

# MATH358 - Advanced Calculus

Based on lectures from Winter 2026 by Prof. John Toth.  
Notes by Louis Meunier

## Contents

1 Differentiation .....	2
1.1 Aside on Tangent Planes .....	7
1.2 Clairault's Theorem .....	7
1.3 Inverse Function Theorem .....	10
1.4 Taylor's Theorem in $\mathbb{R}^n$ and Lagrange Multipliers .....	14
1.5 Lagrange Multipliers .....	16

## §1 DIFFERENTIATION

We say  $\Omega \subset \mathbb{R}^n$  a *domain* if it is open and connected.

→**Definition 1.1** (Differentiation): Let  $f = (f_1, \dots, f_m)^T : \Omega \rightarrow \mathbb{R}^m$ ,  $\Omega$  a domain in  $\mathbb{R}^n$  and  $f_j : \Omega \rightarrow \mathbb{R}$ . We say  $f$  *differentiable* at  $x_0 \in \Omega$  if there exists a linear map  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  such that

$$\lim_{x \rightarrow x_0} \frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} = 0.$$

**Remark 1.1:** Note that the first norm on  $\mathbb{R}^m$ , the second on  $\mathbb{R}^n$ .

**Remark 1.2:** In terms of  $\varepsilon, \delta$ , the definition says that  $\forall \varepsilon > 0, \exists \delta > 0$  such that if  $x \in \Omega \cap B(x_0, \delta)$ , then  $\|f(x) - f(x_0) - L(x - x_0)\| < \varepsilon \|x - x_0\|$ .

→**Theorem 1.1:**  $L$  as above is unique if it exists.

PROOF. Suppose  $L_1 \neq L_2$  both satisfy the definition. Then, for all  $\varepsilon > 0$ , there exists  $\delta > 0$  such that if  $0 < \|x - x_0\| < \delta$ , then

$$\begin{aligned} \|(L_1 - L_2)(x - x_0)\| &\leq \|f(x) - f(x_0) - L_1(x - x_0)\| + \|f(x) - f(x_0) - L_2(x - x_0)\| \\ &\leq \varepsilon \|x - x_0\|, \end{aligned}$$

by differentiability (and the previous remark). In particular,  $\|(L_1 - L_2)u\| < \varepsilon$  for all unit vectors  $u$ , which implies  $\|(L_1 - L_2)u\| = 0$  and thus  $L_1 = L_2$ . ■

→**Definition 1.2:** If  $f$  differentiable at  $x_0$ , we'll write  $Df(x_0) = L$  for the *differential* of  $f$  at  $x_0$ .

→**Proposition 1.1:**  $f$  differentiable at  $x_0$  implies  $f$  continuous at  $x_0$ . In fact,  $f$  is Lipschitz at  $x_0$ .

PROOF. Let  $\delta > 0$  such that  $\|x - x_0\| < \delta$  implies  $\|f(x) - f(x_0) - Df(x_0)(x - x_0)\| < \|x - x_0\|$ , which implies

$$\|f(x) - f(x_0)\| \leq \|Df(x_0)(x - x_0)\| + \|x - x_0\| \leq (\|L\| + 1)\|x - x_0\|,$$

which readily proves the statement. ■

→**Proposition 1.2:**  $f$  differentiable at a point  $x_0$  iff each of its component functions are differentiable at  $x_0$ .

→**Definition 1.3:** For  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ , define the *partial derivative*

$$\frac{\partial f_j}{\partial x_i}(x_1, \dots, x_m) := \lim_{h \rightarrow 0} \frac{[f_j(x_1, \dots, x_i + h, \dots, x_m) - f_j(x_1, \dots, x_i, \dots, x_m)]}{h},$$

if the limit exists.

→ **Proposition 1.3:** Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  be differentiable at  $x_0$ . Then,  $\frac{\partial f_j}{\partial x_i}(x_0)$  exists for each  $i = 1, \dots, n$  and  $j = 1, \dots, m$ , and

$$L = Df(x_0) = \left( \begin{array}{c} \frac{\partial f_j}{\partial x_i}(x_0) \\ \vdots \\ \frac{\partial f_j}{\partial x_i}(x_0) \end{array} \right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}.$$

We call this matrix the Jacobian or derivative of  $f$  at  $x_0$ .

**PROOF.** Write  $L = (a_{ji})$  in the standard basis  $e_1, \dots, e_n$  for  $\mathbb{R}^n$ . Let  $\varepsilon > 0$ , fix some  $i$  with  $1 \leq i \leq n$ , and set  $x := x_0 + he_i$ , with  $|h| < \delta$  sufficiently small. By differentiability,

$$\frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} = \left( \sum_{j=1}^m \left[ \frac{f_j(x) - f_j(x_0)}{h} - a_{ji} \right]^2 \right)^{1/2}.$$

Since the limit as  $h \rightarrow 0$  of the above ratio must be zero, the limit of each term in the summation as  $h \rightarrow 0$  must be zero as well (being a sum of nonnegative terms), i.e.

$$\lim_{h \rightarrow 0} \frac{f_j(x) - f_j(x_0)}{h} = a_{ji} \quad \forall j = 1, \dots, m.$$

But the limit on the left is just  $\frac{\partial f_j}{\partial x_i}(x_0)$ , which proves all of the claims in turn. ■

**Remark 1.3:** This proposition says that  $f$  differentiable at  $x_0$  implies  $\frac{\partial f_j}{\partial x_i}(x_0)$  exists for all  $i, j$ . The converse need not be true. Consider

$$f(x, y) := \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & (x, y) \neq 0 \\ 0 & \text{else} \end{cases}.$$

㊂ **Example 1.1:** Another counterexample as in the previous remark is the function

$$f(x, y) := \begin{cases} \frac{x^2 y}{x^2 + y^2} & (x, y) \neq (0, 0), \\ 0 & (x, y) = (0, 0). \end{cases}$$

*Claim 1:*  $f$  continuous at  $(0, 0)$ . We have, for  $(x, y) \neq (0, 0)$ ,

$$\begin{aligned} |f(x, y) - f(0, 0)| &= \left| \frac{x^2 y}{x^2 + y^2} \right| \\ &= \frac{x^2 |y|}{x^2 + y^2} \\ &\leq |y| \rightarrow 0 \text{ as } (x, y) \rightarrow (0, 0), \end{aligned}$$

so we have continuity indeed.

*Claim 2:*  $\partial_x f, \partial_y f$  exist at the origin, and are equal to zero. Note that  $f(x, 0) = 0$  for  $x \neq 0$ , and  $f(0, 0) = 0$ , so it follows that  $\partial_x f(0, 0) = 0$ . Similarly for  $\partial_y f(0, 0)$ .

*Claim 3:*  $f$  is not differentiable at  $(0, 0)$ . Suppose otherwise. Then,  $L = Df(0, 0) = (0, 0)$ , so

$$\begin{aligned} 0 &= \lim_{(x, y) \rightarrow (0, 0)} \frac{|f(x, y) - f(0, 0) - Df(0, 0)(x, y)|}{\|(x, y)\|} \\ &= \lim_{(x, y) \rightarrow (0, 0)} \frac{|f(x, y)|}{\|(x, y)\|} \\ &= \lim_{(x, y) \rightarrow (0, 0)} \frac{x^2 |y|}{(x^2 + y^2) \cdot \sqrt{x^2 + y^2}} \\ &= \lim_{(x, y) \rightarrow (0, 0)} \frac{x^2 |y|}{(x^2 + y^2)^{3/2}}. \end{aligned}$$

Suppose  $y = x$  in the final term (i.e., we approach the limit on a diagonal), and  $x > 0$ , then this ratio simplifies

$$\frac{x^3}{(2x^2)^{3/2}} = \frac{1}{2^{3/2}} \neq 0,$$

so we have a contradiction.

We can get a partial converse, however, if we assume continuity.

↪ **Theorem 1.2:** Let  $f = (f_1, \dots, f_m) : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Suppose each  $\frac{\partial f_j}{\partial x_i}$  is continuous at some  $x^0 \in \Omega$ . Then,  $f$  is differentiable at  $x^0$ .

**PROOF.** We use MVT, and suppose  $n = 2, m = 1$  for simplicity of notation, so that  $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ . We write  $x = (x_1, x_2) \in \Omega, x^0 = (x_1^0, x_2^0)$ . Let  $\varepsilon > 0$ . By assumption, there exists a  $\delta > 0$  such that

$$\|y - x^0\| < \delta \Rightarrow \left| \frac{\partial f}{\partial x_i}(y) - \frac{\partial f}{\partial x_i}(x^0) \right| \leq \frac{\varepsilon}{2}, \quad i = 1, 2.$$

We write

$$\begin{aligned} f(x) - f(x^0) &= f(x_1, x_2) - f(x_1^0, x_2) + f(x_1^0, x_2) - f(x_1^0, x_2^0) \\ (\text{MVT, coordinate-wise}) \quad &= \frac{\partial f}{\partial x_1}(z_1, x_2)(x_1 - x_1^0) + \frac{\partial f}{\partial x_2}(x_1^0, z_2)(x_2 - x_2^0), \end{aligned}$$

for some  $z_1$  between  $x_1$  and  $x_1^0$  and some  $z_2$  between  $x_2$  and  $x_2^0$ . Thus,

$$\begin{aligned} f(x) - f(x^0) - Df(x^0)(x - x^0) &= f(x) - f(x^0) - (\partial_{x_1} f(x^0), \partial_{x_2} f(x^0)) \cdot (x - x^0) \\ &= \frac{\partial f}{\partial x_1}(z_1, x_2)(x_1 - x_1^0) + \frac{\partial f}{\partial x_2}(x_1^0, z_2)(x_2 - x_2^0) \\ &\quad - \frac{\partial f}{\partial x_1}(x^0)(x_1 - x_1^0) - \frac{\partial f}{\partial x_2}(x^0)(x_2 - x_2^0) \\ &= [\partial_{x_1} f(z_1, x_2) - \partial_{x_1} f(x_1^0, x_2^0)](x_1 - x_1^0) \\ &\quad + [\partial_{x_2} f(x_1^0, z_2) - \partial_{x_2} f(x_1^0, x_2^0)](x_2 - x_2^0). \end{aligned}$$

By choice of  $z_1, z_2$  and for  $(x_1, x_2)$  in  $B(x^0, \delta)$ , we know  $(z_1, x_2) \in B(x^0, \delta)$  and  $(x_1^0, z_2) \in B(x^0, \delta)$  as well, so we can appeal to continuity. In addition, it's clear that  $|x_i - x_i^0| \leq \|x - x^0\|$ . Thus, using continuity, we find

$$|f(x) - f(x^0) - Df(x^0)(x - x^0)| \leq \left( \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \right) \|x - x^0\| = \varepsilon \|x - x^0\|,$$

so dividing both sides by  $\|x - x^0\|$  immediately gives the result. ■

→ **Definition 1.4:** Suppose  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  has continuous  $\frac{\partial f}{\partial x_i}$  at all points in  $\Omega$ . Then, we say  $f$  is *continuously differentiable* (in  $\Omega$ ), and we write  $f \in C^1(\Omega)$ .

**Remark 1.4:** Continuity of partial derivatives is sufficient, but not necessary, for differentiability. For instance,

$$f(x, y) = \begin{cases} \frac{x^2 y^2}{x^2 + y^4} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}.$$

On readily computes  $\partial_x f(0, 0) = \partial_y f(0, 0) = 0$ , but along the parabola  $x = t^2, y = t$  ( $t \neq 0$ ),

$$\partial_x f(t^2, t) = \frac{1}{2},$$

so  $\partial_x f$  can't be continuous. However,  $f$  is still differentiable at  $(0, 0)$ : we claim  $L = 0$ , then

$$\frac{|f(x, y) - f(0, 0) - L(x, y)|}{\|(x, y)\|} = \frac{|f(x, y)|}{(x^2 + y^2)^{\frac{1}{2}}} = \frac{x^2 y^2}{(x^2 + y^4)(x^2 + y^2)^{\frac{1}{2}}} \leq \frac{y^2}{|x^2 + y^2|^{\frac{1}{2}}} \leq |y| \underset{(x, y) \rightarrow 0}{\rightarrow} 0.$$

→**Proposition 1.4** (Basic Properties of Differentiation):

1. If  $f, g : \Omega \rightarrow \mathbb{R}^m$  both differentiable at  $x^0 \in \Omega$ , then so is  $F = f + g$ , and

$$D(f + g)(x^0) = Df(x^0) + Dg(x^0).$$

2. If  $f, g : \Omega \rightarrow \mathbb{R}^m$  both differentiable at  $x^0 \in \Omega$ , then so is  $F = fg : \Omega \rightarrow \mathbb{R}$ , and

$$DF(x^0) = f(x^0)Dg(x^0) + g(x^0)Df(x^0).$$

3.  $f, g : \Omega \rightarrow \mathbb{R}$  both differentiable at  $x^0$  with  $g(x^0) \neq 0$ , then so is  $F = \frac{f}{g}$ , and

$$DF(x^0) = \frac{DF(x^0)}{g(x^0)} - \frac{f(x^0)Dg(x^0)}{g^2(x^0)}.$$

4. (Chain Rule) Given  $f : \Omega \subset \mathbb{R}^n \rightarrow \tilde{\Omega} \subset \mathbb{R}^m$  and  $g : \tilde{\Omega} \rightarrow \mathbb{R}^k$ , with  $f$  differentiable at  $x^0$  and  $g$  differentiable at  $y^0 = f(x^0)$ , then  $H = g \circ f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$  is differentiable at  $x^0$ , and

$$DH(x^0) = Dg(y^0) \cdot Df(x^0),$$

in which one should read the “.” as matrix multiplication.

**PROOF.** 1., 2., 3. left as an exercise. We prove 4., the Chain Rule, for it is realistically the most interesting. Set  $L := Dg(y_0) \cdot Df(x_0)$ , and we'll write  $y = f(x)$  (so in particular  $y_0 = f(x_0)$ , as in the statement). We need to show

$$\lim_{x \rightarrow x_0} \frac{\|H(x) - H(x_0) - L(x - x_0)\|}{\|x - x_0\|} = 0.$$

Let us work the numerator:

$$\begin{aligned} H(x) - H(x_0) - L(x - x_0) &= g(y) - g(y_0) - Dg(y_0)Df(x_0)(x - x_0) \\ &= g(y) - g(y_0) - Dg(y_0)(y - y_0) \\ &\quad + Dg(y_0)(y - y_0) - Dg(y_0)Df(x_0)(x - x_0) \\ &= g(y) - g(y_0) - Dg(y_0)(y - y_0) \\ &\quad + Dg(y_0)(f(x) - f(x_0) - Df(x_0)(x - x_0)). \end{aligned}$$

This means

$$\begin{aligned} \|H(x) - H(x_0) - L(x - x_0)\| &\leq \overbrace{\|g(y) - g(y_0) - Dg(y_0)(y - y_0)\|}^{=: (A)} \\ &\quad + \overbrace{\|Dg(y_0)\| \|f(x) - f(x_0) - Df(x_0)(x - x_0)\|}^{=: (B)}. \end{aligned}$$

By differentiability of  $f$  at  $x_0$ ,  $(B) \rightarrow 0$  as  $\|x - x_0\| \rightarrow 0$ . We also have that, since  $f$  differentiable it is Lipschitz continuous, there is some  $C > 0$  such that for  $\|x - x_0\|$  sufficiently small,

$$(A) = \|y - y_0\| \cdot \frac{(A)}{\|y - y_0\|} \leq C\|x - x_0\| \frac{A}{\|y - y_0\|}.$$

By differentiability of  $g$ , the ratio  $\frac{\|A\|}{\|y-y_0\|} \rightarrow 0$  as  $\|y-y_0\| \rightarrow 0$ . By continuity of  $f$ ,  $\|y-y_0\| = \|f(x) - f(x_0)\|$  will become small as  $\|x-x_0\| \rightarrow 0$ , so that we have in all  $\frac{A}{\|x-x_0\|} \rightarrow 0$  as  $\|x-x_0\| \rightarrow 0$ . ■

**Exercise 1.1:** Let  $f$  differentiable in  $\mathbb{R}^2$  and  $g(r, \theta) := (r \cos \theta, r \sin \theta)$  with  $(r, \theta) \in (0, \infty) \times [0, 2\pi)$ . Let  $F(r, \theta) = f(g(r, \theta))$ . Compute  $\frac{\partial F}{\partial \theta}$  and  $\frac{\partial F}{\partial r}$ .

### §1.1 Aside on Tangent Planes

Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  differentiable on  $\Omega$ . Then  $Df(x) = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) =: \nabla f(x)$ , called the *gradient* of  $f$ . Let  $S := \{(x, z) \in \Omega \times \mathbb{R} : z = f(x)\}$  be the *graph* of  $f$ . Then, for  $x^0 \in \mathbb{R}$ ,

$$T_{x^0} S = \left\{ (x, z) \in \mathbb{R}^n \times \mathbb{R} : z = f(x^0) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(x^0)(x_j - x_j^0) \right\}$$

is the *tangent plane* to  $S$  at  $x^0$ .

To see this, let  $v \in \mathbb{R}^n$  be a unit vector and  $x \in \Omega$ . Define  $g(t) := f(x + tv)$  for  $f : \Omega \rightarrow \mathbb{R}$  differentiable (for  $t$  sufficiently small,  $x + tv$  remains in  $\Omega$  by openness). We find

$$g'(t) = \langle \nabla f(x + tv), v \rangle$$

for  $t$  sufficiently small.

→ **Proposition 1.5:** Suppose  $\nabla f(x) \neq 0$ . Then,  $\nabla f(x)$  points in the direction of steepest increase of  $f$ .

**PROOF.** For  $v$  a unit vector, the *directional derivative* in the direction of  $v$  is  $D_v f(x) = \langle \nabla f(x), v \rangle = \|\nabla f(x)\| \cos(\theta)$  where  $\theta$  the angle between  $\nabla f(x)$  and  $v$ . This is maximized when  $\theta = 0$ , i.e. when  $v = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$ . ■

We can rewrite the graph  $S$  as the level set  $\{(x, z) \in \Omega \times \mathbb{R} \mid g(x, z) = 0\}$  where  $g(x, z) := z - f(x)$ . Heuristically,  $\nabla g(x_0, z_0)$  should be *normal* to the surface  $S$  at  $(x_0, z_0)$  (for steepest increase). As such, we define

$$T_{(x_0, z_0)} S := \{\nabla g(x_0, z_0) \cdot (x - x_0, z - z_0) = 0\}.$$

Note that

$$\nabla g(x_0, z_0) = (-\partial_{x_1} f(x_0), \dots, -\partial_{x_n} f(x_0), 1),$$

so that

$$T_{(x_0, z_0)} = \{z - z_0 = \nabla f(x_0) \cdot (x - x_0)\},$$

which gives the definition from above.

### §1.2 Clairault's Theorem

Here, the question is, given  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  twice differentiable, when can we exchange order of second-order partial derivatives, i.e. when is

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}, \quad \forall i, j = 1, \dots, n?$$

We need to establish first a generalization of the mean-value theorem. First, note that if

$$\gamma : (a, b) \rightarrow \mathbb{R}^n, \quad g : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$$

are two differentiable functions with  $\gamma((a, b)) \subset \Omega$ , then by the chain rule, if we put  $H(t) := g(\gamma(t))$ ,

$$\frac{\partial H}{\partial t} = Dg(\gamma(t)) \cdot D\gamma(t), \quad D\gamma(t) = (\gamma'_1(t), \dots, \gamma'_n(t)).$$

→ **Theorem 1.3** (Mean-Value Theorem): Let  $B \subset \mathbb{R}^n$  be a ball and  $f : B \rightarrow \mathbb{R}$  be differentiable for all  $x \in B$ . Then, for any  $x, y \in B$ , there exists  $z \in B$  such that

$$f(x) - f(y) = Df(z) \cdot (x - y).$$

In particular,  $|f(x) - f(y)| \leq \|Df(z)\| \|x - y\|$ .

**PROOF.** Let  $x, y \in B$  fixed and let  $\gamma(t) := tx + (1 - t)y$  for  $t \in [0, 1]$ . We see that  $\gamma(t) \in B$  for all  $t \in [0, 1]$ , and that  $D\gamma(t) = x - y$ . Set  $F(t) := f(\gamma(t))$  (i.e., we restrict  $f$  to its values along the straight line along  $x$  and  $y$ ), noting  $F : \mathbb{R} \rightarrow \mathbb{R}$ . So, by 1-dimensional mean-value theorem, there is some  $t^* \in [0, 1]$  such that

$$\begin{aligned} f(x) - f(y) &= F(1) - F(0) = F'(t^*) = Df\left(\underbrace{t^*x + (1 - t^*)y}_{=: z \in B}\right) \cdot D\gamma(t) \\ &= Df(z) \cdot (x - y). \end{aligned}$$

■

Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  differentiable. Remember that  $Df : \Omega \rightarrow \mathbb{R}^{m \times n}$ .

→ **Definition 1.5:** We say  $f$  twice differentiable at  $x$  if  $Df$  exists locally to  $x$  and  $Df$  is differentiable at  $x$ . We write

$$D^2f = D(Df),$$

and similarly

$$D^k f := D(D^{k-1}f)$$

with an analogous definition.

→ **Definition 1.6:** Given  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , we see that  $f \in C^k(\Omega)$  for  $k \in \mathbb{Z}_+$  if all the partial derivatives to order  $k$  exist and are continuous in  $\Omega$ .

→ **Definition 1.7:** If  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  twice differentiable, the *Hessian matrix* is given by

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

**Exercise 1.2:** Let  $f(x, y) := \begin{cases} \frac{(xy)(x^2-y^2)}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$  and compute  $H_f(x, y)$ .

→ **Theorem 1.4 (Clairault):** Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  be twice differentiable at  $x \in \Omega$ . Then,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \quad \forall i, j = 1, \dots, n.$$

→ **Corollary 1.1:** If  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  are all continuous at  $x \in \Omega$  for  $i, j = 1, \dots, n$ , then  $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ .

**PROOF.** (Of Clairault's) It's enough to consider  $n = 2$ . Fix  $(x, y) \in \Omega$ , and note that for  $s, t \in \mathbb{R}$  sufficiently small,  $(x + s, y + t) \in \Omega$  by openness. Set

$$\begin{aligned} \Delta(s, t) &:= f(x + s, y + t) - f(x, y + t) - f(x + s, y) + f(x, y) \\ &= g_t(x + s) - g_t(x), \quad g_t(u) := f(u, y + t) - f(u, y). \end{aligned}$$

By the mean-value theorem, there is some  $\xi_{s,t}$  between  $x$  and  $x + s$  such that

$$\Delta(s, t) = \frac{\partial g_t}{\partial x}(\xi_{s,t}) \cdot s = \left[ \frac{\partial f}{\partial x}(\xi_{s,t}, y + t) - \frac{\partial f}{\partial x}(\xi_{s,t}, y) \right] s. \quad (\ddagger)$$

By assumption,  $\frac{\partial f}{\partial x}$  is differentiable at  $(x, y)$ , so

$$\frac{\partial f}{\partial x}(z_1, z_2) = \frac{\partial f}{\partial x}(x, y)(z_1 - x) + \frac{\partial^2 f}{\partial x^2}(x, y)(z_2 - y) + E_1(z_1, z_2), \quad (\dagger)$$

where

$$\frac{|E_1(z_1, z_2)|}{\sqrt{(z_1 - x)^2 + (z_2 - y)^2}} \rightarrow 0, \quad \text{as } (z_1, z_2) \rightarrow (x, y).$$

Evaluating  $(\dagger)$  at  $(z_1, z_2) = (\xi_{s,t}, y + t)$  and  $(\xi_{s,t}, y)$ , and plugging into  $(\ddagger)$  yields

$$\Delta(s, t) = \left( \frac{\partial^2 f}{\partial y \partial x}(x, y)t + E_1(\xi_{s,t}, y + t) - E_1(\xi_{s,t}, y) \right) s.$$

Let  $s = t$  and let  $t \rightarrow 0$ . We claim that

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \lim_{s=t \rightarrow 0} \frac{\Delta(s, t)}{st} = \frac{\partial^2 f}{\partial x \partial y}(x, y).$$

The first equality is obvious from the assumptions on the error terms. On the other hand, we can switch the order of the middle terms in  $\Delta(s, t)$  and write

$$\begin{aligned}\Delta(s, t) &= f(x + s, y + t) - f(x, y + t) - f(x + s, y) + f(x, y) \\ &= h_s(y + t) - h_s(y), \quad h_s(u) := f(x + s, u) - f(x, u).\end{aligned}$$

Repeating the same argument as above with  $g_t$ , we get that

$$\Delta(s, t) = \left( \frac{\partial^2}{\partial x \partial y}(x, y)s + E_2(x + s, \eta_{s,t}) - E_2(x, \eta_{s,t}) \right)t,$$

where  $\eta_{s,t}$  lies between  $y$  and  $y + t$ , and

$$|E_2(x + s, \eta_{s,t})| \leq |s^2 + t^2|, \quad |E_2(x, \eta_{s,t})| \leq \sqrt{s^2 + t^2}.$$

Setting  $s = t$  here, we get

$$\lim_{\substack{s, t \rightarrow 0 \\ s=t}} \frac{\Delta(s, t)}{st} = \frac{\partial^2}{\partial x \partial y}(x, y).$$

This proves the claim. ■

### §1.3 Inverse Function Theorem

→ **Theorem 1.5** (In 1D): If  $f : (a, b) \rightarrow (c, d)$  is differentiable with  $f'(x) > 0$ , then there exists  $g : (c, d) \rightarrow (a, b)$  differentiable such that  $y = f(x) \Leftrightarrow x = g(y)$  (i.e.  $x = g(y)$ ).

In higher dimensions, we recall some preliminaries before proving.

→ **Theorem 1.6:** Let  $(X, d)$  a complete metric space and  $f : X \rightarrow X$  a contraction mapping, with  $d(f(x_2), f(x_1)) \leq \alpha d(x, y)$  for all  $x, y \in X$  for some  $0 < \alpha < 1$ . Then, there exists a unique  $x_0 \in X$  such that  $f(x_0) = x_0$ .

We will write  $M_n := \{n \times n \text{ matrices}\} \cong \mathbb{R}^{n^2}$ , and  $\|A\| := \sqrt{\sum_{i,j=1}^n a_{ij}^2}$  where  $A := (a_{ij}) \in M_n$ . We use

$$\mathrm{GL}(n) := \{A \in M_n : \det(A) \neq 0\} = \det^{-1}(\mathbb{R} \setminus \{0\}), \quad \det : M_n \rightarrow \mathbb{R}.$$

Remark that since  $\mathbb{R} \setminus \{0\}$  is open, and the map  $\det$  is continuous (it can be written as a polynomial in the entries  $a_{ij}$ 's of the matrix  $A$ ), we know that  $\mathrm{GL}(n)$  an open subset of  $M_n$ .

Consider the map

$$f : \mathrm{GL}(n) \rightarrow \mathrm{GL}(n), \quad f(A) := A^{-1}.$$

→ **Lemma 1.1:**  $\mathrm{GL}(n) \subset M_n$  open and  $f \in C^k$  for all  $k = 1, 2, \dots$

**PROOF.** We already proved the first statement in our remarks above.

Let  $A(j|i)$  be  $(n - 1) \times (n - 1)$  matrix with its  $j$ th row and  $i$ th columns deleted, then recall

$$\text{adj}(A) = ((-1)^{i+j} \det A(j|i)).$$

By Cramer's formula from linear algebra,

$$f(A) = A^{-1} \frac{1}{\det(A)} \text{adj}(A),$$

which is in  $C^k$  since  $\det(A)$  is a polynomial in the coefficients of  $A$  and  $\det(A) \neq 0$ . ■

**Theorem 1.7** (Inverse Function Theorem): Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  be  $C^1$ . Let  $x_0 \in \Omega$  and assume  $Df(x_0) \in \text{GL}(n)$ . Then, there exist domains  $U$  and  $V$  of  $x_0$  and  $f(x_0)$  resp. such that  $f(U) = V$  and  $f|_U$  has a  $C^1$  inverse map  $f^{-1} : V \rightarrow U$ . Moreover, for any  $y \in V$  and  $x = f^{-1}(y)$ ,  $Df^{-1}(y) = [Df(x)]^{-1}$ .

**Remark 1.5:** By the first lemma above, if  $f \in C^k$ ,  $k \geq 1$ , we get the same regularity for  $f^{-1}$ .

**PROOF.** By translation, it's enough to assume  $x_0 = f(x_0) = y_0 = 0$  and  $Df(x_0) = \text{Id}$  by replacing  $f$  with  $[Df(0)]^{-1}f$ , so we have a mapping

$$f : \Omega \rightarrow \mathbb{R}^n, \quad f(0) = 0, Df(0) = \text{Id}.$$

Fix  $y \in V$  and set

$$g_y(x) := y + x - f(x),$$

remark that

$$g_y(x) = x \Leftrightarrow y = f(x),$$

so it suffices to show  $g_y$  as a mapping  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  is a contraction mapping, and

$$Dg_y(0) = \text{Id} - \text{Id} = 0.$$

If  $f \in C^1(U)$ , then  $g_y \in C^1(U)$  so that  $Dg_y \in C^0(U)$  (similar if  $f \in C^k \Rightarrow g_y \in C^k$ ). Since  $Dg_0 \in C^0(U)$ , there exists some  $\delta > 0$  sufficiently small such that  $\|Dg_0(x)\| \leq \frac{1}{2}$ , for all  $x \in B_\delta(0)$ . By mean-value theorem, there exists some  $z \in B_\delta(0)$  such that

$$\begin{aligned} \|g_0(x)\| &= \left\| g_0(x) - \underbrace{g_0(0)}_{=0} \right\| \\ &\leq \|Dg_0(z)\| \|x\| \\ &\leq \frac{\|x\|}{2} < \frac{\delta}{2}, \end{aligned}$$

which implies we can view

$$g_0 : B_\delta(0) \rightarrow B_{\delta/2}(0).$$

It follows that

$$g_y : B_\delta(0) \rightarrow B_\delta(0), \quad \forall y \in B_{\delta/2}(0),$$

using the fact  $g_y = y + g_0$  and the triangle inequality. By MVT once again for any  $x, x' \in B_\delta(0)$ , there exists  $y \in B_{\delta/2}(0)$  such that

$$\begin{aligned}\|g_y(x) - g_y(x')\| &= \|g_0(x) - g_0(x')\| \\ &\leq \|Dg_0(y)\| \|x - x'\| \\ &\leq \frac{\|x - x'\|}{2}\end{aligned}$$

hence  $g_y : B_\delta \rightarrow B_\delta$  is a contraction mapping. By the fixed-point theorem, there exists a unique point  $x \in B_\delta(0)$  such that  $g_y(x) = x \Leftrightarrow y = f(x)$ . That is, there exists an inverse map  $f^{-1} : B_{\delta/2}(0) \rightarrow B_\delta(0)$ . Moreover, for any  $x, x' \in B_\delta(0)$ ,

$$\begin{aligned}\|x - x'\| &\leq \|f(x) - f(x')\| + \|g_0(x) - g_0(x')\| \\ &\leq \|f(x) - f(x')\| + \frac{1}{2}\|x - x'\|,\end{aligned}$$

i.e.

$$\|x - x'\| \leq 2\|f(x) - f(x')\|.$$

From here, we know that for  $y, y' \in B_{\delta/2}(0)$ ,

$$\|f^{-1}(y) - f^{-1}(y')\| \leq 2\|y - y'\| \Rightarrow f^{-1} \in C^0(B_{\delta/2}(0)).$$

Next, we need to show that  $Df^{-1}(y)$  exists for  $y \in B_{\delta/2}(0)$  for small  $\delta > 0$ . Since  $Df(0) \in \text{GL}(n)$ , we know  $Df(x) \in \text{GL}(n)$  if  $x \in B_\delta(0)$  (possible after shrinking  $\delta > 0$ ). Set

$$W := f^{-1}(B_{\delta/2}(0)),$$

and choose  $R > 0$  suff. small so that

$$\overline{B_R(0)} \subset W.$$

Since  $[Df]^{-1} \in C^0(\overline{B_R})$  and  $\overline{B_R}(0)$  is compact,

$$\|[Df(x)]^{-1}\| \leq K, x \in \overline{B_r(0)}.$$

Then, given  $y, y' \in B_{\delta/2}(0)$  and with  $x = f^{-1}(y), x' = f^{-1}(y')$ , we find

$$\begin{aligned}\frac{\|f^{-1}(y) - f^{-1}(y') - [Df(x')]^{-1}(y - y')\|}{\|y - y'\|} &= \frac{\|x - x' - [Df(x')]^{-1}(f(x) - f(x'))\|}{\|f(x) - f(x')\|} \\ &= \frac{\|x - x'\|}{\|f(x) - f(x')\|} \frac{\|[Df(x')]^{-1}(f(x) - f(x') - Df(x')(x - x'))\|}{\|x - x'\|} \\ &\leq 2K \frac{\|f(x) - f(x') - Df(x')(x - x')\|}{\|x - x'\|},\end{aligned}$$

which converges to zero by differentiability of  $f$ . This proves the claim  $Df^{-1}(y) = [Df(x)]^{-1}$  where  $y = f(x)$ . ■

**Remark 1.6:** The inverse function theorem is *local*. In general we can't expect to find a single global inverse. For instance, let

$$f(x, y) := (e^y \cos(x), e^y \sin(x)).$$

One easily verifies

$$\det(Df(x, y)) = e^{-y} \neq 0.$$

However,

$$f(x + 2k\pi, y) = f(x, y), \forall k \in \mathbb{Z},$$

so there is certainly no hope of a global inverse, for  $f$  is not even injective.

↪ **Theorem 1.8** (Implicit Function Theorem): Let  $F : \Omega \subset \mathbb{R}_x^n \times \mathbb{R}_y^m \rightarrow \mathbb{R}_y^m$  be a  $C^k$  map. Denote  $X = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ , and let  $X_0 = (x_0, y_0) \in \Omega$  with  $F(X_0) = 0$ . Writing  $F = (F_1, \dots, F_m)$ , assume that

$$D_y F(X_0) = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}(X_0)$$

is invertible. Then, there exist neighborhoods  $U$  and  $V$  of  $x_0 \in \mathbb{R}^n$  and  $y_0 \in \mathbb{R}^m$  resp. and a unique  $C^k$  map  $f : U \rightarrow V$  such that

$$F(x, f(x)) = 0, \quad \forall x \in U.$$

In other words, the level set of  $F$  is locally to  $x_0$  the graph of some function  $f$  of the same regularity as  $F$ .

PROOF. Define  $G : \Omega \rightarrow \mathbb{R}^n \times \mathbb{R}^m$  by

$$G(x, y) := (x, F(x, y)).$$

Obviously  $G$  is  $C^k$ . We can apply the inverse function theorem to  $G$  near  $X_0$ ; indeed,

$$DG(X_0) = \begin{pmatrix} I_{n \times n} & 0 \\ D_x F(X_0) & D_y F(X_0) \end{pmatrix},$$

which means

$$\det DG(X_0) = \det D_y F(X_0) \neq 0,$$

by assumption. Thus there exist neighborhoods  $W_1, W_2$  of  $X_0, (x_0, 0)$  respectively (since  $(x_0, 0) = G(X_0)$ ) for which  $G^{-1}$  exists (and is  $C^k$ ) from  $W_2 \rightarrow W_1$ . Then, there are neighborhoods  $U \subset \mathbb{R}^n$  of  $x_0$  and  $V \subset \mathbb{R}^m$  of  $y_0$  such that  $U \times V \subset W_1$ ; set  $Z = G(U \times V)$  (which is also open, with  $Z \subset W_2$ ). Thus we can view

$$G : U \times V \rightarrow Z, \quad G^{-1} : Z \rightarrow U \times V,$$

which are both  $C^k$  maps. Since  $G(x, y) = (x, F(x, y))$ , we know that  $G^{-1}(x, w) = (x, H(x, w))$  for all  $(x, w) \in Z$ . Here,  $H : Z \rightarrow V$  is  $C^k$  since  $G$  is. Thus,

$$(x, F(x, H(x, w))) = G(x, H(x, w)) = (x, w),$$

using the definition of  $G$  in the first equality and the inverse fact in the second line.  
Thus, it follows that

$$F(x, H(x, w)) = w, \quad \forall (x, w) \in Z,$$

thus taking  $f(x) := H(x, 0)$  gives the proof. ■

↪**Corollary 1.2:** Let  $F : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  be a  $C^k(\Omega)$  function. Let  $X = (x', y) \in \mathbb{R}^{n-1} \times \mathbb{R}$  and suppose  $(x'_0, y_0) \in \Omega$  with  $\frac{\partial f}{\partial y}(x'_0, y_0) \neq 0$ . Then, there exist neighborhoods  $U$  and  $V$  of  $x'_0 \in \mathbb{R}^{n-1}$  and  $y_0 \in \mathbb{R}$  and a unique  $C^k(U)$  function  $f : U \rightarrow V$  such that

$$\{F(x', y) = 0\} = \{y = f(x')\}, \quad (x', y) \in U \times V.$$

↪**Theorem 1.9 (Morse Lemma):** Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  be a  $C^k$  function with  $k \geq 3$ . Let  $0 \in \Omega$  be a critical point, i.e.  $\nabla f(0) = 0$ . Assume further  $f(0) = 0$  and  $\nabla^2 f(0)$  is invertible. There exist open sets  $U, V$  of  $0 \in U \cap V$  and  $g \in C^{k-2}(U)$ ,  $g : U \rightarrow V$  with  $g^{-1} : V \rightarrow U$ ,  $g^{-1} \in C^2(V)$ , such that

$$f(g(y)) = y_{\ell+1}^2 + \cdots + y_n^2 - (y_1^2 + \cdots + y_\ell^2),$$

for some  $\ell \in \mathbb{Z} \cap [0, n]$ .

## §1.4 Taylor's Theorem in $\mathbb{R}^n$ and Lagrange Multipliers

Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^{k+1}(\Omega)$ . Let  $x_0 \in \Omega$  and  $|t|$  small. Consider

$$g(t) := f\left(x_0 + t \frac{x - x_0}{\|x - x_0\|}\right), \quad x \neq x_0, \quad g(0) = f(x_0).$$

Since  $x_0 \in \Omega$  and  $\Omega$  open,  $x_0 + t \frac{x - x_0}{\|x - x_0\|} \in \Omega$  for  $t$  sufficiently small. By Taylor in 1-dimension,

$$g(t) = g(0) + g'(0)t + \frac{g''(0)t^2}{2!} + \cdots + \frac{g^{(k)}(0)t^k}{k!} + R_k(g)(t), \quad \frac{|R_k(g)(t)|}{|t|^k} \leq M |t| \text{ as } t \rightarrow 0.$$

**Remark 1.7:** This expansion suggests that one should be able to weaken the  $C^{k+1}$  assumption on  $g$  if we only require

$$\lim_{t \rightarrow 0} \frac{|R_k(g)(t)|}{|t|^k} = 0.$$

Indeed, we only really need  $g \in C^{k-1}$  and  $D(D^{k-1})g(0)$  existing.

To get the Taylor expansion for  $f(x)$  around  $x_0$ , we set  $t = |x - x_0|$  and apply chain rule to  $g(t)$ . First, we compute  $g^{(j)}(0)$ ; we get

$$\begin{aligned} g(0) &= f(x_0), \\ g(\|x - x_0\|) &= f(x). \end{aligned}$$

By chain rule,

$$g'(0) = \sum_{j=1}^n \frac{\partial f}{\partial x_i}(x_0) \frac{x_j - x_j^0}{\|x - x_0\|} = \nabla f(x_0)^T \left( \frac{x - x_0}{\|x - x_0\|} \right).$$

Similarly,

$$\begin{aligned} g''(t) &= \frac{d}{dt} g'(t) \\ &= \frac{d}{dt} \left( \nabla f \left( x_0 + t \frac{x - x_0}{\|x - x_0\|} \right)^T \frac{x - x_0}{\|x - x_0\|} \right) \\ &= \left( D^2 f \left( x_0 + t \frac{x - x_0}{\|x - x_0\|} \right) \frac{x - x_0}{\|x - x_0\|} \right) \cdot \frac{x - x_0}{\|x - x_0\|} \\ &= \left( \frac{x - x_0}{\|x - x_0\|} \right)^T D^2 f \left( x_0 + t \frac{x - x_0}{\|x - x_0\|} \right) \frac{x - x_0}{\|x - x_0\|} \\ &= \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \left( x_0 + t \frac{x - x_0}{\|x - x_0\|} \right) \frac{(x_i - x_i^0)(x_j - x_j^0)}{\|x - x_0\|^2}, \end{aligned}$$

so that

$$g''(0) = \left[ \frac{x - x_0}{\|x - x_0\|} \right]^T D^2 f(x_0) \cdot \left[ \frac{x - x_0}{\|x - x_0\|} \right].$$

Similar computation can be used to compute  $g^{(\ell)}$  for  $\ell$  up to  $k$ . In general, we find that

$$g^{(\ell)}(0) = \sum_{i_1, \dots, i_\ell=1}^n \frac{\partial^\ell f}{\partial x_{i_1} \cdots \partial x_{i_\ell}}(x_0) \frac{x_{i_1} - x_{i_1}^0}{\|x - x_0\|} \cdots \frac{x_{i_\ell} - x_{i_\ell}^0}{\|x - x_0\|}.$$

Moreover, since  $t^\ell = \|x - x_0\|^\ell$ , the term in the Taylor expansion of  $g$  corresponding to  $g^{(\ell)}(0) \frac{t^\ell}{\ell!}$  becomes  $\frac{1}{\ell!}$  times the above, with the denominator  $\|\dots\|$  terms cancelling. This leads to the following theorem.

**Theorem 1.10** (Taylor in  $\mathbb{R}^n$ ): Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^{k+1}(\Omega)$ . Then, for any  $x^0 \in \Omega$ ,

$$f(x) = f(x^0) + \sum_{j=1}^k \frac{1}{j!} \left[ \sum_{i_1, \dots, i_j=1}^n \frac{\partial^j f}{\partial x_{i_1} \cdots \partial x_{i_j}}(x_{i_1} - x_{i_1}^0) \cdots (x_{i_j} - x_{i_j}^0) \right] + R_k(f)(x),$$

where

$$\frac{|R_k(f)(x)|}{\|x - x^0\|^k} \leq M_k \|x - x^0\| \quad \text{as } x \rightarrow x^0.$$

**Remark 1.8:** The  $k$ -th order Taylor polynomial of  $f$  is the quantity

$$P_k(f)(x) = f(x_0) + \sum_{j=1}^k \frac{1}{j!} \left[ \sum_{i_1, \dots, i_j=1}^n \frac{\partial^j f}{\partial x_{i_1} \cdots \partial x_{i_j}} (x_{i_1} - x_{i_1}^0) \cdots (x_{i_j} - x_{i_j}^0) \right]$$

before the remainder, and is a “good approximation” to  $f(x)$  for  $x$  near  $x^0$ , provided

$$\lim_{x \rightarrow x^0} \frac{|R_k(f)(x)|}{\|x - x^0\|^k} = 0.$$

If we just make this assumption on the remainder term, we get the following more general result with weaker assumptions:

↪ **Theorem 1.11:** Let  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  be  $C^{k-1}(\Omega)$  and assume  $D^{k-1}f$  is differentiable at  $x^0 \in \Omega$ . Then the  $k$ th order Taylor expansion for  $f$  about  $x^0$  from the previous theorem still holds, but with now

$$\lim_{x \rightarrow x^0} \frac{|R_k(f)(x)|}{\|x - x^0\|^k} = 0.$$

**Remark 1.9:** Of course, we lose the rate of decay of the remainder, but we gain fewer assumptions needed on  $f$ .

④ **Example 1.2:** Let  $f(x, y) = e^x + \sin(xy)$ . Then

$$f(0, 0) = 1,$$

$$f_x(0, 0) = 1, \quad f_y(0, 0) = 0,$$

$$f_{xx}(0, 0) = 1, \quad f_{xy}(0, 0) = 1$$

$$f_{yx}(0, 0) = 1, \quad f_{yy}(0, 0) = 0.$$

Thus,

$$f(x, y) = 1 + x + \frac{x^2}{2} + xy + R_2(f)(x, y),$$

with

$$\left| \frac{R_2(f)(x, y)}{x^2 + y^2} \right| \leq M \sqrt{x^2 + y^2}.$$

## §1.5 Lagrange Multipliers

The basic problem is, given  $f, g : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$  both  $C^1(\Omega)$ , putting  $\Sigma := \{x \in \Omega : g(x) = 0\}$ , to find

$$\max f|_{\Sigma}, \quad \min f|_{\Sigma}.$$

We call  $g : \Omega \rightarrow \mathbb{R}$  the “constraint function” (i.e., we are doing constrained optimization of  $f$  subject to  $g$ ).

↪**Theorem 1.12:** Let  $f, g$  as above, and assume  $Dg(x^0) \neq 0$  for all  $x^0 \in \Sigma$ . Then, if  $f|_{\Sigma}$  has a max or min at  $x_0 \in \Sigma$ , then there exists  $\lambda \in \mathbb{R}$  such that

$$\nabla f(x_0) = \lambda \nabla g(x_0).$$