

# Notes on Differential Calculus

## Contents

1. Differentiability	1
2. Chain rule	2
3. Partial derivatives	3
3.1. Directional derivatives	5
3.2. Differentiation on manifolds *	7
4. Mean value theorem	8
5. Inverse and implicit function theorems	9
6. Taylor's theorem	11
7. Critical points and extrema	12

## 1. Differentiability

**Definition 1.1:** Let  $f : (a, b) \rightarrow \mathbb{R}^n$ , and let  $f_i = \pi_i \circ f$  be its components. Then,  $f$  is differentiable at  $t_0 \in (a, b)$  if the following limit exists.

$$f'(t_0) = \lim_{h \rightarrow 0} \frac{f(t_0 + h) - f(t_0)}{h}.$$

*Remark:* The vector  $f'(t_0)$  represents the tangent to the curve  $f$  at the point  $f(t_0)$ . The full tangent line is the parametric curve  $f(t) + f'(t_0)(t - t_0)$ .

**Definition 1.2:** Let  $U \subseteq \mathbb{R}^n$  be open, and let  $f : U \rightarrow \mathbb{R}^m$ . Then,  $f$  is differentiable at  $x \in U$  if there exists a linear transformation  $\lambda : \mathbb{R}^n \rightarrow \mathbb{R}^m$  such that

$$\lim_{h \rightarrow 0} \frac{f(x + h) - f(x) - \lambda h}{\|h\|} = 0.$$

The derivative of  $f$  at  $x$  is denoted by  $\lambda = Df(x)$ .

*Remark:* In a neighbourhood of  $x$ , we may approximate

$$f(x + h) \approx f(x) + Df(x)(h).$$

*Remark:* The statement that this quantity goes to zero means that each of the  $m$  components must also go to zero. For each of these limits, there are  $n$  axes along which we can let  $h \rightarrow 0$ . As a result, we obtain  $m \times n$  limits, which allow us to identify the  $m \times n$  components of the matrix representing the linear transformation  $\lambda$  (in the standard basis). These are the partial derivatives of  $f$ , and the matrix of  $\lambda$  is the Jacobian matrix of  $f$  evaluated at  $x$ .

*Example:* Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map. By choosing  $\lambda = T$ , we see that  $T$  is differentiable everywhere, with  $DT(x) = T$  for every choice of  $x \in \mathbb{R}^n$ . This is made obvious by the fact that

the best linear approximation of a linear map at some point is the map itself; indeed, the ‘approximation’ is exact.

**Lemma 1.1:** If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $x \in \mathbb{R}^n$ , with derivative  $Df(x)$ , then

1.  $f$  is continuous at  $x$ .
2. The linear transformation  $Df(x)$  is unique.

*Proof:* We prove the second part. Suppose that  $\lambda, \mu$  satisfy the requirements for  $Df(x)$ ; it can be shown that  $\lim_{h \rightarrow 0} (\lambda - \mu)h / \|h\| = 0$ . Now, if  $\lambda v \neq \mu v$  for some non-zero vector  $v \in \mathbb{R}^n$ , then

$$\lambda v - \mu v = \frac{\lambda(tv) - \mu(tv)}{\|tv\|} \cdot \|v\| \rightarrow 0,$$

a contradiction. ■

## 2. Chain rule

**Exercise I:** Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transformation. Then, there exists  $M > 0$  such that for all  $x \in \mathbb{R}^n$ , we have

$$\|Tx\| \leq M\|x\|.$$

*Solution:* Set  $v_i = T(e_i)$  where  $e_i$  are the standard unit basis vectors of  $\mathbb{R}^n$ . Then,

$$\|Tx\| = \left\| \sum_i x_i v_i \right\| \leq \sum_i \|x_i v_i\| \leq \max \|v_i\| \sum_i |x_i|.$$

Since each  $|x_i| \leq \|x\|$ , set  $M = n \max \|v_i\|$  and write

$$\|Tx\| \leq \max \|v_i\| \sum_i |x_i| \leq \max \|v_i\| \cdot n\|x\| = M\|x\|.$$

**Theorem 2.1** (Chain Rule): Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  where  $f$  is differentiable at  $a \in \mathbb{R}^n$  and  $g$  is differentiable at  $f(a) \in \mathbb{R}^m$ . Then,  $g \circ f$  is differentiable, with  $D(g \circ f)(a) = Dg(f(a)) \circ Df(a)$ . Note that this means that the Jacobian matrices simply multiply.

*Proof:* Set  $b = f(a) \in \mathbb{R}^m$ ,  $\lambda = Df(a)$ ,  $\mu = Dg(f(a))$ . Define

$$\begin{aligned} \varphi : \mathbb{R}^n &\rightarrow \mathbb{R}^m, & \varphi(x) &= f(x) - f(a) - \lambda(x - a), \\ \psi : \mathbb{R}^m &\rightarrow \mathbb{R}^k, & \psi(y) &= g(y) - g(b) - \mu(y - b). \end{aligned}$$

We claim that

$$\lim_{x \rightarrow a} \frac{g \circ f(x) - g \circ f(a) - \mu \circ \lambda(x - a)}{\|x - a\|} = 0.$$

Write the numerator as

$$g \circ f(x) - g \circ f(a) - \mu \circ \lambda(x - a) = \psi(f(x)) + \mu(\varphi(x)).$$

Note that

$$\lim_{x \rightarrow a} \frac{\varphi(x)}{\|x - a\|} = 0, \quad \lim_{y \rightarrow b} \frac{\psi(y)}{\|y - b\|} = 0.$$

Thus, find  $M > 0$  such that

$$\|\mu(\varphi(x))\| \leq M\|\varphi(x)\|$$

for all  $x \in \mathbb{R}^n$ , hence

$$\lim_{x \rightarrow a} \frac{\|\mu(\varphi(x))\|}{\|x - a\|} \leq \lim_{\{x \rightarrow a\}} \frac{M\|\varphi(x)\|}{\|x - a\|} = 0.$$

Now write

$$\lim_{f(x) \rightarrow b} \frac{\psi(f(x))}{\|f(x) - b\|} = 0,$$

hence for any  $\varepsilon > 0$ , there is a neighbourhood of  $b$  on which

$$\|\psi(f(x))\| \leq \varepsilon\|f(x) - b\| = \varepsilon\|\varphi(x) + \lambda(x - a)\|.$$

Apply the triangle inequality and find  $M' > 0$  such that

$$\|\psi(f(x))\| \leq \varepsilon\|\varphi(x)\| + \varepsilon M'\|x - a\|.$$

Thus,

$$\lim_{x \rightarrow a} \frac{\|\psi(f(x))\|}{\|x - a\|} \leq \lim_{x \rightarrow a} \frac{\varepsilon\|\varphi(x)\|}{\|x - a\|} + \varepsilon M' = \varepsilon M'.$$

Since  $\varepsilon > 0$  was arbitrary, this limit is zero, completing the proof. ■

### 3. Partial derivatives

**Definition 3.1:** Let  $U \subseteq \mathbb{R}^n$  be open, and let  $f : U \rightarrow \mathbb{R}$ . The partial derivative of  $f$  with respect to the coordinate  $x_j$  at some  $a \in U$  is defined by the following limit, if it exists.

$$\frac{\partial f}{\partial x_j}(a) = \lim_{h \rightarrow 0} \frac{f(a + he_j) - f(a)}{h}.$$

**Lemma 3.1:** If  $f : U \rightarrow \mathbb{R}$  is differentiable at a point  $a \in \mathbb{R}^n$ , then

$$Df(a)(x_1, \dots, x_n) = x_1 \frac{\partial f}{\partial x_1}(a) + \dots + x_n \frac{\partial f}{\partial x_n}(a).$$

*Example:* Consider

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto \begin{cases} (xy)/(x^2 + y^2), & \text{if } (x, y) \neq (0, 0) \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

Note that  $f$  is not differentiable at  $(0, 0)$ ; it is not even continuous there. However, both partial derivatives of  $f$  exist at  $(0, 0)$ .

**Lemma 3.2:** If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $a \in \mathbb{R}^n$ , then the matrix representation of  $Df(a)$  in the standard basis is given by

$$[Df(a)] = \left[ \frac{\partial f_i}{\partial x_j}(a) \right]_{ij}.$$

**Lemma 3.3:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be differentiable at  $a \in \mathbb{R}^n$ , and let  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  be differentiable at  $f(a) \in \mathbb{R}^m$ . Then, the matrix representation of  $D(g \circ f)(a)$  in the standard basis is the product

$$[D(g \circ f)(a)] = [Dg(f(a))][Df(a)] = \left[ \sum_{\ell=1}^m \frac{\partial g_i}{\partial y_\ell}(f(a)) \frac{\partial f_\ell}{\partial x_j}(a) \right]_{ij}.$$

In other words,

$$\frac{\partial}{\partial x_j}(g \circ f)_{i(a)} = \sum_{\ell=1}^m \frac{\partial g_i}{\partial y_\ell}(f(a)) \frac{\partial f_\ell}{\partial x_j}(a).$$

*Example:* Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be differentiable, and let  $\Gamma(f) = \{(x, y, f(x, y)) : x, y \in \mathbb{R}\}$  be the graph of  $f$ . Now, let  $\gamma : [-1, 1] \rightarrow \Gamma(f)$  be a differentiable curve, represented by

$$\gamma(t) = (g(t), h(t), f(g(t), h(t))).$$

Then, we can compute the derivative

$$\gamma'(a) = \left( g'(a), h'(a), g'(a) \frac{\partial f}{\partial x} + h'(a) \frac{\partial f}{\partial y} \Big|_{(g(a), h(a))} \right)$$

**Exercise II:** Consider the inner product map,  $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ . What is its derivative?

*Solution:* We treat the inner product as a map  $g : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ , which acts as

$$\langle \mathbf{x}, \mathbf{y} \rangle := g(x_1, \dots, x_n, y_1, \dots, y_n) = x_1 y_1 + \dots + x_n y_n.$$

Now, note that

$$\frac{\partial g}{\partial x_i} = y_i, \quad \frac{\partial g}{\partial y_i} = x_i.$$

Thus,

$$\begin{aligned} Dg(\mathbf{a}, \mathbf{b})(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^n x_i \frac{\partial g}{\partial x_i}(\mathbf{a}, \mathbf{b}) + \sum_{i=1}^n y_i \frac{\partial g}{\partial y_i}(\mathbf{a}, \mathbf{b}) \\ &= \sum_{i=1}^n x_i b_i + \sum_{i=1}^n y_i a_i \\ &= \langle \mathbf{x}, \mathbf{b} \rangle + \langle \mathbf{y}, \mathbf{a} \rangle. \end{aligned}$$

In other words, the matrix representation of the derivative of the inner product map at the point  $(\mathbf{a}, \mathbf{b})$  is given by  $[\mathbf{b}^\top \mathbf{a}^\top]$ .

**Exercise III:** Let  $\gamma : \mathbb{R} \rightarrow \mathbb{R}^n$  be a differentiable curve. What is the derivative of the real map  $t \mapsto \|\gamma(t)\|^2$ ?

*Solution:* We write this map as  $t \mapsto \langle \gamma(t), \gamma(t) \rangle$ . Consider the scheme

$$\mathbb{R} \rightarrow \mathbb{R}^{2n} \rightarrow \mathbb{R}, \quad t \mapsto \begin{pmatrix} \gamma(t) \\ \gamma(t) \end{pmatrix} \mapsto \langle \gamma(t), \gamma(t) \rangle.$$

Pick a point  $t \in \mathbb{R}$ , whence the derivative of the map at  $t$  is

$$\begin{pmatrix} \gamma(t)^\top & \gamma(t)^\top \end{pmatrix} \begin{pmatrix} \gamma'(t) \\ \gamma'(t) \end{pmatrix} = 2\langle \gamma(t), \gamma'(t) \rangle.$$

*Remark:* Consider the surface  $S^{n-1} \subset \mathbb{R}^n$ , and pick an arbitrary differentiable curve  $\gamma : \mathbb{R} \rightarrow S^{n-1}$ . Now, the tangent vector  $\gamma'(t)$  is tangent to the sphere  $S^{n-1}$  at any point  $\gamma(t)$ . We claim that this tangent drawn at  $\gamma(t)$  is always perpendicular to the position vector  $\gamma(t)$ . This is made trivial by our exercise: the map  $t \mapsto \|\gamma(t)\|^2 = 1$  is a constant map since  $\gamma$  is a curve on the unit sphere. This means that it has zero derivative, forcing  $\langle \gamma(t), \gamma'(t) \rangle = 0$ .

### 3.1. Directional derivatives

**Definition 3.1.1:** Let  $U \subseteq \mathbb{R}^n$  be open, and let  $f : U \rightarrow \mathbb{R}$ . The directional derivative of  $f$  along a direction  $v \in \mathbb{R}^n$  at a point  $a \in U$  is defined by the following limit, if it exists.

$$\nabla_v f(a) = \lim_{h \rightarrow 0} \frac{f(a + hv) - f(a)}{h}.$$

*Example:* Consider

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto \begin{cases} x^3/(x^2 + y^2), & \text{if } (x, y) \neq (0, 0) \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

Note that  $f$  is not differentiable at  $(0, 0)$ . However, all directional derivatives of  $f$  exist at  $(0, 0)$ . Indeed, consider a direction  $(\cos \theta, \sin \theta)$ , and examine the limit

$$\lim_{t \rightarrow 0} \frac{1}{t} [f(t \cos \theta, t \sin \theta) - f(0, 0)] = \cos^3 \theta.$$

**Definition 3.1.2:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable. The gradient of  $f$  is defined as the map

$$\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x \mapsto \left[ \frac{\partial f}{\partial x_i}(x) \right]_i.$$

*Remark:* The gradient at a point  $x \in \mathbb{R}^n$  is thought of as a vector. In contrast, the derivative is thought of as a linear transformation. Otherwise, we see that  $\nabla f(x) = [Df(x)]$ .

**Definition 3.1.3:** Let  $C^1(\mathbb{R}^n)$  be the set of real-valued differentiable functions on  $\mathbb{R}^n$ . Fix a point  $a \in \mathbb{R}^n$ , then fix a tangent vector  $v \in \mathbb{R}^n$ . Then, the map

$$\nabla_v : C^1(\mathbb{R}^n) \rightarrow \mathbb{R}, \quad f \mapsto Df(a)(v)$$

is a linear functional. The quantity  $\nabla_v f$  is called the directional derivative of  $f$  in the direction  $v$  at the point  $a$ .

*Remark:* We can represent  $\nabla_v$  as the operator

$$\nabla_v(\cdot) = D(\cdot)(a)(v) = \sum_i v_i \frac{\partial}{\partial x_i} \Big|_a = v \cdot \nabla(\cdot).$$

**Lemma 3.1.1:** The directional derivatives  $\nabla_v$  form a vector space called the tangent space, attached to the point  $a \in \mathbb{R}^n$ . This can be identified with the vector space  $\mathbb{R}^n$  by the natural map  $\nabla_v \mapsto v$ . The standard basis can be informally denoted by the vectors

$$\nabla_{e_1} := \frac{\partial}{\partial x_1}, \quad \dots, \quad \nabla_{e_n} := \frac{\partial}{\partial x_n}.$$

### 3.2. Differentiation on manifolds \*

**Definition 3.2.1:** A homeomorphism is a continuous, bijective map whose inverse is also continuous.

**Lemma 3.2.1:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be continuous. Denote the graph of  $f$  as

$$\Gamma(f) = \{(x, f(x)) : x \in \mathbb{R}^n\}.$$

Then,  $\Gamma(f)$  is a smooth manifold.

*Proof:* Consider the homeomorphism

$$\varphi : \Gamma(f) \rightarrow \mathbb{R}^n, \quad (x, f(x)) \mapsto x.$$

This is clearly bijective, continuous (restriction of a projection map), with a continuous inverse (from the continuity of  $f$ ). Call this homeomorphism  $\varphi$  a coordinate map on  $\Gamma(f)$ . ■

**Definition 3.2.2:** Let  $f : M \rightarrow \mathbb{R}$  where  $M$  is a smooth manifold, with a coordinate map  $\varphi : M \rightarrow \mathbb{R}^n$ . We say that  $f$  is differentiable at a point  $a \in M$  if  $f \circ \varphi^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable at  $\varphi(a)$ .

**Definition 3.2.3:** Let  $f : M \rightarrow \mathbb{R}$  where  $M$  is a smooth manifold, let  $\varphi : M \rightarrow \mathbb{R}^n$  be a coordinate map, and let  $a \in M$ . Let  $\gamma : \mathbb{R} \rightarrow M$  be a curve such that  $\gamma(0) = a$ , and further let  $\gamma$  be differentiable in the sense that  $\varphi \circ \gamma : \mathbb{R} \rightarrow \mathbb{R}^n$  is differentiable. The directional derivative of  $f$  at  $a$  along  $\gamma$  is defined as

$$\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0} = \lim_{h \rightarrow 0} \left. \frac{f(\gamma(t+h)) - f(\gamma(t))}{h} \right|_{t=0}.$$

Note that we are taking the derivative of  $f \circ \gamma : \mathbb{R} \rightarrow \mathbb{R}$  in the conventional sense.

**Lemma 3.2.2:** Let  $\gamma_1$  and  $\gamma_2$  be two curves in  $M$  such that  $\gamma_1(0) = \gamma_2(0) = a$ , and

$$\left. \frac{d}{dt} \varphi \circ \gamma_1(t) \right|_{t=0} = \left. \frac{d}{dt} \varphi \circ \gamma_2(t) \right|_{t=0}.$$

In other words,  $\gamma_1$  and  $\gamma_2$  pass through the same point  $a$  at  $t = 0$ , and have the same velocities there. Then, the directional derivatives of  $f$  at  $a$  along  $\gamma_1$  and  $\gamma_2$  are the same.

**Definition 3.2.4:** Let  $M$  be a smooth manifold, and let  $a \in M$ . Consider the following equivalence relation on the set of all curves  $\gamma$  in  $M$  such that  $\gamma(0) = a$ .

$$\gamma_1 \sim \gamma_2 \implies \frac{d}{dt}\varphi \circ \gamma_1(t)|_{t=0} = \frac{d}{dt}\varphi \circ \gamma_2(t)|_{t=0}.$$

Each resultant equivalence class of curves is called a tangent vector at  $a \in M$ . Note that all these curves in a particular equivalence class pass through  $a$  with the same velocity vector.

The collection of all such tangent vectors, i.e. the space of all curves through  $a$  modulo the equivalence relation which identifies curves with the same velocity vector through  $a$ , is called the tangent space to  $M$  at  $a$ , denoted  $T_a M$ .

*Remark:* Each tangent vector  $v \in T_a M$  acts on a differentiable function  $f : M \rightarrow \mathbb{R}$  yielding a (well-defined) directional derivative at  $a$ .

$$v : C^1(M) \rightarrow \mathbb{R}, \quad f \mapsto \frac{d}{dt}f(\gamma_{v(t)})|_{t=0}.$$

Thus, the tangent space represents all the directions in which taking a derivative of  $f$  makes sense.

*Remark:* The tangent space  $T_a M$  is a vector space. Upon fixing  $f$ , the map  $Df(a) : T_a M \rightarrow \mathbb{R}$ ,  $v \mapsto v f(a)$  is a linear functional on the tangent space.

*Remark:* Given a tangent vector  $v \in T_a M$ , it can be identified with its corresponding velocity vector in  $\mathbb{R}^n$ . Thus, the tangent space  $T_a M$  can be identified with the geometric tangent plane drawn to the manifold  $M$  at the point  $a$ .

## 4. Mean value theorem

Consider a differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , and fix  $a \in \mathbb{R}^n$ . Define the functions

$$g_i : \mathbb{R} \rightarrow \mathbb{R}, \quad g_i(x) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n).$$

Then, each  $g_i$  is differentiable, with

$$g_i'(x) = \frac{\partial f}{\partial x_i}(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n).$$

By applying the Mean Value Theorem on some interval  $[c, d]$ , we can find  $\alpha \in (c, d)$  such that  $g_i(d) - g_i(c) = g_i'(\alpha)(d - c)$ . In other words,

$$f(\dots, d, \dots) - f(\dots, c, \dots) = \frac{\partial f}{\partial x_i}(\dots, \alpha, \dots)(d - c).$$

**Theorem 4.1:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $a \in \mathbb{R}^n$ . Then,  $f$  is differentiable at  $a$  if all the partial derivatives  $\partial f / \partial x_j$  exist in a neighbourhood of  $a$  and are continuous at  $a$ .

*Proof:* Without loss of generality, let  $m = 1$ . We claim that



$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} \left\| f(a+h) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i \right\| = 0.$$

Examine

$$\begin{aligned} f(a+h) - f(a) &= f(a_1 + h_1, \dots, a_n + h_n) - f(a_1, \dots, a_n) \\ &= f(a_1 + h_1, \dots, a_n + h_n) - f(a_1 + h_1, \dots, a_{n-1} + h_{n-1}, a_n) + \\ &\quad f(a_1 + h_1, \dots, a_{n-1} + h_{n-1}, a_n) - f(a_1 + h_1, \dots, a_{n-1}, a_n) + \\ &\quad \dots \\ &\quad f(a_1 + h_1, a_2, \dots, a_n) - f(a_1, \dots, a_n) \\ &= \frac{\partial f}{\partial x_n}(c_n) h_n + \dots + \frac{\partial f}{\partial x_1}(c_1) h_1. \end{aligned}$$

The last step follows from the Mean Value Theorem. As  $h \rightarrow 0$ , each  $c_i \rightarrow a$ . Thus,

$$\begin{aligned} \frac{1}{\|h\|} \left\| f(a+h) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i \right\| &= \frac{1}{\|h\|} \left\| \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right) h_i \right\| \\ &\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right| \frac{|h_i|}{\|h\|} \\ &\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right|. \end{aligned}$$

Taking the limit  $h \rightarrow 0$ , observe that  $(\partial f / \partial x_i)(c_i) \rightarrow (\partial f / \partial x_i)(a)$  by the continuity of the partial derivatives, completing the proof. ■

**Corollary 4.1.1:** All polynomial functions on  $\mathbb{R}^n$  are differentiable.

**Theorem 4.2:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable with continuous partial derivatives, and let  $a \in \mathbb{R}^n$  be a point of local maximum. Then,  $Df(a) = 0$ .

*Proof:* We need only show that each

$$\frac{\partial f}{\partial x_i}(a) = 0.$$

This must be true, since  $a$  is also a local maximum of each of the restrictions  $g_i$  as defined earlier. ■

## 5. Inverse and implicit function theorems

**Theorem 5.1** (Inverse function theorem): Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be continuously differentiable on a neighbourhood of  $a \in \mathbb{R}^n$ , and let  $\det(Df(a)) \neq 0$ . Then, there exist neighbourhoods  $U$  of  $a$  and  $W$  of  $f(a)$  such that the restriction  $f : U \rightarrow W$  is invertible. Furthermore,  $f^{-1}$  is continuous on  $U$  and differentiable on  $U$ .

**Lemma 5.2:** Consider a continuously differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , and let  $M$  denote the surface defined by the zero set of  $f$ . Then,  $M$  can be represented as the graph of a differentiable function  $h : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  at those points where  $Df \neq 0$ .

*Proof:* Without loss of generality, suppose that  $\partial f / \partial x_n \neq 0$  at some point  $a \in M$ . It can be shown that the map

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x \mapsto (x_1, x_2, \dots, x_{n-1}, f(x))$$

is invertible in a neighbourhood  $W$  of  $a$ , with a continuous and differentiable inverse of the form

$$G : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad u \mapsto (u_1, u_2, \dots, u_{n-1}, g(u)).$$

Since  $F \circ G$  must be the identity map on  $W$ , we demand

$$(x_1, x_2, \dots, x_{n-1}, f(x_1, x_2, \dots, x_{n-1}, g(x))) = (x_1, x_2, \dots, x_{n-1}, x_n).$$

Thus, the zero set of  $f$  in this neighbourhood of  $a$  satisfies  $x_n = 0$ , hence

$$f(x_1, x_2, \dots, x_{n-1}, g(x_1, x_2, \dots, x_{n-1}, 0)) = 0.$$

In other words, the part of the surface  $M$  in the neighbourhood of  $a$  is precisely the set of points

$$(x_1, x_2, \dots, x_{n-1}, g(x_1, x_2, \dots, x_{n-1}, 0)).$$

Simply set

$$h : \mathbb{R}^{n-1} \rightarrow \mathbb{R}, \quad x \mapsto g(x_1, x_2, \dots, x_{n-1}, 0),$$

whence the surface  $M$  is locally represented by the graph of  $h$ . ■

*Remark:* Note that by using

$$f(x_1, \dots, x_{n-1}, h(x_1, \dots, x_{n-1})) = 0$$

on the surface, we can use the chain rule to conclude that for all  $1 \leq i < n$ , we have

$$\frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a) \frac{\partial h}{\partial x_i}(a_1, \dots, a_{n-1}) = 0.$$

**Theorem 5.3** (Implicit function theorem): Let  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  be continuously differentiable in an open set containing  $(a, b)$ , with  $f(a, b) = 0$ . Let  $\det (\partial f^j / \partial x_{n+k}(a, b)) \neq 0$ . Then, there exists an open set  $U \subset \mathbb{R}^n$  containing  $a$ , an open set  $V \subset \mathbb{R}^m$  containing  $b$ , and a differentiable function  $g : U \rightarrow V$  such that  $f(x, g(x)) = 0$ .

*Remark:* The condition on the determinant can be rephrased as  $\text{rank } Df(a, b) = m$ .

**Theorem 5.4:** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be continuously differentiable, and let  $M$  be the surface defined by its zero set. Furthermore, let  $\nabla f(a) \neq 0$  for some  $a \in M$ ; thus,  $M$  can be locally represented by a graph on  $\mathbb{R}^{n-1}$ . Then,  $\nabla f(a)$  is normal to the tangent vectors drawn at  $a$  to  $M$ ; in fact, the perpendicular space of  $\nabla f(a)$  is precisely the tangent space  $T_a M$ .

*Proof:* Consider a tangent vector drawn at  $a$  to  $M$ , represented by the differentiable curve  $\gamma : \mathbb{R} \rightarrow M$ ,  $\gamma(0) = a$ ; note that we use the identification  $\gamma'(0) = v \in \mathbb{R}^n$ . Then, calculate

$$\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0} = Df(\gamma(0))(\gamma'(0)) = Df(a)(v).$$

On the other hand, we have  $f(\gamma(t)) = 0$  identically. Thus,

$$v \cdot \nabla f(a) = Df(a)(v) = 0$$

as claimed. ■

## 6. Taylor's theorem

**Theorem 6.1** (Clairaut): Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  have continuous second order partial derivatives. Then,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

**Theorem 6.2** (Taylor): Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  have continuous second order partial derivatives, and let  $(x_0, y_0) \in \mathbb{R}^2$ . Then, there exists  $\varepsilon > 0$  such that for all  $\|(x - x_0, y - y_0)\| < \varepsilon$ ,

$$\begin{aligned} f(x, y) = & f(x_0, y_0) + \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) \\ & + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (x - x_0)^2 + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} (y - y_0)^2 \\ & + \frac{\partial^2 f}{\partial x \partial y} (x - x_0)(y - y_0) + R(x, y), \end{aligned}$$

where as  $(x, y) \rightarrow (x_0, y_0)$ , the remainder term vanishes as

$$\frac{|R(x, y)|}{\|(x - x_0, y - y_0)\|^2} \rightarrow 0.$$

All partial derivatives here are evaluated at  $(x_0, y_0)$ .

*Proof:* This follows from applying the Taylor's Theorem in one variable to the real function  $g : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto f((1 - t)(x_0, y_0) + t(x, y))$ . ■

## 7. Critical points and extrema

**Definition 7.1:** We say that  $a \in \mathbb{R}^n$  is a critical point of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  if all  $\partial f / \partial x^j = 0$  there.

**Lemma 7.1:** All points of extrema of a differentiable function are critical points.

*Proof:* We already know that  $Df(a) = 0$  where  $a$  is either a point of maximum or minimum. ■

*Example:* In order to find a point of extrema of a  $C^2$ -smooth function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , we first identify a critical point  $(x_0, y_0)$ . Next, we must find a neighbourhood of  $(x_0, y_0)$  which contains no other critical points – to do this, apply Taylor's Theorem. Indeed, we see that

$$f(x, y) = f(x_0, y_0) + A(x - x_0)^2 + 2B(x - x_0)(y - y_0) + C(y - y_0)^2 + R_2.$$

For non-degeneracy of solutions, we demand  $AC - B^2 \neq 0$ , i.e. at  $(x_0, y_0)$ , we want

$$\left[ \frac{\partial^2 f}{\partial x \partial y} \right]^2 \neq \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2}.$$

If  $AC - B^2 > 0$  and  $\partial^2 f / \partial x^2 > 0$ , then we have found a point of minima; if  $\partial^2 f / \partial x^2 < 0$ , then we have found a point of maximum. If  $AC - B^2 < 0$ , then we have found a saddle point.

*Example:* Suppose that we wish to maximize the function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , given an equation of constraint  $g = 0$ , where  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Using the method of Lagrange multipliers, we look for solutions of the system

$$\begin{cases} \nabla f(x,y) + \lambda \nabla g(x,y) = 0 \\ g(x,y) = 0. \end{cases}$$