Mathematics Department Stanford University
Math 61CM/DM – Inner products

Recall the definition of an inner product space; see Appendix A.8 of the textbook.

Definition 1 *An inner product space V is a vector space over $\mathbb{R}$ with a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that*

1. *(Positive definiteness) $\langle x, x \rangle \geq 0$ for all $x \in V$, with $\langle x, x \rangle = 0$ if and only if $x = 0$.*
2. *(Linearity in first slot) $\langle (\lambda x + \mu y), z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$ for all $x, y, z \in V$, $\lambda, \mu \in \mathbb{R}$,*
3. *(Symmetry) $\langle x, y \rangle = \langle y, x \rangle$.*

One often writes $x \cdot y = \langle x, y \rangle$ for an inner product. The standard dot product on $\mathbb{R}^n$ is an example of an inner product:

$$x \cdot y = \sum_{j=1}^{n} x_j y_j, \; x = (x_1, \ldots, x_n), \; y = (y_1, \ldots, y_n);$$

another one is, on $V = C([0,1])$ (continuous real valued functions on $[0,1]$) which we do not 'officially know' yet,

$$\langle f, g \rangle = \int_0^1 f(x) g(x) \, dx.$$

There is an extension of the definition when the underlying field is $\mathbb{C}$; the only change is that symmetry is replaced by *Hermitian symmetry*, namely $\langle x, y \rangle = \overline{\langle y, x \rangle}$, where the bar denotes complex conjugate. Complex examples are $\mathbb{C}^n$:

$$x \cdot y = \sum_{j=1}^{n} x_j \overline{y_j}, \; x = (x_1, \ldots, x_n), \; y = (y_1, \ldots, y_n);$$

with the complex conjugate needed for

$$x \cdot x = \sum_{j=1}^{n} x_j \overline{x_j} = \sum_{j=1}^{n} |x_j|^2 \geq 0,$$

$= 0$ if and only if $x = 0$ (i.e. $x_j = 0$ for all j), and for $V = C([0,1]; \mathbb{C})$ (continuous complex valued functions on $[0,1]$),

$$\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} \, dx.$$

Note that when the field is $\mathbb{R}$, symmetry plus linearity in the first slot give linearity in the second slot as well. (If the field is $\mathbb{C}$, they give conjugate linearity in the second slot, i.e. $\langle z, (\lambda x + \mu y) \rangle = \overline{\lambda} \langle z, x \rangle + \overline{\mu} \langle z, y \rangle$ for all $x, y, z \in V$, $\lambda, \mu \in \mathbb{C}$.) This linearity also gives $\langle 0, x \rangle = 0$ for all $x \in V$, as follows by writing $0 = 0 \cdot 0$ (with the first 0 on the right hand side being the real number, all others are vectors, which we temporarily denote by 0_V for clarity):

$$\langle 0_V, v \rangle = \langle 0 \cdot 0_V, v \rangle = 0 \langle 0_V, v \rangle = 0, \; v \in V,$$

and by symmetry then

$$\langle v, 0_V \rangle = \langle 0_V, v \rangle = 0,$$

with Hermitian symmetry working similarly in the complex case.

In inner product spaces one defines the *norm* (which is just a word at this moment) by

$$\|x\| = \sqrt{\langle x, x \rangle},$$

with the square root being the non-negative square root of a non-negative number (the latter being the case by positive definiteness). Note that $\|x\| = 0$ if and only if $x = 0$. We also note a useful property of $\|.\|$:

$$\|cv\|^2 = \langle cv, cv\rangle = c\langle v, cv\rangle = c^2\langle v, v\rangle = |c|^2\|v\|^2,\ c \in \mathbb{R},\ v \in V,$$

so

$$\|cv\| = |c|\,\|v\|. \tag{1}$$

This property of $\|.\|$ is called absolute homogeneity (of degree 1). The same statement, (1), is valid if the field is $\mathbb{C}$, but in that case the proof is

$$\|cv\|^2 = \langle cv, cv\rangle = c\langle v, cv\rangle = c\bar{c}\langle v, v\rangle = |c|^2\|v\|^2,\ c \in \mathbb{C},\ v \in V.$$

One concept that is tremendously useful in inner product spaces is orthogonality:

Definition 2 *Suppose V is an inner product space. For $v, w \in V$ we say that v is orthogonal to w if $\langle v, w\rangle = 0$.*

Note that $\langle v, w\rangle = 0$ if and only of $\langle w, v\rangle = 0$, so v is orthogonal to w if and only if w is orthogonal to v – so we often say simply that v and w are orthogonal.

As an illustration of its use, let's prove *Pythagoras' theorem*:

Lemma 1 *Suppose V is an inner product space, $v, w \in V$ and v and w are orthogonal. Then*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 = \|v - w\|^2.$$

Proof: Since $v - w = v + (-w)$, the statement about $v - w$ follows from the statement for $v + w$ and $\| - w\| = \|w\|$. Now,

$$\langle v + w, v + w\rangle = \langle v, v + w\rangle + \langle w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle = \langle v, v\rangle + \langle w, w\rangle$$

by the orthogonality of v and w, proving the result. □

One use of orthogonality is the following:

Lemma 2 *Suppose $v, w \in V$, $w \neq 0$. Then there exist unique $v_\parallel, v_\perp \in V$ such that $v = v_\parallel + v_\perp$, $v_\parallel = cw$ for some $c \in \mathbb{R}$ (or $\mathbb{C}$ if the field is $\mathbb{C}$) and $\langle v_\perp, w\rangle = 0$; see Figure 1.*

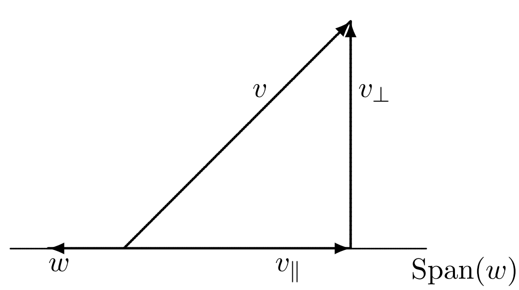

Figure 1: The decomposition $v = v_\parallel + v_\perp$.

Proof: If $v = v_{\|} + v_{\perp}$ then taking the inner product with w and using $v_{\|} = cw$ we deduce

$$\langle v, w\rangle = \langle cw, w\rangle + \langle v_{\perp}, w\rangle = c\|w\|^2,$$

so as $w \neq 0$,

$$c = \frac{\langle v, w\rangle}{\|w\|^2}.$$

Thus, $v_{\|} = cw$ and $v_{\perp} = v - cw$, giving uniqueness.

On the other hand, if we *let*

$$c = \frac{\langle v, w\rangle}{\|w\|^2},\ v_{\|} = cw,\ v_{\perp} = v - cw,$$

then $v_{\perp} + v_{\|} = v$ and $v_{\|} = cw$ are satisfied, so we merely need to check $\langle v_{\perp}, w\rangle = 0$. But

$$\langle v_{\perp}, w\rangle = \langle v, w\rangle - c\langle w, w\rangle = \langle v, w\rangle - \frac{\langle v, w\rangle}{\|w\|^2}\|w\|^2 = 0,$$

so the desired vectors $v_{\perp}$ and $v_{\|}$ indeed exist. □

One calls $v_{\|}$ the *orthogonal projection* of v to the span of w.

In order to make this a useful tool, we need to be able to estimate the inner product using the norm. This is achieved by the *Cauchy-Schwarz inequality.*

Lemma 3 (Cauchy-Schwarz) *In an inner product space V,*

$$|\langle v, w\rangle| \leq \|v\|\|w\|,\ v, w \in V.$$

Remark 1 *In the case of real-valued function spaces, Cauchy-Schwarz says explicitly that*

$$\Big|\int_0^1 f(x)g(x)\,dx\Big| \leq \sqrt{\int_0^1 f(x)^2\,dx}\sqrt{\int_0^1 g(x)^2\,dx}.$$

Proof: If $w = 0$, then both sides vanish, so we may assume $w \neq 0$. Write $v = v_{\|} + v_{\perp}$ as in Lemma 2, so

$$v_{\|} = cw,\ c = \frac{\langle v, w\rangle}{\|w\|^2}.$$

Then by Pythagoras' theorem, using $\langle v_{\|}, v_{\perp}\rangle = c\langle w, v_{\perp}\rangle = 0$,

$$\|v\|^2 = \|v_{\|}\|^2 + \|v_{\perp}\|^2 \geq \|v_{\|}\|^2 = |c|^2\|w\|^2 = \frac{|\langle v, w\rangle|^2}{\|w\|^2}.$$

Multiplying through by $\|w\|^2$ and taking the non-negative square root completes the proof of the lemma. □

A useful consequence of the Cauchy-Schwarz inequality is the triangle inequality for the norm:

Lemma 4 *In an inner product space V,*

$$\|v + w\| \leq \|v\| + \|w\|.$$

Proof: One only needs to prove the equivalent estimate where one takes the square of both sides:

$$\|v+w\|^2 \le \|v\|^2 + 2\|v\|\,\|w\| + \|w\|^2.$$

But

$$\|v+w\|^2 = \langle v+w, v+w\rangle = \|v\|^2 + \langle v,w\rangle + \langle w,v\rangle + \|w\|^2 \le \|v\|^2 + 2\|v\|\,\|w\| + \|w\|^2,$$

where the last inequality is the Cauchy-Schwarz inequality. □

In general, if one has a vector space V, one defines the notion of a norm on it as follows:

Definition 3 *Suppose V is a vector space. A norm on V is a map*

$$\|.\| : V \to \mathbb{R}$$

such that

1. *(positive definiteness) $\|v\| \ge 0$ for all $v \in V$, and $v = 0$ if and only if $\|v\| = 0$.*
2. *(absolute homogeneity) $\|cv\| = |c|\,\|v\|$, $v \in V$, and c a scalar (so $c \in \mathbb{R}$ or $c \in \mathbb{C}$, depending on whether V is real or complex),*
3. *(triangle inequality) $\|v+w\| \le \|v\| + \|w\|$ for all $v, w \in V$.*

Thus, Lemma 4 shows that the map $\|.\| : V \to \mathbb{R}$ we defined on an inner product space is indeed a norm in this sense, i.e. our use of the word norm was justified.

Some examples of a normed vector space where the norm does not come from an inner product are:

1. $V = \mathbb{R}^n$, and $\|(x_1, \dots, x_n)\|_1 = \sum_{j=1}^n |x_j|$. (The triangle inequality follows from that on $\mathbb{R}$, the others are immediate.)
2. $V = \mathbb{R}^n$, and $\|(x_1, \dots, x_n)\|_\infty = \max\{|x_j| :\ j = 1, \dots, n\}$. (The triangle inequality comes from the fact that for each j, $|x_j + y_j| \le |x_j| + |y_j|$, so this is also true for any j maximizing $|x_j + y_j|$, so using that j, $\|x+y\|_\infty = |x_j + y_j| \le |x_j| + |y_j| \le \|x\|_\infty + \|y\|_\infty$.)

Norms will play an important role when we get to analysis.

Before actually turning to inner products, let us discuss sums of subspaces, returning to arbitrary underlying fields.

Definition 4 *If Z is a vector space, V, W subspaces, $V + W = \{v + w :\ v \in V, w \in W\} \subset Z$.*

One easily checks that $V + W$ is a subspace of Z.

Definition 5 *One says that such a sum $V + W$ in Z is* direct *if $V \cap W = \{0\}$. In this case, one writes $V + W = V \oplus W$.*

Given a subspace V of Z, another subspace W is called complementary to V if $V + W = Z$, where the sum is direct.

Note that W complementary to V is equivalent to V complementary to W by symmetry of the definition.

Lemma 5 *If V is a subspace of Z, and W is complementary to V, then for any $z \in Z$ there exist unique $v \in V$, $w \in W$ such that $v + w = z$.*

Proof: Existence of v, w as desired follows from $V + W = Z$. On the other hand, if $v + w = v' + w'$ for some $v, v' \in V$, $w, w' \in W$ then $v - v' = w' - w$, and the left hand side is in V, the right hand side is in W, so they are both in $V \cap W = \{0\}$. Thus, $v = v'$, $w = w'$ as desired. $\square$

Since bases will play an important role from now on, *from this point on we assume that all vector spaces under consideration are finite dimensional.* Some of the results below have more sophisticated infinite dimensional analogues though.

Note that any subspace V of a vector space Z has a complementary subspace. Indeed, let $\{v_1, \ldots, v_k\}$ be a basis of V; complete this to a basis $\{v_1, \ldots, v_n\}$ of Z, $n \geq k$, and let $W = \mathrm{Span}\{v_{k+1}, \ldots, v_n\}$. Then $V + W = Z$ since the v_j form a basis, while $V \cap W = \{0\}$ since otherwise $\sum_{j=1}^{k} c_j v_j = \sum_{j=k+1}^{n} d_j v_j$ for some choice of c_j, d_j, not all 0, and rearranging and using the linear independence of the v_j provides a contradiction.

We also have:

Lemma 6 *If V is a subspace of Z and W is complementary to V, then* $\dim V + \dim W = \dim Z$.

Proof: Let $\{v_1, \ldots, v_k\}$ be a basis of V, $\{w_1, \ldots, w_l\}$ a basis of W. We claim that $\{v_1, \ldots, v_k, w_1, \ldots, w_l\}$ is a basis of Z, hence $\dim Z = k + l = \dim V + \dim W$. To see this claim, note that $\mathrm{Span}\{v_1, \ldots, v_k, w_1, \ldots, w_l\} = Z$ since every element z of Z can be written as $v + w$, $v \in V$, $w \in W$, and v, resp. w, are linear combinations of the corresponding basis vectors. Moreover, if $\sum_{j=1}^{k} c_j v_j + \sum_{i=1}^{l} d_i w_i = 0$ for some choice of c_j, d_i, not all zero, then rearranging gives $\sum_{j=1}^{k} c_j v_j = -\sum_{i=1}^{l} d_i w_i \in V \cap W$, so both vanish, which contradicts either the linear independence of the v_j or those of the w_i. $\square$

Now, if Z is an inner product space (hence the field is $\mathbb{R}$, though $\mathbb{C}$ would work similarly) and V is a subspace, one lets

$$V^{\perp} = \{w \in Z : \; v \in V \Rightarrow v \cdot w = 0\}.$$

With this definition it is immediate that $V \cap V^{\perp} = 0$: if $v \in V \cap V^{\perp}$, then $v \cdot v = 0$, thus $v = 0$. Proceeding as in Section 1.8 of the textbook, one shows that $V + V^{\perp} = Z$, so in particular any $z \in Z$ can be uniquely written as $z = v + w$, $v \in V$, $w \in V^{\perp}$. Thus, in an inner product space there are canonical complements, $V^{\perp}$ (called orthocomplement); in a general spaces there are many choices, none of which is preferred.

Now, if V is an inner product space and $e_1, \ldots, e_n$ is an *orthonormal basis* of V, i.e. $e_i \cdot e_j = 0$ if $i \neq j$, $e_i \cdot e_i = 1$, then it is very easy to express any $v \in V$ as the linear combination of the basis vectors. Namely, we know that one can write

$$v = \sum_{j=1}^{n} c_j e_j$$

for some choice of $c_j \in \mathbb{R}$; taking the inner product with e_i gives

$$v \cdot e_i = \sum_{j=1}^{n} c_j (e_j \cdot e_i) = c_i,$$

i.e. $c_i = v \cdot e_i$.

We postpone for now the existence of orthonormal bases, since for $\mathbb{R}^n$ the standard one is orthonormal; however, this can easily be shown in the same manner bases are constructed by considering a maximal orthonormal subset of a vector space – note that an orthonormal collection of vectors is automatically linearly independent, as follows by taking the inner product with the various vectors. (Later on, in Section 3.5, the Gram-Schmidt procedure will produce an orthonormal basis from any given basis.)

Now consider linear maps $T : V \to W$ where V, W are inner product spaces. If $e_1, \ldots, e_n$, resp. $f_1, \ldots, f_n$ are orthonormal bases of V, resp. W, then the matrix of T in this basis is very easy to find: recall that the ij entry is a_{ij} if $Te_j = \sum_{i=1}^m a_{ij} f_i$. Thus, by the above argument (applied in W),

$$a_{ij} = f_i \cdot Te_j.$$

We claim that there is a unique linear map S such that $Tv \cdot w = v \cdot Sw$ for all $v \in V$, $w \in W$. To see uniqueness, notice that the matrix of S relative to the respective orthonormal bases has ij entry $e_i \cdot Sf_j$, while that of T has lk entry $f_l \cdot Te_k$. If S has the desired property, $e_i \cdot Sf_j = Sf_j \cdot e_i = f_j \cdot Te_i$, so the ij entry of the matrix of S is the ji entry of the matrix of T, hence is determined by T. This also gives existence: if S is defined to have ij matrix entry $f_j \cdot Te_i$, so

$$S \sum_{j=1}^m x_j f_j = \sum_{j=1}^m x_j \sum_{i=1}^n (f_j \cdot Te_i) e_i,$$

then expanding vectors v, w in the respective bases $v = \sum v_i e_i$, $w = \sum w_j f_j$,

$$v \cdot Sw = \sum_{i=1}^n \sum_{j=1}^m v_i w_j (f_j \cdot Te_i) = Tv \cdot w.$$

The map S is called the *adjoint* or *transpose* of T, denoted by T^T or T^*.

Note that if $S = T^T$ then $S^T = T$, directly from the defining property of the adjoint.

The immediate property of T^T and T is the following:

Lemma 7 *We have* $N(T^T) = (\operatorname{Ran} T)^\perp$.

Proof: We have

$$w \in (\operatorname{Ran} T)^\perp \iff w \cdot Tv = 0 \text{ for all } v \in V \iff T^T w \cdot v = 0 \text{ for all } v \in V.$$

But the last statement is equivalent to $T^T w = 0$, with one implication being immediate, and for the other taking $v = T^T w$ shows $\|T^T w\|^2 = 0$, so $T^T w = 0$. This is exactly the statement that $w \in N(T^T)$ as claimed. □

This lemma can be applied with T^T in place of T, yielding $N(T) = \operatorname{Ran}(T^T)^\perp$. These give:

$$\dim \operatorname{Ran}(T^T) = \dim V - \dim N(T) = \dim \operatorname{Ran}(T),$$

where the last equality follows from the rank-nullity theorem. This is exactly the equality of the column-rank and the row-rank of T.