Linear Algebra: Lecture 10

More on inner-product spaces. Topics: orthogonal complements, the Projection Theorem, orthogonal projections.

Orthogonal complements.

We assume throughout that $\mathcal{V}$ is an inner-product space with inner product $\langle \cdot, \cdot \rangle$.

DEFINITION 10.1. *The orthogonal complement of $S \subseteq \mathcal{V}$ is*

$$S^{\perp} = \{v \in \mathcal{V} : \langle v, s \rangle = 0 \text{ for all } s \in S\}.$$

Here, $S^{\perp}$ is read as "S perp." The "perp" symbol $\perp$ is used to denote orthogonality, e.g., $v \perp w$ means $\langle v, w \rangle = 0$.

Note that $v \in S \cap S^{\perp}$ if and only if $v = 0$. Indeed, the "if" part is trivial, and the "only if" follows from the observation that if $v \in S \cap S^{\perp}$, then $\langle v, v \rangle = 0$.

PROPOSITION 10.2. *$S^{\perp}$ is a subspace of $\mathcal{V}$.*

PROPOSITION 10.3. *If $\mathcal{V}$ is finite-dimensional and $\mathcal{S} \subseteq \mathcal{V}$ is a subspace, then $\dim \mathcal{S} + \dim \mathcal{S}^{\perp} = \dim \mathcal{V}$. Moreover, $\left(\mathcal{S}^{\perp}\right)^{\perp} = \mathcal{S}$.*

Proof. Suppose that $\dim \mathcal{V} = n$ and let $\{u_1, \ldots, u_k\}$ and $\{v_1, \ldots, v_\ell\}$ be orthonormal bases of $\mathcal{S}$ and $\mathcal{S}^{\perp}$, respectively. Then $\{u_1, \ldots, u_k\} \cup \{v_1, \ldots, v_\ell\}$ is orthonormal and, therefore, linearly independent in $\mathcal{V}$, and it follows that $k + \ell \leq n$. If $k + \ell < n$, then there is a $w \in \mathcal{V}$ such that $w \notin \operatorname{span}\{\{u_1, \ldots, u_k\} \cup \{v_1, \ldots, v_\ell\}\}$. Then $z = w - \sum_{i=1}^{k} \langle w, u_i \rangle u_i - \sum_{i=1}^{\ell} \langle w, v_i \rangle v_i \neq 0$. However, $\langle z, u_j \rangle = \langle w, u_j \rangle - \langle w, u_j \rangle = 0$ for $1 \leq j \leq k$ and $\langle z, v_j \rangle = \langle w, v_j \rangle - \langle w, v_j \rangle = 0$ for $1 \leq j \leq \ell$. It follows that $z \in \mathcal{S} \cap \mathcal{S}^{\perp}$ and, consequently, that $z = 0$. This is a contradiction, and we conclude that $k + \ell = n$.

To complete the proof, note that, clearly, $\mathcal{S} \subseteq \left(\mathcal{S}^{\perp}\right)^{\perp}$. On the other hand, if $w \in \left(\mathcal{S}^{\perp}\right)^{\perp}$, then $z = w - \sum_{i=1}^{k} \langle w, u_i \rangle u_i$ is also in $\left(\mathcal{S}^{\perp}\right)^{\perp}$. However, one easily verifies as above that $\langle z, u_i \rangle = 0$ for $1 \leq j \leq k$, and it follows that $z \in \mathcal{S}^{\perp}$ as well. Consequently, $z = 0$ and $w \in \mathcal{S}$. ∎

We now work toward the fundamentally important *Projection Theorem.* This is easy to establsh in finite-dimensional spaces using orthonormal bases. However, it is much more instructive and not much harder to develop the result in spaces of arbitrary dimension. We begin with a necessary topological digression.

Complete subspaces.

The following pertains in any normed vector space regardless of whether the norm derives from an inner product.

DEFINITION 10.4.

- *A point $v \in \mathcal{V}$ is a* <u>limit point</u> *of $S \subseteq \mathcal{V}$ if there is a sequence $\{v_j\}_{j=1}^{\infty} \subseteq S$ such that $\lim_{j\to\infty} v_j = v$.*
- *A set $S \subseteq \mathcal{V}$ is* <u>closed</u> *if it contains all of its limit points.*
- *A sequence $\{v_j\}_{j=1}^{\infty} \subseteq \mathcal{V}$ is* <u>Cauchy</u> *if, for every $\epsilon > 0$, there is an n such that $\|v_j - v_k\| < \epsilon$ whenever j, $k \geq n$.*
- *A set $S \subseteq \mathcal{V}$ is* <u>complete</u> *if every Cauchy sequence $\{v_j\}_{j=1}^{\infty}$ in S converges to a point in S.*

Note that if $\mathcal{V}$ is finite-dimensional, then all norms are equivalent, and it follows that the above definitions are *norm-independent*, i.e., the defined properties hold in one norm if and only if they hold in every norm.

Also, It is easily verified that a convergent sequence is Cauchy.

PROPOSITION 10.5. *A complete set is closed.*

Proof. Suppose that $S \subseteq \mathcal{V}$ is complete. Let v be a limit point of S, and suppose that $\{v_k\}_{j=1}^{\infty} \subseteq S$ converges to v. Then $\{v_k\}_{j=1}^{\infty}$ is Cauchy and, since S is complete, converges to a point in S. Since a convergent sequence can have only one limit, that point must be v; hence, $v \in S$. □

PROPOSITION 10.6. *A closed subset of a complete set is complete.*

Proof. Suppose that T is a closed subset of a complete set S, and let $\{v_j\}_{j=1}^{\infty}$ be a Cauchy sequence in T. Since S is complete, there exists a $v \in S$ such that $\lim_{j\to\infty} v_j = v$. Then v is a limit point of T, and, since T is closed, it follows that $v \in T$. □

PROPOSITION 10.7. *If $\mathcal{S}$ is a finite-dimensional subspace of $\mathcal{V}$, then $\mathcal{S}$ is complete.*

Proof. Suppose that $\{v_j\}_{j=1}^{\infty} \subseteq \mathcal{S}$ is Cauchy. Let $\{u_1, \ldots, u_k\}$ be a basis for $\mathcal{S}$, and write $v_j = \sum_{i=1}^{k} \alpha_{ji} u_i$ for each j. Since $\{v_j\}$ is Cauchy, it is, in particular, Cauchy in the norm $\|\cdot\|_{\infty}$, defined by $\|w\|_{\infty} = \max_{1\leq i\leq k} |\beta_i|$ for $w = \sum_{i=1}^{k} \beta_i u_i \in \mathcal{V}$. It follows that $\{\alpha_{ji}\}_{j=1}^{\infty}$ is Cauchy for each i and, therefore, $\{\alpha_{ji}\}_{j=1}^{\infty}$ converges to some α_i for each i. Then $\{v_j = \sum_{i=1}^{k} \alpha_{ji} u_i\}$ converges to $v = \sum_{i=1}^{k} \alpha_i u_i \in \mathcal{S}$. □

In contrast, infinite-dimensional subspaces need not be complete or even closed.

Example. In $C[0,1]$, set $\mathcal{S} = \{f \in C[0,1] : \ f(\frac{1}{2}) = 0\}$ and suppose that the inner product on $C[0,1]$ is $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$. Consider $\{f_j\}_{j=2}^\infty$, where, for each $j \geq 2$, f_j is the piecewise-linear function passing through $(0,1)$, $(\frac{1}{2} - \frac{1}{j}, 1)$, $(\frac{1}{2}, 0)$, $(\frac{1}{2} + \frac{1}{j}, 1)$, and $(1,1)$. Clearly, $\{f_j\}_{j=2}^\infty \subset \mathcal{S}$. Let $f \in C[0,1]$ be defined by $f(x) = 1$ for $0 \leq x \leq 1$. Then

$$\|f_j - f\|^2 = \int_0^1 (f_j(x) - 1)^2\,dx = \int_{1/2-1/j}^{1/2+1/j} (f_j(x) - 1)^2\,dx \leq \frac{2}{j}.$$

Thus $\lim_{j\to\infty} \|f_j - f\| = 0$, and we have that f is a limit point of $\mathcal{S}$. However, $f \notin \mathcal{S}$, and so $\mathcal{S}$ is not closed.

The Projection Theorem.

Theorem 10.8. (Projection Theorem) *If $\mathcal{S}$ is a complete subspace of an inner-product space $\mathcal{V}$, then every $v \in \mathcal{V}$ can be written uniquely as $v = u + w$ for $u \in \mathcal{S}$ and $w \in \mathcal{S}^\perp$.*

Proof. Suppose that $v \in \mathcal{V}$ is given. If $v \in \mathcal{S}$, then the result is trivial, so assume that $v \notin \mathcal{S}$. Since $\mathcal{S}$ is complete and, therefore, closed, we must have

$$\mu \equiv \inf_{u\in\mathcal{S}} \|v - u\| > 0. \tag{10.1}$$

(Otherwise, v would be a limit point of $\mathcal{S}$ and, therefore, in $\mathcal{S}$.) Note that $\|v - u\| \geq \mu$ for all $u \in \mathcal{S}$. Let $\{u_j\} \subseteq \mathcal{S}$ be such that $\lim_{j\to\infty} \|v - u_j\| = \mu$.

We claim that $\{u_j\}$ is Cauchy. Indeed, suppose that $\epsilon > 0$ is given. Choose n such that $\mu^2 \leq \|v - u_j\|^2 < \mu^2 + \epsilon^2/4$ whenever $j \geq n$. Then for j, $\ell \geq n$, the Parallelogram Law gives

$$\|(v - u_j) - (v - u_\ell)\|^2 + \|(v - u_j) + (v - u_\ell)\|^2 = 2\|v - u_j\|^2 + 2\|v - u_\ell\|^2,$$

whence

$$\|u_j - u_\ell\|^2 = 2\|v - u_j\|^2 + 2\|v - u_\ell\|^2 - 4\|v - \frac{(u_j + u_\ell)}{2}\|^2 < 4(\mu^2 + \epsilon^2/4) - 4\mu^2 = \epsilon^2,$$

and we have $\|u_j - u_\ell\| < \epsilon$.

Since $\mathcal{S}$ is complete and $\{u_j\}$ is Cauchy, there is a $u \in \mathcal{S}$ such that $\lim_{j\to\infty} u_j = u$. Set $w = v - u$. Note that $\|w\| = \|v - u\| = \lim_{j\to\infty} \|v - u_j\| = \mu$.

We claim that $w \in \mathcal{S}^\perp$. Indeed, suppose that $\langle w, s\rangle \neq 0$ for some $s \in \mathcal{S}$. Then for $t \in \mathbb{R}$, $u + ts \in \mathcal{S}$ and

$$\|v - (u + ts)\|^2 = \|w - ts\|^2 = \|w\|^2 - t\{\langle w, s\rangle + \langle s, w\rangle\} + t^2\|s\|^2.$$

By choosing t such that sign $t = \text{sign}\,\{\langle w, s\rangle + \langle s, w\rangle\}$ and $|t| > 0$ is sufficiently small, we can make the right-hand side less than $\|w\|^2 = \mu^2$, contradicting (10.1).

We now have the desired representation $v = u + w$ for $u \in \mathcal{S}$ and $w \in \mathcal{S}^\perp$. To show that this representation is unique, suppose that $v = \hat{u} + \hat{w}$ for $\hat{u} \in \mathcal{S}$ and $\hat{w} \in \mathcal{S}^\perp$. Then $u + w = v = \hat{u} + \hat{w}$, and we have $u - \hat{u} = \hat{w} - w$. Since $u - \hat{u} \in \mathcal{S}$ and $\hat{w} - w \in \mathcal{S}^\perp$, it follows that both sides are in $\mathcal{S} \cap \mathcal{S}^\perp$ and, hence, are zero. □

The following corollary extends the second conclusion of Proposition 10.3.

COROLLARY 10.9. *If $\mathcal{S}$ is a complete subspace of an inner-product space, then $(\mathcal{S}^\perp)^\perp = \mathcal{S}$.*

Proof. We know that $(\mathcal{S}^\perp)^\perp$ is a subspace and clearly have $\mathcal{S} \subseteq (\mathcal{S}^\perp)^\perp$. Suppose that $\mathcal{S} \neq (\mathcal{S}^\perp)^\perp$, and let $v \in (\mathcal{S}^\perp)^\perp$ be such that $v \notin \mathcal{S}$. Then the Projection Theorem gives $v = u + w$ for $u \in \mathcal{S}$ and $w \in \mathcal{S}^\perp$. But since $v \in (\mathcal{S}^\perp)^\perp$ and $u \in \mathcal{S} \subseteq (\mathcal{S}^\perp)^\perp$, we also have that $w = v - u \in (\mathcal{S}^\perp)^\perp$. It follows that $w = 0$ and $v = u \in \mathcal{S}$. □

COROLLARY 10.10. *If $\mathcal{S}$ is a finite-dimensional subspace of an inner-product space $\mathcal{V}$, then every $v \in \mathcal{V}$ can be uniquely written as $v = u + w$, where $u \in \mathcal{S}$ and $w \in \mathcal{S}^\perp$.*

Proof. It follows from Proposition 10.7 that $\mathcal{S}$ is complete, and the corollary follows from the Projection Theorem. □

The following can be regarded as a general statement of the Pythagorean Theorem.

PROPOSITION 10.11. *If $v = u + w$ for $u \in \mathcal{S}$ and $w \in \mathcal{S}^\perp$, then $\|v\|^2 = \|u\|^2 + \|w\|^2$.*

Proof. The result follows immediately from the orthogonality of u and w. □

Orthogonal projection.

We begin with the definition and a useful proposition.

DEFINITION 10.12. *If $v = u + w$ for $u \in \mathcal{S}$ and $w \in \mathcal{S}^\perp$, then u is the orthogonal projection of v onto $\mathcal{S}$.*

PROPOSITION 10.13. *If $\mathcal{S}$ is finite-dimensional and $\{u_1, \ldots, u_k\}$ is an orthonormal basis for $\mathcal{S}$, then the orthogonal projection of v onto $\mathcal{S}$ is given by $u = \sum_{i=1}^k \langle v, u_i\rangle\, u_i$. Moreover, we have Bessel's Inequality: $\|v\|^2 \geq \sum_{i=1}^k |\langle v, u_i\rangle|^2$, and $\|v\|^2 = \sum_{i=1}^k |\langle v, u_i\rangle|^2$ if and only if $v = \sum_{i=1}^k \langle v, u_i\rangle\, u_i \in \mathcal{S}$.*

Proof. Clearly, $u = \sum_{i=1}^{k} \langle v, u_i \rangle u_i \in \mathcal{S}$. Also, $w = v - \sum_{i=1}^{k} \langle v, u_i \rangle u_i$ satifies $\langle w, u_j \rangle = \langle v, u_j \rangle - \langle v, u_j \rangle = 0$ for each j, so $w \in \mathcal{S}^{\perp}$. It follows from the uniqueness part of the Projection Theorem that u is the orthogonal projection of v onto $\mathcal{S}$. With Proposition 10.11, Bessel's inequality follows immediately: $\|v\|^2 = \|u\|^2 + \|w\|^2 \geq \|u\|^2 = \sum_{i=1}^{k} |\langle v, u_i \rangle|^2$, and $\|v\|^2 = \sum_{i=1}^{k} |\langle v, u_i \rangle|^2$ if and only if $w = 0$, i.e., $v = u = \sum_{i=1}^{k} \langle v, u_i \rangle u_i \in \mathcal{S}$. □

More generally, suppose that $\mathcal{S}$ is a complete subspace of an inner product space $\mathcal{V}$ so that orthogonal projection onto $\mathcal{S}$ is defined on all of $\mathcal{V}$. Let $P(v)$ denote the orthogonal projection of $v \in \mathcal{V}$. It is easily verified that $P : \mathcal{V} \to \mathcal{V}$ is a linear transformation. Note that for every $v \in \mathcal{V}$, we have that $v = P(v) + (I - P)(v)$, where I is the identity transformation on $\mathcal{V}$ and $(I - P)(v) \in \mathcal{S}^{\perp}$.

Proposition 10.14. *Orthogonal projection onto a complete subspace $\mathcal{S}$ satistifes the following:*

(a) $P^2 = P$;

(b) $\mathcal{R}(P) \equiv \{u \in \mathcal{V} : \ u = P(v) \text{ for some } v \in \mathcal{V}\} = \mathcal{S}$;

(c) $\langle P(v), w \rangle = \langle v, P(w) \rangle$ *for all* v, $w \in \mathcal{V}$.

Moreover, P is uniquely characterized by these properties, i.e., if $Q : \mathcal{V} \to \mathcal{V}$ is a linear transformation satisfying these properties, then $Q = P$.

Proof. Noting that $P(v) \in \mathcal{S}$ for all $v \in \mathcal{V}$ and that P acts as the identity transformation on $\mathcal{S}$, we immediately have (a) and also that $\mathcal{R}(P) \subseteq \mathcal{S}$. On the other hand, if $u \in \mathcal{S}$, then $P(u) = u$. It follows that $\mathcal{S} \subseteq \mathcal{R}(P)$, and (b) holds.

To show (c), write $v = P(v) + (I - P)(v)$ and $w = P(w) + (I - P)(w)$. Recalling that $(I - P)(v)$ and $(I - P)(w)$ are in $\mathcal{S}^{\perp}$, we have that

$$\langle P(v), w \rangle = \langle P(v), P(w) + (I - P)(w) \rangle = \langle P(v), P(w) \rangle = \langle P(v) + (I - P)(v), P(w) \rangle = \langle v, P(w) \rangle,$$

which verifies (c).

To complete the proof, suppose that Q is a linear transformation satisfying (a)–(c). For $v \in \mathcal{V}$, we have that $v = Q(v) + (I - Q)(v)$. We claim that $(I - Q)(v) \in \mathcal{S}^{\perp}$. Indeed, it follows from (a) and (b) that Q acts as the identity transformation on $\mathcal{S}$, and, with (c), we have for $u \in \mathcal{S}$ that

$$\langle (I - Q)(v), u \rangle = \langle v, u \rangle - \langle Q(v), u \rangle = \langle v, u \rangle - \langle v, Q(u) \rangle = \langle v, u \rangle - \langle v, u \rangle = 0.$$

Then $v = Q(v) + (I - Q)(v)$ with $Q(v) \in \mathcal{S}$ and $(I - Q)(v) \in \mathcal{S}^{\perp}$, and it follows from the uniqueness part of the Projection Theorem that $Q(v) = P(v)$. □

Remarks. A linear transformation P satisfying property (a) is said to be *idempotent*. If P satisfies (a) and (b), then P is a *projection* onto $\mathcal{S}$. Note that this projection is characterized by algebraic properties alone. If (a), (b), and (c) hold, then P is an *orthogonal projection* onto $\mathcal{S}$, with the geometric notion of orthogonality brought in through the inner product $\langle \cdot, \cdot \rangle$.

Example 10.15. *Suppose that* $v \in \mathbb{R}^{\mathrm{n}}$. *The* annihilators *of* v *are the matrices* $A \in \mathbb{R}^{\mathrm{n\times n}}$ *such that* $Av = 0$. *Denote the set of all annihilators of* v *by* $\mathcal{A}(v)$. *It is easily verified that* $\mathcal{A}(v)$ *is a subspace of* $\mathbb{R}^{\mathrm{n\times n}}$. *In the trivial case* $v = 0$, *we have that* $\mathcal{A}(v) = \mathbb{R}^{\mathrm{n\times n}}$, *and so we assume that* $v \neq 0$.

Suppose that the inner product on $\mathbb{R}^{\mathrm{n\times n}}$ *is the Frobenius inner product* $\langle \cdot, \cdot \rangle$, *given by* $\langle A, B\rangle = \sum_{i,j} a_{ij}b_{ij} = \mathrm{trace}\left\{AB^T\right\}$. *We claim that orthogonal projection onto* $\mathcal{A}(v)$ *with respect to this inner product is given by* $P(A) = A\left[I - \frac{vv^T}{v^Tv}\right]$ *for* $A \in \mathcal{A}(v)$. *To show this, we only need to verify that properties* (a)–(c) *hold for this* P. *For* $A \in \mathbb{R}^{\mathrm{n\times n}}$, *we have*

$$P^2(A) = A\left[I - \frac{vv^T}{v^Tv}\right]\left[I - \frac{vv^T}{v^Tv}\right] = A\left[I - 2\frac{vv^T}{v^Tv} + \frac{vv^T}{v^Tv}\frac{vv^T}{v^Tv}\right] = A\left[I - \frac{vv^T}{v^Tv}\right] = P(A),$$

and so (a) *holds. To show that* (b) *holds, note first that if* $A \in \mathbb{R}^{\mathrm{n\times n}}$, *then* $P(A)v = A\left[I - \frac{vv^T}{v^Tv}\right]v = A[v - v] = 0$, *and it follows that* $\mathcal{R}(P) \subseteq \mathcal{A}(v)$. *On the other hand, if* $A \in \mathcal{A}(v)$, *then* $P(A) = A\left[I - \frac{vv^T}{v^Tv}\right] = A$, *and so* $\mathcal{A}(v) \subseteq \mathcal{R}(P)$. *We conclude that* $\mathcal{A}(v) = \mathcal{R}(P)$, *i.e., that* (b) *holds. Finally, we have that*

$$\langle P(A), B\rangle = \text{trace}\left\{A\left[I - \frac{vv^T}{v^Tv}\right]B^T\right\} = \text{trace}\left\{A\left(B\left[I - \frac{vv^T}{v^Tv}\right]\right)^T\right\} = \langle A, P(B)\rangle\,,$$

and so (c) *holds as well.*

Remark. Projections onto subspaces of $\mathbb{R}^{\mathrm{n\times n}}$ are fundamentally important for developing and analyzing *quasi-Newton methods* for numerically solving systems of nonlinear equations and optimization problems. The particular projection in Example 10.15 is used to derive and analyze *Broyden's method*, the most widely used quasi-Newton method for general nonlinear systems.