp V^\perp$, we have $x_\perp = 0$ and $x = x_{/\!/} \in \overline{V}$. Conversely, each $x \in \overline{V}$ is the limit of vectors $x_n \in V$ orthogonal to $V^\perp$. By continuity of inner products, for any $w \in V^\perp$, we have $\langle x, w \rangle = \lim_{n\to\infty} \langle x_n, w \rangle = 0$. Hence $x \in (V^\perp)^\perp$.

Part III. If $V^\perp = \{0\}$, then $\overline{V} = (V^\perp)^\perp = \mathcal{H}$, so V is dense. Conversely, if V is dense and $x \in V^\perp$, then the continuous map $v \mapsto \langle v, x \rangle$ vanishes on the dense set V, so must be identically zero. The only vector orthogonal to all others is the zero vector, so $V^\perp = \{0\}$.

Part IV. TODO

Part V. TODO

Part VI. TODO □

Linear Functionals

Theorem 3, (Riesz Representation)**.** *In a complex Hilbert space $\mathcal{H}$, every continuous linear functional $\varphi : \mathcal{H} \to \mathcal{C}$ is an inner product. That is, there is a unique $v \in \mathcal{H}$ such that $\varphi(\cdot) = \langle \cdot, v \rangle$. Furthermore, $\|\varphi\|_{op} = \|v\|$.*

Proof. This result should not be surprising; in finite dimensions, every linear functional analogously has the form $x \overset{\varphi}{\mapsto} c^T x$.

(1) Uniqueness. If $\varphi(\cdot) = \langle \cdot, w \rangle = \langle \cdot, v \rangle$, then $\langle \cdot, v - w \rangle \equiv 0$. Only the zero vector is orthogonal to all others, so we must have $v = w$.

(2) Existence. Assume $\varphi \not\equiv 0$ or the claim is obvious. By continuity, the proper subspace $\ker \varphi \preccurlyeq \mathcal{H}$ is closed. By Property 1, the orthogonal complement contains a nonzero vector $w \in (\ker \varphi)^\perp$. Let $\|w\| = 1$. Since $w \notin \ker \varphi$, we have $\varphi(w) \neq 0$. Now, for any $x \in \mathcal{H}$, the vector $x - \frac{\varphi(x)}{\varphi(w)} w$ lies in $\ker \varphi$, so must be orthogonal to w. Taking inner products, $\langle x, w \rangle - \frac{\varphi(x)}{\varphi(w)} = 0$. Then $\varphi(x) = \left\langle x, \overline{\varphi(w)} w \right\rangle$, as desired.

(3) Norm. Let $\varphi(\cdot) = \langle \cdot, v \rangle$ and recall $\|\varphi\|_{op} = \sup \frac{|\varphi(x)|}{\|x\|}$. For $x \in \mathcal{H}$, Cauchy-Schwarz gives $|\varphi(x)| = |\langle x, v \rangle| \leq \|x\| \|v\|$; thus $\|\varphi\|_{op} \leq \|v\|$. For the reverse direction, notice $\|\varphi\|_{op} \geq |\varphi(v)|/\|v\| = \|v\|$. $\square$

Reproducing Kernel Hilbert Spaces

Riesz representation is fundamental to the theory of reproducing kernels in machine learning. Let $\mathcal{H} = \{f : \mathcal{X} \to \mathbb{R}\}$ be a Hilbert space of functions. The evaluation functional $\varphi_x : f \mapsto f(x)$ is always linear. If the evaluation functional $\varphi_x : f \mapsto f(x)$ has the additional property of being continuous, then $\mathcal{H}$ is called a **reproducing kernel Hilbert space**. Intuitively, this means that

In an RKHS, functions close in norm are pointwise close.

This explains why the Lebesgue space $L^2(\mathbb{R})$ is *not* an RKHS, since L^2 contains equivalence classes of functions which differ on arbitrary null sets. Examples of reproducing spaces include $\mathbb{R}^n$ (interpreted as functions $[n] \to \mathbb{R}$) and the space ℓ_2 of square-summable sequences.

Definition 4. A **positive kernel** is a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with the following properties:

I. (Symmetric) $k(x, y) = k(y, x)$ for all $x, y \in \mathcal{X}$
II. (Positive) For any $x = (x_k)_{k=1}^n \subset \mathcal{X}$ and scalars $\alpha = (\alpha_k)_{k=1}^n \in \mathbb{R}^n$,

$$\alpha^T K(x)\alpha = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) \geq 0$$

In an RKHS, Riesz representation manifests itself in the following form:

Property 2, (Reproducing Property)**.** *To every reproducing kernel Hilbert space $\mathcal{H}$, there corresponds a unique positive kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that*

$$f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{H}} \qquad \forall\, x \in \mathcal{X}, f \in \mathcal{H}$$

Proof. By hypothesis, evaluation $\varphi_x : f \mapsto f(x)$ is a continuous linear functional for any $x \in \mathcal{X}$. Riesz representation gives unique $g_x \in \mathcal{H}$ s.t.

$$\varphi_x(f) = f(x) = \langle f, g_x \rangle_H \qquad \forall\, x \in \mathcal{X}, f \in \mathcal{H}$$

Consider the kernel defined by $k(x, \cdot) = g_x(\cdot)$ for all $x \in \mathcal{X}$. By applying the above property to the functions $g_x, g_y \in \mathcal{H}$, we see that this kernel inherits symmetry from the inner product on $\mathcal{H}$,

$$k(x, y) = g_x(y) = \langle g_x, g_y \rangle_{\mathcal{H}} = \langle g_y, g_x \rangle_{\mathcal{H}} = g_y(x) = k(y, x)$$

Similarly, positivity arises from linearity and the positive-definite property of inner products:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j) = \left\langle \sum_{i=1}^n \alpha_i g_{x_i}, \sum_{i=1}^n \alpha_i g_{x_i} \right\rangle_{\mathcal{H}} \geq 0$$

□

Application: Feature Maps and the Representer Theorem

Above, we've gone from a reproducing space $\mathcal{H}$ to a kernel $k(\cdot,\cdot)$. Conversely, (Schölkopf et al. 2001) show how to construct a unique reproducing kernel Hilbert space starting from only the kernel. Crucially, if $\mathcal{X}$ is any arbitrary set, without defined structure, we can nevertheless think of the pair $(\mathcal{X}, k)$ as (a subset of) a Hilbert space.

The following **representer theorem** shows that a large class of regularized optimization problems on RKHS have solutions that can be expressed as kernel expansions in terms of the training data. Suppose we are given the following components:

- nonempty set $\mathcal{X}$
- positive kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with corresponding RKHS $\mathcal{H}$
- training sample $(x_1, y_1), \dots, (x_n, y_n) \in (\mathcal{X} \times \mathbb{R})$
- strictly increasing $g : [0, +\infty] \to \mathbb{R}$
- arbitrary cost function $c : (X \times \mathbb{R}^2)^m \to \mathbb{R} \cup \{\infty\}$
- function class $\mathcal{F} = \{f(\cdot) = \sum_{k=1}^{\infty} \beta_k k(\cdot, z_k) \mid \beta_k \in \mathbb{R}, z_k \in \mathcal{X}, \|f\|_{\mathcal{H}} < \infty\} \subset \mathbb{R}^{\mathcal{X}}$

Theorem 4, (Nonparametric Representer Theorem)**.** *With the notation above, any $\tilde{f} \in \mathcal{F}$ minimizing the regularized risk functional*

$$f = \arg\min_{f\in\mathcal{F}} \left[c\Big(\big\{ \big(x_n, y_n, f(x_n)\big) \big\}_{n=1}^{N} \Big) + g\big(\|f\|_{\mathcal{H}}\big) \right] \tag{1}$$

admits a representation of the form $f(\cdot) = \sum_{n=1}^{N} \alpha_n k(\cdot, x_n)$.

Proof. Given $(x_1, \dots, x_N)$, consider the orthogonal decomposition of any $f \in \mathcal{F}$ into components parallel and orthogonal to $\overline{\operatorname{span}}\{\phi(x_1), \dots \phi(x_N)\}$,

$$f = v + \sum_{n=1}^{N} \alpha_n \phi(x_n) \quad \text{where} \quad \langle v, \phi(x_n)\rangle_{\mathcal{H}} = 0 \quad \text{for } n = 1, \dots, N$$

The reproducing property gives $f(x_j) = \langle f, \phi(x_j)\rangle$, so application of f to a data point x_k yields

$$f(x_k) = \Big\langle v + \sum_{n=1}^{N} \alpha_n \phi(x_n), \phi(x_k) \Big\rangle = \cancel{\langle v, \phi(x_k)\rangle_{\mathcal{H}}} + \sum_{n=1}^{N} \alpha_n \left\langle \phi(x_n), \phi(x_k) \right\rangle_{\mathcal{H}}$$

Since $f(x_k)$ does not depend on v, neither does the first term in (Eqn. 1). For the second term, since v is orthogonal to $\sum_{n=1}^{N} \alpha_n \phi(x_n)$ and g is strictly monotonic,

$$\begin{aligned} g(\|f\|_{\mathcal{H}}) &= g\Big(\|v + \sum_{n=1}^{N} \alpha_i \phi(x_i)\|_{\mathcal{H}} \Big) \\ &= g\Big(\sqrt{\|v\|_{\mathcal{H}}^2 + \|\sum_{n=1}^{N} \alpha_n \phi(x_n)\|_{\mathcal{H}}^2} \Big) \geq g\Big(\sum_{n=1}^{N} \alpha_n \phi(x_n) \Big) \end{aligned}$$

□

Hilbert Space Adjoints

Theorem 5, (Existence of Adjoints)**.** *For every continuous linear transformation $T \in BC(\mathcal{H}_1, \mathcal{H}_2)$, there is a unique **adjoint** $T^* \in BC(\mathcal{H}_2, \mathcal{H}_1)$ with*

$$\langle Tx, y \rangle_{\mathcal{H}_2} = \langle x, T^*y \rangle_{\mathcal{H}_1} \quad \forall x \in \mathcal{H}_1, y \in \mathcal{H}_2$$

Proof. Concisely, the adjoint sends $y \in \mathcal{H}_2$ to the unique Riesz representation $T^*y \in \mathcal{H}_1$ of the continuous linear functional $x \mapsto \langle Tx, y \rangle_{\mathcal{H}_2}$.

(1) Existence/Uniqueness. Fix $y \in \mathcal{H}_2$. To figure out what $T^*y \in \mathcal{H}_1$ should be, consider the continuous linear functional $\varphi_y : \mathcal{H}_1 \to \mathcal{C}$ given by $\varphi_y(x) = \langle Tx, y \rangle_{\mathcal{H}_2}$. By Riesz representation, there is a unique vector $T^*y \in \mathcal{H}_1$ such that $\varphi_y(\cdot) = \langle T\cdot, y \rangle_{\mathcal{H}_2} = \langle \cdot, T^*y \rangle_{\mathcal{H}_1}$.

(2) Linearity. $\varphi_{\alpha y}(x) = \overline{\alpha} \langle Tx, y \rangle = \overline{\alpha} \langle x, T^*y \rangle = \langle x, \alpha T^*y \rangle$
$\varphi_{y_1+y_2}(x) = \langle Tx, y_1 \rangle + \langle Tx, y_2 \rangle = \langle x, T^*y_1 + T^*y_2 \rangle$

(3) Continuity. The adjoint is a bounded linear map. Observe that

$$\begin{aligned} \|T^*y\|^2 &= |\langle T^*y, T^*y \rangle| = |\langle TT^*y, y \rangle| && \text{(adjoint)} \\ &\le \|TT^*y\|\|y\| \le \|T\|_{op}\|T^*y\|\|y\| && \text{(Cauchy-Schwarz)} \end{aligned}$$

Thus $\|T^*y\| \le \|T\|_{op}\|y\|$ for all $y \in \mathcal{H}_2$, and $\|T^*\|_{op} \le \|T\|_{op}$. □

Property 3, (Properties of Adjoints)**.** *Let $T \in BC(\mathcal{H}_1, \mathcal{H}_2)$.*

I. *$(T^*)^* = T$ and $\|T\|_{op} = \|T^*\|_{op}$*
II. *T is an isometry if and only if $T^*T = \mathrm{Id}_{\mathcal{H}_1}$.*
III. *T is an isomorphism if and only if $T^*T = \mathrm{Id}_{\mathcal{H}_1}$ and $TT^* = \mathrm{Id}_{\mathcal{H}_2}$.*
IV. *If $S \in BC(\mathcal{H}_2, \mathcal{H}_3)$, then $(ST)^* = T^*S^*$.*

Proof. These properties follow from careful application of the definition.

Part I. Fix $x \in \mathcal{H}_1$. By definition, $(T^*)^*x \in \mathcal{H}_2$ represents the continuous linear functional $y \mapsto \langle T^*y, x \rangle_{\mathcal{H}_1}$ on $\mathcal{H}_2$. Since $\langle T^*y, x \rangle_{\mathcal{H}_1} = \langle y, Tx \rangle_{\mathcal{H}_2}$ by construction of T^*, we have $(T^*)^*x = Tx$. In proving Theorem 5, we showed $\|T^*\|_{op} \le \|T\|_{op}$. We now have $\|T\|_{op} = \|(T^*)^*\|_{op} \le \|T^*\|_{op}$.

Part II. An isometry is a distance-preserving map. A continuous linear map $T \in BC(\mathcal{H}_1, \mathcal{H}_2)$ preserves distances precisely when $T^*T = \mathrm{Id}_{\mathcal{H}_1}$, since the adjoint condition is

$$\langle Tx_1, Tx_2 \rangle_{\mathcal{H}_2} = \langle x_1, T^*Tx_2 \rangle_{\mathcal{H}_1} = \langle x_1, x_2 \rangle_{\mathcal{H}_1} \iff T^*Tx_2 = x_2$$

Part III. An isomorphism between inner product spaces is defined to be an invertible isometry. From Parts I & II, the claim is obvious.

Part IV. TODO □

Example 1. The adjoint of the projection map $\pi_V : \mathcal{H} \to V$ onto a closed subspace $V \preccurlyeq \mathcal{H}$ is the inclusion map $\iota_V : V \to \mathcal{H}$. For any $x \in \mathcal{H}, v \in V$,

$$\langle x, v \rangle = \langle x_{/\!/}, v \rangle + \langle x_\perp, v \rangle = \langle x_{/\!/}, v \rangle$$

where $x = x_{/\!/} + x_\perp$ is the orthogonal decomposition of x with respect to V. Since $\pi_V(x) = x_{/\!/}$, the above says $\langle \pi_V(x), v \rangle_V = \langle x, \iota_V(v) \rangle_{\mathcal{H}}$. As a linear operator, $\pi_V : \mathcal{H} \to \mathcal{H}$ is self adjoint, since $\pi_V(v) = v$ for any $v \in V$.

References

[1] Terrence Tao. Amplification, arbitrage, and the tensor power trick. Blog, September 2007. URL https://terrytao.wordpress.com/2007/09/05/amplification-arbitrage-and-the-tensor-power-trick/.

[2] Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A generalized representer theorem. In *International conference on computational learning theory*, pages 416–426. Springer, 2001.

[3] Terrence Tao. *An Epsilon of Room, I.* American Mathematical Society, 2010.