CALCULUS

1. Basic Facts on Multivariate Calculus

↪ If $x, y \in \mathbb{R}^d$, we will denote by $[x,y]$ the closed segment delimited by x and y, i.e., the set of all points $(1-t)x + ty$ for $0 \leq t \leq 1$. One denotes by $[x,y)$, $(x,y]$ and (x,y) the semi-open or open segments with appropriate strict inequality for t. Similarly to the notation for open intervals, whether (x,y) denotes an open segment or a pair of points will always be clear from the context.

↪ The derivative of a differentiable function $f : t \mapsto f(t)$ from an interval $I \subset \mathbb{R}$ to $\mathbb{R}$ will be denoted by ∂f, or $\partial_t f$ if the variable t is well identified. Its value at $t_0 \in I$ is denoted either as $\partial f(t_0)$ or even $\partial f|_{t=t_0}$. Higher derivatives are denoted as $\partial^{(k)} f$, $k \geq 0$, with the usual convention $\partial^{(0)} f = f$. Note that the notation such as $f', f'', f^{(3)}$ will never refer to derivatives. In the following, U is an open subset of $\mathbb{R}^d$. If f is a function from U to $\mathbb{R}^m$, we let $f^{(i)}$ denote the i^{th} component of f, so that:

$$f(x) = \begin{pmatrix} f^{(1)}(x) \\ \vdots \\ f^{(m)}(x) \end{pmatrix}, \quad \text{for } x \in U. \tag{4.1}$$

If $d=1$, and f is differentiable, the derivative of f at x is the column vector of the derivatives of its components. In symbols:

$$\partial f(x) = \begin{pmatrix} \partial f^{(1)}(x) \\ \vdots \\ \partial f^{(m)}(x) \end{pmatrix}, \quad \text{for } x \in U. \tag{4.2}$$

For $d \geq 1$ and $j \in \{1, \dots, d\}$, the j^{th} partial derivative of f at x is given by:

$$\partial_j f(x) = \partial\left(t \mapsto f(x + tc_j)\right)\Big|_{t=0} \in \mathbb{R}^m, \tag{4.3}$$

where $c_1, \dots, c_d$ form the canonical basis of $\mathbb{R}^d$. If the notation for the variables on which f depends is well understood from the context, we will alternatively use $\partial_{x_j} f$. For example, if $f : (\alpha, \beta) \mapsto f(\alpha, \beta)$, we will prefer $\partial_\alpha f$ to $\partial_1 f$. The differential of f at x is the linear mapping from $\mathbb{R}^d$ to $\mathbb{R}^m$ represented by the matrix:

$$df(x) = \left[\partial_1 f(x), \dots, \partial_d f(x)\right]. \tag{4.4}$$

It is defined so that, for all $h \in \mathbb{R}^d$:

$$df(x)h = \partial(t \mapsto f(x+th))\Big|_{t=0} \, , \tag{4.5}$$

where the right-hand side is the directional derivative of f at x in the direction h. Note that, if $f : \mathbb{R}^d \to \mathbb{R}$ (i.e., if $m=1$), $df(x)$ is a row vector. If f is differentiable on U and $df(x)$ is continuous as a function of x, one says that f is continuously differentiable, or C^1.

Differentials obey the product rule and the chain rule. If $f, g : U \to \mathbb{R}$, then:

$$d(fg)(x) = f(x)dg(x) + g(x)df(x). \tag{4.6}$$

If $f : U \to \mathbb{R}^m$, $g : \tilde{U} \subset \mathbb{R}^k \to U$, then:

$$d(f \circ g)(x) = df(g(x))\, dg(x) \, . \tag{4.7}$$

If $d = m$ (so that $df(x)$ is a square matrix), we let $\underline{\nabla} \cdot f(x) = \mathrm{tr}(df(x))$, the divergence of f.

The Euclidean gradient of a differentiable function $f : U \to \mathbb{R}$ is $\underline{\nabla} f(x) = df(x)^T$. More generally, one defines the gradient of f with respect to a tensor field $x \mapsto A(x)$ taking values in $\mathcal{S}_d^{++}$, as the vector $\underline{\nabla}_A f(x)$ that satisfies the following relation:

$$df(x)h = \underline{\nabla}_A f(x)^T A(x) h \, , \ \forall h \in \mathbb{R}^d \implies \underline{\nabla}_A f(x) = A(x)^{-1} df(x)^T . \tag{4.8}$$

In particular, the Euclidean gradient is associated with $A(x) = \mathrm{Id}_{\mathbb{R}^d}$ for all x. With some abuse of notation, we will denote $\underline{\nabla}_A f = A^{-1} \underline{\nabla} f$ when A is a fixed matrix, therefore identified with the constant tensor field $x \mapsto A$.

↪ We here compute, as an illustration and because they will be useful later, the differential of the determinant and the inversion in matrix spaces. Recall that, if $A = [a_1, \dots, a_d] \in \mathcal{M}_d$ is a $d \times d$ matrix, with $a_1, \dots, a_d \in \mathbb{R}^d$ and $\det(A)$ is a d-linear form $\delta(a_1, \dots, a_d)$ which vanishes when two columns coincide and such that $\delta(e_1, \dots, e_d) = 1$. In particular δ changes signs when two of its columns are inverted. It follows from this that:

$$\begin{aligned} \partial_{a_{ij}} \det(A) &= \delta(a_1, \dots, a_{i-1}, e_j, a_{i+1}, \dots, a_d) \\ &= (-1)^{i-1} \delta(e_j, a_1, \dots, a_{i-1}, \dots, a_d) = (-1)^{i+j} \det\left[A^{(ij)}\right] , \end{aligned} \tag{4.9}$$

where $A^{(ij)}$ is the matrix A with row i and column j removed. We therefore find that the differential of $A \mapsto \det(A)$ is the mapping given by:

$$H \mapsto \mathrm{tr}\left(\mathrm{cof}(A)^T H\right), \tag{4.10}$$

where $\mathrm{cof}(A)$ is the matrix composed of co-factors $(-1)^{i+j} \det\left[A^{(ij)}\right]$. As a consequence, if A is invertible, then the differential of $\log|\det(A)|$ is the mapping:

$$H \mapsto \mathrm{tr}\left[\det(A)^{-1} \mathrm{cof}(A)^T H\right] = \mathrm{tr}(A^{-1} H). \tag{4.11}$$

Consider now the function $I(A) = A \mapsto A^{-1}$ defined on $GL_d(\mathbb{R})$, which is an open subset of $M_d(\mathbb{R})$. Using the relation $AI(A) = Id_{\mathbb{R}^d}$ and the product rule, we get :

$$A[dI(A)H] + HI(A) = 0 \iff dI(A)H = -A^{-1}HA^{-1}. \tag{4.12}$$

↪ Higher-order partial derivatives $\partial_{i_k} \cdots \partial_{i_1} f : U \to \mathbb{R}^m$ are defined by iterating the definition of first-order derivatives, namely :

$$\partial_{i_k} \cdots \partial_{i_1} f(x) = \partial_{i_k}(\partial_{i_{k-1}} \cdots \partial_{i_1} f)(x). \tag{4.13}$$

If all order k partial derivatives of f exists and are continuous, one says that f is k-times continuously differentiable, or C^k and, when true, the order in which the derivatives are taken does not matter. In this case, one typically groups derivatives with the same order using a power notation, writing, for example :

$$\partial_1\partial_2\partial_1 f \equiv \partial_1^2\partial_2 f, \text{ for a } C^3 \text{ function}. \tag{4.14}$$

If f is C^k, its k^{th} differential at x is a symmetric k-multilinear map that can also be iteratively defined, for $h_1, \dots, h_k \in \mathbb{R}^d$, by :

$$d^k f(x)(h_1, \dots, h_k) = d[d^{k-1} f(x)(h_1, \dots, h_{k-1})]h_k \in \mathbb{R}^m. \tag{4.15}$$

It is related to partial derivatives through the relation :

$$d^k f(x)(h_1, \dots, h_k) = \sum_{i_1, \dots, i_k = 1}^{d} h_1^{(i_1)} \cdots h_k^{(i_k)} \partial_{i_k} \cdots \partial_{i_1} f(x). \tag{4.16}$$

When $m=1$ and $k=2$, one denotes by $\nabla^2 f(x) = (\partial_i\partial_j f(x), i,j = 1, \dots, n)$ the symmetric matrix formed by partial derivatives of order 2 of f at x. It is called the Hessian of f at x and satisfies :

$$h_1^T \nabla^2 f(x) h_2 = d^2 f(x)(h_1, h_2). \tag{4.17}$$

The Laplacian of f is the trace of $\nabla^2 f$ and denoted by Δf.

↪ Taylor's theorem, in its integral form, generalizes the fundamental theorem of calculus to higher derivatives. It expresses the fact that, if f is C^k on U and $x, y \in U$ are such that the closed segment $[x, y]$ is included in U, then, letting $h = y - x$, we get :

- $$f(x+h) = f(x) + df(x)h + \frac{1}{2} d^2 f(x)(h,h) + \dots + \frac{1}{(k-1)!} d^{k-1} f(x)(h, \dots, h)$$
$$+ \frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} d^k f(x+th)(h, \dots, h)\, dt. \tag{4.18}$$

The last term (remainder) can also be written as :

$$\frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1} d^k f(x+th)(h,\dots,h)\,dt = \frac{1}{k!}\,\frac{\int_0^1 (1-t)^{k-1} d^k f(x+th)(h,\dots,h)\,dt}{\int_0^1 (1-t)^{k-1}\,dt}\,. \tag{4.19}$$

If f takes scalar values, then $d^k f(x+th)(h,\dots,h)$ is real and the intermediate value theorem implies that there exists some $z \in [x,y]$ such that :

- $$f(x+h) = f(x) + df(x)h + \frac{1}{2}d^2 f(x)(h,h) + \dots + \frac{1}{(k-1)!}d^{k-1}f(x)(h,\dots,h) + \frac{1}{k!}d^k f(z)(h,\dots,h)\,. \tag{4.20}$$

This is not true if f takes vector values. However, for any M such that $|d^k f(z)| \leq M$ for $z \in [x,y]$ (such M's always exist because f is C^k), one has :

$$\frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1} d^k f(x+th)(h,\dots,h)\,dt \leq \frac{M}{k!}|h|^k\,. \tag{4.21}$$

Eq. (4.18) can be written as :

- $$f(x+h) = f(x) + df(x)h + \frac{1}{2}d^2 f(x)(h,h) + \dots + \frac{1}{k!}d^k f(x)(h,\dots,h) + \frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1}\left[d^k f(x+th)(h,\dots,h) - d^k f(x)(h,\dots,h)\right]dt\,. \tag{4.22}$$

Let $\varepsilon_x(r)$ be defined as below. Since $d^k f$ is continuous, $\varepsilon_x(r)$ tends to 0 when $r \to 0$ and we have :

$$\varepsilon_x(r) = \max\left\{ |d^k f(x+h) - d^k f(x)| : |h| \leq r \right\}\,,$$

$$\Rightarrow \quad \int_0^1 (1-t)^{k-1}\left|d^k f(x+th)(h,\dots,h) - d^k f(x)(h,\dots,h)\right| dt \leq \frac{|h|^k}{k!}\varepsilon_x(|h|)\,,$$

$$\begin{aligned}\therefore \quad f(x+h) &= f(x) + df(x)h + \frac{1}{2}d^2 f(x)(h,h) + \dots + \frac{1}{k!}d^k f(x)(h,\dots,h) + \frac{|h|^k}{k!}\varepsilon_x(|h|) \\ &= f(x) + df(x)h + \frac{1}{2}d^2 f(x)(h,h) + \dots + \frac{1}{k!}d^k f(x)(h,\dots,h) + O(|h|^k)\end{aligned} \tag{4.23}$$

2. Basic Facts on Topology

↪ The open balls in $\mathbb{R}^d$ will be denoted:

$$\mathcal{B}(x,r) = \{y \in \mathbb{R}^d : |y-x| < r\}, \tag{4.24}$$

with $x \in \mathbb{R}^d$ and $r > 0$. The closed balls are denoted by $\overline{\mathcal{B}}(x,r)$ and contain all y's such that $|y-x| \leq r$. A set $U \subset \mathbb{R}^d$ is open if and only if for any $x \in U$, there exists $r > 0$ such that $\mathcal{B}(x,r) \subset U$. A set $\Gamma \subset \mathbb{R}^d$ is closed if its complement, denoted by:

$$\Gamma^c = \{x \in \mathbb{R}^d : x \notin \Gamma\}, \tag{4.25}$$

is open. The topological interior of a set $A \subset \mathbb{R}^d$ is the largest open set included in A. It will be denoted either by $\mathring{A}$ or $\mathrm{int}(A)$. A point x belongs to $\mathring{A}$ if and only if $\mathcal{B}(x,r) \subset A$ for some $r > 0$.

➜ The closure of A is the smallest closed set that contains A and will be denoted either $\bar{A}$ or $cl(A)$. A point x belongs to $\bar{A}$ if and only if $\mathcal{B}(x,r) \cap A \neq \emptyset$ for all $r > 0$. Alternatively, x belongs to $\bar{A}$ if and only if there exists a sequence (x_k) that converges to x with $x_k \in A$ for all k.

➜ A compact set in $\mathbb{R}^d$ is a set Γ such that any sequence of points in Γ contains a subsequence that converges to some point in Γ. An alternative definition is that, whenever Γ is covered by a collection of open sets, there exists a finite subcollection that still covers Γ. One can show that compact subsets of $\mathbb{R}^d$ are exactly its bounded and closed subsets.

➜ A metric space is a space $\mathcal{R}$ equipped with a distance, i.e., a function $\rho : \mathcal{R} \times \mathcal{R} \to [0, +\infty)$ that satisfies the following three properties:

$$\bullet \begin{cases} \forall x,y \in \mathcal{R} : \rho(x,y) = 0 \Leftrightarrow x = y, \\ \forall x,y \in \mathcal{R} : \rho(x,y) = \rho(y,x), \\ \forall x,y,z \in \mathcal{R} : \rho(x,y) + \rho(y,z). \end{cases} \tag{4.26}$$

The last equation in Eq. (4.26) is the triangle inequality. The norm of the difference between two points: $\rho(x,y) = |x-y|$, is a distance on $\mathbb{R}^d$. The definition of open and closed subsets in metric spaces is the same as above, with $\rho(x,y)$ replacing $|x-y|$, and one says that (x_n) converges to x if and only if $\rho(x_n,x) \to 0$.

⊛ Compact subsets are also defined in the same way, but are not necessarily characterized as bounded and closed.

//