Click here to view the PDF version.

Revision and preliminaries

This chapter provides some revision on concepts you should already know from previous classes. We will need these throughout the course. It is strongly advised you read in carefully in your own time to remind yourself.

Table of Contents

Taylor expansion and approximation

Given an infinitely differentiable function f its **Taylor series** at the point a is the infinite sum

$$\begin{aligned} f(x) &= f(a) + f'(a) \cdot (x-a) + \frac{1}{2!} f''(a) \cdot (x-a)^2 + \frac{1}{3!} f^{(3)}(a)(x-a)^3 + \cdots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a)(x-a)^n \end{aligned} \tag{1}$$

where $f^{(n)}$ is a short-hand for differentiating f n number of times.

Big O notation

Truncating this sum early (i.e. not summing all the way to infinity) is a frequent approximation technique often termed **Taylor approximation**.

Here we choose the example of a **second-degree approximation**, i.e. truncating the sum at $n = 2$ when we take at most two derivatives. We would thus get an approximation

$$f(x) \approx f(a) + (x-a)\, f'(a) + \frac{1}{2} f''(a)\, (x-a)^2.$$

Comparing with (1) we see

$$f(x) = f(a) + (x-a)\, f'(a) + \frac{1}{2} f''(a)\, (x-a)^2 + R(x) \tag{2}$$

where $R(x) = \sum_{n=3}^{\infty} \frac{1}{n!} f^{(n)}(a)(x-a)^n$ is usually called the **remainder term**.

As $x \to a$ the remainder $R(x)$ vanishes. More precisely it becomes small *at least as fast as* $(x-a)^3$. The idea is that as x gets closer to a then $|x-a|$ < 1, such that $|x-a|^n < |x-a|^3$ for all $n > 3$. While some of the derivatives $f^{(n)}(a)$ may be large, there still is a point when x is so close to a, such that $|x-a| \ll 1$ and still $f^{(n)}(a)|x-a|^n < |x-a|^3$ for all $n > 3$.

Mathematically one writes as:

> There exist positive constants $M, \delta > 0$ such that
>
> $$|R(x)| \leq M|x-a|^3 \qquad \forall 0 < |x-a| < \delta,$$

or more compactly using **big O notation** as

> $$R(x) = O((x-a)^3) \qquad \text{as } x \to a.$$

Employing this within (2) we arrive at

$$f(x) = f(a) + (x-a)\, f'(a) + \frac{1}{2} f''(a)\, (x-a)^2 + O((x-a)^3) \tag{3}$$

This **big O notation** is thus a mathematically precise way of saying that there are more terms in the expansion (3) that we don't show but their order is at most $(x-a)^3$. Or more generally

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a)(x-a)^k + O((x-a)^{n+1}) \tag{4}$$

Lagrange form of the reminder

An important result of the study of Taylor approximations is, that the remainder term can in fact be also expressed in terms of the next term of the Taylor approximation itself. Considering for example the general approximation (4) we can find a ξ between x and a — i.e. assuming $x < a$ a number $\xi \in (x, a)$ — such that

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a)(x-a)^k + \frac{1}{(n+1)!} f^{(n+1)}(\xi)\,(x-a)^{n+1}$$

holds. This is the so-called **Lagrange form** of the remainder. Note, how the red part is the next term of the Taylor polynomial, just with the derivative evaluated at a (generally unknown) position ξ.

Second-degree approximation (continued)

To return to our example of a second-degree approximation and assuming $x < a$ we would thus be able to write

$$f(x) = f(a) + (x-a) \cdot f'(a) + \frac{1}{2} f''(a) \cdot (x-a)^2 + \frac{1}{6} f'''(\xi)(x-a)^3$$

for some $\xi \in (x, a)$.

Typical Taylor variants

In the above example we considered the expansion of $f(x)$ around a point $x = a$ leading to an expansion of the form

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a)(x-a)^k + O(|x-a|^{n+1}).$$

Another common setting is to expand a function $f(x+h)$ about x. Replacing x by $x+h$ in the above expression directly yields

$$f(x+h) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x)\, h^k + O(|h|^{n+1})$$

Second-degree approximation (continued)

Again in our second-degree example we would obtain

$$f(x+h) = f(x) + h\, f'(x) + \frac{1}{2} f''(x)\, h^2 + O(|h|^3)$$

or using the Lagrange remainder

$$f(x+h) = f(x) + h\, f'(x) + \frac{1}{2} f''(x)\, h^2 + \frac{1}{6} f'''(\xi)\, h^3$$

for some $\xi \in (x, x+h)$.

Vector spaces

From a physical point of view the concept of a vector space V boils down to a well-defined set of objects, which are closed under linear combination. That is to say for two vectors $\mathbf{v} \in V$ and $\mathbf{w} \in V$ and real scalars $\alpha, \beta \in \mathbb{R}$ we want to have that any linear combination is also a vector, i.e.

$$\alpha\mathbf{v} + \beta\mathbf{w} \in V.$$

Provided that ask the scalar multiplication $\alpha\mathbf{v}$ and the vector addition $\mathbf{v} + \mathbf{w}$ to satisfy a few basic axioms, such as

- $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ ($+$ is commutative)
- $\alpha\mathbf{v} + \alpha\mathbf{w} = \alpha(\mathbf{v} + \mathbf{w})$ and $\alpha\mathbf{v} + \beta\mathbf{v} = (\alpha + \beta)\mathbf{v}$ ($+$ and scalar multiplication are distributive)
- ... (see Wikipedia for the full list)

we get a range of important properties, such as:

- Concept of linear independence, i.e. n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent if the only scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ to obtain

$$\mathbf{0} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$

are the zeros $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

- Existance of a **basis**, i.e. a set of n linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ such that all elements of V can be generated as linear combinations of these vectors, i.e. for all $\mathbf{w} \in V$ we can find scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$ such that

$$\mathbf{w} = \sum_{k=1}^{n} \alpha_k \mathbf{v}_k \tag{5}$$

Scalar product and norm

Frequently vector spaces feature an additional operation, the **inner product** or **scalar product**. For example for two Euclidean vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the scalar product is

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i \, y_i$$

Notable are the properties for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$:

- $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ *(Symmetry)*
- *Linearity:*

$$(\alpha \mathbf{x} + \beta \mathbf{y}) \cdot \mathbf{z} = \alpha \mathbf{x} \cdot \mathbf{z} + \beta \mathbf{y} \cdot \mathbf{z}$$

- $\mathbf{x} \cdot \mathbf{x} \geq 0$ *(Non-negativity)*

Derived from the scaler product is the **vector norm**, measuring the length of a vector:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Euclidean vector spaces

In this lecture when talking about vectors $\mathbf{v}$ without specifying any further details we usually refer to elements of the Euclidean vector space $\mathbb{R}^n$, that is elements

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

Polynomials as vector spaces

Any polynomial $p(x)$ of degree less than n can be written as

$$p(x) = \sum_{k=1}^{n} a_k x^k \tag{6}$$

where $a_k \in \mathbb{R}$. Comparing with (5) immediately suggests the set of all polynomials of degree less than n to be a vector space. This indeed can be shown to be the case.

- In comparing with (5) we notice the **monomials** $1 = x^0, x = x^1, x^2, x^3, \dots, x^n$ to be a possible choice for a basis.
- Similar to Euclidean vector spaces this is not the only choice of basis and in fact many families of polynomials are known, which are frequently employed as basis functions (e.g. Lagrange polynomials, Chebyshev polynomials, Hermite polynomials, ...)
- One basis we will discuss in the context of polynomial interpolation are Lagrange polynomials, which have the form

$$\begin{aligned} L_{i}(x) &= \prod_{\substack{j=1\\ j\neq i}}^{n+1} \frac{x - x_j}{x_i - x_j} \\ &= \frac{(x-x_1)(x-x_2)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_{n+1})}{(x_i-x_1)(x_i-x_2)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)} \end{aligned}$$

  for $i = 1, \dots, n+1$.

Taylor expansions of multi-dimensional functions

Scalar-valued functions

We consider a function $f : \mathbb{R}^n \to \mathbb{R}$. It's **second-order Taylor expansion** around $\mathbf{a} \in \mathbb{R}^n$ can be written as

$$f(\mathbf{x}) = f(\mathbf{a}) + (\nabla f(\mathbf{a}))^T\,(\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T\,\mathbf{H}_f(\mathbf{a})\,(\mathbf{x}-\mathbf{a}) + O(\|\mathbf{x}-\mathbf{a}\|^3),$$

where we introduced the gradient of f at $\mathbf{x} = \mathbf{a}$

$$\nabla f(\mathbf{a}) = \begin{pmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{x}=\mathbf{a}} \\ \left.\frac{\partial f}{\partial x_2}\right|_{\mathbf{x}=\mathbf{a}} \\ \vdots \\ \left.\frac{\partial f}{\partial x_n}\right|_{\mathbf{x}=\mathbf{a}} \end{pmatrix} \in \mathbb{R}^n$$

which is a vector of first partial derivatives. Similarly the Hessian of f at $\mathbf{x} = \mathbf{a}$ is the matrix of all second partial derivatives

$$\mathbf{H}_f(\mathbf{a}) = \begin{pmatrix} \left.\frac{\partial^2 f}{\partial x_1 \partial x_1}\right|_{\mathbf{x}=\mathbf{a}} & \left.\frac{\partial^2 f}{\partial x_1 \partial x_2}\right|_{\mathbf{x}=\mathbf{a}} & \cdots & \left.\frac{\partial^2 f}{\partial x_1 \partial x_n}\right|_{\mathbf{x}=\mathbf{a}} \\ \left.\frac{\partial^2 f}{\partial x_2 \partial x_1}\right|_{\mathbf{x}=\mathbf{a}} & \left.\frac{\partial^2 f}{\partial x_2 \partial x_2}\right|_{\mathbf{x}=\mathbf{a}} & \cdots & \left.\frac{\partial^2 f}{\partial x_2 \partial x_n}\right|_{\mathbf{x}=\mathbf{a}} \\ \vdots & \vdots & & \vdots \\ \left.\frac{\partial^2 f}{\partial x_n \partial x_1}\right|_{\mathbf{x}=\mathbf{a}} & \left.\frac{\partial^2 f}{\partial x_n \partial x_2}\right|_{\mathbf{x}=\mathbf{a}} & \cdots & \left.\frac{\partial^2 f}{\partial x_n \partial x_n}\right|_{\mathbf{x}=\mathbf{a}} \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Example

We consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined as

$$f(\mathbf{x}) = f(\,(x_1, x_2, x_3)^T\,) = x_1^3 + x_1 x_3^2 + x_2^2 x_3 + 3x_3^2$$

and compute the second-order Taylor expansion at

$$\mathbf{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

First we compute the gradient $\nabla f(\mathbf{a})$

$$\nabla f(\mathbf{a}) = \begin{pmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{x}=\mathbf{a}} \\ \left.\frac{\partial f}{\partial x_2}\right|_{\mathbf{x}=\mathbf{a}} \\ \left.\frac{\partial f}{\partial x_3}\right|_{\mathbf{x}=\mathbf{a}} \end{pmatrix} = \begin{pmatrix} 3a_1^2 + a_3^2 \\ 2a_2 a_3 \\ 2a_1 a_3 + a_2^2 + 6a_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 8 \end{pmatrix}$$

Next the Hessian matrix $\mathbf{H}_f(\mathbf{a})$

$$\mathbf{H}_f(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial^2 f}{\partial x_1 \partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial^2 f}{\partial x_1 \partial x_3}\Big|_{\mathbf{x}=\mathbf{a}} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial^2 f}{\partial x_2 \partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial^2 f}{\partial x_2 \partial x_3}\Big|_{\mathbf{x}=\mathbf{a}} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial^2 f}{\partial x_3 \partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial^2 f}{\partial x_3 \partial x_3}\Big|_{\mathbf{x}=\mathbf{a}} \end{pmatrix}$$

$$= \begin{pmatrix} 6a_1 & 0 & 2a_3 \\ 0 & 2a_3 & 2a_2 \\ 2a_3 & 2a_2 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 6 \end{pmatrix}$$

Combining the results we get

$$f(\mathbf{x}) = f(\mathbf{a}) + \begin{pmatrix} 4 \\ 0 \\ 8 \end{pmatrix}^T (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \begin{pmatrix} 6 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 6 \end{pmatrix} (\mathbf{x} - \mathbf{a}) + O(\|\mathbf{x} - \mathbf{a}\|^3),$$

which simplifies (after some algebra) to

$$f(\mathbf{x}) = 4 - \begin{pmatrix} 10 \\ 0 \\ 0 \end{pmatrix}^T \mathbf{x} + \mathbf{x}^T \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{pmatrix} \mathbf{x} + O(\|\mathbf{x} - \mathbf{a}\|^3),$$

Vector-valued function

Now we consider a vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. It's **first-order Taylor expansion** around $\mathbf{a} \in \mathbb{R}^n$ can be written as

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \mathbf{J}_\mathbf{f}(\mathbf{a})\ (\mathbf{x} - \mathbf{a}) + O(\|\mathbf{x} - \mathbf{a}\|^2).$$

In this the **Jacobian matrix** $\mathbf{J}_\mathbf{f}(\mathbf{x}) \in \mathbb{R}^{m \times n}$ is the collection of all partial derivatives of $\mathbf{f}$, i.e.

$$\mathbf{J}_\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_1}{\partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \cdots & \frac{\partial f_1}{\partial x_n}\Big|_{\mathbf{x}=\mathbf{a}} \\ \frac{\partial f_2}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_2}{\partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \cdots & \frac{\partial f_2}{\partial x_n}\Big|_{\mathbf{x}=\mathbf{a}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_m}{\partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \cdots & \frac{\partial f_m}{\partial x_n}\Big|_{\mathbf{x}=\mathbf{a}} \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Example

We consider the function $\mathbf{f}: \mathbb{R}^3 \to \mathbb{R}^2$ defined as

$$\begin{cases} f_1(x_1, x_2, x_3) = -x_1 \cos(x_2) - 1 \\ f_2(x_1, x_2, x_3) = x_1\, x_2 + x_3 \end{cases}$$

and compute its first-order Taylor expansion at

$$\mathbf{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

The Jacobian is

$$\begin{aligned} \mathbf{J}_{\mathbf{f}}(\mathbf{a}) &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_1}{\partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_1}{\partial x_3}\Big|_{\mathbf{x}=\mathbf{a}} \\ \frac{\partial f_2}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_2}{\partial x_2}\Big|_{\mathbf{x}=\mathbf{a}} & \frac{\partial f_2}{\partial x_3}\Big|_{\mathbf{x}=\mathbf{a}} \end{pmatrix} \\ &= \begin{pmatrix} -\cos(a_2) & a_1 \sin(a_2) & 0 \\ a_2 & a_1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \end{aligned}$$

such that we arrive at

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} (\mathbf{x} - \mathbf{a}) + O(\|\mathbf{x} - \mathbf{a}\|^2)$$

Again some algebra this is simplified to

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} -3 \\ 2 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mathbf{x} + O(\|\mathbf{x} - \mathbf{a}\|^2)$$

Triangle inequality

An important inequality between numbers and vectors is the **triangle inequality**. For example between two vectors $\mathbf{x}$, $\mathbf{y}$ it reads:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

Similarly between real numbers $\alpha_1, \alpha_2 \in \mathbb{R}$:

$$|\alpha_1 + \alpha_2| \leq |\alpha_1| + |\alpha_2|.$$

Notably, this generalises to arbitrarily long sums:

$$\left|\sum_{i=1}^{n} \alpha_i\right| \leq \sum_{i=1}^{n} |\alpha_i|$$

Numerical analysis