Linear algebra review

Daniel Hsu (COMS 4771)

Euclidean spaces

For each natural number n, the n-dimensional *Euclidean space* is denoted by $\mathbb{R}^n$, and it is a *vector space* over the *real field* $\mathbb{R}$ (i.e., $\mathbb{R}^n$ is a *real vector space*). Vectors $v_1, \dots, v_k \in \mathbb{R}^n$ are *linearly dependent* if there exist $c_1, \dots, c_k \in \mathbb{R}$, not all zero, such that $c_1 v_1 + \cdots + c_k v_k = 0$. If $v_1, \dots, v_k \in \mathbb{R}^n$ are not linearly dependent, then we say they are *linearly independent*. The *span* of $v_1, \dots, v_k$, denoted by $\operatorname{span}\{v_1, \dots, v_k\}$, is the space of all *linear combinations* of $v_1, \dots, v_k$, i.e., $\operatorname{span}\{v_1, \dots, v_k\} = \{c_1 v_1 + \cdots + c_k v_k : c_1, \dots, c_k \in \mathbb{R}\}$. The span of a collection of vectors from $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$, which is itself a real vector space in its own right. If $v_1, \dots, v_k \in \mathbb{R}^n$ are linearly independent, then they form an *(ordered) basis* for $\operatorname{span}\{v_1, \dots, v_k\}$. In this case, for every vector $u \in \operatorname{span}\{v_1, \dots, v_k\}$, there is a unique choice of $c_1, \dots, c_k \in \mathbb{R}$ such that $u = c_1 v_1 + \cdots + c_k v_k$.

We agree on a special ordered basis $e_1, \dots, e_n$ for $\mathbb{R}^n$, which we call the *standard coordinate basis* for $\mathbb{R}^n$. This ordered basis defines a coordinate system, and we write vectors $v \in \mathbb{R}^n$ in terms of this coordinate system, as $v = (v_1, \dots, v_n) = \sum_{i=1}^n v_i e_i$. The *(Euclidean) inner product* (or *dot product*) on $\mathbb{R}^n$ will be written either using the transpose notation, $u^\top v$, or the angle bracket notation, $\langle u, v \rangle$. In terms of their coordinates $u = (u_1, \dots, u_n)$ and $v = (v_1, \dots, v_n)$, we have

$$\langle u, v \rangle = \sum_{i=1}^n u_i v_i .$$

The *(Euclidean) norm* will be written as $\|v\|_2 = \sqrt{\langle v, v \rangle}$. The inner product satisfies the *Cauchy-Schwarz inequality*,

$$\langle u, v \rangle \leq \|u\|_2 \|v\|_2, \quad u, v \in \mathbb{R}^n,$$

as well as the *polarization identity*

$$\langle u, v \rangle = \frac{\|u + v\|_2^2 - \|u - v\|_2^2}{4}, \quad u, v \in \mathbb{R}^n.$$

The vectors $e_1, \dots, e_n$ are *orthogonal*, i.e., $\langle e_i, e_j \rangle = 0$ for $i \neq j$. Each e_i is also a *unit vector*, i.e., $\|e_i\|_2 = 1$. A collection of orthogonal unit vectors is said to be *orthonormal*, so the basis $e_1, \dots, e_n$ is orthonormal.

Linear maps

Linear maps $A \colon \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ are written as *matrices* in $\mathbb{R}^{m \times n}$, using the standard coordinate bases in the respective Euclidean spaces:

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}.$$

The *adjoint* $A^\top \colon \mathbb{R}^m \to \mathbb{R}^n$ of a linear map $A \colon \mathbb{R}^n \to \mathbb{R}^m$ is written using the transpose notation:

$$\langle u, Av \rangle = \langle A^\top u, v \rangle, \quad u \in \mathbb{R}^m; v \in \mathbb{R}^n.$$

In matrix notation, we also have

$$A^\top = \begin{bmatrix} A_{1,1} & \cdots & A_{m,1} \\ \vdots & \ddots & \vdots \\ A_{1,n} & \cdots & A_{n,m} \end{bmatrix}.$$

Note that $(A^\top)^\top = A$. Composition of linear maps $A\colon \mathbb{R}^n \to \mathbb{R}^m$ and $B\colon \mathbb{R}^p \to \mathbb{R}^n$ is obtained by *matrix multiplication*: $C = AB$, where

$$C_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}, \quad i = 1, \dots, m; j = 1, \dots, p.$$

The adjoint of the composition AB is the composition of the adjoints in reverse order: $(AB)^\top = B^\top A^\top$. In the context of matrix multiplication, vectors $v \in \mathbb{R}^n$ shall be regarded as *column vectors*, so

$$Av = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n A_{1,j} v_j \\ \vdots \\ \sum_{j=1}^n A_{m,j} v_j \end{bmatrix}.$$

If the j-th column of A is $a_j \in \mathbb{R}^m$ (for each $j = 1, \dots, n$), then $Av = \sum_{j=1}^n v_j a_j$. Note that this is consistent with the transpose notation for inner products $u^\top v$. If the i-th row of A is $a_i^\top$ for some $a_i \in \mathbb{R}^n$ (for each $i = 1, \dots, m$), then $Av = (a_1^\top v, \dots, a_m^\top v)$. The *outer product* of vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ refers to $uv^\top \in \mathbb{R}^{m \times n}$:

$$uv^\top = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \vdots & \ddots & \vdots \\ u_m v_1 & \cdots & u_m v_n \end{bmatrix}.$$

Fundamental subspaces

With every linear map $A\colon \mathbb{R}^n \to \mathbb{R}^m$, we associate four fundamental *subspaces*: range(A), range$(A^\top)$, null(A), and null$(A^\top)$. The *range* of a linear map $A\colon \mathbb{R}^n \to \mathbb{R}^m$, denoted range$(A)$, is the subspace $\{Av : v \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$. Its dimension is the *rank* of A, denoted rank(A). When $A \in \mathbb{R}^{m \times n}$ is regarded as a matrix, the range of A is the same as the *column space* of A (so if the columns of A are the vectors $a_1, \dots, a_n \in \mathbb{R}^n$, range$(A) = \text{span}\{a_1, \dots, a_n\}$). The *row space* of A is the column space of $A^\top$. The *null space* of A, denoted null(A), is the subspace $\{v \in \mathbb{R}^n : Av = 0\} \subseteq \mathbb{R}^n$. We always have

$$\text{rank}(A) = \text{rank}(A^\top)$$

and

$$n = \dim(\text{null}(A)) + \text{rank}(A).$$

In particular, if rank$(A) = n$, which is equivalent to the columns of A being linearly independent, then null$(A) = \{0\}$. The subspaces range(A) and null$(A^\top)$ are *orthogonal*, written range$(A) \perp$ null$(A^\top)$, meaning every $u \in$ range(A) and $v \in$ null$(A^\top)$ have $\langle u, v \rangle = 0$. Similarly, range$(A^\top)$ and null(A) are orthogonal.

We can obtain an orthonormal basis $v_1, v_2, \dots$ for range(A) using the *Gram-Schmidt orthogonalization process*, which is given as follows. Let $a_1, \dots, a_n$ denote the columns of A. Then for $i = 1, 2, \dots$: (1) if all a_j are zero, then stop; (2) pick a non-zero a_j; (3) let $v_i := a_j / \|a_j\|_2$; (4) replace each a_j with $a_j - \langle v_i, a_j \rangle v_i$.

Linear operators

A *linear operator* $A\colon \mathbb{R}^n \to \mathbb{R}^n$ (where range(A) is a subspace of the domain) is represented by a *square* matrix $A \in \mathbb{R}^{n \times n}$. We say A is *singular* if $\dim(\text{null}(A)) > 0$; if $\dim(\text{null}(A)) = 0$, we say A is *non-singular*.

The *identity* map is denoted by I (or sometimes I_n to emphasize that it is the identity operator $I\colon \mathbb{R}^n \to \mathbb{R}^n$ for $\mathbb{R}^n$), and I is clearly non-singular. Its matrix representation is

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

i.e., every *diagonal* entry is 1 and every *off-diagonal* entry is 0. (A matrix is *diagonal* if all of its off-diagonal entries are 0.) A linear operator is non-singular if and only if it has an *inverse*, denoted A^{-1}, that satisfies $AA^{-1} = A^{-1}A = I$. (So a synonym for *non-singular* is *invertible*.) A linear operator A is *self-adjoint* (or equivalently, its matrix representation is *symmetric*) if $A = A^\top$. If A and B are non-singular, then so is their composition AB; the inverse of AB is $(AB)^{-1} = B^{-1}A^{-1}$. Also, if A is non-singular, then so is $A^\top$, and its inverse is denoted by $A^{-\top}$.

A linear operator A is *orthogonal* if $A^\top$ is its inverse, i.e., $A^\top = A^{-1}$. From the matrix equation $A^\top A = I$, we see that if $a_1, \dots, a_n \in \mathbb{R}^n$ are the columns of A, then A is orthogonal if and only if the vectors $a_1, \dots, a_n$ are orthonormal. If A is orthogonal, then for any vector $v \in \mathbb{R}^n$, we have $\|Av\|_2 = \|v\|_2$ (*Parseval's identity*).

A linear operator $A\colon \mathbb{R}^n \to \mathbb{R}^n$ is a *projection operator* (or *projector*) if $AA = A$ (i.e., A is *idempotent*), $Ax = x$ for all $x \in \operatorname{range}(A)$, and every $x \in \mathbb{R}^n$ can be written uniquely as $x = y + z$ for some $y \in \operatorname{range}(A)$ and $z \in \operatorname{null}(A)$ (i.e., $\mathbb{R}^n$ is the *direct sum* of $\operatorname{range}(A)$ and $\operatorname{null}(A)$, written $\mathbb{R}^n = \operatorname{range}(A) \oplus \operatorname{null}(A)$). A *projector* A is an *orthogonal projection operator* (or *orthoprojector*) if $A = A^\top$. Every subspace of $\mathbb{R}^n$ has a unique orthoprojector. A generic way to obtain the orthoprojector Π for a subspace S is to start with an orthonormal basis $u_1, u_2, \dots$ for S, and then form the sum of outer products $\Pi := \sum_i u_i u_i^\top$. For any orthoprojector Π, we have the *Pythagorean theorem*

$$\|v\|_2^2 = \|\Pi v\|_2^2 + \|(I - \Pi)v\|_2^2, \quad v \in \mathbb{R}^n.$$

In particular, for any $v \in \mathbb{R}^n$ and $u \in \operatorname{range}(\Pi)$,

$$\|u - v\|_2^2 = \|\Pi v - u\|_2^2 + \|v - \Pi v\|_2^2.$$

Put another way, the *orthogonal projection* of $v \in \mathbb{R}^n$ is the closest point in $\operatorname{range}(\Pi)$ to v.

Determinants

The determinant of $A \in \mathbb{R}^{n \times n}$, written $\det(A)$, is defined by

$$\det(A) = \sum_\sigma \operatorname{sgn}(\sigma) \prod_{i=1}^n A_{i,\sigma(i)},$$

where the summation is over all permutations σ of $\{1, \dots, n\}$, and $\operatorname{sgn}(\sigma)$ is the *sign* of the permutation σ (which takes value either 1 or -1). When the n^2 entries of the matrix A are viewed as formal variables, the determinant can be regarded as a degree-n multivariate polynomial in these variables.

A linear operator A is non-singular if and only if $\det(A) \neq 0$.

Eigenvectors and eigenvalues

A scalar λ is an *eigenvalue* of a linear operator $A\colon \mathbb{R}^n \to \mathbb{R}^n$ if there is a non-zero vector $v \neq 0$ such that $Av = \lambda v$. This vector v is an *eigenvector* corresponding to the eigenvalue λ. Note that corresponding eigenvectors are not unique; if v is an eigenvector corresponding to λ, then so is cv for any $c \neq 0$. Every linear operator has eigenvalues. To see this, observe the following equivalences:

$$
\begin{aligned}
&\lambda \text{ is an eigenvalue of } A \\
\Leftrightarrow\ &\text{there exists } v \neq 0 \text{ such that } Av = \lambda v \\
\Leftrightarrow\ &\text{there exists } v \neq 0 \text{ such that } (A - \lambda I)v = 0 \\
\Leftrightarrow\ &A - \lambda I \text{ is singular} \\
\Leftrightarrow\ &\det(A - \lambda I) = 0.
\end{aligned}
$$

The function $\lambda \mapsto \det(A - \lambda I)$ is a degree-n polynomial in λ, and hence it has n roots (where some roots may be repeated, and some may be complex).[1] The determinant of $A \in \mathbb{R}^{n\times n}$ is the product of its n eigenvalues.

An important special case is when A is self-adjoint (i.e., A is symmetric). In this case, all n of its eigenvalues $\lambda_1, \dots, \lambda_n$ are real, and all corresponding eigenvectors are vectors in $\mathbb{R}^n$. In particular, there is a collection of n corresponding eigenvectors $v_1, \dots, v_n$, where v_i corresponds to λ_i, such that $v_1, \dots, v_n$ are orthonormal, and $A = \sum_{i=1}^n \lambda_i v_i v_i^\top$. When all of the eigenvalues are non-negative, we say A is *positive semi-definite* (*psd*); when all of the eigenvalues are positive, we say A is *positive definite*.

The *trace* of a matrix $A \in \mathbb{R}^{n\times n}$, denoted $\operatorname{tr}(A)$, is the sum of its diagonal entries. The sum of the n eigenvalues of A is equal to the trace of A. For symmetric matrices A, this fact can be easily deduced from the fact that $\operatorname{tr}\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ is linear. Indeed, let $v_1, \dots, v_n$ be orthonormal eigenvectors corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$ of A. Then

$$\operatorname{tr}(A) = \operatorname{tr}\left(\sum_{i=1}^n \lambda_i v_i v_i^\top\right) = \sum_{i=1}^n \lambda_i \operatorname{tr}(v_i v_i^\top) = \sum_{i=1}^n \lambda_i.$$

The last step uses the fact that $\operatorname{tr}(v_i v_i^\top) = \operatorname{tr}(v_i^\top v_i) = v_i^\top v_i = 1$. This is a special case of a more general identity: if A and B are matrices such that both AB and BA are square, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

[1] If you are not a fan of determinants, you may prefer the approach from http://www.axler.net/DwD.pdf to deduce the existence of eigenvalues.