

## § 0.1 Digression on the Notation Used

In this chapter (and from now on, mostly), we will use a notation which is called *abstract index notation* with the *Einstein summation convention*. This is usually abbreviated in common literature as the *Einstein index notation*. We will give here a brief explanation of how this notation actually works, and why it's so useful in shortening mathematical expressions. Let  $\mathcal{V}$  be a vector space and  $\mathcal{V}^*$  be the dual space associated with  $\mathcal{V}$ . Then we can write the elements  $v \in \mathcal{V}$ ,  $\varphi \in \mathcal{V}^*$  with respect to some basis as follows

$$\begin{aligned} v &= (v_1, v_2, \dots, v_n) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ \varphi &= (\varphi_1, \varphi_2, \dots, \varphi_n) = (\varphi_1 \quad \varphi_2 \quad \dots \quad \varphi_n) \end{aligned} \tag{1}$$

The first notation is the *ordered tuple* notation, meanwhile the second notation is the usual column/row notation for vectors utilized in linear algebra. In Einstein notation we will have that

$$\begin{aligned} v &\longrightarrow v^i \\ \varphi &\longrightarrow \varphi_i \end{aligned} \tag{2}$$

Where the vector in the space will be indicated with a raised index (*index, not power!*) and the covector with a lower index, where the index will span all the values  $i = 1, \dots, \dim(\mathcal{V})$ .

Let's represent the scalar product in Einstein notation. Let's say that we want to write the scalar product  $\langle v, v \rangle$

$$\langle v, v \rangle = \sum_{i=1}^{\dim(\mathcal{V})} v_i v_i \longrightarrow v_i v^i \tag{3}$$

Note how we have omitted the sum over the repeated index. Now one might ask why it's not written as  $v_i v_i$  (or  $v^i v^i$ , since  $v \in \mathcal{V}$ ), and this is easily explained introducing the matrix  $g_{ij}$ , which is the matrix of the scalar product.

Applying this matrix to  $v^i$  we have  $g_{ij} v^i$ . Note how the low index  $j$  is free and  $i$  is being summed over, hence is a dummy index, this means that the result must have a lower index  $j$  for consistency. So we can write  $v_j = g_{ij} v^i$ , and due to the lower index we already know that this is a covector, i.e. a linear functional  $\mathcal{V} \longrightarrow \mathbb{F}$ , hence it will “eat” a vector and “spew” a scalar (with no indices!). Feeding to this covector the vector  $v^j$  we have finally

$$\langle v, v \rangle = v_j v^j = g_{ij} v^i v^j \tag{4}$$

Where, algebraically we have “omitted” the definition of  $\iota(\cdot) = \langle v, \cdot \rangle$ , which is the canonical isomorphism between  $\mathcal{V}$  and  $\mathcal{V}^*$ .

With this definition we have defined what mathematically are called *musical isomorphisms*, applications which raise and lower indexes. Ironically, this operation is called *index gymnastics*, since we're raising and lowering indices. Thanks to these conventions operations with matrices (and tensors) become much much easier. Let  $a_j^i$  and  $b_j^i$  be two  $n \times n$  matrices over the ordered field  $\mathbb{F}$ . The multiplication of these two matrices will simply be

$$c_j^i = a_k^i b_j^k \tag{5}$$

Note how the  $k$  index gets “eaten”. This mathematical cannibalism is called *contraction of the index*  $k$ . So, the trace for a matrix  $a_j^i$  will be

$$\text{tr}(a) = a_i^i \quad (6)$$

And now comes the tricky part. In order to write determinants we need to define a symbol, the so called *Levi-Civita symbol*,  $\epsilon_{i_1 \dots i_n}$ . In three dimensions it's  $\epsilon_{ijk}$ , and follows the following rules

$$\epsilon_{ijk} = \begin{cases} 1 & \text{even permutation of the indices} \\ -1 & \text{uneven permutation of the indices} \\ 0 & i = j \vee j = k \vee k = i \end{cases} \quad (7)$$

In  $n$  dimensions, it becomes

$$\epsilon_{i_1 \dots i_n} = \begin{cases} 1 & \text{even permutation of indices} \\ -1 & \text{uneven permutation of indices} \\ 0 & i_i = i_j \text{ for some } i, j \end{cases} \quad (8)$$

It's obvious by definition that this weird entity is completely antisymmetrical and unitary, and therefore it's perfect for representing permutations (it's also known as permutation symbol for a reason). Therefore, remembering the definition of the determinant we can write, for an  $n \times n$  matrix  $a_j^i$

$$\det(a) = \epsilon_{i_1 \dots i_n} a^{1i_1} a^{2i_2} \dots a^{ni_n} = \epsilon_{i_1 \dots i_n} g^{ji_1} a_j^1 g^{ki_2} a_k^2 \dots g^{li_n} a_{i_n}^n \quad (9)$$

If  $\dim(\mathcal{V}) = 3$ , we can therefore immediately define the cross product of two vectors as follows

$$\mathbf{c} = \mathbf{v} \times \mathbf{w} \longrightarrow c^i = g^{ij} \epsilon_{jkl} v^k w^l \quad (10)$$

(Note how we had to raise the index  $i$ ).

From now on, we will start to use Greek letters for indices and Latin letters for labels, in order to avoid confusions, simply look again at the formula for the determinant, it's much clearer this way. In fact, letting  $\mu, \nu, \dots$  be our indices and  $i, j, k, \dots$  our labels, we can write, for a matrix  $A_\nu^\mu$

$$\det(A) = A = \epsilon_{\mu_1 \dots \mu_n} g^{\nu\mu_1} g^{\sigma\mu_2} \dots g^{\zeta\mu_n} A_\nu^1 A_\sigma^2 \dots A_\zeta^n \quad (11)$$

See? Much clearer, at least in my opinion.

Now we might want to understand how to write norms with this notation. For the usual Euclidean norm it's quite easy. So we can easily write

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \longrightarrow \sqrt{v_\mu v^\mu} \quad (12)$$

In case we have a vector function  $f^\mu(x^\nu)$ , the following notation will be used

$$\|\mathbf{f}(\mathbf{x})\| \longrightarrow \sqrt{f_\mu f^\mu(x^\nu)} \quad (13)$$

Or, for the sum of two functions  $g^\mu(x^\nu) \pm f^\mu(y^\nu)$

$$\|\mathbf{g}(\mathbf{x}) \pm \mathbf{f}(\mathbf{y})\| \longrightarrow \sqrt{g_\mu g^\mu(x^\nu) + f_\mu f^\mu(y^\nu) \pm 2g_\mu(x^\nu) f^\mu(y^\nu)} \quad (14)$$

Or

$$\|\mathbf{g}(\mathbf{x}) \pm \mathbf{f}(\mathbf{y})\| \longrightarrow \sqrt{g_{\mu\nu}(g^\mu(x^\gamma) \pm f^\mu(y^\gamma))(g^\nu(x^\gamma) \pm f^\nu(y^\gamma))}$$

A shorthand notation can be created by directly using the norm symbol, but with the contracted index in the upper or lower position as follows

$$\|\mathbf{g}(\mathbf{x}) \pm \mathbf{f}(\mathbf{y})\| \longrightarrow \|g^\mu(x^\nu) \pm f^\mu(y^\nu)\|_\mu \quad (15)$$

For  $p$ -norms we have to watch out for a little detail. We have to add a square root in order to “fix” the squaring of every element. So we get

$$\|\mathbf{v}\|_p \longrightarrow \sqrt[p]{(v_\mu)^{\frac{p}{2}}(v^\mu)^{\frac{p}{2}}} = \left((v_\mu)^{\frac{p}{2}}(v^\mu)^{\frac{p}{2}}\right)^{\frac{1}{p}} \quad (16)$$

**Theorem 0.1.**  $((v_\mu v^\mu)^{p/2})^{1/p}$  is wrong

*Proof.* It’s easy to see why it doesn’t work by expanding the sum on  $\mu$

$$\begin{aligned} (v_\mu v^\mu)^p &= (v_1 v^1 + v_2 v^2 + \dots + v_n v^n)^p \\ (v_\mu)^{\frac{p}{2}}(v^\mu)^{\frac{p}{2}} &= ((v_1 v^1)^p + (v_2 v^2)^p + \dots + (v_n v^n)^p) \end{aligned} \quad (17)$$

□

Moreover, it’s time to bring down some formal rules for the usage of this notation

**Theorem 0.2** (Rules for Index Calculus in Einstein Notation). 1. *Free indices must be consistent in both sides of the equation. I.e.  $a_\nu^\mu b_{\mu\gamma\delta} = c_{\nu\gamma\delta}$ .  $a_\nu^\mu b_{\mu\gamma\delta} \neq c_{\nu\delta}^\gamma$ ,  $a_\nu^\mu b_{\mu\gamma\delta} \neq c_{\nu\gamma\sigma}$*

2. *An index can be repeated only two times per factor and must be contracted diagonally. I.e.  $a^\mu b_\mu f_\gamma^\delta = c_\gamma^\delta$  is defined correctly,  $a_\mu b_\mu$ ,  $a^\mu b^\mu$  or  $a^\mu b_\mu f_\gamma^\mu$  are ill defined*

3. *Dummy indices can be replaced at will, since they don’t contribute to the “index equation”*

### §§ 0.1.1 Differential Operators

Differential operators will be defined formally in the next sections, but for now we will simply explain how they actually work with this notation (and what are the advantages of such), alongside the usual boldface notation.

We will begin by defining the derivative along the coordinate vectors (usually indicated with  $x^\mu$ ). We will use the differential operator *del* ( $\partial$ ).

This operator will be used as follows

1. If there is no ambiguity for the coordinate system, the derivative alongside the coordinates  $x^\mu$  will be indicated as  $\partial_\mu$
2. In case of ambiguity, something will be added in order to distinguish the operators. I.e. let  $(x^\mu, y^\nu)$  be our coordinate system, then we will have  $\partial_{x^\mu}$  or  $\partial_{y^\nu}$
3. In every single case, even the differential operator *must follow the index calculus rules*

Now let  $f(x^\mu)$  be some (scalar, there are no free indices) function of the coordinates. The derivative (or gradient, it will soon be defined properly) can be written in various ways. In boldface notation it's usual to indicate this as  $\nabla f$ , which can be translated as follows

$$\nabla f \longrightarrow \partial_\mu f = \frac{\partial f}{\partial x^\mu} = f_{,\mu} \quad (18)$$

Note how in the RHS it's obvious that this quantity must be a vector due to the free index. The last one is the *comma notation* for derivation, used for compacting (even more) the notation (Also check how in the second notation, even if the index is raised, it behaves as a lower index, we will check deeply this part in the section on differential forms).

Now comes the fun part. Higher order derivatives.

For the same function, we can define the *Hessian matrix* (the matrix of second derivatives)  $\mathbf{Hf}$ , simply applying two times the  $\partial$  operator

$$\mathbf{Hf} \longrightarrow \partial_\nu \partial_\mu f = \partial_{\mu\nu}^2 f = \partial_{\mu\nu} f = \frac{\partial^2 f}{\partial x^\mu \partial x^\nu} = f_{,\mu\nu} \quad (19)$$

Derivatives of order  $> 2$  can then be defined recursively.

Now we might ask, what if we have a vector field  $F^\mu$ ? Nothing changes. We simply have to remember to not repeat indices in order not to represent scalar products.

We have  $\mathbf{JF}$  as the *Jacobian matrix* of  $F^\mu$ , basically the derivative matrix which in Einstein notation, as before, has a quite obvious nature

$$\mathbf{JF} \longrightarrow \partial_\nu F^\mu = \frac{\partial F^\mu}{\partial x^\nu} = F^{\mu}_{,\nu} \quad (20)$$

*And so on, and so on...*<sup>1</sup>

Let's now define the divergence and curl operators. Take now a vector field  $g^\mu$ . We then have

$$\begin{aligned} \nabla \cdot \mathbf{g} &\longrightarrow g_{\mu\nu} g^{\mu\delta} \partial_\delta g^\nu = \partial_\mu g^\mu = \frac{\partial g^\mu}{\partial x^\mu} = g^{\mu}_{,\mu} \\ \nabla \times \mathbf{g} &\longrightarrow \epsilon_{\mu\nu\sigma} g^{\nu\delta} \partial_\delta g^\sigma = \epsilon_{\mu\nu\sigma} \partial^\nu g^\sigma = \epsilon_{\mu\nu\sigma} \frac{\partial g^\sigma}{\partial x_\mu} = \epsilon_{\mu\nu\sigma} g^{\sigma,\nu} \end{aligned} \quad (21)$$

And therefore, defining the Laplacian as  $\nabla^2 = \nabla \cdot \nabla$ , we will simply have, for whatever function  $h$

$$\nabla^2 h \longrightarrow g^{\mu\nu} \partial_\nu \partial_\mu h = \partial^\mu \partial_\mu h = \frac{\partial^2 h}{\partial x^\mu \partial x_\mu} = h^{\mu}_{,\mu} \quad (22)$$

Note how the operator  $\partial^\mu$  appears. This can be seen as a derivation along the covector basis ( $x_\mu = g_{\mu\nu} x^\nu$ ).

Now, we can go back to our mathematical rigor.

---

<sup>1</sup>It's quite fun to dive into the dumpster of Einstein notation, isn't it?

## § 0.2 Curves in $\mathbb{R}^n$

**Definition 0.2.1** (Scalar Field). A *scalar field* is a function  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  where  $A$  is an open set

**Definition 0.2.2** (Vector Field). A *vector field* is a function  $f^\mu : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  where  $A$  is an open set

**Definition 0.2.3** (Continuity). A scalar field  $f : A \rightarrow \mathbb{R}$  is said to be continuous in a point  $p^\mu \in A$  if

$$\forall \epsilon > 0 \exists \delta_p : \|x^\mu - p^\mu\| < \delta \implies |f(x^\mu) - f(p^\mu)| < \epsilon \quad (23)$$

A vector field  $f^\mu : A \rightarrow \mathbb{R}^m$  is said to be continuous instead if

$$\forall \epsilon > 0 \exists \delta_p : \|x^\nu - p^\nu\|_\nu < \delta \implies \|f^\mu(x^\nu) - f^\mu(p^\nu)\|_\mu < \epsilon \quad (24)$$

If this function is continuous  $\forall p^\mu \in A$ , then the vector field is said to be part of the space  $C(A)$ , with  $A \subseteq \mathbb{R}^n$

**Definition 0.2.4** (Canonical Scalar Product). Let  $x^\mu, y^\mu \in \mathbb{R}^n$ , the *canonical scalar product* is a bilinear application  $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  where, if the components of the two vectors are defined as  $x^\mu, y^\mu$ , is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x^i y^i \rightarrow x_\mu y^\mu \quad (25)$$

It's easy to see that the canonical scalar product induces the euclidean norm as follows

$$\|\mathbf{v}\| = \|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_\mu v^\mu} \quad (26)$$

**Definition 0.2.5** (Curves in  $\mathbb{R}^n$ ). A *curve* is an application  $\varphi : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$ .

The function  $\varphi^\mu(t) = p^\mu$ , with  $t \in [a, b]$  is called the *parametric representation* of the curve.

Remembering how indexes work in this notation, we already know that this application can be represented with an ordered  $n$ -tuple or a vector in  $\mathbb{R}^n$

**Definition 0.2.6** (Regular Curves). A curve  $\varphi^\mu(t)$  is said to be continuous if all its components are continuous. A curve is said to be regular iff

$$\begin{aligned} \varphi^\mu(t) &\in C^1([a, b]) \\ \sqrt{\varphi_\mu(t) \varphi^\mu(t)} &\neq 0 \quad t \in (a, b) \end{aligned} \quad (27)$$

A curve is said to be *piecewise regular* if it's not regular in  $[a, b]$  but it's regular in a finite number of subsets  $[a_n, b_n] \subset [a, b]$

**Definition 0.2.7** (Homotopy of Curves). Let  $\gamma^\mu, \eta^\mu$  be two curves from a set  $[a, b]$ ,  $[c, d]$  respectively. These two curves are said to be *homotopic* to one another, and it's indicated as  $\gamma^\mu \sim \eta^\mu$  iff

$$\exists h : [c, d] \xrightarrow{\sim} [a, b], h \in C([c, d]), h^{-1} \in C([a, b]), h(s) > h(t) \text{ for } s > t : \eta^\mu = \gamma^\mu \circ h \quad (28)$$

**Definition 0.2.8** (Tangent Vector). The tangent vector of a regular curve is defined as the following vector.

$$T^\mu(t) = \frac{\dot{\gamma}^\mu}{\sqrt{(\dot{\gamma}_\mu \dot{\gamma}^\mu)}} \quad (29)$$

Where with  $\dot{\gamma}^\mu(t)$  we indicate the derivative of  $\gamma^\mu$  with respect to the only variable  $t$ .

**Definition 0.2.9** (Tangent Line). A curve  $\gamma^\mu : [a, b] \rightarrow \mathbb{R}^n$  is said to have *tangent line* at a point  $t_0 \in [a, b]$  if it's regular, therefore the line will have parametric equations

$$p^\mu(t) = \gamma^\mu(t_0) + \dot{\gamma}^\mu(t_0)(t - t_0) \quad (30)$$

**Definition 0.2.10** (Length of a Curve). The *length* of a curve  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  is defined as follows

$$L_\gamma := \int_a^b \sqrt{\dot{\gamma}_\mu(t) \dot{\gamma}^\mu(t)} dt \quad (31)$$

*Remark.* In  $\mathbb{R}^2$ , if a curve is defined in polar coordinates, it will appear as follows

$$\rho = \rho(\theta), \quad \theta \in [\theta_0, \theta_1] \quad (32)$$

Its length will be given from the following integral

$$L_\rho := \int_{\theta_0}^{\theta_1} \sqrt{(\rho'(\theta))^2 + (\rho(\theta))^2} d\theta \quad (33)$$

The graph of a function  $f : [a, b] \rightarrow \mathbb{R}$ ,  $f \in C^1(a, b)$  can also be parametrized from a curve  $\varphi^\mu(t)$ , where

$$\varphi^\mu(t) \rightarrow (t, f(t)) \quad (34)$$

Its length will be then calculated with the following integral

$$L_\varphi := \int_a^b \sqrt{1 + (\dot{f}(x))^2} dx \quad (35)$$

**Definition 0.2.11** (Curviline Coordinate). Let  $\varphi : [a, b] \rightarrow \mathbb{R}^n$ , we can define a function  $s(t)$  as follows

$$s(t) = \int_a^t \sqrt{\dot{\varphi}_\mu(\tau) \dot{\varphi}^\mu(\tau)} d\tau \quad (36)$$

Then

$$ds = \sqrt{\dot{\varphi}_\mu(t) \dot{\varphi}^\mu(t)} dt \quad (37)$$

And the length of a curve can also be indicated as follows

$$L_\varphi = \int_\varphi ds \quad (38)$$

**Definition 0.2.12** (Curvature, Normal Vector). The *curvature* of a curve is defined as follows

$$\kappa(s) = \sqrt{T_\mu(s) T^\mu(s)} = \sqrt{\ddot{\varphi}_\mu(s) \ddot{\varphi}^\mu(s)} \quad (39)$$

(Note that  $\|\varphi'(s)\| = 1$ ) The *normal vector* is similarly defined as

$$N^\mu(s) = \frac{\dot{T}^\mu(s)}{\kappa(s)} \quad (40)$$

**Definition 0.2.13** (Simple Curve, Closed Curve). A *simple curve* is an injective application  $\gamma : [a, b] \rightarrow \mathbb{R}^n$ . A curve is said to be closed iff  $\gamma^\mu(a) = \gamma^\mu(b)$

**Theorem 0.3** (Jordan Curve). Let  $\gamma^\mu$  be a simple and closed curve in  $\mathbb{R}^2$  or  $\mathbb{C}$  (note that  $\mathbb{C} \simeq \mathbb{R}^2$ ), then the set  $\{\gamma\}^c$  is defined as follows

$$\{\gamma\}^c = \{\gamma\}^\circ \cup \text{extr}(\{\gamma\}) \quad (41)$$

Note that  $\{\gamma\} \subset \mathbb{R}^2$  is the image of the application  $\gamma$  and  $\text{extr}(\{\gamma\})$  is the set of points that lay outside of the closed curve.

In  $\mathbb{C}$  everything that was said about curves holds, however one must watch out for the definition of modulus, for a curve  $\gamma^\mu \in \mathbb{C}$  we will have

$$|\dot{\gamma}(t)| = \sqrt{(\Re'(\gamma))^2 + (\Im'(\gamma))^2} = \sqrt{\bar{\gamma}(t)\gamma'(t)} \quad (42)$$

### § 0.3 Differentiability in $\mathbb{R}^n$

**Definition 0.3.1** (Directional Derivative). Let  $A \subseteq \mathbb{R}^n$  be an open set, and  $f : A \rightarrow \mathbb{R}$ . The function is said to be *derivable* with respect to the direction  $v^\mu \in A$  at a point  $p^\mu \in A$ , if the following limit is finite

$$\partial_{v^\mu} f(p^\mu) = \lim_{h \rightarrow 0} \frac{f(p^\mu + hv^\mu) - f(p^\mu)}{h} \quad (43)$$

If  $v^\mu = x^\mu$  then this is called a *partial derivative*, and it will be indicated in the following ways

$$\frac{\partial f}{\partial x^\mu} = \partial_\mu f = \partial_{x^\mu} f \quad (44)$$

**Definition 0.3.2** (Differentiability). A scalar field  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , with  $A$  open, is said to be *differentiable* in a point  $p^\mu \in A$  if and only if there exists a linear application  $a_\mu(p^\mu) = a_\mu$ , such that the following limit is finite

$$\lim_{\sqrt{h_\mu h^\mu} \rightarrow 0} \frac{R(p^\mu + h^\mu)}{\sqrt{h_\mu h^\mu}} = 0 \quad (45)$$

Where we define the function  $R$  as follows

$$R(p^\mu + h^\mu) = f(p^\mu + h^\mu) - (f(p^\mu) + a_\mu h^\mu) \quad (46)$$

This means that

$$f(p^\mu + h^\mu) = f(p^\mu) + a_\mu h^\mu + \mathcal{O}(\sqrt{h_\mu h^\mu}) \quad (47)$$

**Theorem 0.4** (Consequences of Differentiability). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable scalar field in every point of  $A$ , then

1.  $f \in C(A)$
2.  $f$  is differentiable in  $A$ , and  $a_\mu = \partial_\mu$ , where is the vector differential operator, composed by the partial derivatives

3.  $f$  has directional derivatives in  $A$  and the following equation holds

$$\partial_{v^\mu} f(p^\nu) = \partial_\mu f(p^\nu) v^\mu$$

4.  $\partial_\mu f$  indicates the maximum and minimum growth of the function  $f$

5. There exist a tangent hyperplane to the graphic of the function at the point  $(p^\mu, f(p^\mu)) \in \mathbb{R}^{n+1}$  and has the following equation

$$x^{n+1} = f(p^\mu) + \partial_\mu f(p^\nu)(x^\mu - p^\mu)$$

*Proof.* 1.  $f$  differentiable in  $A$  implies  $f \in C(A)$

$$\lim_{\sqrt{h_\mu h^\mu} \rightarrow 0} f(p^\mu + h^\mu) = \lim_{\sqrt{h_\mu h^\mu} \rightarrow 0} (f(p^\mu) + a_\mu h^\mu + \mathcal{O}(\sqrt{h_\mu h^\mu})) = f(p^\mu)$$

2.  $f$  differentiable in  $A$  implies  $f$  derivable in  $A$

$$\partial_i f(p^\mu) = \lim_{h \rightarrow 0} \frac{f(p^\mu + h e^i) - f(p^\mu)}{h} = \lim_{h \rightarrow 0} \frac{a^i h + \mathcal{O}(h)}{h} = a^i \in \mathbb{R}$$

Then

$$R(p^\mu + h^\mu) = f(p^\mu + h^\mu) - f(p^\mu) + \partial_\mu f(p^\nu) h^\mu = \mathcal{O}(\sqrt{h_\mu h^\mu})$$

3.  $\partial_{v^\mu} f = \partial_\mu f v^\mu$

$$\partial_{v^\mu} f(p^\nu) = \lim_{h \rightarrow 0} \frac{f(p^\mu + h v^\mu) - f(p^\mu)}{h} = \lim_{h \rightarrow 0} \frac{h \partial_\mu f(p^\nu) v^\mu + \mathcal{O}(h)}{h} = \partial_\mu f(p^\nu) v^\mu$$

4.  $\partial_\mu f$  indicates the direction of maximum growth.

For Cauchy-Schwartz, we have

$$\sqrt{\partial_{v^\mu} f \partial^{v^\mu} f} = \sqrt{\partial_\mu f v^\mu \partial_\nu f v^\nu} \leq \sqrt{\partial_\mu f \partial^\mu f} \sqrt{v_\nu v^\nu}$$

□

**Theorem 0.5** (Continuous Differentiation). *Let  $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , with  $A$  open. If  $f \in C^1(A)$  (i.e. the derivatives of  $f$  are continuous), then  $f$  is differentiable in  $A$ , the vice versa is also true*

*Proof.* We can write the following equation

$$f(p^\mu + h^\mu) = f(p^1 + h^1, \dots, p^n) - f(p^1, \dots, p^n) + \dots + f(p^1, \dots, p^n + h^n) - f(p^1, \dots, p^n) \quad (48)$$

For Lagrange, we will have

$$f(p^i + h^i) = h^i \partial_i f(p^1, \dots, p^i, \dots, p^n) = h^i \partial_i f(c_i) \quad (49)$$

Therefore

$$\lim_{\sqrt{h_\mu h^\mu} \rightarrow 0} \frac{|f(p^\mu + h^\mu) - f(p^\mu) - \partial_\mu f h^\mu|}{\sqrt{h_\mu h^\mu}} \leq \lim_{h \rightarrow 0} \sum_{i=1}^n |\partial_i f(c_i) - \partial_i f(p^\mu)| \frac{|h^i|}{\sqrt{h_\mu h^\mu}} = 0 \quad (50)$$

Therefore  $\partial_i f(p^\mu)$  is continuous and the function is differentiable. □



**Theorem 0.6** (Differentiability of Vector Fields, Jacobian Matrix). *Let  $f^\mu : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a vector field and  $A$  an open set, then the function  $f^\mu$  is differentiable iff exists a matrix  $J_\nu^\mu \in M_{nm}(\mathbb{R})$  such that*

$$\lim_{\sqrt{h_\mu h^\mu} \rightarrow 0} \frac{\|f^\mu(p^\nu + h^\nu) - f^\mu(p^\nu) - J_\nu^\mu h^\nu\|_\mu}{\sqrt{h_\mu h^\mu}} \quad (51)$$

Or, equivalently

$$f_\mu(p^\nu + h^\nu) = f^\mu(p^\nu) + J_\nu^\mu h^\nu + \mathcal{O}(\sqrt{h_\mu h^\mu}) \quad (52)$$

The then  $J_\nu^\mu$  is the matrix of partial derivatives of the vector field, called the Jacobian matrix of the vector field  $f^\mu$ , and can be calculated as follows

$$J_\nu^\mu(p^\sigma) = \partial_\nu f^\mu(p^\sigma) \quad (53)$$

**Theorem 0.7** (Composite Derivation). *Let  $f^\mu : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  and  $g^\nu : B \subseteq \mathbb{R}^k \rightarrow \mathbb{R}^p$  be two differentiable functions in  $p^\sigma \in A$ ,  $f^\mu(p^\sigma) \in B$  and  $A, B$  open sets, then  $h^\nu = g^\nu \circ f^\mu$  is differentiable, and*

$$\partial_\sigma h^\nu = \partial_\mu g^\nu(f^\mu) \partial_\sigma f^\mu \quad (54)$$

Since  $\mu = 1, \dots, k$ ,  $\nu = 1, \dots, p$ ,  $\sigma = 1, \dots, n$  it's obvious that  $\partial_\sigma h^\nu \in M_{p,n}(\mathbb{R})$ ,  $\partial_\mu g^\nu \in M_{p,k}(\mathbb{R})$ ,  $\partial_\sigma f^\mu \in M_{k,n}(\mathbb{R})$ .

*Proof.* We have that  $(g^\nu \circ f^\mu) = g^\nu(f^\mu(p^\sigma))$ . Then  $g^\nu$  is differentiable at  $f^\mu(p^\sigma)$  if

$$g^\nu(f^\mu(p^\sigma + s^\sigma)) = g^\nu(f^\mu(p^\sigma)) + \partial_\sigma (g^\nu \circ f^\mu)(f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)) + \mathcal{O}(\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu)$$

Since  $f^\mu$  is differentiable, we have

$$\begin{aligned} g^\nu(f^\mu(p^\sigma + s^\sigma)) &= g^\nu(f^\mu(p^\sigma)) + \partial_\sigma (g^\nu \circ f^\mu) \partial_\sigma f^\mu + \\ &+ \partial_\sigma (g^\nu \circ f^\mu) \mathcal{O}(\sqrt{s_\mu s^\mu}) + \mathcal{O}(\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu) \end{aligned}$$

Then, we must prove that

$$\lim_{\sqrt{s_\mu s^\mu} \rightarrow 0} \frac{\partial_\sigma (g^\nu \circ f^\mu) \mathcal{O}(\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu)}{\sqrt{s_\mu s^\mu}} = 0$$

But

$$\frac{\partial_\sigma (g^\nu \circ f^\mu) \mathcal{O}(\sqrt{s_\mu s^\mu})}{\sqrt{s_\mu s^\mu}} \rightarrow 0$$

And

$$\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu \leq \sqrt{\partial^\sigma f_\mu \partial_\sigma f^\mu} \sqrt{s_\mu s^\mu} + \mathcal{O}(\sqrt{s_\mu s^\mu}) \leq C \sqrt{s_\mu s^\mu}$$

Therefore

$$\frac{\mathcal{O}(\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu)}{\sqrt{s_\mu s^\mu}} = \frac{\mathcal{O}(\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu)}{\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu} \frac{\|f^\mu(p^\sigma + s^\sigma) - f^\mu(p^\sigma)\|_\mu}{\sqrt{s_\mu s^\mu}} \rightarrow 0$$

□

## § 0.4 Differentiability in $\mathbb{C}$

**Definition 0.4.1** (Differentiability). A function  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$  with  $G$  open, is said to be *differentiable* or *derivable* at a point  $a \in G$  if exists finite the following limit

$$\left. \frac{df}{dz} \right|_a = f'(a) = \lim_{z \rightarrow a} \frac{f(z) - f(a)}{z - a} \quad (55)$$

As usual, if this holds  $\forall a \in G$ , the function is derivable in  $G$

**Theorem 0.8.** *If  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$  is derivable in  $a \in G$ , then  $f$  is continuous in  $a$*

*Proof.*

$$\lim_{z \rightarrow a} (f(z) - f(a)) = \lim_{z \rightarrow a} (z - a) \lim_{z \rightarrow a} \frac{f(z) - f(a)}{z - a} = 0$$

□

**Theorem 0.9** (Some Simple Rules). *Let  $f, g : G \subset \mathbb{C} \rightarrow \mathbb{C}$*

1.  $(f \pm g)'(z) = f'(z) \pm g'(z)$
2.  $(fg)'(z) = f'(z)g(z) + f(z)g'(z)$
3.  $(f/g)'(z) = f'(z)/g(z) - f(z)g'(z)/g^2(z) \quad \forall z \in G : g(z) \neq 0$
4.  $f(z) = c \implies f'(z) = 0$
5.  $f(z) = z^n \implies f'(z) = nz^{n-1}$

**Theorem 0.10** (Composite Function Derivation). *Let  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$  and  $g : F \subset \mathbb{C} \rightarrow \mathbb{C}$ , where  $f(G) \subset F$ . If  $f$  is derivable in  $a \in G$  and  $g$  is derivable in  $f(a) \in F$ , then  $g \circ f$  is derivable, and its derivative is calculated as follows*

$$\left. \frac{d}{dz} (g \circ f) \right|_a = \left. \frac{dg}{dz} \right|_{f(a)} \left. \frac{df}{dz} \right|_a = g'(f(a))f'(a) \quad (56)$$

*Proof.* Since  $G$  is open,  $\exists B_r(a) \subset G$ . Therefore, taking a sequence  $(z)_n \in B_r(a) : \lim_{n \rightarrow \infty} (z)_n = a$ . Letting  $f(z_n) \neq a$ , we can directly write in the definition of derivative

$$\lim_{n \rightarrow \infty} \frac{(g \circ f)(z_n) - (g \circ f)(a)}{z_n - a} = (g \circ f)'(a) = g'(f(a))f'(a)$$

Thus, rewriting the function inside the limit

$$\frac{(g \circ f)(z_n) - (g \circ f)(a)}{z_n - a} = \frac{(g \circ f)(z_n) - (g \circ f)(a)}{f(z_n) - f(a)} \frac{f(z_n) - f(a)}{z_n - a} \rightarrow 0$$

Since  $f$  is continuous in  $a \in G$

□

**Theorem 0.11** (Inverse Function Derivation). *Let  $f : G \subset \mathbb{C} \xrightarrow{\sim} \mathbb{C}$  be a bijective continuous map, with  $f^{-1}(w) = z$ . If  $f(a) \neq 0$  and it's derivable at that same point, we have*

$$\left. \frac{df^{-1}}{dw} \right|_{f(a)} = \frac{1}{f'(a)} \quad (57)$$

*Proof.* Since  $f$  is bijective and continuous we can write

$$\left. \frac{df^{-1}}{dw} \right|_{f(a)} = \lim_{w \rightarrow f(a)} \frac{f^{-1}(w) - f^{-1}(f(a))}{w - f(a)} = \lim_{z \rightarrow a} \frac{z - a}{f(z) - f(a)} = \frac{1}{f'(a)}$$

□

### §§ 0.4.1 Holomorphic Functions

**Definition 0.4.2** (Holomorphic Function). A function  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$  is said to be *holomorphic* in its domain  $G$  if  $G$  is open, and

$$\forall z \in G \exists \frac{df}{dz} = f'(z) \quad (58)$$

It is indicated as  $f \in H(G)$ . It's easy to demonstrate that this set is a vector space.

**Theorem 0.12** (Cauchy-Riemann Equation). *Let  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$ , where  $G$  is open and  $f \in H(G)$ . Then, if we write  $z = x + iy$*

$$\begin{cases} \Re(f(z)) = u(x, y) \\ \Im(f(z)) = v(x, y) \end{cases} \quad (59)$$

*We have that the function is holomorphic if and only if*

$$\begin{cases} \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} = 0 \\ \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = 0 \end{cases} \quad (60)$$

*Alternatively, it can be written as follows*

$$\frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} = 0 \quad (61)$$

**Definition 0.4.3** (Wirtinger Derivatives). Before demonstrating the previous theorem, we define the *Wirtinger derivatives* as follows.

Let  $z \in \mathbb{C}$ ,  $z = x + iy$  and  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$ .

$$\begin{cases} \frac{\partial f}{\partial z} = \partial f(z) = \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) f(z) \\ \frac{\partial f}{\partial \bar{z}} = \bar{\partial} f(z) = \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) f(z) \end{cases} \quad (62)$$

Then, the Cauchy-Riemann equations will be equivalent to the following equation

$$\frac{\partial f}{\partial \bar{z}} = \bar{\partial} f(z) = 0 \quad (63)$$

*Proof.* Let  $f(z) = u(x, y) + iv(x, y) : G \subset \mathbb{C} \rightarrow \mathbb{C}$  be a differentiable function in a point  $z_0$ , then as we defined, we have that  $f \in H(B_\epsilon(z_0))$ , and therefore

$$\left. \frac{df}{dz} \right|_{z_0} = \lim_{h \rightarrow 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

And, therefore, along the imaginary axis and the real axis, we have

$$\begin{aligned} \lim_{\Re(h) \rightarrow 0} \frac{f(z_0 + \Re(h)) - f(z_0)}{\Re(h)} &= \left. \frac{\partial f}{\partial x} \right|_{z_0} \\ \lim_{\Im(h) \rightarrow 0} \frac{f(z_0 + i\Im(h)) - f(z_0)}{i\Im(h)} &= \frac{1}{i} \left. \frac{\partial f}{\partial y} \right|_{z_0} = -i \left. \frac{\partial f}{\partial y} \right|_{z_0} \end{aligned}$$

Due to the continuity of the derivative ( $f \in H(B_\epsilon(z_0))$ ) we must have an equality between these limits

$$\frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} = 2 \frac{\partial f}{\partial \bar{z}} = 0, \therefore f \in H(B_\epsilon(z_0)) \implies \frac{\partial f}{\partial \bar{z}} = 0$$

But, since  $f(z) = u(x, y) + iv(x, y)$ , we will have that

$$\begin{aligned} \frac{\partial f}{\partial \bar{z}} &= \frac{\partial}{\partial \bar{z}} (u(x, y) + iv(x, y)) = \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) (u(x, y) + iv(x, y)) \\ &= \frac{1}{2} \left( \frac{\partial u}{\partial x} + i \frac{\partial u}{\partial y} + i \frac{\partial v}{\partial x} - \frac{\partial v}{\partial y} \right) = 0 \\ \therefore \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} &= i \frac{\partial v}{\partial x} + i \frac{\partial u}{\partial y} \end{aligned}$$

Rewriting the previous equation in a system, we immediately get back the Cauchy-Riemann equations

$$\begin{aligned} \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} &= 0 \\ \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} &= 0 \end{aligned}$$

□

**Definition 0.4.4** (Whole Function). A function  $f : \mathbb{C} \rightarrow \mathbb{C}$  is said to be *whole* iff  $f \in H(\mathbb{C})$

**Definition 0.4.5** (Singular Point). Let  $f : G \subset \mathbb{C} \rightarrow \mathbb{C}$  be function such that if  $D = B_\epsilon(z_0) \setminus \{z_0\}$  and  $f \in H(D)$ , then  $z_0$  is said to be a *singular point* of  $f$

For functions  $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{C}$  every theorem already stated for curves in  $\mathbb{R}^n$  with  $n = 2$  holds, since  $\mathbb{C} \simeq \mathbb{R}^2$ . The only thing that should be checked thoroughly is that

$$f(t) = \begin{pmatrix} \Re(f(t)) \\ \Im(f(t)) \end{pmatrix} \in \mathbb{C}|_{\mathbb{R}}$$

Is written as

$$f(t) = \Re(f(t)) + i\Im(f(t)) \in \mathbb{C}$$

## § 0.5 Surfaces

**Definition 0.5.1** (Regular Surface). Let  $K \subset \mathbb{R}^2$ ,  $K = \overline{E}$  where  $E$  is an open and connected subset. A *regular surface* in  $\mathbb{R}^3$  is an application

$$r^\mu : K \longrightarrow \mathbb{R}^3$$

Such that

1.  $r^\mu \in C^1(K)$ , i.e.  $\exists \partial_\nu r^\mu \in C(K)$
2.  $r^\mu$  is injective in  $K$
3.  $\text{rank}(\partial_\nu r^\mu) = 2$

The image  $\text{Im}(r^\mu) = \Sigma \subset \mathbb{R}^3$  is then defined by the following parametric equations

$$r^\mu(u, v) = \begin{cases} x(u, v) = r^1(u, v) \\ y(u, v) = r^2(u, v) \\ z(u, v) = r^3(u, v) \end{cases} \quad (64)$$

The third condition can be rewritten as follows

$$\epsilon^\mu_{\nu\sigma} \partial_1 r^\nu \partial_2 r^\sigma = \epsilon^\mu_{\nu\sigma} \partial_1 r^\nu \partial_2 r^\sigma \neq 0 \quad \forall (u, v) \in K^\circ \quad (65)$$

*Remark.* A function  $f \in C^1(K)$  defines automatically a surface with parametric equations  $r^\mu(u, v) = (u, v, f(u, v))$ . This surface is always regular since  $\epsilon_{\mu\nu\sigma} \partial_1 r^\nu \partial_2 r^\sigma = (-2u, -2v, 1) \neq 0 \quad \forall (u, v) \in K$

**Definition 0.5.2** (Coordinate Lines). The curves obtained fixing one of the two variables are called *coordinate lines* in the surface  $\Sigma$ . We have therefore, for a parametric surface  $r^\mu(u, v)$  and two fixed values  $\tilde{u}, \tilde{v} \in I \subset \mathbb{R}$

$$\begin{aligned} x_1^\mu(t) &= r^\mu(t, \tilde{v}) \\ x_2^\mu(t) &= r^\mu(\tilde{u}, t) \end{aligned} \quad (66)$$

*Example 0.5.1.* The sphere centered in a point  $p_0^\mu \in \mathbb{R}^3$ ,  $p_0^\mu = (x_0, y_0, z_0)$  with radius  $R \geq 0$  has the following parametric equations

$$\begin{aligned} x &= x_0 + R \sin(u) \cos(v) \\ y &= y_0 + R \sin(u) \sin(v) \\ z &= z_0 + R \cos(u) \end{aligned} \quad (67)$$

With  $(u, v) \in [0, \pi] \times [0, 2\pi]$ . It's a regular surface, since

$$\|\epsilon^\mu_{\nu\sigma} \partial_1 r^\nu \partial_2 r^\sigma\|_\mu = R^2 \sin(u) > 0 \quad \forall (u, v) \in [0, \pi] \times [0, 2\pi]$$

**Definition 0.5.3** (Curve on a Surface). Let  $\gamma : [a, b] \subset \mathbb{R} \longrightarrow K \subset \mathbb{R}^3$  be a regular curve, and  $\mathbf{r} : K \longrightarrow \Sigma$ , with the following parametric equations

$$\gamma^\mu(t) = \begin{cases} u = u(t) \\ v = v(t) \end{cases} \quad (68)$$

The regular curve  $p^\mu(t) = r^\mu(u(t), v(t))$  has  $\text{Im } p^\mu \subset \Sigma$ . If it passes for a point  $p_0^\mu = (u_0, v_0)$  it has tangent line

$$p^\mu(t) = p_0^\mu + \dot{r}^\mu(t)(t - t_0) = p_0^\mu + \partial_u r^\mu(u(t), v(t))\dot{u}(t) + \partial_v r^\mu(u(t), v(t))\dot{v}(t) \quad (69)$$

The line is contained inside the following plane

$$\det \begin{pmatrix} (x - x_0) & (y - y_0) & (z - z_0) \\ \partial_1 r^1 & \partial_1 r^2 & \partial_3 r^1 \\ \partial_2 r^1 & \partial_2 r^2 & \partial_3 r^3 \end{pmatrix} \quad (70)$$

For a *cartesian surface*, i.e. the surface generated from the graph of a function  $f(x, y)$ , the tangent plane will be

$$z = f(x_0^\mu) + \partial_\mu f(x_0^\nu)(x^\mu - x_0^\mu) \quad (71)$$

**Definition 0.5.4** (Normal Vector). The *normal vector* to a surface  $\Sigma$ ,  $n^\mu(u, v)$  is the vector

$$n^\mu(u, v) = \frac{1}{\sqrt{\epsilon_{\nu\sigma}^\mu \epsilon_\mu^{\delta\gamma} \partial_1 r^\nu \partial_2 r^\sigma \partial_1 r_\delta \partial_2 r_\gamma}} \epsilon_{\nu\sigma}^\mu \partial_u r^\nu \partial_v r^\sigma \quad (72)$$

For a cartesian surface we have

$$n^\mu(x, y) = \frac{1}{\sqrt{1 + \partial_\mu f \partial^\mu f}} \begin{pmatrix} -\partial_1 f \\ -\partial_2 f \\ 1 \end{pmatrix}^\mu \quad (73)$$

**Definition 0.5.5** (Implicit Surface). Let  $F : A \subset \mathbb{R}^3 \rightarrow \mathbb{R}$  be a function such that  $F \in C^1(A)$ , letting  $\Sigma := \{x^\mu \in \mathbb{R}^3 \mid F(x^\mu) = 0\}$ . If  $x_0^\nu \in \Sigma$  and  $\partial_\mu F(x_0^\nu) \neq 0$ ,  $\Sigma$  coincides locally to a cartesian surface, and the equation of the tangent plane at the point  $x_0^\nu$  is the following

$$\partial_\mu F(x_0^\nu)(x^\mu - x_0^\mu) = 0 \quad (74)$$

**Definition 0.5.6** (Metric Tensor). Let  $ds$  be the curviline coordinate of some curve  $\gamma^\mu$  inside a regular surface  $\Sigma$ . Then we have that

$$s^\mu(t) = r^\mu(u(t), v(t)) \quad (75)$$

And therefore

$$ds^2 = dr_\mu dr^\mu = \partial_1 r^\mu \partial_1 r_\mu (dx^1)^2 + 2\partial_1 r^\mu \partial_2 r_\mu dx^1 dx^2 + \partial_2 r^\mu \partial_2 r_\mu (dx^2)^2 \quad (76)$$

In compact form, we can write

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (77)$$

And, the *metric tensor*  $g_{\mu\nu}$  can defined as follows

$$g_{\mu\nu} = \partial_\mu r^\sigma \partial_\nu r_\sigma \quad (78)$$

Or, in matrix notation

$$g_{\mu\nu} = \begin{pmatrix} \partial_1 r^\mu \partial_1 r_\mu & \partial_1 r^\mu \partial_2 r_\mu \\ \partial_2 r^\mu \partial_1 r_\mu & \partial_2 r^\mu \partial_2 r_\mu \end{pmatrix}_{\mu\nu} \quad (79)$$

In usual mathematical notation we have

$$g_{\mu\nu} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}_{\mu\nu} \quad (80)$$

And it's called the *first fundamental quadratic form* in the language of differential geometry. Then, we can write

$$ds^2 = E (dx^1)^2 + 2F dx^1 dx^2 + G (dx^2)^2 \quad (81)$$

## § 0.6 Optimization

### §§ 0.6.1 Critical Points

**Theorem 0.13** (Fermat). *Let  $p^\nu \in A$  be a point of local minimal or maximal for the function  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  with  $f \in C^1(A)$ . If  $f$  is differentiable in  $p^\nu$  we have*

$$\partial_\mu f(p^\nu) = 0 \quad (82)$$

*The point  $p^\nu$  satisfying this condition is then called a stationary point or a critical point for the function  $f$*

*Proof.* Let  $v^\mu \in A$  be a direction. The function  $g(t) = f(p^\mu + tv^\mu)$  has a point of local maximal or minimal for  $t = 0$ . Then

$$F'(0) = \partial_{v^\mu} f(p^\nu) = \partial_\mu f(p^\nu) v^\mu = 0 \implies \partial_\mu f(p^\nu) = 0 \quad (83)$$

□

**Definition 0.6.1** (Hessian Matrix). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , and let  $f \in C^1(A)$ , then we define the *Hessian matrix* as the matrix of the second partial derivatives of the function  $f$

$$\partial_\mu \partial_\nu f(x^\gamma) = \partial_{\mu\nu} f(x^\gamma) \begin{pmatrix} \partial_{11} f & \cdots & \partial_{1n} f \\ \vdots & \ddots & \vdots \\ \partial_{n1} f & \cdots & \partial_{nn} f \end{pmatrix}_{\mu\nu} (x^\gamma) \quad (84)$$

**Theorem 0.14** (Schwarz). *Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2(A)$ , then*

$$\partial_{\mu\nu} f = \partial_{\nu\mu} f \quad (85)$$

**Definition 0.6.2** (Nature of Critical Points). Let  $p^\gamma$  be a critical point for a function  $f \in C^1(A)$ . Then

1.  $\partial_{\mu\nu} f(p^\gamma)$  is definite positive, then  $p^\gamma$  is a local minimum
2.  $\partial_{\mu\nu} f(p^\gamma)$  is definite negative, then  $p^\gamma$  is a local maximum
3.  $\partial_{\mu\nu} f(p^\gamma)$  is indefinite, then  $p^\gamma$  is a saddle point

**Theorem 0.15.** *Here is a list of some rules in order to determine the definition of the matrix  $\partial_{\mu\nu} f$ . Let  $v^\mu \in A$  be a direction, and  $p^\gamma \in A$  a critical point of the function  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  then*

1. If  $\partial_{\mu\nu} f(p^\gamma) v^\mu v^\nu > 0 \quad \forall v^\mu \in A \implies \partial_{\mu\nu} f(p^\gamma)$  positive definite
2. If  $\partial_{\mu\nu} f(p^\gamma) v^\mu v^\nu < 0 \quad \forall v^\mu \in A \implies \partial_{\mu\nu} f(p^\gamma)$  negative definite
3. If  $\partial_{\mu\nu} f(p^\gamma) v^\mu v^\nu \geq 0 \quad \forall v^\mu \in A \implies \partial_{\mu\nu} f(p^\gamma)$  semi-positive definite
4. If  $\partial_{\mu\nu} f(p^\gamma) v^\mu v^\nu \leq 0 \quad \forall v^\mu \in A \implies \partial_{\mu\nu} f(p^\gamma)$  semi-negative definite
5. If  $v^\mu \neq w^\mu$  are two directions, and  $\partial_{\mu\nu} f(p^\gamma) v^\mu v^\nu > 0 \wedge \partial_{\mu\nu} f(p^\gamma) w^\mu w^\nu < 0 \implies \partial_{\mu\nu} f(p^\gamma)$  indefinite

**Theorem 0.16** (Sylvester's Criteria). Let  $A_\nu^\mu \in M_{nn}(\mathbb{R})$ , and  $(A_k)_\nu^\mu$  be the reduced matrix with order  $k \leq n$ , then

1.  $\det_{\mu\nu} ((A_k)_\nu^\mu) > 0 \implies A_\nu^\mu$  positive definite
2.  $(-1)^k \det_{\mu\nu} ((A_k)_\nu^\mu) > 0 \implies A_\nu^\mu$  negative definite
3. If  $\det_{\mu\nu} ((A_{2k})_\nu^\mu) < 0$  or if  $\det_{\mu\nu} ((A_{2k+1})_\nu^\mu) < 0 \wedge \det_{\mu\nu} ((A_{2n+1})_\nu^\mu) > 0$  for  $k \neq n$ , then  $A_\nu^\mu$  is indefinite

**Theorem 0.17** (Compact Weierstrass). Let  $f : K \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ ,  $f \in C(K)$ , with  $K$  a compact set, then

$$\exists p^\mu, q^\mu \in K : \min_K(f) = f(p^\mu) \leq f(x^\mu) \leq \max_K(f) = f(q^\mu) \quad \forall x^\mu \in K \quad (86)$$

*Proof.* Being  $K$  a compact set, we have that every sequence  $(p^\mu)_n$  converges inside the set, therefore, letting  $(p^\mu)_n$  being a minimizing sequence for  $f$ . Then there exist a converging subsequence  $(p^\mu)_{n_k}$  such that

$$f(p_{n_k}^\mu) \rightarrow f(p^\mu)$$

But, since  $(p^\mu)_n$  is a minimizing sequence, we have

$$f(p^\mu) = \min_K(f)$$

By definition of minimizing sequence. Analogously, one can define a maximizing sequence and obtain the same result for the maximum of the function in  $K$   $\square$

**Theorem 0.18** (Closed Weierstrass). Let  $f : L \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ . If  $L = \bar{L}$  and  $f \in C(L)$  is a coercitive function, i.e.

$$\lim_{\sqrt{x_\mu x^\mu} \rightarrow \infty} f(x^\mu) = +\infty \quad (87)$$

Then

$$\exists x^\mu \in L : \min_L(f) = f(x^\mu) \quad (88)$$

*Proof.* Let  $(p^\mu)_n$  be a minimizing sequence for  $f$  in  $L$ . If this sequence wasn't limited, we would have that  $\sqrt{(p^\mu)_n (p^\mu)_n} \rightarrow \infty$ , and therefore

$$\inf_L(f) = \lim_{n \rightarrow \infty} f(p_n^\mu) = +\infty \quad \nexists$$

Therefore  $(p^\mu)_n$  must be limited, and the proof is the same as in the case of a compact set.  $\square$



**Theorem 0.19** (Topology and Functions). *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C(\mathbb{R}^n)$ . Then*

$$\begin{aligned} \{x^\mu \in \mathbb{R}^n \mid f(x^\mu) < a \in \mathbb{R}\} \\ \{x^\mu \in \mathbb{R}^n \mid f(x^\mu) > b \in \mathbb{R}\} \end{aligned} \quad (89)$$

*Are open sets in  $\mathbb{R}^n$  with the standard topology, and*

$$\begin{aligned} \{x^\mu \in \mathbb{R}^n \mid f(x^\mu) \leq a \in \mathbb{R}\} \\ \{x^\mu \in \mathbb{R}^n \mid f(x^\mu) \geq b \in \mathbb{R}\} \end{aligned} \quad (90)$$

*Are closed sets*

### §§ 0.6.2 Convexity and Implicit Functions

**Definition 0.6.3** (Convex Set). A set  $A \subset \mathbb{R}^n$  is said to be *convex* if

$$\lambda x^\mu + (1 - \lambda)y^\mu \in A \quad \forall x^\mu, y^\mu \in A, \forall \lambda \in [0, 1] \quad (91)$$

Analogously, a function  $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$  is said to be *convex*, if

$$f(\lambda x^\mu + (1 - \lambda)y^\mu) \leq \lambda f(x^\mu) + (1 - \lambda)f(y^\mu) \quad \forall x^\mu, y^\mu \in A, \forall \lambda \in [0, 1] \quad (92)$$

The function  $f$  is also known as a *sublinear* function

Also, the set

$$E_f = \{(x^\mu, \lambda) \in A \times \mathbb{R} \mid f(x^\mu) \leq \lambda\} \quad (93)$$

Is convex

**Theorem 0.20** (Convexity). *Let  $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ .*

1.  $f$  convex in  $A \implies f \in C(A)$
2.  $f$  differentiable in  $A \implies f$  convex  $\iff f(x^\mu) \geq f(p^\mu) + \langle \nabla f(p^\mu), x^\mu - p^\mu \rangle$
3.  $f \in C^2(A) \implies f$  convex  $\iff \partial_{\mu\nu} f(x^\gamma)$  is positive semidefinite

**Definition 0.6.4** (Matrix Infinite Norm). Let  $A_\nu^\mu(x^\gamma) \in \mathcal{V} \rightarrow M_{mn}(\mathbb{F})$ , where  $\dim(\mathcal{V}) = n$ . We can define a norm for this space as follows

$$\|A_\nu^\mu\|_\infty = \sqrt{m} \sqrt{\max_\mu \sup_{x^\gamma \in \mathcal{V}} A_\nu^{(\mu)} A_{(\mu)}^\nu(x^\gamma)} \quad (94)$$

**Theorem 0.21** (Average Value). *Let  $f^\mu : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ , with  $f \in C^1(A)$ ,  $A$  an open set and  $K \subset A$  a compact convex subset, then*

$$\|f^\mu(x^\nu) - f^\mu(y^\nu)\|_\mu \leq \|\partial_\nu f^\mu\|_\infty \|x^\nu - y^\nu\|_\nu \quad (95)$$

*Proof.* Let  $r^\nu(t) = (1 - t)y^\nu + tx^\nu$  be a smooth parametrization of a segment connecting the two points  $x^\nu, y^\nu$ , then

$$\begin{aligned} \|f^\mu(r^\nu(1)) - f^\mu(r^\nu(0))\|_\mu^2 &\leq \partial^\nu f_\mu \partial_\nu f^\mu(r^\nu(t)) \leq \sup_{r^\gamma} (\partial^\nu f_\mu \partial_\nu f^\mu(r^\gamma)) \|x^\nu - y^\nu\|_\nu^2 \\ &\leq m \max_\mu \sup_\gamma \left( \partial^\nu f_{(\mu)} \partial_\nu f^{(\mu)}(x^\gamma) \right) \|x^\nu - y^\nu\|_\nu^2 \end{aligned}$$

Therefore

$$\begin{aligned} \|f^\mu(x^\nu) - f^\mu(y^\nu)\|_\mu &\leq \sqrt{m} \sqrt{\max_\mu \sup_\gamma (\partial^\nu f_{(\mu)} \partial_\nu f^{(\mu)})} \|x^\nu - y^\nu\|_\nu = \\ &= \|\partial_\nu f^\mu\|_\infty \|x^\nu - y^\nu\|_\nu \quad \forall x^\nu, y^\nu \in K \end{aligned}$$

□

**Theorem 0.22** (Implicit Functions, Dini). *Let  $f^\mu : A \subseteq \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ , where  $f^\mu \in C^1(A)$ . Also let  $(x_0^\nu, y_0^\gamma) \in A$  such that*

$$\begin{aligned} f^\mu(x_0^\nu, y_0^\gamma) &= 0 \\ \det_{\mu\nu} \left( \frac{\partial f^\mu}{\partial y^\gamma} \right) &\neq 0 \end{aligned}$$

Then

$$\exists B_\epsilon(x_0^\nu) = I \subset \mathbb{R}^m, \quad B_\epsilon(y_0^\gamma) = J \subset \mathbb{R}^n : f^\mu(x^\nu, y^\gamma) = 0 \quad \forall (x^\nu, y^\gamma) \in I \times J$$

Has a unique solution  $y^\gamma = g^\gamma(x^\nu) \in J$ , with  $g^\gamma \in C^1(I)$ , and

$$\frac{\partial g^\gamma}{\partial x^\nu} = - \left( \frac{\partial f^\mu}{\partial y^\gamma} \right)^{-1} \frac{\partial f^\mu}{\partial x^\nu} \quad (96)$$

*Proof.* Let  $B_\mu^\gamma = \left( \partial_{y_0^\gamma} f^\mu \right)^{-1}$ , then we know that

$$f^\mu(x^\nu, y^\gamma) = 0 \iff B_\mu^\gamma f^\mu(x^\nu, y^\sigma) = 0 \iff G^\gamma(x^\nu, y^\sigma) = y^\gamma - B_\mu^\gamma f^\mu(x^\nu, y^\sigma) = 0$$

We have therefore

$$\begin{aligned} G^\gamma(x^\nu, g^\sigma(x^\nu)) &= g^\gamma(x^\nu) - B_\mu^\gamma f^\mu(x^\nu, g^\sigma(x^\nu)) = g^\gamma(x^\nu) \quad \forall x^\nu \in \overline{B}_r(x_0^\nu) = I \\ \frac{\partial G^\gamma}{\partial y^\sigma} &= \delta_\sigma^\gamma - B_\mu^\gamma \frac{\partial f^\mu}{\partial y^\sigma} \\ \frac{\partial G^\gamma}{\partial y_0^\sigma} &= \delta_\sigma^\gamma - B_\mu^\gamma \frac{\partial f^\mu}{\partial y^\sigma} = \delta_\sigma^\gamma - \delta_\sigma^\gamma = 0 \end{aligned}$$

Now take  $(X, d) = (C(I, J), \|\cdot\|_\infty)$ , with  $J = \overline{B}_\epsilon(y_0^\gamma)$ , and define an application  $H : X \longrightarrow X$  such that

$$H^\gamma(w^\sigma(x^\nu)) = G^\gamma(x^\nu, w^\sigma(x^\nu))$$

We need to demonstrate that this application is a contraction, i.e. that  $\exists! g^\gamma(x^\nu) : f^\mu(x^\nu, g^\gamma(x^\nu)) = 0 \quad \forall (x^\nu, y^\gamma) \in I \times J$

$$\begin{aligned} \|H^\gamma(w^\sigma(x^\nu)) - y_0^\gamma\|_\gamma &= \|G^\gamma(x^\nu, w^\sigma(x^\nu)) - y_0^\gamma\|_\gamma \leq \\ &\leq \|G^\gamma(x^\nu, w^\sigma(x^\nu)) - G^\gamma(x^\nu, y_0^\sigma)\|_\gamma + \|G^\gamma(x^\nu, y_0^\sigma) - G^\gamma(x_0^\nu, y_0^\sigma)\|_\gamma \leq \\ &\leq \left\| \frac{\partial G^\gamma}{\partial y^\sigma} \right\|_\infty \|w^\sigma(x^\nu) - y_0^\sigma\|_\sigma + \|G^\gamma(x^\nu, y_0^\sigma) - G^\gamma(x_0^\nu, y_0^\sigma)\|_\gamma \leq \epsilon \end{aligned}$$

Since  $\|G^\gamma(x^\nu, y_0^\gamma) - G^\gamma(x_0^\nu, y_0^\gamma)\|_\gamma \leq \epsilon/2$  and  $\|w^\sigma(x^\nu) - y_0^\sigma\|_\sigma \leq \epsilon$ ,  $\forall (x^\nu, y^\gamma) \in I \times J$ , we have

$$\left\| \frac{\partial G^\gamma}{\partial y^\sigma} \right\|_\infty \leq \frac{1}{2}$$

Therefore

$$\|H^\gamma(w^\gamma(x^\nu)) - H^\gamma(v^\gamma(x^\nu))\|_\gamma \leq \frac{1}{2}\|w^\gamma(x^\nu) - v^\gamma(x^\nu)\|_\gamma$$

I.e.  $H$  is a contraction in  $C(I, J)$ .

Due to the differentiability of  $f^\mu$  we can write

$$\begin{aligned} \forall \epsilon > 0 \exists \eta_\epsilon : \|h^\nu\|_\nu, \|k^\gamma\|_\gamma \leq \eta_\epsilon \implies & \|f^\mu(x^\nu, g^\gamma(x^\nu) + k^\gamma) - f^\mu(x^\nu, g^\gamma(x^\nu)) - \partial_{x^\nu} f^\mu h^\nu - \partial_{y^\gamma} f^\mu k^\gamma\|_\mu \leq \\ & \leq \epsilon(\|h^\nu\|_\nu + \|k^\gamma\|_\gamma) \end{aligned}$$

Letting  $k^\gamma = g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu)$  we have by definition

$$f^\mu(x^\nu + h^\nu, g^\gamma(x^\nu) + k^\gamma) - f^\mu(x^\nu, g^\gamma(x^\nu)) = 0$$

And therefore, putting  $\partial_{y^\gamma} f^\mu = \delta_\gamma^\mu$

$$\|g^\gamma(x^\nu + h^\nu) + \partial_{x^\nu} f^\mu h^\nu\|_\mu \leq \epsilon \left( \|h^\nu\|_\nu + \|g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu)\|_\gamma \right)$$

Letting  $\epsilon = 1/2$  we have  $\|g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu)\|_\gamma \leq \eta_{1/2}$ , and we have

$$\begin{aligned} \|g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu)\|_\gamma & \leq \|g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu) - \partial_{x^\nu} f^\gamma h^\nu\|_\gamma + \|\partial_{x^\nu} f^\mu h^\nu\|_\gamma \\ & \leq \frac{1}{2} \left( \|h^\nu\|_\nu + \|g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu)\|_\gamma \right) + \|\partial_{x^\nu} f^\mu\|_\infty \|h^\nu\| \end{aligned}$$

Which implies

$$\|g^\gamma(x^\nu + h^\nu) - g^\gamma(x^\nu)\|_\gamma \leq \|h^\nu\|_\nu (1 + 2\|\partial_{x^\nu} f^\mu\|_\infty)$$

Which implies that  $g^\gamma(x^\nu)$  is continuously differentiable in  $I$ . Whenever  $\partial_{y^\gamma} f^\mu \neq \delta_\gamma^\mu$  we can find a transformed function  $\tilde{f}^\mu$  such that  $\partial_{y^\gamma} \tilde{f}^\mu = \delta_\gamma^\mu$   $\square$

### §§ 0.6.3 Lagrange Multipliers

**Definition 0.6.5** (Vinculated Critical Points). Let  $f : K \subset A \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$  with  $A$  open and  $f \in C^1(A)$ . Let  $\partial K = \bigcup_{k \leq n} \gamma_k$  with  $\gamma_k : [a_k, b_k] \rightarrow \mathbb{R}^2$ . The critical points of  $f|_{\partial K}$  are found in  $P_{ik} = \gamma_k(t_i) \in \partial K$ , for which

$$\frac{d}{dt} f(P_{ik}) = 0$$

**Definition 0.6.6** (Argmax, Argmin). Let  $f^\mu : A \rightarrow \mathbb{R}^n$  a function which reach its maximum in  $x_i^\nu \in A$   $i = 1, \dots, m$  and its minimum at  $y_j^\nu \in A$   $j = 1, \dots, k$ . Then we can define

$$\begin{aligned} \text{Arg max}_A(f) &:= \{x_1^\nu, \dots, x_m^\nu\} \\ \text{Arg min}_A(f) &:= \{y_1^\nu, \dots, y_k^\nu\} \end{aligned} \tag{97}$$

**Theorem 0.23** (Lagrange Multipliers). Let  $f, g : A \rightarrow \mathbb{R}$ ,  $f, g \in C^1(A)$ ,  $A \subseteq \mathbb{R}^n$  open, and  $\mathcal{M} = \{x^\mu \in A \mid g(x^\mu) = 0\}$  and let  $x_0^\mu \in \mathcal{M} : \partial_m u g(x_0^\mu) \neq 0$ , then  $x_0^\mu \in \text{Arg max}_{\mathcal{M}} f \vee \text{Arg min}_{\mathcal{M}} f$  if it's a free critical point of the Lagrangian

$$\mathcal{L}(x^\mu, \lambda) = f(x^\mu) - \lambda g(x^\mu) \quad (x^\mu, \lambda) \in A \times \mathbb{R} \tag{98}$$

I.e.  $\exists \lambda_0 \in \mathbb{R} : (x_0^\mu, \lambda_0)$  solves

$$\begin{cases} \partial_\mu f(x^\nu) = \lambda \partial_\mu g(x^\nu) \\ g(x^\mu) = 0 \end{cases} \quad (99)$$

Or, that

$$\text{rank} \begin{pmatrix} \partial_\mu f(x_0^\nu) \\ \partial_\mu g(x_0^\nu) \end{pmatrix} = 1$$

*Proof.* Let  $\partial_n g \neq 0$ , then we can see  $\mathcal{M}$  as a graph of a regular implicit function of  $g$ ,  $h : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}$ , where

$$g(x^\mu, h(x^\mu)) = 0 \quad \forall x^\mu \in B_r(\tilde{x}_0^\mu) \subset \mathbb{R}^{n-1}$$

Letting  $\varphi : (-\epsilon, \epsilon) \longrightarrow B_r(x_0^\mu)$  a smooth curve, such that  $\varphi^\mu(0) = x_0^\mu$ , we have that  $\psi^\nu(t) = (\varphi^\mu(t), h(t)) \in \mathcal{M}$  is the parameterization of a smooth curve that passes through  $x_0^\mu \in \mathcal{M}$ . We have

$$\begin{aligned} \frac{d}{dt} f(\psi^\nu(0)) &= \partial_\mu f \dot{\phi}^\mu(0) + \partial_n f \dot{h}(\phi^\mu(0)) = \partial_\nu f(x_0^\mu) s^\nu \\ \frac{d}{dt} g(\psi^\nu(0)) &= \partial_\nu g(x_0^\mu) s^\nu \end{aligned}$$

With  $s^\nu = \dot{\psi}^\nu(0)$ , therefore  $\partial_\nu f|_{\psi^\nu(0)} \partial_\nu g$  □

**Theorem 0.24** (Generalized Lagrange Multiplier Method). *Let  $f, g_i : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ ,  $0 < i < n$ ,  $f, g_i \in C^1(A)$  with  $A$  an open set, let  $\mathcal{M} := \{x^\nu \in A \mid g(x^\nu) = 0\}$ . Take  $x_0^\nu \in \mathcal{M}$  such that*

$$\text{rank} \partial_\nu g^\mu(x_0^\gamma) = k$$

*Then  $x_0^\nu$  is a critical point for  $f|_{\mathcal{M}}$ , and it's a free critical point for the Lagrangian  $\mathcal{L}$*

$$\mathcal{L}(x^\gamma, \lambda^\mu) = f(x^\gamma) - \lambda_\nu g^\nu(x^\gamma)$$

I.e.  $\exists (x_0^\gamma, \lambda_0^\nu) \in A \times \mathbb{R}$  solution of the system

$$\begin{cases} \partial_\nu f(x^\gamma) = \partial_\nu g^\mu(x^\gamma) \lambda^\nu \\ g^\nu(x^\gamma) = 0 \end{cases}$$

Alternatively, one can check that

$$\text{rank}(A) = \begin{pmatrix} \partial_\mu f(x_0^\gamma) \\ \partial_\nu g^\mu(x^\gamma) \end{pmatrix} = k$$