Review Session 1 – Linear Algebra

References/suggested reading

(i) *Introduction to Linear Algebra*, by Gilbert Strang.

(ii) *Linear Algebra Done Right*, by Sheldon Axler.

(iii) *Numerical Linear Algebra*, by Trefethen and Bau.

1 Linear dependence, rank, and null space

Recall a (real) vector space V is a set of objects called vectors, equipped with two operations: addition and multiplication by real scalars. These operations obey standard axioms (e.g., associativity, commutativity, distributivity of scalar multiplication) and behave as one would expect. The standard example is $\mathbb{R}^d$. Supposing V is a real vector space, let's review some of the basic terminology:

1. Vectors $v_1, \ldots, v_k \in V$ are said to be *linearly independent* if $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0 \implies \alpha_1 = \cdots = \alpha_k = 0$. We can write this more compactly using matrix notation: the columns of $A = (v_1, \ldots, v_k)$ are linearly independent if $A\alpha = 0 \implies \alpha = 0$. Equivalently, $A\alpha \neq 0$ for all nonzero α.
2. The *dimension* of a vector space V is the maximum number of linearly independent vectors in V.
3. The *span* $\mathrm{Span}\{v_1, \ldots, v_n\}$ of a set of vectors $v_1, \ldots, v_n \in V$ is the linear subspace $\{\sum_{i=1}^n \alpha_i v_i : \alpha_i \in \mathbb{R}\}$. For any $v \in \mathrm{Span}\{v_1, \ldots, v_n\}$, $\exists \alpha \in \mathbb{R}^n$ such that $v = \sum_{i=1}^n \alpha_i v_i$.
4. The *basis* of a vector space V is a set of linearly independent vectors $v_1, \ldots, v_k$ with $\mathrm{Span}\{v_1, \ldots, v_k\} = V$.
5. The *column rank* of a matrix $A \in \mathbb{R}^{m \times n}$ is the dimension of the span of its columns $a_{:1}, \ldots, a_{:n}$. We call this span the *column space* of A. The *row rank* and *row space* are analogous. We'll see below that for square matrices, row rank $=$ column rank, and thus in that setting we refer to both of these as just the rank.
6. The *null space* or kernel of a matrix $A \in \mathbb{R}^{m \times n}$, written $\mathrm{Null}(A)$, is the set of vectors that A maps to 0: $\mathrm{Null}(A) := \{v \in \mathbb{R}^n : Av = 0\}$. $\mathrm{Null}(A)$ is a linear subspace of $\mathbb{R}^n$.

Example 1.1

Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$.

1. We claim $\mathrm{rank}(A) = 2$. We have $\alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0$ iff $\alpha + 2\beta = 0$ and $2\alpha + \beta = 0$. Solving the first equation gives $\alpha = -2\beta$ and plugging that into the second gives $\beta = 0$ so that $\alpha = 0$.
2. $\mathrm{Null}(A) = \{0\}$ because if there were any nonzero $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathrm{Null}(A)$, then we'd have a linear combination of the columns that is equal to zero. We just showed above this is not the case (see also the rank-nullity theorem below).

Lemma 1.2

For square matrices, row rank = column rank.

Proof. Let A be an $n \times m$ matrix with column rank r. Let $c_{:1}, \dots, c_{:r}$ be a basis for the column space of A. Each column $a_{:j}$ of A is a linear combination $\sum_{k=1}^{r} c_{:k} b_{kj}$ of the basis vectors. Let C be the matrix with columns $c_{:1}, \dots, c_{:r}$ and B be the matrix with columns $b_{:1}, \dots, b_{:m}$. In terms of these matrices, we have $A = CB$.

Now consider $A^T = B^T C^T$. Each column of A^T is a linear combination of columns of B^T, so each row of A is a linear combination of rows of B. Since B has r rows, the row rank of A is at most r.

Suppose the row rank of A were $s < r$. Then the column rank of A^T is s. Applying the same argument as above in reverse shows that the column rank of A is at most s. Then, we have $r \leq s < r$, a contradiction. Thus, we must have the row rank of A is r. ■

Theorem 1.3 (rank-nullity theorem)

If $A \in \mathbb{R}^{m \times n}$, then $\dim(\mathrm{Null}(A)) + \mathrm{rank}(A) = n$ where $\mathrm{rank}(A)$ denotes the column rank.

Proof. Suppose $\dim(\mathrm{Null}(A)) = k$ and that $\mathrm{Null}(A)$ has basis $\{u_1, \dots, u_k\}$. Then, we can extend this basis to a basis of $\mathbb{R}^n$: $\{u_1, \dots, u_k, v_1, \dots, v_{n-k}\}$. We claim $\{Av_1, \dots, Av_{n-k}\}$ is a basis for the column space of A (which would complete the proof since then $\mathrm{rank}(A) = n - k$). First, we show $\{Av_1, \dots, Av_{n-k}\}$ span the column space of A. We have if $w \in \mathbb{R}^n$ is an arbitrary vector, then there exist $a_1, \dots, a_k, b_1, \dots, b_{n-k} \in \mathbb{R}$ such that

$$\begin{aligned} w &= a_1 u_1 + \cdots + a_k u_k + b_1 v_1 + \cdots + b_{n-k} v_{n-k} \\ Aw &= a_1 A u_1 + \cdots + a_k A u_k + b_1 A v_1 + \cdots + b_{n-k} A v_{n-k} \\ &= b_1 A v_1 + \cdots + b_{n-k} \end{aligned}$$

Next, we show $\{Av_1, \dots, Av_{n-k}\}$ are linearly independent. If $c_1 A v_1 + \cdots + c_{n-k} A v_{n-k} = 0$, then, $A(c_1 v_1 + \cdots + c_{n-k} v_{n-k}) = 0 \implies c_1 v_1 + \cdots + c_{n-k} v_{n-k} \in \mathrm{Null}(A)$. Thus, there exists $d_1, \dots, d_k \in \mathbb{R}$ such that

$$\begin{aligned} & c_1 v_1 + \cdots + c_{n-k} v_{n-k} = d_1 u_1 + \cdots + d_k u_k \\ \implies & c_1 v_1 + \cdots + c_{n-k} v_{n-k} - d_1 u_1 - \cdots - d_k u_k = 0 \\ \implies & c_1 = \cdots = c_{n-k} = d_1 = \cdots = d_k = 0. \end{aligned}$$

Thus, $\{Av_1, \dots, Av_{n-k}\}$ is a basis for the column space of A meaning $\mathrm{rank}(A) = n - k$. ■

2 Invertibility

A matrix A is called *right-invertible* if there exists a matrix B such that $AB = \mathrm{Id}$ and left-invertible if there exists C such that $CA = \mathrm{Id}$. We call B a right inverse and C a left inverse. We say A is invertible if it is both right and left invertible – in that case, $B = C$:

$$B = (CA)B = C(AB) = C.$$

Theorem 2.1

A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible iff it is left or right invertible.

Proof. Using what we've learned so far, it suffices to show that A has a left inverse iff it has linearly independent columns (then, the same statement can be derived for right-inverses and linearly independent rows). If A has a left inverse, $CA = \mathrm{Id}$, then for any vector $\alpha \neq 0$, $CA\alpha = \alpha$ so that $A\alpha \neq 0$. Thus, A has linearly independent columns.

Next, we show that if A has linearly independent columns, then it has a left inverse. Let $a_{:1}, \dots, a_{:n}$ be the columns of A. Then, by virtue of $\mathrm{Span}\{a_{:1}, \dots, a_{:n}\} = \mathbb{R}^n$, there is some linear combination of the columns of A that make each basis vector or for each $k \in \{1, \ldots, n\}$, there is a vector $\alpha_k \in \mathbb{R}^n$ such that $A\alpha_k = e_k$, the k-th basis vector. Then, the matrix with k-th row α_k is a left-inverse to A:

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \begin{pmatrix} a_{:1} & \cdots & a_{:n} \end{pmatrix} = \begin{pmatrix} a_1^T \alpha_1 & \cdots & a_n^T \alpha_1 \\ \vdots & \ddots & \vdots \\ a_1^T \alpha_n & \cdots & a_n^T \alpha_n \end{pmatrix} = \begin{pmatrix} A\alpha_1 \\ \vdots \\ A\alpha_n \end{pmatrix} = \mathrm{Id}_n .$$

Thus, since a square matrix has linearly independent rows iff it has linearly independent columns, then it is right invertible iff it is left invertible. ■

Theorem 2.1 (multiplying by an invertible matrix doesn't change rank)

If A is $m \times n$ and B is an $n \times n$ invertible matrix, then $\mathrm{rank}(AB) = \mathrm{rank}(A)$.

Proof. A and AB are both $m \times n$ so by rank-nullity it suffices to show $\dim(\mathrm{Null}(A)) = \dim(\mathrm{Null}(AB))$. But, $v \in \mathrm{Null}(A)$ iff $B^{-1}v \in \mathrm{Null}(AB)$. So, the dimensions of these null spaces are equal. ■

Remark 2.2. $A \in \mathbb{R}^{n \times n}$ is invertible if and only if $\mathrm{Null}(A) = \{\mathbf{0}_n\}$. More generally, see here for a list of ways to check if a matrix is invertible.

3 Orthogonality and orthogonalization

Recall an *inner product space* $(V, \langle \cdot, \cdot \rangle)$ is a vector space V equipped with an *inner product* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$. The standard example here for $V = \mathbb{R}^n$ is the Euclidean inner product $\langle v, w \rangle = v^T w$. The *norm* of a vector is then defined as $\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{v^T v}$ which is here the usual 2-norm. Let's run with this standard example for the rest of this section.

One tool that can be very useful when thinking about linear independence is orthogonalization. The *Gram-Schmidt* process takes a sequence of vectors $v_1, \ldots, v_k \in \mathbb{R}^n$ and gives you an orthogonal basis $u_1, \ldots, u_\ell$ for $\mathrm{Span}\{v_1, \ldots, v_k\}$. The idea is to let u_i be v_i minus its projection onto $u_1, \ldots, u_{i-1}$ so that $v_i \in \mathrm{Span}\{u_1, \ldots, u_i\}$ and u_i is orthogonal to $u_1, \ldots, u_{i-1}$. Here is the construction in a little more detail.

Let $\mathrm{proj}_u(v) = \frac{u^T v}{\|u\|^2} \cdot u$. Then, $\mathrm{proj}_u(v)$ is the projection of v onto u: i.e., it has the same direction as u and has the same inner product with u that v does. We can check this easily: $u^T \mathrm{proj}_u(v) = \frac{u^T v}{\|u\|^2} u^T u = u^T v$. Now let's construct our sequence $u_1, \ldots, u_\ell$. For ease of bookkeeping we'll construct a sequence $\tilde{u}_1, \ldots, \tilde{u}_k$ where some of the $\tilde{u}_i$ are zero, then discard all the zero vectors.

Let $\tilde{u}_1 = v_1$ and $\tilde{u}_i = v_i - \sum_{k=1}^{i-1} \mathrm{proj}_{\tilde{u}_j}(v_i)$. We'll show that $\tilde{u}_1, \ldots, \tilde{u}_n$ are pairwise orthogonal and have the same span as $v_1, \ldots, v_n$ using induction. The base case is obvious. For the inductive step, assume that $\tilde{u}_1, \ldots, \tilde{u}_{i-1}$ have the same span as $v_1, \ldots, v_{i-1}$.

- **Orthogonality:** We'll show that for $j < i$, $\tilde{u}_j$ is orthogonal to $\tilde{u}_i$. Combining this with our inductive hypothesis gives pairwise orthogonality of $\tilde{u}_1, \ldots, \tilde{u}_i$. We have

$$\tilde{u}_j^T \tilde{u}_i = \tilde{u}_j^T v_i - \sum_{k=1}^{i-1} \tilde{u}_j^T \mathrm{proj}_{\tilde{u}_j}(v_i) = \tilde{u}_i^T v_i - \tilde{u}_j^T \mathrm{proj}_{\tilde{u}_j}(v_i) = 0.$$

  The second equality holds since $\tilde{u}_j$ is orthogonal to $\tilde{u}_k$ for $k \neq j < i$ by our inductive hypothesis. The last equality holds because $u^T \mathrm{proj}_u(v) = u^T v$ for any u, v.

- **Span:** We'll show that we can express each of $v_1, \ldots, v_i$ as linear combinations of the other $\tilde{u}_1, \ldots, \tilde{u}_i$. By our inductive hypothesis, it suffices to show that we can express v_i as a linear combination of $\tilde{u}_1, \ldots, \tilde{u}_i$ and $\tilde{u}_i$ as a linear combination of $\tilde{u}_1, \ldots, \tilde{u}_{i-1}, v_i$. The definition $v_i = \tilde{u}_i + \sum_{k=1}^{i-1} \tilde{\alpha}_{ik} u_k$ where $\tilde{\alpha}_{ik} = v_i^T u_k$ gives us both.

Some of the $\tilde{u}_i$ we computed may have been zero, so we can discard them to compute a new sequence $u_1, \ldots, u_\ell = \tilde{u}_{k_1}, \ldots, \tilde{u}_{k_\ell}$ where $k_1, \ldots, k_\ell$ are the indices of the nonzero $\tilde{u}$.

This algorithm gives us more than an orthogonal basis for $\mathrm{Span}\{v_1, \ldots, v_n\}$. It gives us a matrix decomposition $A := (v_1, \ldots, v_n) = QR$ where the columns of Q are orthonormal and R is upper triangular (also called the *QR decomposition* of A). Let $\alpha_{i1}, \ldots, \alpha_{i\ell} = \tilde{\alpha}_{ik_1}, \ldots, \tilde{\alpha}_{ik_\ell}$ to match u.

$$\begin{aligned}
\begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix} &= \begin{pmatrix} \tilde{u}_1 & \cdots & \tilde{u}_k \end{pmatrix} \begin{pmatrix} 1 & \tilde{\alpha}_{12} & \tilde{\alpha}_{13} & \cdots \tilde{\alpha}_{1k} & \\ 0 & 1 & \tilde{\alpha}_{23} & \cdots & \tilde{\alpha}_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \\
&= \begin{pmatrix} u_1 & \cdots & u_\ell \end{pmatrix} \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1k} \\ 0 & 1 & \alpha_{23} & \cdots & \alpha_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} \frac{u_1}{\|u_1\|} & \cdots & \frac{u_\ell}{\|u_\ell\|} \end{pmatrix}}_{Q_1} \underbrace{\begin{pmatrix} \|u_1\| & \|u_1\|\alpha_{12} & \|u_1\|\alpha_{13} & \cdots & \|u_1\|\alpha_{1k} \\ 0 & \|u_2\| & \|u_2\|\alpha_{23} & \cdots & \|u_2\|\alpha_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \|u_\ell\| \end{pmatrix}}_{R_1}
\end{aligned}$$

If $k = n$ and $v_1, \ldots, v_n$ span $\mathbb{R}^n$, then Q is an $n \times n$ matrix with orthonormal columns and R_1 is an $n \times n$ upper triangular matrix with positive entries on the diagonal. If $v_1, \ldots, v_k$ don't span $\mathbb{R}^n$, we can get a similar decomposition. Let $Q_1^\perp$ be an orthogonal basis for the r-dimensional subspace W orthogonal to $\mathrm{Span}\{v_1, \ldots, v_k\}$ (recall this means $w \in W$ iff $\langle w, v_i \rangle = 0$ for all i). Then,

$$(v_1, \ldots, v_k) = \underbrace{\begin{pmatrix} Q_1 & Q_1^\perp \end{pmatrix}}_{Q} \underbrace{\begin{pmatrix} R_1 \\ 0_{r \times k} \end{pmatrix}}_{R}.$$

We call this the QR decomposition of the matrix $A = (v_1, \ldots, v_k)$. Q is an $n \times n$ matrix with orthonormal columns and R is an $n \times k$ upper triangular matrix. Notice that for full-rank square A, $Q = Q_1$ and $R = R_1$.

So far I've been using the phrase 'an $n \times n$ matrix with orthonormal columns' whereas it is conventional to use the term 'orthogonal matrix'. The conventional definition of an orthogonal matrix is a matrix Q such that $Q^T Q = QQ^T = \mathrm{Id}$. It's straightforward to show that these definitions are equivalent. We have $(Q^T Q)_{ij} = q_{:i}^T q_{:j}$ so that $Q^T Q = \mathrm{Id}$ iff Q has orthonormal columns. Because Q^T is a left inverse of Q, it is also a right inverse $QQ^T = \mathrm{Id}$. To get another equivalence, observe that $QQ^T = \mathrm{Id}$ iff the columns of Q^T, which are the rows of Q, are orthonormal.

One common application of the QR decomposition is to solve linear systems and least square problems. For a brief summary of how that works, let's consider the system $Ax = b$ where $A = QR$. Geometrically, we're asking whether b is in the span of the columns of A. Because Q is invertible with inverse Q^T, $Ax = b$ iff $Rx = Q^T b$. If A is not full rank, R will have some zero rows, and if $Q^T b$ is nonzero in those rows, then the equation has no solution. If $Q^T b$ is zero in all rows where R is, then we can compute solution easily through back substitution.

4 Determinants

Recall the determinant of a set of vectors $v_1, \ldots, v_n \in \mathbb{R}^n$, written $\det(v_1, \ldots, v_n)$, is the signed volume of the parallelotope with sides $v_1, \ldots, v_n$ (this is a mathematical generalization of the notion of volume of a parallelepiped for $n = 3$). It's common to talk about the determinant $\det(A)$ of an $n \times n$ matrix. This means the determinant $\det(a_{:1}, \ldots, a_{:n})$ of the columns of A.

The determinant tends to come up in two roles:

1. When we reparametrize, or do a change of variables, in an integral by letting $x = \phi(u)$, we need to know the volume of $\phi(U)$ where U is an infinitesimal cube. Linearly approximating ϕ, this amounts to finding the volume of a parellelepiped which means the determinant should come into play. For the core competency exam, two common situations where you need to do this are (i) when computing the pdf of a transformation of a random vector and (ii) when computing the Fisher information of a transformation of a parameter.

2. When we want to know whether a set of vectors $v_1, \dots, v_n$ is linearly independent. The volume of the parallelotope with columns $v_1, \dots, v_n$ is zero iff it is flat in some direction. The direction of flatness is a projection of the columns that is equal to zero – it indicates linear dependence. In fact, $\det(v_1, \dots, v_n) = 0$ iff $v_1, \dots, v_n$ are linearly dependent, so we're often interested only in whether $\det$ is zero or nonzero. We'll prove this property when we have a definition of $\det$.

4.1 Determinants in $\mathbb{R}^2$

Let's start by defining it in $\mathbb{R}^2$ because the notation is easier. The determinant must satisfy three properties:

1. $\det$ is multilinear: $\det(u, v + \alpha w) = \det(u, v) + \alpha \det(u, w)$.
2. $\det$ is alternating: $\det(u, v) = -\det(v, u)$.
3. $\det(\mathrm{Id}) = 1$.

In fact, these properties uniquely define the determinant. We'll use them to derive a formula for the determinant, but first we'll need to prove one property.

Lemma 4.1

$\det(v, v) = 0$.

Proof. This follows from the alternating property. $\det(v, v) = -\det(v, v)$. ■

Now, we'll derive a formula for the determinant of a matrix $A \in \mathbb{R}^{2\times 2}$. Let e_1, e_2 be the columns of the identity. and let $a_{:1} = a_{11}e_1 + a_{21}e_2$ and $a_{:2} = a_{12}e_1 + a_{22}e_2$. Then, we have

$$\begin{aligned} \det(a_{:1}, a_{:2}) &= \det(a_{11}e_1 + a_{21}e_2, a_{12}e_1 + a_{22}e_2) \\ &= a_{11}\det(e_1, a_{12}e_1 + a_{22}e_2) + a_{21}\det(e_2, a_{12}e_1 + a_{22}e_2) \\ &= a_{11}a_{12}\det(e_1, e_1) + a_{12}a_{22}(e_1, e_2) + a_{12}a_{12}\det(e_2, e_1) + a_{21}a_{22}\det(e_2, e_2) \\ &= a_{11}a_{22} - a_{21}a_{12}, \end{aligned}$$

where we use the third and second properties of the determinant to plug in $\det(e_1, e_2) = 1 = -\det(e_2, e_1)$ and our lemma to plug in $\det(e_1, e_1) = \det(e_2, e_2) = 0$.

Now, let's generalize this to $\mathbb{R}^n$.

4.2 Determinants in $\mathbb{R}^n$

The properties of our determinant generalize to:

1. $\det$ is multilinear: $\det(u_1, \dots, u_i + \alpha v_i, \dots, u_n) = \det(u_1, \dots, u_i, \dots, u_n) + \alpha \det(v_1, \dots, v_i, \dots, u_n)$.
2. $\det$ is alternating: $\det(u_1, \dots, u_{i-1}, u_j, u_{i+1}, \dots, u_{j-1}, u_i, u_{j+1}, \dots, u_n) = -\det(u_1, \dots, u_n)$ (in other words, if we swap vectors u_i and u_j, we flip the sign of the determinant).
3. $\det(\mathrm{Id}) = 1$.

Lemma 4.2

$u_i = u_j \implies \det(u_1, \dots, u_n) = 0$.

Proof. We can imagine swapping the i-th and j-th columnn which does not change the value of the determinant, but at the same time by the alternating property, flips its sign. Thus, we must have $\det(u_1, \dots, u_n) = -\det(u_1, \dots, u_n) = 0$. ■

Lemma 4.3

For $i \neq j$, $\det(u_1, \dots, u_{i-1}, u_i + \alpha u_j, u_{i+1}, \dots, u_n) = \det(u_1, \dots, u_n)$.

Proof.

$$\begin{aligned}\det(u_1, \dots, u_{i-1}, u_i + \alpha u_j, u_{i+1}, \dots, u_n) &= \det(u_1, \dots, u_n) + \alpha \det(u_1, \dots, u_{i-1}, u_j, u_{i+1}, \dots, u_{j-1}, u_j, u_{j+1}, \dots, u_n) \\ &= \det(u_1, \dots, u_n) + 0 \quad \text{(by the previous lemma)}\end{aligned}$$

∎

Proposition 4.4

If $u_1, \dots, u_n$ are linearly dependent, then $\det(u_1, \dots, u_n) = 0$.

Proof. If $u_1, \dots, u_n$ are linearly dependent, then $\exists j$ such that $u_j = \sum_{i \neq j} \alpha_i u_i$. WLOG, let $j = 1$. Then, $\det(u_1, \dots, u_n) = \sum_{i=2}^n \alpha_i \det(u_i, u_2, \dots, u_i, \dots, u_n) = \sum_{i=2}^n \alpha_i \cdot 0 = 0$, where we use the previous two lemmas. ∎

Now, we're ready to work out a formula for the determinant of an $n \times n$ matrix. Let $e_1, \dots, e_n$ be the columns of the $n \times n$ identity matrix. Then, expanding each $a_{:i}$ into a linear combination of basis vectors, we have:

$$\begin{aligned}\det(A) &= \det(a_{:1}, \dots, a_{:n}) \\ &= \det\left(\sum_{i_1=1}^n a_{i_1,1} e_{i_1}, a_{:2}, \dots, a_{:n}\right) \\ &= \sum_{i_1=1}^n a_{i_1,1} \det(e_{i_1}, a_{:2}, \dots, a_{:n}) \\ &= \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n a_{i_1,1} \cdots a_{i_n,n} \det(e_{i_1}, \dots, e_{i_n}).\end{aligned}$$

We have that $\det(e_{i_1}, \dots, e_{i_n}) = 0$ if $i_j = i_k$ for some $j \neq k$ – in other words, unless $i_1, \dots, i_n$ is a permutation of the numbers $1, \dots, n$. So, we just need to sum the above summand over the set S_n of permutations of the n numbers $\{1, \dots, n\}$. Let $\sigma \in S_n$ be a permutation. The *sign* of the permutation, $\mathrm{sgn}(\sigma)$, is 1 if the number of interchanges taking σ to $\{1, \dots, n\}$ is even and -1 if it is odd. Recall here a basic fact of combinatorics that every permutation of n objects can be written as the composition of interchanges (i.e., swapping consecutive objects one at a time) and that any such composition is always even or odd in length depending on the permutation. But, this is the same rule we use when interchanging columns in the determinant: $\mathrm{sgn}(\sigma) = \det(e_{\sigma_1}, \dots, e_{\sigma_n})$. Thus, our determinant formula is:

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \cdot a_{\sigma_1,1} \cdots a_{\sigma_n,n}.$$

Let's think about why we might expect this formula. First, multilinear functions are products, so we'd expect to have a product of entires of $a_{i:}$. Second, we often sum a function over all permutations of its arguments to get a permutation invariant version of the function. Summing a function times an alternating function of the permutation over all permutations seems like an obvious way to get an alternating function.

Now, let's work out a few more properties of the determinant:

Proposition 4.5

$\det(AB) = \det(A)\det(B)$.

Proof. When we derived the formula for $\det(A)$ above, we wrote each column of A as a sum of the columns of Id. We then used multilinearity and the alternating property to transform this into a sum $\sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)a_{\sigma_1,1}\cdots a_{\sigma_n,n}$. Here, in a similar manner, we're going to write each column of AB as a sum of columns of A and use a similar process.

$$
\begin{aligned}
AB &= \left[\sum_{i_1=1}^n a_{:1}b_{i_1,1} \quad \ldots \quad \sum_{i_n=1}^n a_{:,i_n}b_{i_n,n}\right] \\
\det(AB) &= \det\left(\sum_{i_1=1}^n a_{:1}b_{i_1,1}, \ldots, \sum_{i_n=1}^n a_{:,i_n}b_{i_n,n}\right) \\
&= \sum_{i_1=1}^n \cdots \sum_{i_1=1}^n b_{i_1,1}\cdots b_{i_n,n}\det(a_{:,i_1},\ldots,a_{:,i_n}) \\
&= \sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\det(A)\prod_{i=1}^n b_{\sigma_i,i} \\
&= \det(A)\det(B)
\end{aligned}
$$

■

Proposition 4.6

$\det(A^T) = \det(A)$.

Proof. This one relies on the property that the set S_n of permutations on n indices is a group under composition. Each element $\sigma\in S_n$ has a unique inverse $\sigma^{-1}\in S_n$ such that $\sigma(\sigma_i^{-1}) = \sigma^{-1}(\sigma_i) = i$. We'll also need the property that $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\sigma^{-1})$ (which is clear from writing σ as a composition of interchanges and then realizing σ^{-1} is the same composition in reverse). So, we have

$$
\begin{aligned}
\det(A) &= \sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\prod_{i=1}^n a_{\sigma_i,i} \\
&= \sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\prod_{i=1}^n a_{\sigma(\sigma^{-1}(i)).\sigma^{-1}(i)} \quad \text{(reordering the product)} \\
&= \sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\prod_{i=1}^n a_{i,\sigma^{-1}(i)} \\
&= \sum_{\sigma\in S_n}\mathrm{sgn}(\sigma^{-1})\prod_{i=1}^n a_{i,\sigma^{-1}(i)} \\
&= \sum_{\sigma'\in S_n}\mathrm{sgn}(\sigma')\prod_{i=1}^n a_{i,\sigma'_i} \quad \text{(writing } \sigma' \text{ for } \sigma^{-1} \text{ and reordering the sum)}
\end{aligned}
$$

■

One consequence of the above proposition is that because the determinant is invariant under the column operation $a_{:i}\leftarrow a_{:i}+\alpha a_{:j}$, the determinant must also be invariant under the row operation $a_{i:}\leftarrow a_{i:}+\alpha a_{j:}$.

Proposition 4.7

$\det(A) = 0$ iff $a_{:1},\ldots,a_{:n}$ are linearly dependent.

Proof. We already showed the 'if' direction earlier, so we just need to show the 'only if' part. If A has linearly independent columns, then these columns span $\mathbb{R}^n$ so that A has a QR decomposition $A = QR$ where Q is a matrix with orthonormal columns and R is upper triangular with positive entries on the diagonal. Then, $\det(A) = \det(Q)\det(R)$.

We first compute $\det(Q)$. Since $Q^TQ = \mathrm{Id}$, we must have $\det(Q^T)\det(Q) = \det(\mathrm{Id}) = 1$. But, we also have $\det(Q^T) = \det(Q)$ so that $\det(Q)^2 = 1$ and thus $\det(Q) = \pm 1$.

Next, we compute $\det(R)$. Consider the formula $\det(R) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) r_{\sigma_1,1} \cdots r_{\sigma_n,n}$. Because $r_{ij} = 0$ for $i < j$,. the term associated with a permutation σ is nonzero only if $\sigma_i \geq i$ for all $i \in 1, \ldots, n$. Only the identity permutation $\sigma : \sigma_i = i$ satisfies this constraint. Can you see why? Thus, $\det(R) = r_{11} \cdots r_{nn}$, the product of the diagonal elements. Because these are all positive, $\det(R) > 0$.

Using these properties, we have $\det(A) = \pm \det(R) \neq 0$. Thus, we have shown the contrapositive of $\det(A) = 0$ if $a_{:1}, \ldots, a_{:n}$ are linearly dependent. ■

Computationally, it's often more efficient to use other methods like QR decomposition to determine whether a set of vectors is linearly dependent, but the determinant will be a powerful theoretical tool. We'll see it applied to the question of existence of eigenvalues and eigenvectors in the next section.

5 Eigenvalues and eigenvectors

5.1 Existence of eigenvalues and eigenvectors

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. Suppose there exists a vector $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that

$$Av = \lambda v.$$

Then, v is called an **eigenvector** of A corresponding to **eigenvalue** λ. Typically, we assume the eigenvector v is nonzero.

If λ and v are corresponding eigenvalue and eigenvector, then $(A - \lambda \operatorname{Id})v = 0$. From the previous section, we know there exists v such that $(A - \lambda \operatorname{Id})v = 0$ iff $\det(A - \lambda \operatorname{Id}) = 0$. Applying the formula for the determinant, we see that $\det(A - \lambda \operatorname{Id})$ is an n-th degree polynomial in λ and therefore has between 1 and n distinct roots λ_i by the fundamental theorem of algebra. These are the eigenvalues of A. Note that they may be complex in general.

It's tempting to ask whether A having exactly k distinct eigenvalues implies it has exactly k distinct eigenvectors. That isn't quite the right question. If v_i is an eigenvector associated with an eigenvalue λ_i, then so is αv_i for any scalar α, so there is an infinite scale-family of eigenvectors associated with every eigenvalue. Even if we constrain all eigenvalues to have unit length, we still get infinite families in some cases. Consider the identity matrix Id. $\operatorname{Id} v = 1 \cdot v$ for all vectors v so every vector in $\mathbb{R}^n$ is an eigenvector of Id associated with the eigenvalue 1. Instead, we can say the following.

Theorem 5.1

If $v_1, \ldots, v_k$ are eigenvectors associated with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then $v_1, \ldots, v_k$ are linearly independent.

Proof. Let j be the maximal j such that $v_1, \ldots, v_j$ are linearly independent. If $j < k$, then there exists α such that $v_{j+1} = \sum_{i=1}^{j} \alpha_i v_i$. Multiplying both sides by A, we get $\lambda_{j+1} v_{j+1} = \sum_{i=1}^{j} \lambda_i \alpha_i v_i$. Expanding v_{j+1}, we get $\lambda_{j+1} \sum_{i=1}^{j} \alpha_i v_i = \sum_{i=1}^{j} \lambda_i \alpha_i v_i$ and therefore $\sum_{i=1}^{j} (\lambda_{j+1} - \lambda_j) \alpha_i v_i = 0$, which implies that $v_1, \ldots, v_j$ are not linearly independent either, contradicting our assumption. Therefore, we must have $j = k$. ■

This guarantees us a basis of eigenvectors if $A \in \mathbb{R}^{n \times n}$ has n distinct eigenvalues. Behavior is more complex when it has $k < n$. As we saw with the example of the identity matrix, sometimes a matrix with a repeated eigenvalue has a basis of eigenvectors. Let's look at one that doesn't.

Example 5.2

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

We have $\det(A - \lambda \operatorname{Id}) = (1-\lambda)^2$ so A has only one eigenvalue 1. Its eigenvectors satisfy $(A - 1 \cdot \operatorname{Id})v = 0$ or

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Then, y must be zero, so the subspace of eigenvectors is spanned by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

We call the order of the root λ in the polynomial $\det(A - \lambda \operatorname{Id})$ the *algebraic multiplicity* of λ. We call the number of linearly independent eigenvectors with eigenvalue λ the *geometric multiplicity* of λ.

When considering complex matrices $A \in \mathbb{C}^{n \times n}$ and complex eigenvectors $v \in \mathbb{C}^n$ and eigenvalues $\lambda \in \mathbb{C}$, by the fundamental theorem of algebra, $\det(A - \lambda \operatorname{Id})$ has n roots in λ meaning the sum of algebraic multiplicites $\sum_{i=1}^k a_i = n$. If $g_1, \ldots, g_k$ are the geometric multiplicities of the k distinct eigenvalues of A, then $k \leq \sum_{i=1}^k g_i \leq n = \sum_{i=1}^k a_i$. A stronger relationship $1 \leq g_i \leq a_i$ can be proven using an argument like the one in next section's proof of the spectral theorem for symmetric matrices.

Theorem 5.3

If $v_1, \ldots, v_k$ are the eigenvectors of the same eigenvalue λ, then any vector in $\operatorname{Span}\{v_1, \ldots, v_k\}$ is also an eigenvector.

Proof. Let $w = \sum_i \alpha_i v_i$. Then, $Aw = \sum_i \alpha_i A v_i = \sum_i \alpha_i \lambda v_i = \lambda w$. ∎

For this reason we sometimes refer to the *eigenspace* or subspace (of eigenvectors) associated with an eigenvalue. In the next section, we'll look at symmetric matrices, for which it is possible to prove existence of a basis of eigenvectors even when eigenvalues are repeated. In the remainder of this one, we'll discuss a decomposition that exists for matrices which have a basis of eigenvectors.

5.2 Diagonalization

Theorem 5.4

Suppose that $A \in \mathbb{C}^{n \times n}$ has a basis of eigenvectors. Then $A = V \lambda V^{-1}$ where Λ is a diagonal matrix. We call this a *diagonalization* or eigenvalue decomposition of A.

Proof. Let $V = (v_{:1}, \ldots, v_{:n})$ where $v_{:1}, \ldots, v_{:n}$ are eigenvectors of A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Let Λ be a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \ldots, \lambda_n$. Then,

$$AV = (Av_{:1}, \ldots, Av_{:n}) = (\lambda_1 v_{:1}, \ldots, \lambda_n v_{:n}) = V\Lambda.$$

Next, since the eigenvectors are linearly independent, V is invertible, so $A = V \Lambda V^{-1}$. ∎

This result and its implication for symmetric matrices is one of the most important results you will use in statistics. This theorem is at the heart of many different statistical scenarios such as principal component analysis, the analysis of ridge regression, spectral methods for graph data, etc.. We'll use this decomposition to prove a few results.

Corollary 5.5

Suppose A has a basis of eigenvectors. Then, $\det(A) = \prod_{i=1}^n \lambda_i$.

Proof. A has diagonalization $V\Lambda V^{-1}$. Then,

$$\det(A) = \det(V)\det(\Lambda)\det(V^{-1}) = \det(VV^{-1})\det(\Lambda) = 1 \cdot \prod_i \lambda_i,$$

where we use the product rule for determinants and the simplified formula for the determinant of a diagonal matrix. ■

This result is actually true even if A doesn't have a basis of eigenvectors. There are proofs very similar to the one above using matrix decompsitions that exist for all $A \in \mathbb{C}^{n\times n}$. Two options are the Jordan normal form and the Schur decomposition. The second will be discussed briefly in the following section.

Corollary 5.6

Suppose A has a basis of eigenvectors. Then, $\mathrm{Tr}(A) = \sum_{i=1}^n \lambda_i$.

Proof. First, we'll show that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$:

$$\mathrm{Tr}(AB) = \sum_{i=1}^n (AB)_{ii} = \sum_i \sum_k a_{ik} b_{ki} = \sum_k \sum_i b_{ki} a_{ik} = \mathrm{Tr}(BA).$$

Now, we use the diagonalization:

$$\mathrm{Tr}(A) = \mathrm{Tr}((V\Lambda)V^{-1}) = \mathrm{Tr}(V^{-1}V\Lambda) = \mathrm{Tr}(\Lambda) = \sum_i \lambda_i.$$

■

Like the previous result, the above corollary is true even if A doesn't have a basis of eigenvectors, and the proof in that event is similar to the one above.

Corollary 5.7

Suppose A has a basis of eigenvectors. Then, $\mathrm{rank}(A)$ is the number of nonzero eigenvalues λ_i of A.

Proof. $\mathrm{rank}(A) = \mathrm{rank}(\Lambda)$ because multiplying by an invertible matrix doesn't change rank. The nonzero columns of Λ are linearly independent, so $\mathrm{rank}(A)$ is the number of columns containing a nonzero eigenvalue. This is the number of nonzero eigenvalues counted with multiplicity. ■

6 Eigenvalues and eigenvectors of symmetric matrices

A *symmetric matrix* is a matrix $A \in \mathbb{R}^{n\times n}$ such that $A^T = A$. A standard example in statistics is a covariance matrix. Symmetric matrices are nice to work with because they are always diagonalizable with real eigenvalues. We'll prove those properties now.

Theorem 6.1

Symmetric matrices have real eigenvalues.

Proof. Let λ and v be a (complex) eigenvalue and eigenvector pair of a symmetric matrix A. Then taking complex conjugates, $\overline{\lambda}\overline{v} = \overline{\lambda v} = \overline{Av} = A\overline{v}$ because A is real. Thus,

$$\overline{\lambda}\|v\|^2 = v^T\overline{\lambda}\overline{v} = v^T A\overline{v} = (A^T v)^T\overline{v} = (Av)^T\overline{v} = \lambda v^T\overline{v} = \lambda\|v\|^2.$$

Thus, $\lambda = \overline{\lambda}$ meaning λ is real. ■

In fact, we will deduce from the next theorem that the eigenvectors can be taken to be real in the sense that there is an orthogonal basis of real eigenvectors for $\mathbb{R}^n$. One slight subtlety here is that a real symmetric matrix could still have complex eigenvectors *when treated as a complex matrix* (e.g., any complex vector is an eigenvector of the identity matrix for eigenvalue 1). It turns out that this difference is essentially ignorable for real symmetric matrices: regardless of whether you work over $\mathbb{R}$ or $\mathbb{C}$, the eigendimensions of real eigenvectors coincide and there is a basis of real eigenvectors of A for $\mathbb{C}^n$.

Theorem 6.2 (spectral theorem for symmetric matrices)

If $A \in \mathbb{R}^{n\times n}$ is symmetric, then A has an orthogonal basis of eigenvectors.

Proof. First, recall from the last section that A has an orthogonal basis of eigenvectors iff there exists orthogonal Q and diagonal Λ such that $A = Q^T \Lambda Q$. We'll be using this characterization in the proof. We'll prove this result by induction on the dimension n. The base case $n = 1$ is trivial.

Assume the result is true for dimension $n-1$. Let $A \in \mathbb{R}^{n\times n}$ and let λ, u be an eigenvalue and corresponding (normalized) eigenvector of A. Let $V \in \mathbb{R}^{n\times(n-1)}$ have columns that form an orthogonal basis for the space orthogonal to u. We can compute the columns $v_{:1}, \ldots, v_{:,n-1}$ by applying Gram-Schmidt to the sequence of vectors $u, e_1, \ldots, e_n$. $V^T AV$ is then a symmetric $(n-1)\times(n-1)$ matrix and thus has a decomposition $Q\Lambda Q^T$ by the induction hypothesis. Then, $Q^T V^T AVQ = \Lambda_1$, the diagonal matrix of the first $n-1$ eigenvalues. Thus,

$$\begin{pmatrix} u & VQ \end{pmatrix}^T A \begin{pmatrix} u & VQ \end{pmatrix} = \begin{pmatrix} u^T Au & u^T AVQ \\ Q^T V^T Au & Q^T V^T AVQ \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \Lambda_1 \end{pmatrix},$$

where we use the facts that (1) the column space of V and u are orthogonal and (2) V is orthogonal to establish $u^T AV = (A^T u)^T V = (Au)^T V = \lambda u^T V = \lambda \cdot 0 = 0$ and $V^T Au = V^T \lambda u = 0$. Therefore, $\begin{pmatrix} u & VQ \end{pmatrix}$ is a basis of eigenvectors for A. All that remains to check is that it's orthogonal:

$$\begin{pmatrix} u & VQ \end{pmatrix}^T \begin{pmatrix} u & VQ \end{pmatrix} = \begin{pmatrix} u^T u & u^T VQ \\ Q^T V^T u & Q^T V^T VQ \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{Id}_{n-1} \end{pmatrix}.$$

■

You may have noticed that the only place the above proof uses symmetry is to show that the top-right block of

$$\begin{pmatrix} u^T Au & u^T AVQ \\ Q^T V^T Au & Q^T V^T AVQ \end{pmatrix}$$

is zero. If A is *not* symmetric, you can use the same construction and proof to show that any matrix $A \in \mathbb{R}^{n\times n}$ has a decomposition QLQ^T where Q is orthogonal and L is lower triangular. This is known as a *Schur decomposition*.

There is another way of writing this decomposition that can be useful.

Corollary 6.3

If $A \in \mathbb{R}^{n\times n}$ is symmetric, $A = \sum_{i=1}^n \lambda_i q_{:,i} q_{:,i}^T$ where $q_{:,i}$ are orthogonal eigenvectors of A with corresponding eigenvalue λ_i.

Proof.

$$A_{ij} = ((Q\Lambda)Q^T)_{ij} = \sum_k (Q\Lambda)_{ik} Q^T_{kj} = \sum_k \lambda_k q_{ik} (Q^T)_{kj} = \sum_k \lambda_k q_{ik} q_{jk} = \sum_k \lambda_k (q_{:k} q_{:k}^T)_{ij}$$

■

7 Eigenvalues and optimization

Many optimization problems have eigenvalues as optima and eigenvectors as solutions. Here is a simple one.

Theorem 7.1

Let A be a real symmetric $n \times n$ matrix. $\sup_{b \in \mathbb{R}^n : b^T b \neq 0} \frac{b^T A b}{b^T b} = \lambda_{\max}(A)$.

Proof.

$$\frac{b^T A b}{b^T b} = \frac{b^T Q^T \Lambda Q b}{b^T b} = \frac{b^T Q^T \Lambda Q b}{b^T Q^T Q b} = \frac{a^T \Lambda a}{a^T a},$$

where $a := Qb$. Thus, if we define $v := \frac{a}{\sqrt{a^T a}}$, we obtain

$$\frac{b^T A b}{b^T b} = \frac{a^T}{\sqrt{a^T a}} \Lambda \frac{a}{\sqrt{a^T a}} = v^T \Lambda v,$$

with $v^T v = 1$. Hence,

$$\sup_{b : b^T b \neq 0} \frac{b^T A b}{b^T b} = \sup_{v : v^T v = 1} v^T \Lambda v.$$

Now, $\sup_{v : v^T v = 1} v^T \Lambda v = \sup_{v : v^T v = 1} \sum_{i=1}^n \lambda_i v_i^2 = \lambda_{\max}(A)$.

In fact, we claim the $\sup$ is attained iff v is in the eigenspace of $\lambda_{\max}(A)$. This is true since $v \propto Qb$ is a weighting of eigenvectors and the $\sup$ is attained iff all the weights are placed on the eigenvectors corresponding to the top eigenvalue(s). Thus, v is in the eigenspace of $\lambda_{\max}(A)$. Now let's figure out what that makes b.

$$v = \frac{Qb}{\sqrt{(Qb)^T (Qb)}} = \frac{Qb}{\|b\|} \implies \frac{b}{\|b\|} = Q^T v.$$

Thus, the optimum is attained iff b is in the column span of $Q^T V$ where V is a basis for the eigenspace of $\lambda_{\max}$. ■

Corollary 7.2

$\inf_{b \neq 0} \frac{b^T A b}{b^T b} = \lambda_{\min}(A)$.

Corollary 7.3

A matrix $A \in \mathbb{R}^{n \times n}$ is called *nonnegative definite* or *positive semidefinite* if $v^T A v \geq 0$ for all $v \in \mathbb{R}^n$. It is called *positive definite* if the inequality is strict for all $v \in \mathbb{R}^n \backslash \{0\}$. The previous result implies the following.

1. If a symmetric matrix A is nonnegative definite, then all the eigenvalues of A are nonnegative.
2. If a symmetric matrix A is positive definite, then all the eigenvalues of A are positive.

8 Functions of symmetric matrices

The $A = Q \Lambda Q^T$ representation enables us to define and calculate different functions of the matrix A in a natural way. Here are a few useful examples:

1. Square-root of a nonnegative definite matrix: Let A be a nonnegative definite matrix. We can define and calculate

$$A^{1/2} := Q \cdot \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}) \cdot Q^T.$$

   This matrix has the properties that $A^{1/2}$ is symmetric and nonnegative definite, and that $A^{1/2} A^{1/2} = A$.

2. Polynomials of symmetric matrices: applying a polynomial function to a symmetric matrix is equivalent to applying the polynomial to only the eigenvalues of the matrix, i.e. if $p(x)$ is a polynomial in x, we have
$$p(A) = Q \cdot \text{diag}(p(\lambda_1), \ldots, p(\lambda_n)) \cdot Q^T.$$
3. Exponential of a symmetric matrix: we can also define
$$e^A := Q \cdot \text{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n}) \cdot Q^T.$$
Again, this definition of the exponential has many nice properties such as the Taylor expansion:
$$e^A = \mathrm{Id} + A + \frac{A^2}{2!} + \cdots .$$

9 Singular value decomposition

There is a more generalized decomposition for non-symmetric matrices called the *singular value decomposition* (SVD). The eigendecomposition is a special case of the SVD. This decomposition also appears in many statistical problems, such as principal component analysis or clustering.

To start, we can recognize a general matrix $A \in \mathbb{R}^{m \times n}$ as a linear transformation of a vector v in its row space to a vector u in its column space: $Av = u$. The SVD arises from finding an orthogonal basis for the row space that gets transformed into an orthogonal basis for the column space: $Av = \sigma u$. It's not hard to find an orthogonal basis for the row space – the Gram-Schmidt procedure gives us one right away. But in general, there's no reason to expect A to transform that basis to another orthogonal basis. Intuitively, however, the geometric fact that a unit sphere is taken to a hyperellipse by any linear transformation should guide us.

Theorem 9.1 (SVD)

Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ matrix. Then, it can be decomposed as $A = U \Sigma V^T$ where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are both orthogonal, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal (i.e., has non-zero entries only on the diagonal).

The *singular values* are the diagonal entries of Σ with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. The *left singular vectors* of A are the column vectors of U, while the *right singular vectors* of A are the column vectors of V.

We will not prove this existence theorem, although it is similar to our proof of the spectral theorem for symmetric matrices and relies on an induction on the size of A and Gram-Schmidt. The proof may be found in various places, e.g. Lecture 4 of the textbook “Numerical Linear Algebra” by Trefethen and Bau.

Here are some of the relevant facts about SVD. Note there is a striking similarity with many of the properties we derived for eigenvalues and the eigendecomposition.

1. Every matrix $A \in \mathbb{R}^{m \times n}$ has an SVD. Furthermore, the singular values $\{\sigma_j\}$ are uniquely determined and, if A is square, they are all distinct. The left and right singular vectors $\{u_j\}$ and $\{v_j\}$ are uniquely determined up to signs.
2. The eigenvalues are not always real, but the singular values are always nonnegative reals.
3. The SVD is often more conveniently written as $AV = U\Sigma$ or as $Av_i = u_i \sigma_i$ for singular value/vectors σ_i, v_i, u_i.
4. The largest singular value $\sigma_1(A)$ satisfies $\sigma_1(A) = \sup_{v : \|v\|_2 = 1} \|Av\|_2$.
5. The nonzero singular values of A are the square roots of the nonzero eigenvalues of $A^T A$ (or AA^T). If $A = A^T$ is symmetric, then the singular values of A are the absolute values of the eigenvalues of A. Furthermore, the eigenvectors of AA^T are the left singular vectors and the eignvectors of $A^T A$ are the right singular vectors.
6. If $A \in \mathbb{R}^{n \times n}$ is a square matrix, then $|\det(A)| = \prod_{i=1}^n \sigma_i$.
7. $A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T$.

10 Problems

10.1 Previous Core Competency Problems

Problem 1 (2018 Summer Practice Problems, # 18). Suppose Σ is a nonnegative definite matrix of $n \times n$ with real entries and real eigenvalues. Show that $\mathrm{Tr}(\Sigma^2) \geq n \cdot \det(\Sigma)^{2/n}$.

Problem 2 (2020 September Exam, # 8). For every $n \geq 1$, let A_n be an $n \times n$ symmetric matrix with non negative entries. Let $R_n(i) := \sum_{j=1}^n A_n(i,j)$ denote the ith row/column sum of A_n. Assume that

$$\lim_{n\to\infty} \max_{1\leq i\leq n} |R_n(i) - 1| = 0.$$

Let $\lambda_n \geq 0$ denote an eigenvalue with the largest absolute value, and let $\mathbf{x} := (x_1, \ldots, x_n)$ denote its corresponding eigenvector.

(a) Show that $\frac{1}{n}\sum_{i,j=1}^n A_n(i,j) \to 1$.

(b) Show that $\lambda_n |x_i| \leq \max_{1\leq j\leq n} |x_j| R_n(i)$.

(c) Using parts (a) and (b), show that $\lambda_n \to 1$.

Problem 3 (2021 May Exam, # 7). Suppose that $A = (a_{ij})_{1\leq i,j\leq 2}$ is a 2×2 symmetric matrix, with $a_{11} = a_{22} = \frac{3}{4}$ and $a_{12} = a_{21} = \frac{1}{4}$.

1. Find the eigenvalues and eigenvectors of the matrix A.
2. Compute $\lim_{n\to+\infty} a_{12}^{(n)}$, where $a_{ij}^{(n)}$ denotes the (i,j)'s entry of matrix A^n.

Problem 4 (2021 Sept Exam, # 6). Let $A \in \mathbb{R}^{m\times n}$ denote an $m \times n$ matrix with $n < m$. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ denote, respectively, the eigenvalues and eigenvectors of A^TA. What can we say about ALL the eigenvalues and eigenvectors of AA^T? Justify your answer.

10.2 Additional Practice

Problem 5. Let A be a 3×3 real-valued matrix such that $A^TA = AA^T = \mathrm{Id}_3$ and $\det(A) = 1$. Prove that 1 is an eigenvalue of A.

Problem 6. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric $n \times n$ matrix such that $\mathrm{Tr}(A^2) = 0$. Show that $T = \mathbf{0}_{n\times n}$. Hint: use the fact that $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$ for matrices A, B, C.

Problem 7. Let matrices $A, B \in \mathbb{R}^{n\times n}$ have respective eigendecompositions $Q_1D_1Q_1^T$ and $Q_2D_2Q_2^T$ (recall this means each D_i is a diagonal matrix of eigenvalues and each Q_i is an orthogonal matrix). Prove that $Q_1 = Q_2$ if and only if $AB = BA$. You may assume that A, B do not have any repeated eigenvalues.

Problem 8. Let $A = uv^T \in \mathbb{R}^{n\times n}$ be a *rank-one* matrix, i.e. $u, v \in \mathbb{R}^n$. Suppose $u, v \neq \mathbf{0}_n$. Find, with proof, all the eigenvalues of A.

Problem 9 (Heisenberg uncertainty principle). Suppose $A, B \in \mathbb{R}^{n\times n}$ are symmetric matrices satisfying $AB + BA = \mathrm{Id}_n$. Show that for all vectors $v \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$,

$$\max\left\{\frac{\|Av\|_2}{\|v\|_2}, \frac{\|Bv\|_2}{\|v\|_2}\right\} \geq 1/\sqrt{2}.$$

Problem 10. Let $A = (a_{i,j})$ be a $n \times n$ real matrix whose diagonal entries $a_{i,i}$ satisfy $a_{i,i} \geq 1$ for all $i \in 1, \ldots, n$. Suppose also that

$$\sum_{i\neq j} a_{i,j}^2 < 1.$$

Prove that the inverse matrix A^{-1} exists.

Problem 11 (Greshgorin circle theorem). Let $A \in \mathbb{C}^{n\times n}$ with entries a_{ij}. For $i \in \{1, \ldots, n\}$ let R_i be the sum of the absolute values of the non-diagonal entries in the i-th row:

$$R_i := \sum_{j\neq i} |a_{ij}|.$$

Let $D(a_{ii}, R_i) \subseteq \mathbb{C}$ be a closed disc centered at a_{ii} with radius R_i, called a *Gershgorin disc*. Show that every eigenvalue of A lies within at least one of the Gershgorin discs.

Problem 12. Let $A = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$. Find $\lim_{n\to\infty} A^n$. Hint: the eigenvalues of a lower triangular matrix are its diagonal entries.

Problem 13. Let A, B be $n \times n$ matrices. Show that BA and AB have the same eigenvalues if A is invertible

Problem 14. Let $A = (a_{ij})$ be a 2×2 real matrix such that

$$a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2 < \frac{1}{1000}.$$

Prove that $\mathrm{Id}_{2\times 2} + A$ is invertible.

Problem 15. Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric $n \times n$ matrix and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues in decreasing order. Show that

$$\lambda_k \leq \max_{U:\dim(U)=k} \min_{x\in U:\|x\|_2=1} \langle Ax, x\rangle.$$

The maximum above is over all k-dimensional subspaces U of $\mathbb{R}^n$. Hint: form an orthonormal basis of eigenvectors to make U.

Problem 16. For a vector $v \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$, define the map $F : \mathbb{R}^n \to \mathbb{R}^n$ via $F(x) = \operatorname*{argmin}_{z\in \mathrm{Span}(v)} \|z - x\|_2$. Compute F explicitly in terms of v. Is $F : \mathbb{R}^n \to \mathbb{R}^n$ a linear transformation?

Problem 17. Suppose $A \in \mathbb{R}^{n\times n}$ and $A = A^T$ with all eigenvalues of A being positive. Show there exists a matrix B such that $B^2 = A$.

Problem 18. Let $A \in \mathbb{R}^{n\times n}$ and let $\{\sigma_i\}_{i=1}^n$ be the singular values of A. Show that $|\det(A)| = \prod_{i=1}^n \sigma_i$.

Problem 19 (low-rank approximation)**.** Let $A \in \mathbb{R}^{m\times n}$ matrix and for a positive integer $p < \mathrm{rank}(A)$, define $A_p = \sum_{i=1}^p \sigma_i u_i v_i^T$ where σ_i is the i-th (largest) singular value of A, and u_i, v_i are respective left/right singular vectors, i.e. the SVD is $A = U\Sigma V^T$. Then, prove that

$$\sup_{x:\|x\|_2=1} \|(A - A_p)x\|_2 = \sigma_{p+1}.$$

Problem 20. Suppose $P \in \mathbb{R}^{n\times n}$ is a symmetric matrix that satisfies $P^2 = P$, a so-called *idempotent* matrix. Find all the eigenvalues of P with their (algebraic) multiplicities in terms of P.

Problem 21. Suppose that Σ is the covariance matrix of k zero-mean random variables $X_1, \ldots, X_k$, i.e. if $X = (X_1, \ldots, X_k)$ then $\Sigma := \mathbb{E}[XX^T]$. Prove that if Σ is singular, then $X_1, \ldots, X_k$ are linearly dependent almost everywhere.