Notes on the Calculus with
Functions of Several Variables - Part 1

Matteo Barigozzi[*]

Continuity

A general function of several variables is defined on an open set U as

$$\mathbf{f} : U \subset \mathbb{R}^n \to \mathbb{R}^m \quad \text{s.t.} \quad \mathbf{f} : \mathbf{x} \equiv (x_1 \dots x_n) \mapsto (f_1(x_1 \dots x_n) \dots f_m(x_1 \dots x_n))$$

For simplicity we assume $m = 1$.

Definition 1 (Distance in $\mathbb{R}^n$) *Since $\mathbb{R}^n$ is a metric space we can define the Euclidean distance between two point $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ as*

$$d(\mathbf{x}, \mathbf{y}) \equiv ||\mathbf{x} - \mathbf{y}|| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

This is usually called "norm".

Definition 2 (Limit) *Given $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, consider a point $\mathbf{x}_0 \in \bar{U}$ and $b \in \mathbb{R}$, then*

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = b \quad \textit{if} \quad \forall \, \{\mathbf{x}_m\} \subset U \; \textit{s.t.} \; \lim_{m \to \infty} \mathbf{x}_m = \mathbf{x}_0 \; \Rightarrow \; \lim_{m \to \infty} f(\mathbf{x}_m) = b$$

---

[*] matteo.barigozzi@gmail.com

Theorem 1 (Existence of the Limit) *Given* $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ *open, consider a point* $\mathbf{x}_0 \in \bar{U}$ *and* $b \in \mathbb{R}$, *then*

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b \Leftrightarrow \forall\,\varepsilon > 0 \;\; \exists\,\delta > 0 \;\; s.t. \;\; \forall\,\mathbf{x} \in U \;\; ||\mathbf{x} - \mathbf{x}_0|| < \delta \;\Rightarrow\; ||f(\mathbf{x}) - b|| < \varepsilon$$

Theorem 2 (Properties of the Limit) *Given* $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ *open, a point* $\mathbf{x}_0 \in \bar{U}$ *and* $b, b_1, b_2 \in \mathbb{R}$, *then*

$$1) \quad \lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b_1 \quad and \quad \lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b_2 \;\Rightarrow\; b_1 = b_2$$

$$2) \quad \lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b \;\Rightarrow\; \lim_{\mathbf{x}\to\mathbf{x}_0} cf(\mathbf{x}) = cb \;\; \forall\, c \in \mathbb{R}$$

$$3) \quad \lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b_1 \quad and \quad \lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x}) = b_2 \;\Rightarrow\; \lim_{\mathbf{x}\to\mathbf{x}_0} (f+g)(\mathbf{x}) = b_1 + b_2$$

$$4) \quad \lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b_1 \quad and \quad \lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x}) = b_2 \;\Rightarrow\; \lim_{\mathbf{x}\to\mathbf{x}_0} (fg)(\mathbf{x}) = b_1 b_2$$

$$5) \quad \lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = b \neq 0 \;\Rightarrow\; \lim_{\mathbf{x}\to\mathbf{x}_0} \left(\frac{1}{f}\right)(\mathbf{x}) = \frac{1}{b}$$

Definition 3 (Continuity) *Given* $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ *open, consider a point* $\mathbf{x}_0 \in U$, *then* f *is continous at* $\mathbf{x}_0$ *if*

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$$

If f *is continous for any* $\mathbf{x}_0 \in U$ *we say that* $f \in \mathcal{C}^0(U)$.

Theorem 3 (Properties of Continous Functions) *Given* $f : U \to \mathbb{R}$ *and* $g : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ *open and a point* $\mathbf{x}_0 \in U$ *such that* f *and* g *are continous at* $\mathbf{x}_0$, *then*

1. cf *is continous at* $\mathbf{x}_0$, *for any* $c \in \mathbb{R}$;
2. $f + g$ *and* fg *are continous at* $\mathbf{x}_0$;
3. *if* $f(\mathbf{x}_0) \neq 0$, *then* $1/f$ *is continous at* $\mathbf{x}_0$.

Theorem 4 *Consider* $\mathbf{f} : U \to \mathbb{R}^m$*,* $U \subset \mathbb{R}^n$ *open, and* $\mathbf{g} : V \to \mathbb{R}^p$*,* $V \subset R^m$ *open such that* $\mathbf{f}(U) \subset V$ *and* $\mathbf{f} \circ \mathbf{g} : U \to \mathbb{R}^p$*. If* $\mathbf{f}$ *is continous at* $\mathbf{x}_0 \in U$ *and* $\mathbf{g}$ *continous at* $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0) \in V$*, then this implies that* $\mathbf{f} \circ \mathbf{g}$ *is continous at* $\mathbf{x}_0$*.*

Example Consider the function

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

it is not defined in $(0,0)$, can we compute the $\lim_{(x,y)\to(0,0)} f(x,y)$? Remember that $(x,y) \to (0,0)$ is equivalent to say that $||\mathbf{x}|| \to 0$, then

$$|f(x,y)| = \left|\frac{x^2 y}{x^2 + y^2}\right| = \frac{|x|^2|y|}{|x^2 + y^2|} = \frac{|x|^2|y|}{||\mathbf{x}||^2} \leq \frac{||\mathbf{x}||^2||\mathbf{x}||}{||\mathbf{x}||} \to 0$$

This implies that

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0$$

Example Consider the function

$$f(x,y,z) = \frac{x^3 - 2y^3 + z^3}{x^2 + y^2 + z^2}$$

which is not defined in $(0,0,0)$, however we can proceed as in the previous example and we have

$$|f(x,y,z)| \leq \frac{|x|^3 + |2y^3| + |z^3|}{||\mathbf{x}||^2} \leq \frac{4||\mathbf{x}||^3}{||\mathbf{x}||^2} \to 0$$

therefore

$$\lim_{(x,y,z)\to(0,0,0)} f(x,y,z) = 0$$

Theorem 5 *Given $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, consider a point $\mathbf{x}_0 \in U$, then, for any $\mathbf{v} \in \mathbb{R}^n$*

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = b \Leftrightarrow \lim_{t \to 0} f(t\mathbf{v} + \mathbf{x}_0) = b \quad t \in \mathbb{R}$$

To prove that a limit does not exist it is enough to find two directions in $\mathbb{R}^n$ along which the limits are different.

Example Consider the function

$$f(x, y, z) = \frac{x^3 - 2y^2 + z^3}{x^2 + y^2 + z^2}$$

which is not defined in $(0, 0, 0)$, and the limit for $(x, y, z) \to (0, 0, 0)$ does not exist, indeed choose $\mathbf{v}_1 = (1, 0, 0)$ and $\mathbf{v}_2 = (0, 1, 0)$, then

$$\lim_{t \to 0} f(t\mathbf{v}_1) = \lim_{t \to 0} f(t, 0, 0) = 0 \quad \text{but} \quad \lim_{t \to 0} f(t\mathbf{v}_2) = \lim_{t \to 0} f(0, t, 0) = -2$$

Differentiability

Remember the case of functions $f : (a, b) \to \mathbb{R}$, where $(a, b) \in \mathbb{R}$. Then for any $x \in (a, b)$ we define the derivative of f in x_0 as

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{(x - x_0)}$$

if this limit exists the f is differentiable in x_0. In general we can write a first-order approximation of f in a neighborhood of x_0 as

$$f(x) = f(x_0) + \alpha(x - x_0) + R_\alpha((x - x_0))$$

where $\alpha \in \mathbb{R}$ and the last term is the residual that contains terms of order higher than one. Thus

$$R_\alpha((x - x_0)) = f(x) - f(x_0) - \alpha(x - x_0)$$

The residual is defined by

$$\lim_{x \to x_0} \frac{R_\alpha((x - x_0))}{(x - x_0)} = 0$$

which is equivalent to

$$\lim_{x \to x_0} \frac{R_\alpha((x - x_0))}{(x - x_0)} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{(x - x_0)} - \alpha = 0$$

This implies that $\alpha = f'(x_0)$ and it is the angular coefficient of the line tangent to f in x_0. Therefore we can write the following linear approximation

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o((x - x_0)) \quad \text{s.t.} \quad \lim_{x \to x_0} \frac{o((x - x_0))}{(x - x_0)} = 0$$

Notice that we can intepret the derivative as a linear transformation A such that, once x_0 is fixed, $A((x - x_0)) = f'(x_0)(x - x_0)$.

Definition 4 *Given $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, consider a point $\mathbf{x}_0 \in U$, then f is differentiable at $\mathbf{x}_0$ if there exists a linear transformation A such that*

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{||f(\mathbf{x}) - f(\mathbf{x}_0) - A(\mathbf{x} - \mathbf{x}_0)||}{||(\mathbf{x} - \mathbf{x}_0)||} = 0$$

If this limit exists then A is unique. We can write $A = Df(\mathbf{x}_0)$ and we call it the total differential of f. By multplying both sides by $||(\mathbf{x} - \mathbf{x}_0)||$ we get the first order linear approximation

$$f(\mathbf{x}) = f(\mathbf{x} - \mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(||\mathbf{x} - \mathbf{x}_0||) \quad \text{s.t.} \quad \lim_{\mathbf{x} \to \mathbf{x}_0} \frac{o(||\mathbf{x} - \mathbf{x}_0||)}{||(\mathbf{x} - \mathbf{x}_0||} = 0$$

Notice that we can interpret $Df(\mathbf{x})$ in two ways: as a function or as a linear operator

$$\begin{aligned} Df : U \in \mathbb{R}^n &\rightarrow L(\mathbb{R}^n, \mathbb{R}) \\ \mathbf{x} &\mapsto Df(\mathbf{x}) \end{aligned}$$

$$\begin{aligned} Df(x) : \mathbb{R}^n &\rightarrow \mathbb{R} \\ \mathbf{h} &\mapsto Df(\mathbf{x})\mathbf{h} \end{aligned}$$

Once $\mathbf{x}$ is fixed, then $Df(\mathbf{x})$ is a linear transformation and its matrix representation is an n-dimensional object.

All the usual properties of derivatives are satisfied for two differentiable functions f and g:

1) $D(f+g)(\mathbf{x}) = Df(\mathbf{x}) + Dg(\mathbf{x})$

2) $D(\alpha f)(\mathbf{x}) = \alpha Df(\mathbf{x}) \quad \alpha \in \mathbb{R}$

Theorem 6 (Chain Rule) *Consider* $\mathbf{f} : U \rightarrow \mathbb{R}^m$, $U \subset \mathbb{R}^n$ *open,* $\mathbf{f}$ *differentiable at* $\mathbf{x}_0 \in U$, *and given* $\mathbf{g} : V \rightarrow \mathbb{R}^p$, $V \subset R^m$ *open, such that* $\mathbf{f}(U) \subset V \subset$ *and* $\mathbf{g}$ *differentiable at* $\mathbf{f}(\mathbf{x}_0) \in V$. *The function defined as* $\mathbf{F} : U \rightarrow \mathbb{R}^p$ *s.t.* $\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$ *is differentiable at* $\mathbf{x}_0$ *and*

$$D\mathbf{F}(\mathbf{x}_0) = D\mathbf{g}(\mathbf{f}(\mathbf{x}_0))D\mathbf{f}(\mathbf{x}_0)$$

We now need to find an explicit form for the total differential.

Definition 5 (Partial Derivative) *Consider* $\mathbf{f} : U \rightarrow \mathbb{R}^m$, $U \subset \mathbb{R}^n$ *open, and a point* $\mathbf{x} \in \mathbb{R}^n$, *then for any direction defined by a component of the standard basis* $\mathbf{e}_i$, *we can define the i-th partial derivative of* f *in* $\mathbf{x} = (x_1 \dots x_n)$ *as*

$$D_{\mathbf{e}_i} f(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t} = \lim_{t \to 0} \frac{f(x_1 \dots x_i + t \dots x_n) - f(x_1 \dots x_i \dots x_n)}{t} = \frac{\partial f}{\partial x_i}(\mathbf{x}) \quad t \in \mathbb{R}$$

Theorem 7 *Consider* $f : U \rightarrow \mathbb{R}$, $U \subset \mathbb{R}^n$ *open, and assume that* f *is differentiable at* $\mathbf{x} \in U$, *then*

$$Df(\mathbf{x})\mathbf{e}_i = \frac{\partial f}{\partial x_i}(\mathbf{x}) \quad i = 1, \ldots, n$$

Proof: fix i, given that f is differentiable

$$f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x}) = Df(\mathbf{x})(t\mathbf{e}_i) + o(t) = tDf(\mathbf{x})\mathbf{e}_i + o(t)$$

Now by dividing by t and taking the limit for $t \to 0$ we have

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t} = Df(\mathbf{x})\mathbf{e}_i$$

Q.E.D.

Now, given the previous theorem, we have the matrix representation of the total differential which we call gradient:

$$\nabla f(\mathbf{x}) = (Df(\mathbf{x})\mathbf{e}_1 \ldots Df(\mathbf{x})\mathbf{e}_n) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}) \ldots \frac{\partial f}{\partial x_n}(\mathbf{x})\right)$$

If we have a vector valued function $\mathbf{f} : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ open, then $D\mathbf{f}(\mathbf{x})$ has an $m \times n$ matrix representation, called Jacobian $J_{\mathbf{f}}(\mathbf{x})$, such that the generic element is defined as

$$[J_{\mathbf{f}}(\mathbf{x})]_{ij} = \left[\frac{\partial f_i}{\partial x_j}(\mathbf{x})\right] \quad i = 1, \ldots, m \quad j = 1, \ldots, n$$

Definition 6 (Best Linear Approximation) *Consider $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, and assume that f is differentiable at $\mathbf{x}_0 \in U$, then for any $\mathbf{x} \in U$ the best linear approximation of f in a neighborhood of $\mathbf{x}_0$ is*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(||\mathbf{x} - \mathbf{x}_0||)$$

The chain rule can be written explicitly as follows. Assume $\mathbf{f} : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ open, $\mathbf{f}$ differentiable at $\mathbf{x}_0 \in U$, and given $\mathbf{g} : V \to \mathbb{R}^p$, $V \subset R^m$ open, such that $\mathbf{f}(U) \subset V \subset \mathbb{R}^n$ and $\mathbf{g}$ differentiable at $\mathbf{f}(\mathbf{x}_0) \in V$. The function defined as $\mathbf{F} : U \to \mathbb{R}^p \quad \text{s.t.} \quad \mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$ is differentiable at $\mathbf{x}_0$. If we call $(x_1 \dots x_n)$ a generic vector in $\mathbb{R}^n$ and $(y_1 \dots y_m)$ a generic vector in $\mathbb{R}^m$ we have

$$J_{\mathbf{F}}(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(\mathbf{f}(\mathbf{x}_0)) & \dots & \frac{\partial g_1}{\partial y_m}(\mathbf{f}(\mathbf{x}_0)) \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial y_1}(\mathbf{f}(\mathbf{x}_0)) & \dots & \frac{\partial g_p}{\partial y_m}(\mathbf{f}(\mathbf{x}_0)) \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} = J_{\mathbf{g}}(\mathbf{f}(\mathbf{x}_0)) \cdot J_{\mathbf{f}}(\mathbf{x}_0)$$

In the simplest case $m = p = 1$ this becomes

$$F'(\mathbf{x}_0) = g'(f(\mathbf{x}_0)) \nabla f(\mathbf{x}_0)$$

Definition 7 (Directional Derivative) *Consider $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, f differentiable in U, then for any $\mathbf{u} \in \mathbb{R}^n$ such that $||\mathbf{u}|| = 1$ and for any $\mathbf{x} \in U$, we define the derivative of f along the direction $\mathbf{u}$ as*

$$D_{\mathbf{u}} f(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t} = \nabla f(\mathbf{x}) \mathbf{u}$$

Definition 8 *Consider $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, f differentiable in U, then f is continously differentiable if $Df : U \to L(\mathbb{R}^n, \mathbb{R})$ is a continous mapping. In this case we write $f \in \mathcal{C}^1(U)$. This is equivalent to say that*

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; s.t. \;\; \forall \mathbf{x}, \mathbf{y} \in U \;\; ||\mathbf{x} - \mathbf{y}|| < \delta \;\Rightarrow\; ||Df(\mathbf{x}) - Df(\mathbf{y})|| < \varepsilon$$

Theorem 8 *Given $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, $f \in \mathcal{C}^1(U)$ if and only if for any $i = 1, \dots, n$ the functions $\frac{\partial f}{\partial x_i}$ exist and are continous in U.*

Notice that it is not enough to prove the existence of all the partial derivatives to prove the existence of the total differential, we need also the continuity of the partial derivatives. Therefore if f is differentiable then f is continous but the viceversa does not always hold.

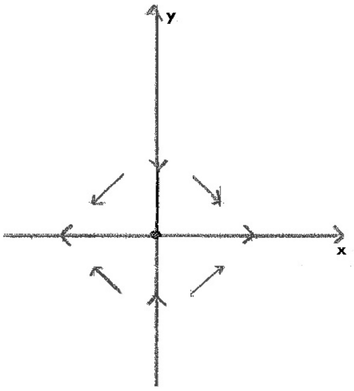

Figure 1:

Example Consider the function

$$f(x,y) = \frac{3}{2}x^2 - \frac{1}{4}y^2 \Rightarrow \nabla f(x,y) = \left(3x\ ,\ -\frac{1}{2}y\right)$$

The gradient is a vector field and we can plot it on the plane (x,y). At every point of the plane the gradient vector points towards the direction of maximum increase of f. This tecnique will be useful when we look for extreme values of the function. In this example we have that (see figure 1)

$$\begin{array}{llllll}
\nabla f(x,y) & \uparrow & \text{if} & x=0 & y<0 \\
\nabla f(x,y) & \downarrow & \text{if} & x=0 & y>0 \\
\nabla f(x,y) & \rightarrow & \text{if} & x>0 & y=0 \\
\nabla f(x,y) & \leftarrow & \text{if} & x<0 & y=0 \\
\nabla f(x,y) & \searrow & \text{if} & x>0 & y>0 \\
\nabla f(x,y) & \swarrow & \text{if} & x<0 & y>0 \\
\nabla f(x,y) & \nwarrow & \text{if} & x<0 & y<0 \\
\nabla f(x,y) & \nearrow & \text{if} & x>0 & y<0
\end{array}$$

Example Consider the function

$$f(x,y) = 4 - 2x^2 - y^2 \Rightarrow \nabla f(x,y) = (-4x\ ,\ -2y)$$

its best linear approximation around the point $\mathbf{x}_0 = (1, 1)$ is

$$l_{(1,1)}(x, y) = f(1, 1) + \nabla f(1, 1)(x - 1 \quad y - 1) = 7 - 4x - 2y$$

In general, the linear approximation $l_{\mathbf{x}_0}(\mathbf{x})$ of f in a neighborhood of $\mathbf{x}_0$ defines the hyperplane $\mathcal{P}$ tangent to f at the point $\mathbf{x}_0$ which satisfies the equation (see the previous example)

$$f(\mathbf{x}) = l_{\mathbf{x}_0}(\mathbf{x}) + o(||\mathbf{x} - \mathbf{x}_0||)$$

or explicitely

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(||\mathbf{x} - \mathbf{x}_0||)$$

For points in a neighborhood of $\mathbf{x}_0$ we can write the best linear approximation of f as

$$\nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = l_{\mathbf{x}_0}(\mathbf{x}) - f(\mathbf{x}_0)$$
$$\underbrace{((\mathbf{x} - \mathbf{x}_0)\ ,\ l_{\mathbf{x}_0}(\mathbf{x}) - f(\mathbf{x}_0))}_{\mathbf{a}}\underbrace{(\nabla f(\mathbf{x}_0)\ ,\ -1)}_{\mathbf{n}} = 0$$

The vector $\mathbf{a}$ lies on the hyperplane $\mathcal{P}$ (a 2-dimensional plane in $\mathbb{R}^3$) and it is orthogonal to the vector $\mathbf{n}$. To draw these vectors we need fix $\mathbf{x}$ in a neighborhood of $\mathbf{x}_0$ and then impose $f(\mathbf{x}_0) = l_{\mathbf{x}_0}(\mathbf{x})$ i.e. we solve

$$\nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = 0$$

In practice for a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ we proceed as follows.
If we are given a point $\mathbf{x}_0$, the plane tangent to f in $\mathbf{x}_0$ is given by the points $\mathbf{x}$ satisfying the linear equation

$$f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = l_{\mathbf{x}_0}(\mathbf{x})$$

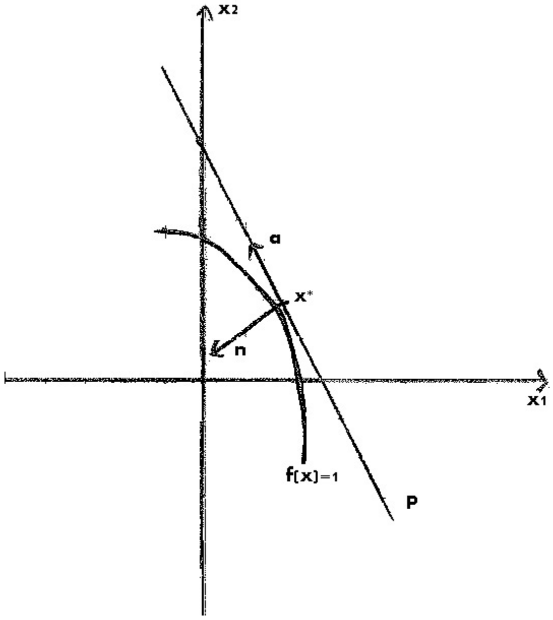

Figure 2:

moreover $l_{\mathbf{x}_0}(\mathbf{x}_0) = f(\mathbf{x}_0)$. This implies that

$$\nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = 0$$

We plot the vector $(\mathbf{x} - \mathbf{x}_0)$ and we have the plane tangent to f in $\mathbf{x}_0$ (which in this case is just a line through $\mathbf{x}_0$). This is the projection of $\mathbf{a}$ in the 2-dimensional space (x_1, x_2). The projection of $\mathbf{n}$ is simply the 2-dimensional vector $\nabla f(\mathbf{x}_0)$. If the calculations are correct we must have that $\mathbf{a} \perp \mathbf{n}$ (see figure 2). Therefore, in the space (x_1, x_2) at the point $\mathbf{x}_0$ the gradient is orthogonal to the function level set $f(\mathbf{x}) = f(\mathbf{x}_0)$.

The same reasononig holds for $n > 2$ but remember that for $n = 2$ the graph of f is in $\mathbb{R}^3$, thus its projection can be plotted (see figure 2), while for $n > 2$ the projection becomes difficult (if $n = 3$) or even impossible ($n > 3$) to plot.

Example Consider the functions

$$f(x_1, x_2) = 2x_1 - 3x_2^2 \qquad \mathbf{g}(t) = (t^3 - \log t\ ,\ \cos t - \sin t)$$

Define $h(t) = f(\mathbf{g}(t))$ then we apply the chain rule

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \nabla f(\mathbf{g}(t)) J_{\mathbf{g}}(t) = (2\ ,\ -6(\cos t - \sin t)) \begin{pmatrix} 3t^2 - \frac{1}{t} \\ -\sin t - \cos t \end{pmatrix} = 6t^2 - \frac{2}{t} - 6(\cos^2 t + \sin^2 t)$$

The Implicit Function Theorem

Example Take the function $f(x, y) = x^2 + y^2$. The level sets $f(x, y) = c$ (we will always take $c = 0$ for simplicity as this does not change anything of what follows) define implicitely $y = \pm\sqrt{1 - x^2}$. However this is not a function since for any x the correspondent y is not uniquely defined. We need to state some conditions in order to be able to write $y = g(x)$. In general given a point (a, b) such that $f(a, b) = 0$ we can solve the equation $f(x, y) = 0$ writing y in terms of x in a neighborhood of (a, b) if $\frac{\partial f}{\partial y}(a, b) \neq 0$.

Intuitively given $f(\mathbf{x}, \mathbf{y}) = 0$ and assuming that we can write $\mathbf{y} = \mathbf{g}(\mathbf{x})$ (difficult to do explicitely!) we can always compute $J_{\mathbf{g}}(\mathbf{x})$ by using the chain rule (easy to do!).
Assume $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ both differentiable, and assume $f(x, y) = f(x, g(x)) = 0$, then, in a generic x_0 such that $f(x_0, g(x_0)) = 0$, we have

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) = \nabla f(x_0, g(x_0))(1\ ,\ g'(x_0)) = 0$$

Therefore

$$g'(x_0) = -\frac{\frac{\partial f}{\partial x}(x_0, g(x_0))}{\frac{\partial f}{\partial y}(x_0, g(x_0))}$$

Theorem 9 (Linear Case) *Consider a linear transformation* $\mathbf{A} : \mathbb{R}^{n+m} \to \mathbb{R}^m$*, and a generic point* $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+m}$ *such that* $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{y} \in \mathbb{R}^m$*. We define* $\mathbf{A} = (\mathbf{A}_x, \mathbf{A}_y)$ *such that* $\mathbf{A}(\mathbf{x}, \mathbf{y}) = \mathbf{A}_x\mathbf{x} + \mathbf{A}_y\mathbf{y}$*. If* $\mathbf{A}_y$ *is invertible then*

$$\forall\, \mathbf{x} \in \mathbb{R}^n\ \ \exists\,!\, \mathbf{y} \in \mathbb{R}^m\ \ \textit{s.t.}\ \ \mathbf{A}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \Rightarrow \mathbf{y} = -\mathbf{A}_y^{-1}\mathbf{A}_x\mathbf{x}$$

Theorem 10 (Dini's Theorem) *Consider a function* $\mathbf{f} : U \to \mathbb{R}^m$*, with* $U \subset \mathbb{R}^{n+m}$ *open and* $\mathbf{f} \in \mathcal{C}^1(U)$*. Define a generic point as* $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+m}$ *such that* $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{y} \in \mathbb{R}^m$*. Assume there exist a point* $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{n+m}$ *such that* $\mathbf{f}(\mathbf{a}, \mathbf{b}) = \mathbf{0}$*. Define the Jacobian of* $\mathbf{f}$ *as* $J_{\mathbf{f}}(\mathbf{a}, \mathbf{b}) = (J_{\mathbf{xf}}(\mathbf{a}, \mathbf{b}), J_{\mathbf{yf}}(\mathbf{a}, \mathbf{b}))$ *and assume that*

$$J_{\mathbf{yf}}(\mathbf{a}, \mathbf{b}) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{pmatrix} (\mathbf{a}, \mathbf{b}) \qquad \text{is invertible}$$

Then there exist $V \subset \mathbb{R}^{n+m}$ *open such that* $(\mathbf{a}, \mathbf{b}) \in V$*, and* $W \subset \mathbb{R}^n$ *open such that* $\mathbf{a} \in W$*. In these sets the following holds*

$$1) \quad \forall \mathbf{x} \in W \;\; \exists!\, \mathbf{y} \;\; s.t. \;\; (\mathbf{x}, \mathbf{y}) \in V \;\; and \;\; \mathbf{f}(\mathbf{x}, \mathbf{y}) = 0$$

$$2) \quad \forall \mathbf{x} \in W \;\; \exists!\, \mathbf{g} \in \mathcal{C}^1(W) \;\; s.t. \;\; \mathbf{y} = \mathbf{g}(\mathbf{x}) \;\; and \;\; \mathbf{f}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = \mathbf{0}$$

$$3) \quad \forall\, (\mathbf{x}, \mathbf{y}) \in V \;\; J_{\mathbf{g}}(\mathbf{x}) = -J_{\mathbf{yf}}^{-1}(\mathbf{x}, \mathbf{g}(\mathbf{x})) J_{\mathbf{xf}}(\mathbf{x}, \mathbf{g}(\mathbf{x}))$$

For simplicity we restate the theorem in the case $m = 1$

Theorem 11 (Dini's Theorem - 2) *Consider a function* $f : U \to \mathbb{R}$*, with* $U \subset \mathbb{R}^{n+1}$ *open and* $f \in \mathcal{C}^1(U)$*. Define a generic point as* $(\mathbf{x}, y) \in \mathbb{R}^{n+1}$ *such that* $\mathbf{x} \in \mathbb{R}^n$ *and* $y \in \mathbb{R}$*. Assume there exist a point* $(\mathbf{a}, b) \in \mathbb{R}^{n+1}$ *such that* $f(\mathbf{a}, b) = 0$*. The gradient of* f *is*

$$\nabla f(\mathbf{a}, b) = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}, \frac{\partial f}{\partial y} \right)$$

and assume that

$$\frac{\partial f}{\partial y} \neq 0$$

Then there exist $V \subset \mathbb{R}^{n+1}$ *open such that* $(\mathbf{a}, b) \in V$, *and* $W \subset \mathbb{R}^n$ *open such that* $\mathbf{a} \in W$. *In these sets the following holds*

$$1) \quad \forall \mathbf{x} \in W \ \ \exists!\, y \ \ s.t. \ \ (\mathbf{x}, y) \in V \ \ and \ \ f(\mathbf{x}, y) = 0$$

$$2) \quad \forall \mathbf{x} \in W \ \ \exists!\, g \in \mathcal{C}^1(W) \ \ s.t. \ \ y = g(\mathbf{x}) \ \ and \ \ f(\mathbf{x}, g(\mathbf{x})) = 0$$

$$3) \quad \forall\, (\mathbf{x}, y) \in V \ \ \frac{\partial g}{\partial x_i}(\mathbf{x}) = -\frac{\frac{\partial f}{\partial x_i}(\mathbf{x}, g(\mathbf{x})}{\frac{\partial f}{\partial y}(\mathbf{x}, g(\mathbf{x})} \quad i = 1, \ldots, n$$

Example Consider the function $f(x, y) = x^2 + y^2 - 1$ and level set describing the unit circle $f(x, y) = 0$. If we take $(a, b) = (1, 0)$ then $f(a, b) = 0$. Given $x < 1$ and close to a can we find y close to b such that $f(x, y) = 0$?

the answer is yes, but the y is not unique, indeed $y = \pm\sqrt{1 - x^2}$ and the hypotheis of the theorem does not hold because $\frac{\partial f}{\partial y}(1, 0) = 0$.

Now take $(a, b) = (x_0, y_0)$ such that $x_0 > 0$ and $y_0 > 0$. Now the solution is unique and is $y = \sqrt{1 - x^2}$ because the other solution $y = -\sqrt{1 - x^2} < 0$ is not close to y_0. In this case $\frac{\partial f}{\partial y}(x_0, y_0) = 2y_0 > 0$ the hypothesis of theorem is satisfied and we can define $g(x) = \sqrt{1 - x^2}$.

Finally,

$$\frac{\mathrm{d}g}{\mathrm{d}x}(x_0) = -\frac{2x_0}{2y_0} = -\frac{x_0}{\sqrt{1 - x_0^2}}$$

By choosing different values of (a, b) we can plot in each point of the plane the function g if we can compute it. In any case we can always compute its gradient (or derivative) and have

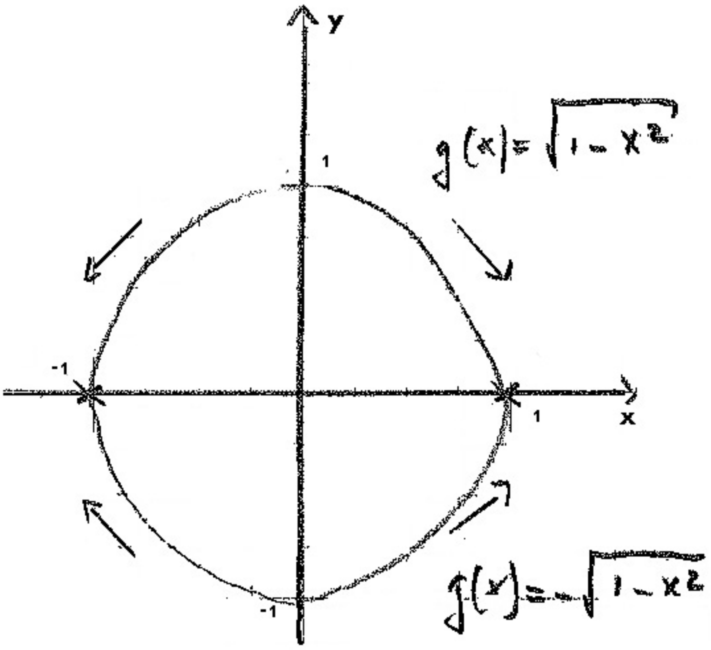

Figure 3:

at least an idea of the vector field (see figure 3).

Example Consider a generic utility function $u : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ such that $u(0,0) = 0$ and $u(x_1, x_2) > 0$ for $x_1 > 0$ and/or $x_2 > 0$. Assume $\frac{\partial u}{\partial x_i} > 0$ i.e. the marginal utilities are positive. Take a point $\mathbf{x}_0$ such that $u(\mathbf{x}_0) = c > 0$. Then if $u \in \mathcal{C}^1(U)$ in every point $\mathbf{x}$ close to $\mathbf{x}_0$ we can write $x_2 = g(x_1)$ (or viceversa) and

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1}(\mathbf{x}) = -\frac{\frac{\partial u}{\partial x_1}(\mathbf{x})}{\frac{\partial u}{\partial x_2}(\mathbf{x})} < 0$$

the above quantity is called Marginal Rate of Substitution between good 1 and good 2 (MRS_{12}). Moreover the vector tangent to u at $\mathbf{x}_0$ is orthogonal to ∇u indeed the previous equation can be written as

$$\nabla u(\mathbf{x}) \underbrace{(1\ ,\ MRS_{12})}_{\mathbf{a}} = 0$$

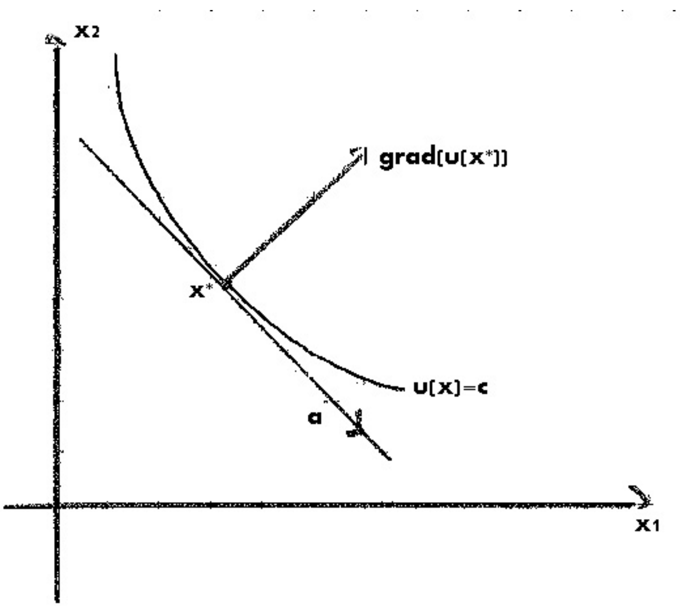

Figure 4:

MRS_{12} tells how to move x_2 once x_1 is changed of 1 unit in order to keep $u(x_1, x_2) = c$. The direction of this change must be orthogonal to the gradient of u (see figure 4).

Main Reference

Rudin, W. *Principles of Mathematical Analysis* McGraw-Hill, Inc. 1976