Linear Algebra: Lecture 11

More on inner-product spaces. Topics: Linear transformations on inner product spaces, linear functionals, the Riesz Representation Theorem, adjoint transformations, linear transformations revisited, matrix rank.

Linear transformations on inner product spaces..

This broad topic can be regarded as a capstone for all of the material developed so far. We assume throughout that $\mathcal{V}$ is an inner-product space with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$, and we begin with arguably the simplest type of linear transformation defined on $\mathcal{V}$.

Linear functionals.

Definition 11.1. *A linear functional is a linear transformation from $\mathcal{V}$ to $\mathbb{F}$.*

Recall from Lecture 9 that a linear transformation T from $\mathcal{V}$ to a normed vector space $\mathcal{W}$ is *bounded* if there exists a C such that $\|T(v)\| \leq C\|v\|$ for all $v \in \mathcal{V}$. Moreover, if T is bounded, then it is continuous everywhere. A linear transformation need not be bounded in general; however, if $\mathcal{V}$ is finite-dimensional, then every linear transformation on $\mathcal{V}$ is bounded.

A linear functional $\ell : \mathcal{V} \to \mathbb{F}$ is bounded if there is a C such that $|\ell(v)| \leq C\|v\|$ for all $v \in \mathcal{V}$. As in the case of general linear transformations, linear functionals are not always bounded.

Example 11.2. *Define $\ell(f) = f(\frac{1}{2})$ for $f \in C[0,1]$, and suppose that the inner product on $C[0,1]$ is $\langle f,g\rangle = \int_0^1 f(x)g(x)\,dx$. Consider $\{f_k\}_{k=2}^{\infty} \subset C[0,1]$, where, for each $k \geq 2$, f_k is the piecewise-linear function passing through $(0,0)$, $(\frac{1}{2}-\frac{1}{k},0)$, $(\frac{1}{2},1)$, $(\frac{1}{2}+\frac{1}{k},0)$, and $(1,0)$. Then $\ell(f_k) = 1$ for each k, while $\|f_k\| = \left(\int_0^1 f_k(x)^2\,dx\right)^{1/2} \leq \sqrt{\frac{2}{k}} \to 0$ as $k \to \infty$. It follows that no inequality $|\ell(f_k)| \leq C\|f_k\|$ can hold for all k, and so ℓ is not bounded.*

There is a particular source of bounded linear functionals on an inner-product space, namely, the vectors in the space themselves. Indeed, for $w \in \mathcal{V}$, we can define $\ell(v) = \langle v,w\rangle$ for $v \in \mathcal{V}$. It is easily verified that ℓ is linear on $\mathcal{V}$. Moreover, the Schwarz Inequality yields $|\ell(v)| \leq \|w\|\|v\|$, and so ℓ is bounded.

Our goal is to establish the converse: if ℓ is a bounded linear functional on $\mathcal{V}$, then there is a $w \in \mathcal{V}$ such that $\ell(v) = \langle v,w\rangle$ for every $v \in \mathcal{V}$. This is the *Riesz Representation Theorem*.

Lemma 11.3. *Suppose that $\mathcal{V}$ is a complete inner-product space and that ℓ is a bounded linear functional on $\mathcal{V}$. Then $\mathcal{N}(\ell)$ is complete.*

Proof. Suppose that $\{v_k\}$ is a Cauchy sequence in $\mathcal{N}(\ell)$. Then $\{v_k\}$ is also Cauchy in $\mathcal{V}$, and since $\mathcal{V}$ is complete, there is a $v \in \mathcal{V}$ such that $\lim_{k\to\infty} v_k = v$. Since ℓ is bounded, it is continuous, and we have $\ell(v) = \lim_{k\to\infty} \ell(v_k) = 0$. Thus $v \in \mathcal{N}(\ell)$, and it follows that $\mathcal{N}(\ell)$ is complete. □

THEOREM 11.4. (RIESZ REPRESENTATION THEOREM) Suppose that $\mathcal{V}$ is a complete inner-product space and that ℓ is a bounded linear functional on $\mathcal{V}$. Then there is a unique $w \in \mathcal{V}$ such that $\ell(v) = \langle v, w\rangle$ for all $v \in \mathcal{V}$.

Proof. We first show the existence of such a w. Note that if $\ell = 0$, i.e., if $\ell(v) = 0$ for all $v \in \mathcal{V}$, then $w = 0$ will do. Suppose that $\ell \neq 0$.

We claim that $\dim \mathcal{N}(\ell)^{\perp} = 1$. To show this, we first note that, since $\ell \neq 0$, there is a $v \in \mathcal{V}$ such that $\ell(v) \neq 0$. Since ℓ is bounded and $\mathcal{V}$ is complete, we have from Lemma 11.3 that $\mathcal{N}(\ell)$ is complete. Then we can apply the Projection Theorem to obtain $v = u + w$, where $u \in \mathcal{N}(\ell)$ and $w \in \mathcal{N}(\ell)^{\perp}$. Since $v \notin \mathcal{N}(\ell)$, we have that $w \neq 0$, and it follows that $\dim \mathcal{N}(\ell)^{\perp} \geq 1$. To show that $\dim \mathcal{N}(\ell)^{\perp} = 1$, suppose that u and v are in $\mathcal{N}(\ell)^{\perp}$. There are α, $\beta \in \mathbb{F}$ that are not both zero and are such that $\alpha\ell(u) + \beta\ell(v) = 0$. Then $\ell(\alpha u + \beta v) = 0$, which is to say that $\alpha u + \beta v \in \mathcal{N}(\ell)$. Since $\alpha u + \beta v \in \mathcal{N}(\ell)^{\perp}$ as well, it follows that $\alpha u + \beta v = 0$. Consequently, u and v are linearly dependent, and we conclude that $\dim \mathcal{N}(\ell)^{\perp} = 1$.

Choose $\hat{w} \in \mathcal{N}(\ell)^{\perp}$ such that $\|\hat{w}\| = 1$, and set $w = \overline{\ell(\hat{w})}\, \hat{w}$. (For convenience, we assume that $\mathbb{F} = \mathbb{C}$, which subsumes the case $\mathbb{F} = \mathbb{R}$.) By Proposition 10.7, $\mathcal{N}(\ell)^{\perp}$ is complete, and we can apply the Projection Theorem to define the orthogonal projection P onto $\mathcal{N}(\ell)^{\perp}$. For $v \in \mathcal{V}$, we have that $P(v) = \langle v, \hat{w}\rangle\, \hat{w}$ by Proposition 10.13; thus $v = P(v) + (I - P)(v) = \langle v, \hat{w}\rangle\, \hat{w} + (I - P)(v)$. Note that $(I - P)(v) \in \left(\mathcal{N}(\ell)^{\perp}\right)^{\perp} = \mathcal{N}(\ell)$ since $\mathcal{N}(\ell)$ is complete. Then

$$\begin{aligned}
\ell(v) &= \ell(P(v) + (I - P)(v)) = \ell(P(v)) = \ell(\langle v, \hat{w}\rangle\, \hat{w}) = \langle v, \hat{w}\rangle\, \ell(\hat{w}) \\
&= \left\langle v, \overline{\ell(\hat{w})}\, \hat{w} \right\rangle = \langle v, w\rangle\,.
\end{aligned}$$

To show that this w is unique, suppose that $\tilde{w} \in \mathcal{V}$ is such that $\ell(v) = \langle v, \tilde{w}\rangle$ for all $v \in \mathcal{V}$. Then $0 = \ell(v) - \ell(v) = \langle v, w\rangle - \langle v, \tilde{w}\rangle = \langle v, w - \tilde{w}\rangle$ for all $v \in \mathcal{V}$. In particular, $0 = \langle w - \tilde{w}, w - \tilde{w}\rangle$, and it follows that $\tilde{w} = w$. $\square$

Adjoint transformations.

We now consider linear transformations from $\mathcal{V}$ to another inner-product space $\mathcal{W}$. We use the same symbols $\langle \cdot, \cdot\rangle$ and $\|\cdot\|$ to denote the inner products and norms on both spaces. We assume that $\mathcal{V}$ is complete throughout, although this is not necessary for Proposition 11.6.

Suppose that $T : \mathcal{V} \to \mathcal{W}$ is a bounded linear transformation. As in Lecture 9,[1] we define

$$\|T\| = \sup_{v \neq 0} \frac{\|T(v)\|}{\|v\|} = \sup_{\|v\|=1} \|T(v)\|.$$

[1] In Lecture 9, we defined $\|T\| = \max_{v\neq 0} \|T(v)\|/\|v\| = \max_{\|v\|=1} \|T(v)\|$, assuming that $\mathcal{V}$ is finite-dimensional. Here, we do not assume this and, consequently, must replace "max" with "sup."

If $\mathcal{V}$ is finite-dimensional, then $\{v \in \mathcal{V} : \ \|v\| = 1\}$ is compact, and "sup" can be replaced by "max" in this definition.

For $w \in \mathcal{W}$, consider a linear functional on $\mathcal{V}$ defined by $\ell(v) = \langle T(v), w\rangle$ for $v \in \mathcal{V}$. With the Schwarz Inequality, we have that

$$|\ell(v)| = |\langle T(v), w\rangle| \leq \|T(v)\| \, \|w\| \leq (\|T\| \, \|w\|) \, \|v\|,$$

so ℓ is a bounded linear functional on $\mathcal{V}$. It follows from the Riesz Representation Theorem that there is a $w_* \in \mathcal{V}$ such that $\ell(v) = \langle v, w_*\rangle$ and, hence, $\langle T(v), w\rangle = \langle v, w_*\rangle$ for all $v \in \mathcal{V}$. Denoting w_* by $T^*(w)$, we have that

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle \tag{11.1}$$

for all $v \in \mathcal{V}$ and all $w \in \mathcal{W}$. The assignment $w \to T^*(w)$ defines a map $T^* : \mathcal{W} \to \mathcal{V}$ called the *adjoint* of T.

PROPOSITION 11.5. *T^* is a linear transformation from $\mathcal{W}$ to $\mathcal{V}$.*

Proof. Suppose that w and $\hat{w}$ are in $\mathcal{W}$ and α and β are scalars. Then for $v \in \mathcal{V}$,

$$\begin{aligned}
\langle v, \alpha T^*(w) + \beta T^*(\hat{w})\rangle &= \overline{\alpha}\,\langle v, T^*(w)\rangle + \overline{\beta}\,\langle v, T^*(\hat{w})\rangle = \overline{\alpha}\,\langle T(v), w\rangle + \overline{\beta}\,\langle T(v), \hat{w}\rangle \\
&= \langle T(v), \alpha w\rangle + \langle T(v), \beta\hat{w}\rangle = \langle T(v), \alpha w + \beta\hat{w}\rangle \\
&= \langle v, T^*(\alpha w + \beta\hat{w})\rangle,
\end{aligned}$$

and it follows that $\alpha T^*(w) + \beta T^*(\hat{w}) = T^*(\alpha w + \beta\hat{w})$. □

The following lemma will be useful in obtaining the result that follows.

LEMMA 11.6. *For $v \in \mathcal{V}$,*

$$\|v\| = \max_{\|u\|=1} |\langle u, v\rangle|.$$

Proof. Suppose that $v \in \mathcal{V}$. If $u \in \mathcal{V}$ is such that $\|u\| = 1$, then $|\langle u, v\rangle| \leq \|u\| \, \|v\| = \|v\|$, so $\|v\| \geq \sup_{\|u\|=1} |\langle u, v\rangle|$. On the other hand, taking $u = v/\|v\|$ gives $|\langle u, v\rangle| = |\langle v/\|v\|, v\rangle| = \|v\|$. It follows that $\|v\| = \sup_{\|u\|=1} |\langle u, v\rangle|$, and we can replace "sup" with "max." □

PROPOSITION 11.7. *$T^* : \mathcal{W} \to \mathcal{V}$ is bounded, and $\|T^*\| = \|T\|$.*

Proof. With Lemma 11.6, we have

$$\|T^*\| = \sup_{\|w\|=1} \|T^*(w)\| = \sup_{\|w\|=1} \max_{\|v\|=1} |\langle v, T^*(w)\rangle| = \sup_{\|w\|=1} \max_{\|v\|=1} |\langle T(v), w\rangle|.$$

Since $|\langle T(v), w\rangle| \le \|T(v)\|\,\|w\| \le \|T\|\,\|v\|\,\|w\|$, it follows that $\|T^*\| \le \|T\|$. On the other hand,

$$\sup_{\|w\|=1} \max_{\|v\|=1} |\langle T(v), w\rangle| \ge \max_{\|v\|=1} |\langle T(v), T(v)/\|T(v)\|\rangle| = \max_{\|v\|=1} \|T(v)\| = \|T\|.$$

□

There are many important examples of adjoint transformations, but the most important for us are provided by matrices. Suppose that $T : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{m}}$ is defined by $T(v) = Av$ for $v \in \mathbb{R}^{\mathrm{n}}$, where $A \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$. Suppose also that the inner product on both $\mathbb{R}^{\mathrm{n}}$ and $\mathbb{R}^{\mathrm{m}}$ is the Euclidean inner product, i.e., $\langle v, w\rangle = v^T w$. Then for $v \in \mathbb{R}^{\mathrm{n}}$ and $w \in \mathbb{R}^{\mathrm{m}}$, we have

$$\langle Av, w\rangle = (Av)^T w = v^T A^T w = \left\langle v, A^T w\right\rangle.$$

It follows that the adjoint transformation is given by $T^*(w) = A^T w$. Note that $A^T \in I\!R^{n\times m}$.

In the complex case, if the transformation is from $\mathbb{C}^{\mathrm{n}}$ to $\mathbb{C}^{\mathrm{m}}$ and $A \in \mathbb{C}^{\mathrm{m}\times\mathrm{n}}$, then

$$\langle Av, w\rangle = (Av)^T \overline{w} = v^T A^T \overline{w} = v^T \left(\overline{A^* w}\right) = \langle v, A^* w\rangle,$$

where $A^* = \overline{A^T}$ is the *conjugate transpose* (or *Hermitian transpose*) of A.

Linear equations revisited.

In general, suppose that $\mathcal{V}$ and $\mathcal{W}$ are inner-product spaces and that T is a bounded linear transformation from $\mathcal{V}$ to $\mathcal{W}$. Consider a *linear equation*

$$T(v) = w, \quad w \in \mathcal{W}. \tag{11.2}$$

This has a solution $v \in \mathcal{V}$ if and only if w is in $\mathcal{R}(T)$. We proceed to characterize $\mathcal{R}(T)$.

LEMMA 11.8. $\mathcal{R}(T)^\perp = \mathcal{N}(T^*)$.

Proof. We have that

$$\begin{aligned}\mathcal{R}(T)^\perp &= \{w \in \mathcal{W} : \langle T(v), w\rangle = 0 \text{ for all } v \in \mathcal{V}\} = \{w \in \mathcal{W} : \langle v, T^*(w)\rangle = 0 \text{ for all } v \in \mathcal{V}\} \\ &= \{w \in \mathcal{W} : T^*(w) = 0\} = \mathcal{N}(T^*).\end{aligned}$$

□

The following provides our fundamental characterization of $\mathcal{R}(T)$.

Theorem 11.9. *If $\mathcal{R}(T)$ is complete, then $\mathcal{R}(T) = \mathcal{N}(T^*)^\perp$.*

Proof. If $\mathcal{R}(T)$ is complete, then Corollary 10.9 implies that $\mathcal{R}(T) = \left(\mathcal{R}(T)^\perp\right)^\perp$, and it follows from Lemma 11.8 that $\mathcal{R}(T) = \mathcal{N}(T^*)^\perp$. □

Corollary 11.10. *If either $\mathcal{V}$ or $\mathcal{W}$ is finite-dimensional, then $\mathcal{R}(T) = \mathcal{N}(T^*)^\perp$.*

Proof. If either $\mathcal{V}$ or $\mathcal{W}$ is finite-dimensional, then so is $\mathcal{R}(T)$, and it follows from Proposition 10.7 that $\mathcal{R}(T)$ is complete. With Theorem 11.9, we conclude that $\mathcal{R}(T) = \mathcal{N}(T^*)^\perp$. □

Remark. With regard to (11.2), a useful interpretation of the condition $\mathcal{R}(T) = \mathcal{N}(T^*)^\perp$ is that $T(v) = w$ has a solution $v \in \mathcal{V}$ if and only if $\langle w, z\rangle = 0$ for all $z \in \mathcal{W}$ such that $T^*(z) = 0$.

Linear equations in $\mathbb{R}^{\mathrm{n}}$ and $\mathbb{R}^{\mathrm{m}}$ revisited.

Suppose that (11.2) is specialized to

$$Ax = b, \quad A \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}, \quad b \in \mathbb{R}^{\mathrm{m}}. \tag{11.3}$$

From Corollary 11.10 and the subsequent remark, we have that $Ax = b$ has a solution if and only if $b \in \mathcal{N}(A^T)^\perp$, i.e., $b^T z = 0$ for all $z \in \mathbb{R}^{\mathrm{m}}$ such that $A^T z = 0$.

Similarly, if $A \in \mathbb{C}^{\mathrm{m}\times\mathrm{n}}$ and $b \in \mathbb{C}^{\mathrm{n}}$, then $Ax = b$ has a solution if and only if $b \in \mathcal{N}(A^*)^\perp$, i.e., $b^T \overline{z} = 0$ for all $z \in \mathbb{C}^{\mathrm{m}}$ such that $A^* z = 0$.

Matrix rank.

Recall that, for A in $\mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ or $\mathbb{C}^{\mathrm{m}\times\mathrm{n}}$, we defined the *column rank* of A to be the number of linearly independent columns of A, which is also the dimension of $\mathcal{R}(A)$. We can similarly define the *row rank* of A to be the number of linearly independent rows of A. However, as we now show, there is no need to distinguish between the row and column ranks of A.

Lemma 11.11. *For $A \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$, $\dim \mathcal{R}(A^T) = \dim \mathcal{R}(A)$. Similarly, for $A \in \mathbb{C}^{\mathrm{m}\times\mathrm{n}}$, $\dim \mathcal{R}(A^*) = \dim \mathcal{R}(A)$.*

Proof. For $A \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$, Proposition 10.3 implies that $m = \dim \mathcal{R}(A) + \dim \mathcal{R}(A)^\perp$. Since $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ by Lemma 11.8, this yields $m = \dim \mathcal{R}(A) + \dim \mathcal{N}(A^T)$. From Theorem 3.6, we have that $\dim \mathcal{R}(A^T) + \dim \mathcal{N}(A^T) = m$, and it follows that $m = \dim \mathcal{R}(A) + m - \dim \mathcal{R}(A^T)$, whence $\dim \mathcal{R}(A^T) = \dim \mathcal{R}(A)$. The proof for $A \in \mathbb{C}^{\mathrm{m}\times\mathrm{n}}$ is similar. □

It follows from the lemma that the number of linearly independent rows of A is equal to the number of linearly independent columns of A; in other words, *the row rank of a matrix is equal to its column rank*. Consequently, we can speak simply of the *rank* of a matrix.

The following are immediately seen for A in $\mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ or $\mathbb{C}^{\mathrm{m}\times\mathrm{n}}$:

- $\operatorname{rank} A = \operatorname{rank} A^T$;
- $\operatorname{rank} A \le \min\{m, n\}$.