Spectral theorems, SVD, and Quadratic forms

This set of notes covers materials from Chapter 8. The non-examinable sections are marked with an asterisk.

Contents

1 Spectral theorem: symmetric matrices

The following theorem is absolutely the main theorem of the course.

Theorem (Spectral theorem, version I)**.** Let A be an $n \times n$ symmetric matrix, then there exists an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^T$.

This section will be devoted to studying some consequences of this theorem. The first observation to make is that $Q^T = Q^{-1}$, so A is in fact diagonalized by Q, so we have

Corollary. Symmetric matrices are diagonalizable over $\mathbb{R}$.

This re-interpretation using diagonalization yields an equivalent formulation of the spectral theorem.

Theorem (Spectral theorem, version II)**.** Let A be an $n \times n$ symmetric matrix, then all eigenvalues of A are real, their algebraic multiplicities and geometric multiplicities agree, and **eigenspaces for distinct eigenvalues are orthogonal**.

Proof. The first two properties is equivalent to A being diagonalizable over $\mathbb{R}$. It follows that $A = SDS^{-1}$, where the columns of S are eigenvectors of A. But S is orthogonal if and only if its columns form an orthonormal basis, which is equivalent to the statement about orthogonality of distinct eigenspaces. □

Theorem (Spectral theorem, version III)**.** Let A be an $n \times n$ symmetric matrix, then there exists orthogonal projections $P_1, \cdots, P_r$ and real numbers $\lambda_1, \cdots, \lambda_r$ such that

- $P_iP_j = 0$ if $i \neq j$.
- $I_n = P_1 + \cdots + P_r$.
- $AP_i = \lambda_i P_i$.

This is the decomposition of the identity into **eigenspace projections**. In these notations, we have

$$A = \lambda_1 P_1 + \cdots + \lambda_r P_r$$

Proof. Let $\lambda_1, \cdots, \lambda_r$ be the distinct eigenvalues of A, which are all real by the spectral theorem. Let P_i be the orthogonal projection onto the eigenspace $\ker(A - \lambda_i I_n)$, then we have $AP_i = \lambda_i P_i$. Moreover, by version II, the different eigenspaces are orthogonal to each other, so $P_i P_j = 0$. Finally, the statement that $I_n = P_1 + \cdots + P_n$ is equivalent to saying there is a basis of eigenvectors. $\square$

Example. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Its characteristic polynomial is $\det(A - \lambda I_3) = -(\lambda - 1)(\lambda^2 - 2\lambda - 1)$, which has roots $\lambda_1 = 1 + \sqrt{2}, \lambda_2 = 1, \lambda_3 = 1 - \sqrt{2}$. They are all real, as expected.

The three eigenspaces are respectively

$$E_1 = \text{span} \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}, \ E_2 = \text{span} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \ E_3 = \text{span} \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}$$

The three vectors are mutually orthogonal, which is expected from version II of the spectral theorem. We can normalize them to form an orthogonal matrix

$$Q = [\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3] = \frac{1}{2} \begin{bmatrix} 1 & -\sqrt{2} & 1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & \sqrt{2} & 1 \end{bmatrix}$$

The statement of the spectral theorem, version I, in this case becomes

$$A = QDQ^T, \quad D = \begin{bmatrix} 1 + \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 - \sqrt{2} \end{bmatrix}$$

Finally, the three eigenspace projectors are

$$P_1 = \vec{u}_1 \vec{u}_1^T = \frac{1}{4} \begin{bmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 2 & \sqrt{2} \\ 1 & \sqrt{2} & 1 \end{bmatrix}$$

$$P_2 = \vec{u}_2 \vec{u}_2^T = \frac{1}{2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$P_1 = \vec{u}_3 \vec{u}_3^T = \frac{1}{4} \begin{bmatrix} 1 & -\sqrt{2} & 1 \\ -\sqrt{2} & 2 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{bmatrix}$$

We indeed have $P_1 + P_2 + P_3 = I_3$. Moreover, it is good to check that the trace of each matrix is equal to 1, the dimension of the eigenspace.

Example. We now do an example with some multiplicities. Let

$$A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$$

The eigenvalues of A are $\lambda_1 = 9$ with algebraic multiplicity 2 and $\lambda_2 = 0$ with algebraic multiplicity 1. The two eigenspaces are

$$E_1 = \ker \begin{bmatrix} -1 & -2 & 2 \\ -2 & -4 & 4 \\ 2 & 4 & -4 \end{bmatrix} = \operatorname{span}\left(\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right)$$

$$E_2 = \ker A = \operatorname{span} \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix}$$

Observe that E_1 is orthogonal to E_2.

We want to find an orthogonal basis $\{\vec{u}_1, \vec{u}_2\}$ for E_1. This can be done by applying Gram–Schmit on the vectors we found, which gives

$$\vec{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \frac{1}{3\sqrt{5}} \begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix}$$

The required diagaonalization of A using an orthogonal matrix is therefore $A = QDQ^T$, where

$$Q = \frac{1}{3\sqrt{5}} \begin{bmatrix} 6 & -2 & -\sqrt{5} \\ 0 & 5 & -2\sqrt{5} \\ 3 & 4 & 2\sqrt{5} \end{bmatrix}, \quad D = \begin{bmatrix} 9 & & \\ & 9 & \\ & & 0 \end{bmatrix}$$

There are many choices for the first two columns of Q. For example, the following is also a possibility for Q

$$Q = \frac{1}{3} \begin{bmatrix} 2 & 2 & -1 \\ -2 & 1 & -2 \\ -1 & 2 & 2 \end{bmatrix}$$

However, D is unique if we require the diagonal entries to be in decreasing order.

Now to verify version III, we need to compute the two projectors. We have

$$P_2 = \vec{u}_3 \vec{u}_3^T = \frac{1}{9} \begin{bmatrix} 1 & 2 & -2 \\ 2 & 4 & -4 \\ -2 & -4 & 4 \end{bmatrix}$$

To compute P_1, we can use the orthogonal basis we have computed, but instead, we apply the general projection formula as an example. Let

$$C = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

then

$$P_1 = C(C^TC)^{-1}C^T = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$$

Observe that $I_3 = P_1 + P_2$, as predicted. Moreover, even though the matrix Q is not unique, the projections P_1 and P_2 are uniquely determined.

We now give some theoretic applications of the spectral theorem.

Definition. The **spectrum** of a square matrix A is the set of its eigenvalues.

Theorem (Borel functional calculus)**.** Let A be an $n \times n$ symmetric matrix. Let f be a real-valued function defined on the spectrum of A, then we can define $f(A)$. Moreover, the usual algebraic operations on functions correspond to the algebraic operations on matrices.

Proof. We use the projection version of the spectral theorem and define

$$f(A) = \sum_{i=1}^{r} f(\lambda_i) P_i$$

In the diagonalization language, this corresponds to

$$f(A) = Qf(D)Q^T, \quad f(D) = \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} \qquad \square$$

This is in parallel with the matrix function we defined in the last chapter using power series. It applies to an arbitrary function, as opposed to those that can be represented by power series, but it applies to a smaller class of matrices. We will expand the applicable class of matrices in the next section to include orthogonal matrices, but one should not expect to be able to apply an arbitrary function to an arbitrary matrix.

Definition. A symmetric matrix is **positive definite/positive semidefinite/negative definite/negative semidefinite** if respectively, its spectrum lies in $\mathbb{R}_{>0}/\mathbb{R}_{\geq 0}/\mathbb{R}_{<0}/\mathbb{R}_{\leq 0}$. It is **indefinite** if it is none of the above, i.e. it has both positive and negative eigenvalues.

Corollary. If A is positive semidefinite, then there exists a symmetric matrix B such that $A = B^2$. If we require B to also be positive semidefinite, then B is unique, which we denote by $\sqrt{A}$.

Proof. Apply the above theorem to the function $f : [0, \infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$. $\square$

If A is an *arbitrary* matrix, then A^TA is positive semidefinite. Indeed, it is symmetric, so it only has real eigenvalues. Let λ be one eigenvalue, so $A^TA\vec{v} = \lambda\vec{v}$ for a non-zero $\vec{v}$. Multiply by $\vec{v}^T$ gives

$$\lambda = \frac{\vec{v}^TA^TA\vec{v}}{\vec{v}^T\vec{v}} = \frac{\|A\vec{v}\|^2}{\|\vec{v}\|^2} \geq 0$$

We can therefore define $|A| = \sqrt{A^TA}$. This commutes with A.

Theorem (Polar decomposition)**.** Let A be a square matrix, then there exists a decomposition $A = U|A|$, where U is orthogonal and $|A|$ is positive definite.

Proof. Define a linear transformation T by setting $T(|A|\vec{v}) = A\vec{v}$ and $T(\vec{w}) = 0$ if $\vec{w} \perp \mathrm{Im}(|A|)$. We need to check that if $|A|\vec{v} = |A|\vec{w}$, then $A\vec{v} = A\vec{w}$. But observe that

$$\||A|\vec{v}\|^2 = \vec{v}^T|A|^T|A|\vec{v} = \vec{v}^TA^TA\vec{v} = \|A\vec{v}\|^2$$

If $|A|\vec{v} = |A|\vec{w}$, then $|A|(\vec{v} - \vec{w}) = 0$, so $A(\vec{v} - \vec{w}) = 0$ by this computation. It follows that T is a well-defined function. Linearity follows easily form this. Moreover, T preserves the length of the vectors in $\mathrm{Im}(|A|)$, so we can re-define T on $\mathrm{Im}(|A|)^\perp$ to make T orthogonal. The matrix associated to T is the U we want. $\square$

We remark that the proof is very easy if A is invertible: simply take $U = A|A|^{-1}$ and check that U is orthogonal. This makes sense in general if we use the Moore–Penrose pseudo-inverse, recalled later. It is also easy if A is symmetric using the Borel functional calculus by considering the identity

$$x = \mathrm{sign}(x)|x|, \quad \mathrm{sign}(x) = \begin{cases} -1 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

We will give an alternative proof later using the singular value decomposition.

2 Spectral theorem: extensions*

The most general class of matrices for which the spectral theorem holds is the following

Definition. A square matrix A is **normal** if $A\bar{A}^T = \bar{A}^T A$.

In particular, orthogonal matrices and real symmetric matrices are normal.

Theorem (Spectral theorem for normal matrices)**.** Let A be a normal matrix, then there exists complex matrices U and D such that $A = UD\overline{U}^T$, with U unitary and D diagonal.

Recall that a square matrix U is unitary if $U\overline{U}^T = I$. This is the appropriate generalization of orthogonal matrices to the complex numbers. The other two versions carry over, except that, for complex vectors, the dot product is defined by $\vec{v} \cdot \vec{w} := \overline{\vec{v}}^T \vec{w}$. Similarly, Borel functional calculus carries over, and we can prove the existence of a polar decomposition.

Since this is a non-examinable section, we will talk briefly about infinite dimensional operators from the perspective of quantum mechanics. Vectors are now called states, and they are written as $|\psi\rangle$. The conjugate transpose of $|\psi\rangle$ is denoted by $\langle\psi|$, so the dot product of two vectors can be written as $\langle\psi|\varphi\rangle$. If $|\psi\rangle$ is a unit vector, then the projection onto its span is $|\psi\rangle\langle\psi|$.

Let $|\psi_1\rangle, \cdots, |\psi_n\rangle$ be an orthonormal basis, so $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. The fact that they span the whole space can be written suggestively as

$$1 = \sum_{i=1}^{n} |\psi_i\rangle\langle\psi_i|$$

which is equivalent to

$$|\varphi\rangle = \sum_{i=1}^{n} |\psi_i\rangle\langle\psi_i|\varphi\rangle$$

So far, all we did was change the notations. Now we will be really vague and only convey a flavour of how things go for infinite dimensional spaces. Let H be a normal operator, and let $\sigma(H) \subseteq \mathbb{C}$ denote its spectrum, then the spectral decomposition theorem is formally

$$1 = \int_{\sigma(H)} dz|z\rangle\langle z|, \quad H = \int_{\sigma(H)} zdz|z\rangle\langle z|$$

Now suppose the vector space is the space of functions on $\mathbb{R}$ and the operator is $Hf(x) = xf(x)$, then there is in some sense a continuous basis of eigenspaces, one for each real number. The projection of f onto an eigenspace $|x\rangle$ is $\langle x|f\rangle = f(x)$, so the spectral theorem formally says

$$|f\rangle = \int_{-\infty}^{\infty} dx|x\rangle\langle x|f\rangle = \int_{-\infty}^{\infty} f(x)|x\rangle dx$$

We are therefore decomposing a function on $\mathbb{R}$ into a sum of functions each defined over a single point. If H is something more sophisticated, say $H = \frac{1}{2m}p^2 + V(x)$, where $p = -i\hbar\frac{d}{dx}$ and $V(x)$ describes some potential energy, then the resulting spectral decomposition is the decomposition of a general state into energy levels. The famous harmonic oscillator has states indexed by non-negative integers n, with the corresponding eigenvalue (i.e. energy) growing linearly in n. This is an instance of a discrete spectrum.

At this point, it should be emphasized that none of the above manipulations make sense, which is the reason we decided to use the physics notations. While infinite dimensional spaces retain some of the flavours of linear algebra in finite dimension spaces, the additional issue of convergence makes the whole picture a lot more intricate. With a serious amount of work, the above discussion can be made precise.

3 Spectral theorem: proofs*

We now give a proof of the spectral theorems stated. First consider the case of symmetric matrices.

Theorem. If A is a real symmetric matrix, then its eigenvalues are real, and eigenvectors with different eigenvalues are orthogonal.

Proof. Suppose λ is a complex eigenvalue, so $A\vec{v} = \lambda\vec{v}$ for some $\vec{v} \neq 0$, then

$$\bar{\vec{v}}^T A\vec{v} = \lambda\bar{\vec{v}}^T \vec{v} = \lambda\|\vec{v}\|^2$$

Taking conjugate transpose of both sides gives

$$\bar{\vec{v}}^T \bar{A}^T \vec{v} = \bar{\lambda}\|\vec{v}\|^2$$

Since A is real and symmetric, $\bar{A}^T = A$, so $\bar{\lambda} = \lambda$, which proves that λ is real.

Suppose $A\vec{v} = \lambda\vec{v}$ and $A\vec{w} = \mu\vec{w}$, where $\lambda \neq \mu$, then

$$\lambda\vec{v}^T \vec{w} = \vec{v}^T A^T \vec{w} = \vec{v}^T A\vec{w} = \mu\vec{v}^T \vec{w}$$

Since $\lambda \neq \mu$, we must have $\vec{v}^T \vec{w} = 0$. $\square$

We now prove the spectral theorem in version III. This is an induction on the size of A. Given an $n \times n$ symmetric matrix A, by the previous theorem and the fundamental theorem of algebra, A has a real eigenvalue λ. Let E_1 be its eigenspace, and let $E_2 = E_1^\perp$. Let P_1, P_2 be the orthogonal projections onto E_1, E_2 respectively, then we have

$$P_1 \neq 0,\ P_1 P_2 = 0,\ P_1 + P_2 = I_n,\ AP_1 = \lambda P_1$$

We now show that $AP_2 = P_2 A$, or equivalently, $AE_1^\perp \subseteq E_1^\perp$. This is the same argument used in the proof of the previous theorem. Indeed, suppose $\vec{w} \in E_1^\perp$, then for all $\vec{v} \in E_1$, we have

$$\vec{v}^T A\vec{w} = \vec{v}^T A^T \vec{w} = (A\vec{v})^T \vec{w} = \lambda\vec{v}^T \vec{w} = 0$$

It follows that $A\vec{w} \in E_1^\perp$.

In other words, we have decomposed $\mathbb{R}^n$ into two orthogonal subspaces, both of which are fixed by A. On the first one, A acts by multiplication by λ. Since the dimension of the second space is strictly less than n, the induction hypothesis shows that it has a spectral decomposition with respect to A. Adding P_1 to this list of projectors gives the required list for the original space.

Everything works for normal operators in the same way provided we can prove the claim that $AE_1^\perp \subseteq E_1^\perp$. The same proof reduces this to showing that $A^T E_1 \subseteq E_1$. Let $\vec{v} \in E_1$, then

$$A(A^T \vec{v}) = A^T A\vec{v} = \lambda A^T \vec{v}$$

so $A^T \vec{v} \in E_1$, as required.

4 Singular Value Decomposition

In this course, we have seen or referred to several matrix decompositions:

LU decomposition : $A = LU$, where L is lower triangular and U is upper triangular.

QR decomposition : $A = QR$, where Q has orthonormal columns and R is upper triangular.

Polar decomposition : $A = U|A|$, where U is orthogonal and $|A|$ is positive semidefinite.

We will introduce one more decomposition in this section: the singular value decomposition. It has many uses both theoretically and in applications. The following video illustration

`http://youtu.be/R9UoFyqJca8`

was produced by the Los Alamos laboratory in 1976. Parts of the animation were used in the 1979 Star Trek movie at around 1:17:40, on the screen behind Spock.

Theorem. Let A be an $m \times n$ matrix, then there exists a decomposition

$$A = U\Sigma V^T$$

where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and Σ is a diagonal $m \times n$ matrix, in the sense that $\Sigma_{ij} = 0$ if $i \neq j$. Let $\sigma_i = \Sigma_{ii}$, then we require the sequence $\sigma_1, \sigma_2, \cdots$ to be non-negative and non-increasing.

Let r be the rank of A, then $\sigma_1, \cdots, \sigma_r$ are exactly the non-zero entries of Σ, and they are called the **singular varies** of A, and the decomposition is the **singular value decomposition**.

Proof. We have seen before that A^TA is positive semidefinite. Let $\lambda_1 > \cdots > \lambda_n \geq 0$ be its eigenvalues. Let $\sigma_i = \sqrt{\lambda_i}$, and let D be the $n \times n$ diagonal matrix with diagonal entries $\sigma_1, \cdot, \sigma_n$. The spectral theorem gives a factorization

$$A^TA = VD^2V^T$$

where V is an orthogonal $n \times n$ matrix. Recall that $\ker(A^TA) = \ker(A)$. Since A^TA is diagonalizable, the algebraic multiplicity of 0 is equal to $\dim\ker(A)$, so there are exactly $r = \operatorname{rank}(A)$ non-zero eigenvalues, namely $\lambda_1, \cdots, \lambda_r$.

Let $\vec{v}_1, \cdots, \vec{v}_n$ be the columns of V, then $\vec{v}_i \cdot \vec{v}_j = \delta_{ij}$ and $A^TA\vec{v}_i = \lambda_i\vec{v}_i$, so

$$(A\vec{v}_i)^T(A\vec{v}_j) = \vec{v}_i^T\lambda_j\vec{v}_j = \lambda_i\delta_{ij} = \sigma_i^2\delta_{ij}$$

For $i = 1, \cdots, r$, let $\vec{u}_i = \frac{1}{\sigma_i}A\vec{v}_i$. The above computation shows that $\vec{u}_1, \cdots, \vec{u}_r$ are orthonormal, so we can extend them to an orthonormal basis of $\mathbb{R}^m$. Let U be the orthogonal matrix formed with them as columns, then by construction

$$AV = U\Sigma$$

where Σ is the $m \times n$ matrix such that $\Sigma_{ii} = \sigma_i$ for $i = 1, \cdots, r$ and all other entries are 0. Multiplying both sides by V^T gives the singular value decomposition. □

The proof gives the following process for finding the SVD of a matrix:

(Step 1) Find the non-zero eigenvalues $\lambda_1 > \cdots > \lambda_r$ of A^TA. The singular values of A are their square roots $\sigma_i = \sqrt{\lambda_i}$. The rank of A is r.

(Step 2) Find an orthonormal basis $\vec{v}_1, \cdots, \vec{v}_n$ of eigenvectors for A^TA. They form the columns of V.

(Step 3) Compute $\frac{1}{\sigma_i}A\vec{v}_i$ for $i = 1, \cdots, r$. They form the first r columns of U.

(Step 4) Extend the r vectors to an orthonormal basis of $\mathbb{R}^m$.

Example. Consider the 3×2 matrix

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}$$

We want to compute its singular value decomposition, following the above algorithm.

(Step 1) We can compute that

$$A^TA = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$$

This will always be a symmetric matrix, so you only need to compute three entries. It is however accidental that the two diagonal entries are equal.

The characteristic equation is $\lambda^2 - 34\lambda + 225 = 0$, which has roots $\lambda_1 = 25$, $\lambda_2 = 9$, so the singular values of A are $\sigma_1 = 5$, $\sigma_2 = 3$. From this, we see that A has rank 2, which one can see directly from A.

(Step 2) A unit eigenvector for $\lambda_1 = 25$ is $\vec{v}_1 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. A unit eigenvector for $\lambda_2 = 9$ is $\vec{v}_2 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$, so

$$V = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

The two eigenvectors correspond to different eigenvalues, so they must be orthogonal, which we can easily verify here.

(Step 3) We now compute

$$\vec{u}_1 = \frac{1}{\sigma_1}A\vec{v}_1 = \frac{1}{5\sqrt{2}}\begin{bmatrix}3 & 2\\ 2 & 3\\ 2 & -2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}$$
$$\vec{u}_2 = \frac{1}{\sigma_2}A\vec{v}_2 = \frac{1}{3\sqrt{2}}\begin{bmatrix}3 & 2\\ 2 & 3\\ 2 & -2\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix} = \frac{1}{3\sqrt{2}}\begin{bmatrix}1\\-1\\4\end{bmatrix}$$

By general theory, they should both have unit lengths and they should be orthogonal.

(Step 4) We need a vector $\vec{u}_3$ so that $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is orthonormal. There are exactly two choices: $\pm\frac{1}{3}\left[\begin{smallmatrix}2\\-2\\-1\end{smallmatrix}\right]$. Choosing the positive sign gives

$$U = \frac{1}{3\sqrt{2}}\begin{bmatrix}3 & 1 & 2\sqrt{2}\\ 3 & -1 & -2\sqrt{2}\\ 0 & 4 & -\sqrt{2}\end{bmatrix}$$

Therefore, the final result is $A = U\Sigma V^T$, where

$$U = \frac{1}{3\sqrt{2}}\begin{bmatrix}3 & 1 & 2\sqrt{2}\\ 3 & -1 & -2\sqrt{2}\\ 0 & 4 & -\sqrt{2}\end{bmatrix}, \quad \Sigma = \begin{bmatrix}5 & 0\\ 0 & 3\\ 0 & 0\end{bmatrix}, \quad V = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

Example. We can start with the transpose of the matrix used last time

$$B = A^T = \begin{bmatrix}3 & 2 & 2\\ 2 & 3 & -2\end{bmatrix}$$

By the computation above, we already know a singular value decomposition is $B = V\Sigma^T U^T$. Instead, we quickly outline the steps in the algorithm applied to B and comment on the differences.

(Step 1) We compute that

$$B^T B = \begin{bmatrix}13 & 12 & 2\\ 12 & 13 & -2\\ 2 & -2 & 8\end{bmatrix}$$

This has eigenvalues $25, 9, 0$, in agreement with our expectations.

(Step 2) A choice of eigenvectors are

$$\vec{v}_1 = \begin{bmatrix}1\\-1\\0\end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix}1\\-1\\4\end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix}2\\-2\\-1\end{bmatrix}$$

Normalizing them recovers the columns of U we found earlier.

(Step 3) We only have to do two computations now because $\lambda_3 = 0$. Doing them recovers the columns of V we found before, as expected.

(Step 4) This step is unnecessary since we already have enough vectors.

One obvious application of the SVD is to solve systems of linear equations. Indeed, if we want to solve $A\vec{x} = \vec{b}$, and $A = U\Sigma V^T$, then

$$\Sigma V^T \vec{x} = U^T \vec{b}$$

Since Σ is diagonal, it is easy to solve for $V^T\vec{x}$ or see that no solution exists. Given $V^T\vec{x}$, $\vec{x}$ can be recovered by left multiplication by V.

We now take a closer look at the case where a solution does not exist. Given a Σ an $m \times n$ diagonal matrix as in the theorem, let Σ^+ denote the $n \times m$ diagonal matrix whose (i,i)-entry is σ_i^{-1} if $\sigma_i \neq 0$ and is 0 otherwise, so Σ^+ would be the inverse of Σ if Σ is invertible matrix. Define $A^+ = V\Sigma^+U^T$. This is called the **Moore–Penrose pseudoinverse**, which we introduced before in the context of least-squared problems.

Theorem. The vector $\vec{x}^* = A^+\vec{b}$ satisfies the following two optimality properties

1. $\|A\vec{x}^* - \vec{b}\| \leq \|A\vec{x} - \vec{b}\|$ for all $\vec{x} \in \mathbb{R}^n$.
2. $\|\vec{x}^*\| \leq \|\vec{x}\|$ if $A\vec{x}^* = A\vec{x}$.

so $A^+\vec{b}$ is the least-squared solution to $A\vec{x} = \vec{b}$ of minimal length.

Proof. The essential point of this proof is that U and V both preserve lengths, so the theorem is reduced to the case when $A = \Sigma$, a diagonal matrix, where the claims can be checked explicitly.

We will first look at the second property. Let $\vec{y}^* = V^T\vec{x}^* = \Sigma^+U^T\vec{b}$ and $\vec{y}^* = V^T\vec{y}$, then the property reduces to showing that $\Sigma\vec{y}^* = \Sigma\vec{y}$ implies $\|\vec{y}^*\| \leq \|\vec{y}\|$. Observe that

$$\Sigma \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

So the first r entries of $\vec{y}^*$ and $\vec{y}$ agree. But $\vec{y}^*$ is in the image of Σ^+, so its last $n-r$ entries are 0, so the required inequality follows.

Now for the least square property, we have

$$\|A\vec{x} - \vec{b}\| = \|\Sigma V^T\vec{x} - U^T\vec{b}\|$$

We want to minimize this over all $\vec{x} \in \mathbb{R}^n$. Let $\vec{y} = V^T\vec{x}$, then $\vec{y}$ also ranges over $\mathbb{R}^n$, so as before, we need to minimize $\|\Sigma\vec{y} - U^T\vec{b}\|$. The image of Σ consists of vectors whose final $n-r$ entries are 0, so the error is minimized if the first r entries of $\Sigma\vec{y}$ and $U^T\vec{b}$ agree, which is the case if $\vec{y} = \Sigma^+U^T\vec{b}$, or correspondingly, when $\vec{x} = A^+\vec{b}$. $\square$

Example. We want to find a least squared solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

This is the matrix we used in the previous example, so we already have $A = U\Sigma V^T$. A solution is therefore

$$\vec{x}^* = V\Sigma^+U^T\vec{b} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}\frac{1}{3\sqrt{2}}\begin{bmatrix} 3 & 3 & 0 \\ 1 & -1 & 4 \\ 2\sqrt{2} & -2\sqrt{2} & -\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{45}\begin{bmatrix} 41 \\ -14 \end{bmatrix}$$

We could have done this using the QR-factorization or even the normal equation, but the singular value decomposition reveals a lot more structure to the matrix A. In fact, after truncating the matrix Σ by setting all but the largest p singular values to 0, we get the optimal approximation of the matrix by a rank p matrix. This idea is used in practice for data compression.

In practice, the SVD is obtained using an iterative scheme similar to the eigenvalue algorithms discussed in the last chapter. The video above shows a two-step process: deflation to tridiagonal form followed by an implicit QR-iteration with shifts. Both are achieved using Givens rotations.

5 Matrix norm and condition number

The goal of this section is to study ways of measuring the size of a matrix. Let A be an $n \times n$ matrix. If we want to say A is small, then there are at least two possible criteria

- The entries of A are all small.
- $\|A\vec{x}\|$ is small compared to $\|\vec{x}\|$.

The following two definitions of sizes are therefore sensible.

Definition. Let A be an $n \times n$ matrix. Its *maximum norm* is

$$\|A\|_\infty := \max_{1\le i,j\le n} |a_{ij}|$$

Its *2-norm* is

$$\|A\|_2 := \max_{\vec{x}\neq 0} \frac{\|A\vec{x}\|}{\|\vec{x}\|}$$

We can give a definition of $\|A\|_\infty$ analogous to the one given for $\|A\|_2$ by defining the maximum norm of a vector to be $\|\vec{x}\|_\infty = \max_{1\le i\le n} |x_i|$. Other definitions of norms of a vector gives different matrix norms, but we will focus on these two in this set of notes.

Theorem. We have the following estimates:

$$\|A\|_\infty \le \|A\|_2 \le \sqrt{n}\|A\|_\infty$$

so the two definitions of norms are equivalent.

Proof. For each i, j,

$$|a_{ij}|^2 \le \sum_{k=1}^{n} |a_{kj}|^2 = \|A\vec{e}_j\|_2^2 \le \|A\|_2^2 \qquad \square$$

which implies the first inequality. Let $M = \|A\|_\infty$, then by the Cauchy–Schwarz inequality

$$\|A\vec{x}\|_2^2 = \sum_{i=1}^{n}\Big(\sum_{j=1}^{n} a_{ij}x_j\Big)^2 \le \sum_{i=1}^{n}\Big(\sum_{j=1}^{n} a_{ij}^2\Big)\Big(\sum_{j=1}^{n} x_j^2\Big) \le nM^2\|\vec{x}\|_2^2$$

which implies the bound $\|A\|_2 \le \sqrt{n}\|A\|_\infty$.

Notice that the control of $\|A\|_2$ by $\|A\|_\infty$ gets weaker as $n \to \infty$, which strongly suggests that for infinite dimensional space, we could have a matrix with bounded entry but which is unbounded on lengths. This is indeed correct, once we make everything precise.

Now observe that $\|A\vec{x}\|^2 = (A\vec{x})^T(A\vec{x}) = \vec{x}^T A^T A\vec{x}$, so $\|A\|_2^2 = \max_{\vec{x}\neq 0} R(A^T A, \vec{x})$, where

$$R(A^T A, \vec{x}) = \frac{\vec{x}^T A^T A\vec{x}}{\vec{x}^T\vec{x}}$$

This is the **Rayleigh quotient** of $A^T A$. It's worth studying it in detail. Let H be a general symmetric matrix, then the spectral theorem says that $H = Q^T DQ$, where Q is orthogonal and D is diagonal, so

$$R(H, \vec{x}) = \frac{\vec{x}^T Q^T DQ\vec{x}}{\vec{x}^T\vec{x}} = \frac{(Q\vec{x})^T D(Q\vec{x})}{(Q\vec{x})^T(Q\vec{x})} = R(D, Q\vec{x})$$

The possible values of $R(D, \vec{y})$ is easy to study. In particular, if H is positive definite with eigenvalues $\lambda_1 > \cdots > \lambda_n > 0$, then this analysis gives $\lambda_n \le R(H, \vec{x}) \le \lambda_1$. In summary, we have the following theorem.

Theorem. Let A be an $n \times n$ matrix, then $\|A\|_2 = \sigma_1$, its largest singular value.

In practice, people often use the Rayleigh quotient to obtain good estimates for the maximal or minimal eigenvalue for a symmetric matrix, which gives a good input to the eigenvalue algorithms discussed in the last chapter. In quantum mechanics, they can be used to estimate the possible energies of a system.

Finally, we introduce a measure of numerical stability of a matrix. This is also important in numerical problems because there are many sources of round-off errors, and it is useful to know if they accumulate. If the matrix is not stable, then the algorithms we used for this course (Gauss–Jordan, Gram–Schmidt, normal equation) will fail terribly, as we have seen in the slides for stable QR-algorithms.

Definition. The **condition number** of an invertible matrix A is $\mathrm{cond}(A) := \sigma_1/\sigma_n = \|A\|_2\|A^{-1}\|_2$.

For large n, the expected value of the condition number of a random matrix is linear in n, which is not too bad by itself. There are two pieces of bad news: 1. the distribution is very spread out, so the tails are quite probable; 2. large matrices showing up in real life are not Gaussian random matrices. In fact, for applications to dimension reduction and data compression, we expect a few large singular values, which tell us the features of the system. They should significantly dominate the rest, which are interpreted as noises. They will therefore have large condition numbers.

6 Quadratic forms

In the final sections, we give a different geometric interpretation of symmetric matrices. Instead of viewing it as a function $\mathbb{R}^n \to \mathbb{R}^n$, we will view it as a function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, or equivalently via currying, as a function $\mathbb{R}^n \to [\mathbb{R}^n \to \mathbb{R}]$. This is responsible for the different transformation rules required for the change of basis formula.

Definition. Let A be an $n \times n$ symmetric matrix. The (symmetric) **bilinear form** associated to A is the function $f_A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $f_A(\vec{v}, \vec{w}) = \vec{v}^T A \vec{w}$. We will typically use the notation $\langle \vec{v}, \vec{w} \rangle_A$ for $f_A(\vec{v}, \vec{w})$.

The **quadratic form** associated to A is the function $q_A : \mathbb{R}^n \to \mathbb{R}$ defined by

$$q(\vec{v}) = \vec{v}^T A \vec{v} = \langle \vec{v}, \vec{v} \rangle_A$$

Definition. A quadratic form q is **positive semidefinite** if $q(\vec{v}) \ge 0$ for all $\vec{v} \in \mathbb{R}^n$. It is **positive definite** if it is positive semidefinite and $q(\vec{v}) = 0$ implies $\vec{v} = 0$.

A quadratic form q is **negative (semi)definite** if $-q$ is positive (semi)definite. It is **indefinite** if it is neither positive semidefinite nor negative semidefinite.

A symmetric matrix A is positive definite if its associated quadratic form is positive definite. Similarly for the other 4 terms.

Example.

1. The matrix $A = I_n$ is positive definite: its bilinear form is

   $$\langle \vec{v}, \vec{w} \rangle_{I_n} = v_1 w_1 + \cdots + v_n w_n$$

   and its quadratic form is

   $$q_{I_n}(\vec{v}) = v_1^2 + \cdots + v_n^2$$

   We have recovered the usual dot product and square of the norm. The theory of quadratic form is in some aspects a study of generalized ways of defining lengths and angles.

2. Let η be the 4×4 diagonal matrix with diagonal entries $-c^2, 1, 1, 1$ for some positive constant c. Its associated quadratic form is

   $$q_\eta(t, x_1, x_2, x_3) = -c^2 t^2 + (x_1^2 + x_2^2 + x_3^2)$$

   This is an indefinite form since $q_\eta(1, 0, 0, 0) < 0$ but $q_\eta(0, 1, 0, 0) > 0$. In special relativity, length is measured using this form.

3. The form $q(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2 = (x_1 + x_2)^2$ is positive semidefinite. It is not positive definite since $q(1, -1) = 0$. It is associated to the matrix $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$.

Quadratic forms and (symmetric) bilinear forms satisfy the following properties

1. Bilinearity: $\langle \lambda\vec{u} + \mu\vec{v}, \vec{w}\rangle = \lambda\langle \vec{u}, \vec{w}\rangle + \mu\langle \vec{v}, \vec{w}\rangle$.
2. Symmetry: $\langle \vec{v}, \vec{w}\rangle = \langle \vec{w}, \vec{v}\rangle$.
3. Homogeneity: $q(\lambda\vec{v}) = \lambda^2 q(\vec{v})$.
4. Polarization: $q(\vec{v} + \vec{w}) - q(\vec{v}) - q(\vec{w}) = 2\langle \vec{v}, \vec{w}\rangle$.

It's worth drawing attention to polarization: in other words, **the quadratic form determines the bilinear form**. We saw this in Chapter 5 when we showed that length-preserving transformations also preserve angles.

Finally, we want to define an appropriate notion of equivalence of quadratic forms, so for example we want to $q_1(x, y) = x^2 - y^2$ and $q_2(x, y) = xy$ to be equivalent, since they are related by a linear change of variables: $q_2(x + y, x - y) = q_1(x, y)$.

Definition. Two symmetric matrices A and B are **equivalent** if there exists an invertible matrix Q such that $B = Q^T AQ$.

For us, the key problem is to classify quadratic forms up to equivalence. This turns out to be much easier than what we did last chapter for the similarity relation. Indeed, by the spectral theorem, we can already reduce A to a diagonal matrix using *only* orthogonal transformations. By multiplying a row of the transformation matrix by a scalar, we can change each diagonal entry by a square, so the end result is that A is equivalent to a diagonal matrix with only $0, 1, -1$ on the diagonal. We will see that this representation is almost unique.

7 Abstract definition*

When we first introduced linear transformations, we gave a definition using matrices and another one using the abstract properties of linearity. We will do a similar thing here.

Definition. A **symmetric bilinear form** in n variables is a function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying (Bilinearity) and (Symmetry) in the above list of properties.

Definition. A **quadratic form** in n variables is a function $q : \mathbb{R}^n \to \mathbb{R}$ satisfying (Homogeneity) and (Polarization), where (Polarization) means the function $f(\vec{v}, \vec{w}) = q(\vec{v} + \vec{w}) - q(\vec{v}) - q(\vec{w})$ is a symmetric bilinear form.

Choose a basis $\{\vec{v}_1, \cdots, \vec{v}_n\}$ for $\mathbb{R}^n$. Let f be a symmetric bilinear form, then it is easy to see that

$$f(\sum_i a_i\vec{v}_i, \sum_j b_j\vec{v}_j) = \sum_{i,j} a_ib_j f(\vec{v}_i, \vec{v}_j) = [a_1 \ \cdots \ a_n] A_f \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

where $A_f = [f(\vec{v}_i, \vec{v}_j)]$ is sometimes called the Gram matrix of f. It follows that with a choice of basis, the abstract definition and the definition in terms of matrices agree. The definition of equivalence using matrices is exactly the correct change of basis transformation for quadratic form.

8 Sylvester's law of inertia

Theorem (Sylvester's law of inertia)**.** Let A be an $n \times n$ symmetric matrix, then there exists a unique pair of non-negative integers (p, q) such that A is equivalent to the matrix

$$\begin{bmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{bmatrix}$$

In particular, if A is positive definite, then $A = LL^T$ for some invertible matrix L (since necessarily $(p, q) = (n, 0)$). This is the **Cholesky decomposition**.

Proof. It follows easily from the spectral theorem that A is equivalent to a diagonal matrix. Multiplying each row of the transformation matrix by an appropriate constant gives the required form. The hard part is uniqueness, namely to show that two matrices with distinct choices of (p, q) are not equivalent.

Suppose A is similar to B, then they have the same rank, since B is obtained from A by multiplying by invertible matrices. It follows that $p + q$ is an invariant. Therefore, we only need to determine p. We show that this is the maximum dimension of a subspace of $\mathbb{R}^n$ on which the quadratic form

$$f(x_1, \cdots, x_n) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2$$

is positive definite. This is a geometric property of a quadratic form, so it does not depend on the equivalence class of the matrix. Therefore, this would prove the theorem.

First, observe that on the subspace V_+ spanned by $\vec{e}_1, \cdots, \vec{e}_p$, the form is indeed positive definite, so the maximum is at least p. Similarly, on the subspace $V_- = \operatorname{span}(\vec{e}_{p+1}, \cdots, \vec{e}_n)$, the form is negative semidefinite. Therefore, if f is positive-definite on a subspace $V \subseteq \mathbb{R}^n$, then $V \cap V_- = \{0\}$, which implies $\dim V \leq n - \dim V_- = p$. $\square$

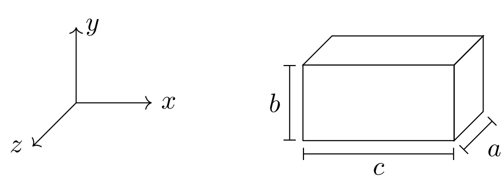
Definition. The **signature** of a quadratic form is the pair (p, q). The form is **non-degenerate** if $p+q = n$.

So for example, the form $-x_0^2 + (x_1^2 + x_2^2 + x_3^2)$ has signature $(3, 1)$, and the form x_0x_1 has signature $(1, 1)$. They are both non-degenerate. From the proof, it's clear that p is the number of positive eigenvalues, and q is the number of negative eigenvalues.

We end with a connection to physics. Given a rigid body $\Omega \subseteq \mathbb{R}^3$, we can consider its rotational moment of inertia around any axis. This is a function q assigning a (positive) number to each direction in $\mathbb{R}^3$. It can be packaged into a 3×3 matrix I_Ω called the *moment of inertia tensor* by physicists. With the terminologies we have, this is a symmetric bilinear form on $\mathbb{R}^3$, which is moreover positive definite.

By the spectral theorem, I_Ω is diagonal in an orthonormal basis. The three vectors in this basis are called *principal axes of rotations*, and the associated eigenvalues are the *principal moments of inertia*. If they are all distinct, then the axes are unique, and rotations around them are equilibrium states. It is stable if and only if the corresponding eigenvalue is not the middle one.

Example. Consider a rectangular box with side lengths $a < b < c$.



The moment of inertia tensor in this reference frame is the matrix

$$I_\Omega = \frac{1}{12} m \begin{bmatrix} b^2 + c^2 & & \\ & a^2 + c^2 & \\ & & a^2 + b^2 \end{bmatrix}$$

So the three axes of symmetries of the box are the three principal axes of rotations, which is expected. Moreover, the rotations around the x and z-axes are stable, and the rotation around the y-axis is unstable. One can experimentally verify this by throwing a box-shaped object (such as your textbook, bound together by a rubber band) up in the air and observe how much it wobbles.