

Lecture 1

Gidon Rosalki
2026-10-13

1. Introduction

Since everyone comes with a different background in mathematics to their masters in CompSci, this course exists to ensure everyone knows a standard foundation.

The course is split into 4 main parts, Linear Algebra, Infinitesimal Calculus, Optimisation, and Probability. Aside from the optimisation section, it will all be content already taught in one's bachelors degree. The purpose is to give us skills that we may use in our research.

2. Linear Algebra

We will be focusing on slightly less abstract concepts than in the bachelors courses. We will only work over the field $\mathbb{R}$, real matrices, and vector spaces, and so on, rather than abstract fields, spotlighting applicative concepts.

3. Matrices

Let there be a matrix $M \in \text{Mat}_{n \times m}(\mathbb{R})$. The vectors will all be column vectors (unless stated otherwise) $v \in \mathbb{R}^n$. We may transpose our matrix $M : M^T \in \text{Mat}_{m \times n}(\mathbb{R})$ defined such that $M_{i,j}^T = M_{j,i}$.

If M is a square matrix: $M \in \text{Mat}_{n \times n}(\mathbb{R})$, it is called **invertible** if $\exists M^{-1} \in M_n(\mathbb{R}) : M \cdot M^{-1} = I_n$.

3.1. Eigenvalues and Eigenvectors

Let there be $M \in \text{Mat}_n(\mathbb{R})$.

Definition (eigenvalue): $\lambda \in \mathbb{R}$ is called an **eigenvalue** of M if $\exists v \in \mathbb{R}^n \neq 0 : M \cdot v = \lambda \cdot v$. The eigenvalues of a matrix are called its *spectrum*.

Definition (eigenvector): Given an eigenvalue λ of matrix M , where $M \cdot v = \lambda \cdot v$, then $v \in \mathbb{R}^n \neq 0$ is called an **eigenvector**.

Definition (eigenspace): The **eigenspace** of λ is defined as: $V_\lambda = \{v \in \mathbb{R}^n : Mv = \lambda v\}$.

It is easy to show that V_λ is a linear subspace of $\mathbb{R}^n$.

$\dim(V_\lambda)$ is the (geometric) multiplicity of λ .

Fact:

$$\sum_{\lambda \text{ eigenvalue of } M} \dim(V_\lambda) \leq n$$

Definition (diagonalisable): M is diagonalisable if there is an invertible matrix P , and a diagonal matrix D such that $M = PDP^{-1}$, where $P, D \in M_n$.

Let v_i be the i -th column of P . Then

$$\begin{aligned} M \cdot v_i &= PDP^{-1} \cdot v_i \\ P \cdot P^{-1} &\Rightarrow = PDe_i \\ &= P \cdot D_{ii} \cdot e_i \\ &= D_{i,i}Pe_i \\ &= D_{i,i}v_i \end{aligned}$$

This implies that $D_{i,i}$ is an eigenvalue, with eigenvector v_i . From this we may conclude:

Theorem: M is diagonalisable $\iff$ there is a basis $v_1, \dots, v_n$ of $\mathbb{R}^n$ of eigenvectors of M .

3.2. Orthogonal Diagonalisation

We will firstly discuss the standard inner product. There are many different inner products, but we will focus on the standard one:

$$x, y \in \mathbb{R}^n \implies \langle x, y \rangle = \sum_{i=1}^n x_i \cdot y_i$$

Let $0 \neq v_1, \dots, v_n \in \mathbb{R}^n$.

Definition (orthogonal): $v_1, \dots, v_n$ are *orthogonal* if $\forall i \neq j \quad \langle v_i, v_j \rangle = 0$

Definition (orthonormal): $v_1, \dots, v_n$ are *orthonormal* if they are orthogonal, and $\forall i \quad \|v_i\| = \sqrt{\langle v_i, v_i \rangle} = 1$ (unit vectors)

Definition (orthogonal matrix): An $n \times n$ matrix V is called *orthogonal* if $V^T \cdot V = I$. Equivalent to:

1. The columns of V are orthonormal
2. The rows of V are orthonormal
3. $V^{-1} = V^T$
4. $\forall x \in \mathbb{R}^n \quad \|Vx\| = \|x\|$

Definition (orthogonally diagonalisable): A matrix M is *orthogonally diagonalisable* if there exist orthogonal matrix V , diagonal matrix D such that $M = V \cdot D \cdot V^T$.

Theorem (Spectral theorem for symmetric matrices): If M is symmetric ($M = M^T$), then M is orthogonally diagonalisable, i.e. there is an orthonormal basis of eigenvectors. In other words $M = V \cdot D \cdot V^T$, where V is orthogonal, and D is diagonal.

Examples:

No eigenvalues: $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$

Two eigenvalues: $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$

$$\lambda_1 = 1, \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\lambda_2 = 2, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The second matrix is diagonalisable, but not orthogonal diagonalisable.

Why is orthonormality so useful? Let us suppose that $v_1, \dots, v_n$ is an orthonormal basis of $\mathbb{R}^n$. This means that there is a unique way to write the vector x as a linear combination of these:

$$x = \sum_{i=1}^n \alpha_i v_i$$

If we suppose

$$y = \sum_{i=1}^n \beta_i v_i$$

$$\implies \langle x, y \rangle = \sum_{i=1}^n \alpha_i \beta_i$$

Similarly: $\|x\|^2 = \sum_{i=1}^n \alpha_i^2$.

Additionally, assuming that $v_1, \dots, v_n$ are eigenvectors, with eigenvalues $\lambda_1, \dots, \lambda_n$, for matrix M :

$$Mx = \sum_{i=1}^n \alpha_i \lambda_i v_i$$

$$\|Mx\|^2 = \sum_{i=1}^n \alpha_i^2 \lambda_i^2$$

3.3. Singular Value Decomposition (SVD)

Definition (SVD): Let $M \in M_{n \times m}(\mathbb{R})$.

1. V is an $m \times m$ orthogonal matrix
2. U is an $n \times n$ orthogonal matrix
3. Σ is diagonal: $\Sigma_{i,j} = 0$ if $i \neq j$, where $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \dots \geq \Sigma_{k,k} \geq 0 : k = \min(m, n)$

$$\text{s.t. } M = U \cdot \Sigma \cdot V^T$$

We denote $\sigma_i = \Sigma_{i,i} : 1 \leq i \leq k$, and call $\sigma_1, \dots, \sigma_k$ the singular values of M .

Let $u_1, \dots, u_n$ be the columns of U , and $v_1, \dots, v_n$ be the columns of V . This allows the following:

$$\begin{aligned} M \cdot v_i &= U \cdot \Sigma \cdot \overbrace{V^T \cdot v_i}^{e_i} \\ &= U \cdot \Sigma \cdot e_i \\ &= U \cdot \sigma_i \cdot e_i \\ &= \sigma_i \cdot U \cdot e_i \\ &= \sigma_i \cdot u_i \end{aligned}$$

Remark: If $m > n \wedge i > n \Rightarrow Mv_i = 0$

Let us now consider finding M^T :

$$M^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

Definition: $U = u_1, \dots, u_n, V = v_1, \dots, v_n$ are called the *singular vectors* of M .

Definition (operator norm): $M \in M_{n \times m}(\mathbb{R})$, $\|M\|_{\text{op}} = \max\{\|M \cdot v\| : v \in \mathbb{R}^m \wedge \|v\| = 1\}$ (may be considered the maximum increase that M can do to a vector).

Claim: $\|M\|_{\text{op}} = \sigma_1$, and v_1 is a vector that attains the maximum.

Proof: Let $v \in \mathbb{R}^m : \|v\| = 1$. We will write $v = \sum_{i=1}^m \alpha_i v_i$. So:

$$\begin{aligned} M \cdot v &= \sum_{i=1}^m \alpha_i (M \cdot v_i) \\ &= \sum_{i=1}^k \alpha_i \cdot \sigma_i u_i \\ 1 = \|v\|^2 &= \sum_{i=1}^m \alpha_i^2 \\ \|Mv\| &= \sum_{i=1}^k \alpha_i^2 \cdot \sigma_i^2 \\ &\leq \sigma_1^2 \cdot \sum_{i=1}^k \alpha_i^2 \\ &\leq \sigma_1^2 \end{aligned}$$

We showed that $\|Mv\| \leq \sigma_1$, and this may be attained by taking $v = v_1$: $Mv_1 = \sigma_1 u_1 \Rightarrow \|Mv_1\| = \sigma_1$.

Reminder: $v \perp v' \Rightarrow \langle v, v' \rangle = 0$

Exercise: $\max\{\|Mv\| : v \in \mathbb{R}^m, \|v\| = 1, v \perp v_1\} = \sigma_2$, v_2 attains the maximum.

3.3.1. SVD Proof

Our strategy is that we will construct an orthonormal basis $v_1, \dots, v_m$ for $\mathbb{R}^m$, such that $Mv_1, \dots, Mv_m$ are orthogonal: $\langle Mv_i, Mv_j \rangle = 0$. The idea is that

$$u_1 = \frac{Mv_1}{\|Mv_1\|}, \quad u_2 = \frac{Mv_2}{\|Mv_2\|}, \quad \dots$$

as long as $Mv_i \neq 0$, and complete it to an orthonormal basis. In particular, $\sigma_i := \|Mv_i\|$.

Proof: We will construct recursively:

1. Let v_1 be a unit vector that maximises

$$\{\|Mv\| : v \in \mathbb{R}^m, \|v\| = 1\}$$

2. Given $v_1, \dots, v_{i-1}$, we will let v_i be a unit vector that maximises

$$\{\|Mv\| : v \in \mathbb{R}^m, \|v\| = 1, v \perp v_1, \dots, v_{i-1}\}$$

We will now prove that $\forall i < j \quad \langle Mv_i, Mv_j \rangle = 0$. Consider $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(t) = \left\| M \left(\frac{v_i + tv_j}{\|v_i + tv_j\|} \right) \right\|^2$.

1. We will observe that $t = 0$ is a local maximum of $f(t)$ since $x_t = \frac{v_i + tv_j}{\|v_i + tv_j\|}$ is a unit vector

$$\text{orthogonal to } v_1, \dots, v_{i-1}. \text{ Therefore } \|Mx_t\|^2 \leq \|Mv_i\|^2 = \left\| \underbrace{Mx_0}_{t=0} \right\|^2$$

Now:

$$\begin{aligned} f(t) &= \left\| M \left(\frac{v_i + tv_j}{\|v_i + tv_j\|} \right) \right\|^2 \\ &= \frac{\|M(v_i + tv_j)\|^2}{\|v_i + tv_j\|^2} \\ &= \frac{\|M(v_i) + tM(v_j)\|^2}{\|v_i\|^2 + 2\langle v_i, tv_j \rangle + \|tv_j\|^2} \\ &= \frac{\|Mv_i\|^2 + 2\langle Mv_i, tMv_j \rangle + \|tMv_j\|^2}{1 + 0 + t^2} \\ &= \frac{\|Mv_i\|^2 + 2t\langle Mv_i, Mv_j \rangle + t^2\|Mv_j\|^2}{1 + t^2} \\ f'(t) &= \frac{\left(2\langle Mv_i, Mv_j \rangle + 2\|Mv_j\|^2 t \right) \cdot (1 + t^2) - 2t \cdot (\dots)}{(1 + t^2)^2} \\ \implies f'(0) &= 2\langle Mv_i, Mv_j \rangle \\ &= 0 \text{ since } t = 0 \text{ is a local maximum} \end{aligned}$$