Chapter 11

Implicit Functions, Curves and Surfaces

11.1 Implicit Function Theorem

Motivation. In many problems, objects or quantities of interest can only be described indirectly or implicitly. It is then important to know when such implicit representations do indeed determine the objects of interest. Examples are

- Implicit representation of functions. An equation of type $F(x, y) = c$ might determine a function $f(x)$ such that $F(x, f(x)) = 0$ but not allow for an explicit calculation of it.

- The previous situation is for instance sometimes encountered when solving ordinary differential equations such as

  $$\alpha(x, y) + \beta(x, y)\dot{y} = 0$$

  if

  $$\begin{cases} \alpha(x, y) = \partial_x F(x, y) \\ \beta(x, y) = \partial_y F(x, y) \end{cases}$$

  for some $F(x, y)$. Then solutions of the ordinary differential equation lie on level sets of F.

- Curves and surfaces are often best described implicitly. Take the circle $\mathbb{S}^1$ as an example

  $$\mathbb{S}^1 = \{x \in \mathbb{R}^2 \,|\, |x|_2^2 = 1\}.$$

These sketchy exampes should serve as motivation for considering the following general problem. Given a function

$$F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m \,,\; (x, y) \mapsto F(x, y) \,,$$

is it possible to find $y : \mathbb{R}^n \to \mathbb{R}^m$ such that $F(x, y(x)) = 0 \,\forall\, x \in \mathbb{R}^n$? Or, in other words, is it possible to describe solutions of $F = 0$ as the graph of a function $y : \mathbb{R}^n \to \mathbb{R}^m$?
This is often too much asked or difficult to tell, but a local version of this questions leads to an easy computable criterion. The local version can be described as follows. Given (x_0, y_0) such that $F(x_0, y_0) = 0$, is it possible to find neighborhoods $U \in \mathcal{U}(x_0)$, $V \in \mathcal{U}(y_0)$ and a map $y : U \to V$ such that $F(x, y(x)) = 0$? Or, in other words, is it possible to desribe the solution set of $F = 0$ about the point (x_0, y_0) as the graph of a function y?
The answer the this last question is contained in the so-called *implicit function theorem*, the proof of which relies on the following basic idea. We shall assume that F is smooth. Then we have that

$$0 = F(x, y) \sim F(x_0, y_0) + D_xF(x_0, y_0)(x - x_0) + D_yF(x_0, y_0)(y - y_0)$$

where we used the notation $D_xf(x_0, y_0) = [\partial_{x_k} F_j]_{1\le j\le m, 1\le k\le n} \in \mathbb{R}^{m\times n}$ and $D_yf(x_0, y_0) = [\partial_{y_k} F_j]_{1\le j,k\le m} \in \mathbb{R}^{m\times m}$. Thus, locally and approximately, the equation looks linear. In that case solving for y as a function of x should be possible whenever $D_yF(x_0, y_0)$ is invertible. This is indeed the case and we shall now give a precise formulation of the result and give a rigorous proof for it.

Definition 11.1.1. (Matrix-Norm)
Let $A \in \mathbb{R}^{m\times m}$. Its *(matrix)-norm* is given by

$$\|A\| := \sup_{0\neq x\in\mathbb{R}^m} \frac{|Ax|}{|x|} = \sup_{|x|=1} |Ax| \,.$$

Remarks 11.1.2. (a) Taking $x = e_k$ and observing that

$$|A_{jk}| = |(Ae_k)_j| \le |Ae_k| \le \|A\||e_k| = \|A\|$$

it is easily seen that $|A_{jk}| \le \|A\| \,\forall\, j, k = 1, \dots, m$
(b) The estimate

$$|Ax|^2 \le \sum_{j=1}^m \Big(\sum_{k=1}^m A_{jk}x_k\Big)^2 \le \sum_{j=1}^m \Big(\sum_{k=1}^m A_{jk}^2 \sum_{k=1}^m x_k^2\Big) = \Big(\sum_{j,k=1}^m A_{jk}^2\Big)|x|^2$$

follows from the Cauchy-Schwarz inequality and readily implies

$$\|A\| \leq \Big(\sum_{j,k=1}^{m} A_{jk}^2\Big)^{\frac{1}{2}}.$$

(c) It holds that $\|AB\| \leq \|A\|\|B\|$ for $A, B \in \mathbb{R}^{m\times m}$ as follows from

$$\|AB\| = \sup_{|x|=1} |ABx| \leq \sup_{|x|=1} \|A\||Bx| = \|A\|\|B\|.$$

(d) The function $\|\cdot\| : \mathbb{R}^{m\times m} \to [0,\infty)$ defines a norm.

Lemma 11.1.3.

$$\mathcal{GL}_m(\mathbb{R}) = \{A \in \mathbb{R}^{m\times m}\,|A \text{ invertible}\} \overset{o}{\subset} \mathbb{R}^{m\times m}.$$

Proof. It is obvious that $\mathcal{GL}_m(\mathbb{R}) = \det^{-1}\big(\mathbb{R}\setminus\{0\}\big)$ and the claim follows from the continuity of the determinant function combined with theorem 4.1.10. $\surd$

We are now ready for the main theorems of this section.

Theorem 11.1.4. *(Implicit Function Theorem)*
Let $\tilde{x} \in \mathbb{R}^n, y \in \mathbb{R}^m$, $U \in \mathcal{U}(\tilde{x}), V \in \mathcal{U}(\tilde{y})$ and

$$F \in \mathrm{C}^1(U\times V, \mathbb{R}^m).$$

Assume that $F(\tilde{x},\tilde{y}) = c$ and that $D_yF(\tilde{x},\tilde{y}) \in \mathcal{GL}_m(\mathbb{R})$.
Then there exists a neighborhood $\tilde{U} \in \mathcal{U}(\tilde{x})$ and a map $y \in \mathrm{C}^1(\tilde{U}, \mathbb{R}^m)$ such that

$$y(\tilde{x}) = \tilde{y} \text{ and } F(x, y(x)) = c\;\forall\, x \in \tilde{U}.$$

In other words, $\big(x, y(x)\big)$ is the unique solution of $F(x,y) = c$ in the given neighborhood of $(\tilde{x},\tilde{y})$. Moreover

$$Dy(x) = -D_yF\big(x,y(x)\big)^{-1}D_xF\big(x,y(x)\big).$$

Theorem 11.1.5. *(Inverse Function Theorem)*
Let $\tilde{y} \in \mathbb{R}^n$, $U \in \mathcal{U}(\tilde{x})$ and assume that

$$f \in \mathrm{C}^1(U, \mathbb{R}^n),\; Df(\tilde{y}) \in \mathcal{GL}_n(\mathbb{R}).$$

Then there exists $\tilde{U} \in \mathcal{U}\big(f(\tilde{y})\big)$ and $g \in \mathrm{C}^1(\tilde{U}, \mathbb{R}^n)$ such that

$$f\big(g(x)\big) = x\;\forall\, x \in \tilde{U}.$$

Also g maps one-to-one onto a neighborhood V of $\tilde{y}$ and

$$g\big(f(y)\big) = y \,\forall\, y \in V\,.$$

The function g is unique in that $g(x)$ is the unique solution of the equation $f(y) = x$ in V. Moreover

$$Dg(x) = Df(y)^{-1} \textit{ for } x = f(y)$$

Proof. Use theorem 11.1.4 setting

$$F(x,y) = x - f(y)\,,\; c = 0\,.$$

Then $D_y F(x,y) = Df(y)$. $\checkmark$

Remark 11.1.6. The inverse function theorem could be used to prove the implicit function theorem as well. Given F as in theorem 11.1.4 define the function f by

$$f : U \to \mathbb{R}^{n+m}\,,\; (x,y) \mapsto \big(x, F(x,y)\big)\,.$$

In this case

$$Df(x,y) = \begin{bmatrix} \mathrm{id}_{\mathbb{R}^n} & 0 \\ D_x F & D_y F \end{bmatrix} \in \mathcal{GL}_{n+m}(\mathbb{R}) \Longleftrightarrow D_y F \in \mathcal{GL}_m(\mathbb{R})\,.$$

The inverse function theorem then implies that f is locally invertible about $(\tilde{x}, \tilde{y})$ and there exists g such that $f\big(g(x,c)\big) = (x,c)$ in appropriate neighborhoods. Observe that $g(x,c) = \big(x, y(x)\big)$ and therefore $f\big(x, y(x)\big) = c$.

Examples 11.1.7. (a) Consider the equation $x^2 + y^2 = 1$. Define the function $F(x,y) = x^2 + y^2$, then $D_y F(x,y) = 2y$. It follows that $D_y F(x,y)$ is invertible if $y \neq 0$. This means that we can locally solve the equation for y as a function of x about any point on the circle with the exception of $(1,0)$ and $(-1,0)$.
(b) Consider

$$f : \mathbb{R}^2 \to \mathbb{R}^2\,,\; (x,y) \mapsto (x^2 - y^2, 2xy)\,.$$

Then

$$Df(x,y) = \begin{bmatrix} 2x & -2y \\ 2y & 2x \end{bmatrix}$$

and $J(x,y) = 4x^2 + 4y^2$. Thus, unless $(x,y) \neq (0,0)$, the assumptions of the inverse function theorem are satisfied.

Proof. (Of theorem 11.1.4)
Since $F(\tilde{x},\tilde{y}) = c$ and F is smooth we have that $F(x,y) \approx c$ in the vicinity of $(\tilde{x},\tilde{y})$. Also observe that

$$\begin{aligned} &F(x,y) + D_yF(x,y)(z-y) = c \\ &\qquad\qquad \iff z = y + D_yF(x,y)^{-1}\big(c - F(x,y)\big) = \widetilde{T}(y)\,. \end{aligned}$$

It thus follows that

$$\begin{aligned} &(x,y) \text{ solves } F(x,y) = c \iff z \text{ is a fixed-point of } \widetilde{T} \\ &\qquad\qquad\qquad\qquad \iff z \text{ is a fixed-point of } T\,. \end{aligned}$$

for

$$T(y) = y + D_yF(\tilde{x},\tilde{y})^{-1}\big(c - F(x,y)\big)\,.$$

If we can prove that T is a contraction in some neighborhood V of $\tilde{y}$ we could use the contraction mapping theorem 8.3.16 to obtain a unique fixed-point in V. We first show that $T : \mathbb{B}(\tilde{y},\delta) \to \mathbb{B}(\tilde{y},\delta)$ is a contraction if $\delta > 0$ is chosen small enough. The contractivity is obtained as follows. Observe that

$$\begin{aligned} |Ty - Tx| &= y - z + D_yF(\tilde{x},\tilde{y})^{-1}\big[F(x,z) - F(x,y)\big] \\ &= D_yF(\tilde{x},\tilde{y})^{-1}\big[F(x,z) - F(x,y) - D_yF(\tilde{x},\tilde{y})(z-y)\big] \\ &= D_yF(\tilde{x},\tilde{y})^{-1}\int_0^1 \big[D_yF\big(x, y + t(y-z)\big) - D_yF(\tilde{x},\tilde{y})\big](z-y)\,dt \end{aligned}$$

which implies, by continuity of D_yF, that

$$|Ty - Tz| \le M|z-y|c(\varepsilon,\delta)\,,\ x \in \mathbb{B}(\tilde{x},\varepsilon)\,,\ y \in \mathbb{B}(\tilde{y},\delta)$$

with a consant $c(\varepsilon,\delta)$ such that $\lim_{\varepsilon,\delta\to 0} c(\varepsilon,\delta) = 0$. Thus, choosing ε and δ small enough, contractivity is obtained. Next we need to make sure that T is a self-map of $\mathbb{B}(\tilde{y},\delta)$. To that end, consider

$$\begin{aligned} |Ty - \tilde{y}| &= y - \tilde{y} + D_yF(\tilde{x},\tilde{y})^{-1}\big[c - F(x,y)\big] \\ &= D_yF(\tilde{x},\tilde{y})^{-1}\big[F(\tilde{x},\tilde{y}) - F(x,y) + D_yF(\tilde{x},\tilde{y})(y-\tilde{y})\big] \\ &\qquad = o(|x-\tilde{x}| + |y-\tilde{y}|) - D_xF(\tilde{x},\tilde{y})(x-\tilde{x}) \end{aligned}$$

since

$$\begin{aligned} F(x,y) &= F(\tilde{x},\tilde{y}) + D_{x,y}F(\tilde{x},\tilde{y})\big((x,y) - (\tilde{x},\tilde{y})\big) \\ &= F(\tilde{x},\tilde{y}) + D_xF(\tilde{x},\tilde{y})(x-\tilde{x}) + D_yF(\tilde{x},\tilde{y})(y-\tilde{y}) \\ &\qquad + o(|x-\tilde{x}| + |y-\tilde{y}|) \end{aligned}$$

as $(x, y) \to (\tilde{x}, \tilde{y})$. Thus it suffices to choose ε such that

$$|D_x F(\tilde{x}, \tilde{y})(x - \tilde{x})| \leq \frac{\delta}{2}$$

and δ such that $o(|x - \tilde{x}| + |y - \tilde{y}|) \leq \frac{\delta}{2}$ which is possible by definition of o. Observe that ε might need to be made smaller once more. Then

$$|T(y) - \tilde{y}| \leq \delta \,\forall\, x \in \mathbb{B}(\tilde{x}, \varepsilon) \,\forall\, y \in \mathbb{E}(\tilde{y}, \delta) \,.$$

which is the desired self-map property. By theorem 8.3.16 we therefore obtain that

$$\forall\, x \in \mathbb{B}(\tilde{x}, \varepsilon) \,\exists\, y = y(x) \in \mathbb{B}(\tilde{y}, \delta) \text{ s.t } F\big(x, y(x)\big) = c \,.$$

Finally

$$0 = F\big(x, y(x)\big) - F(\tilde{x}, \tilde{y}) = D_x F(\tilde{x}, \tilde{y})(x - \tilde{x}) + D_y F(\tilde{x}, \tilde{y})(y(x) - \tilde{y}) + R(x)$$
$$\text{with } R(x) = o(|x - \tilde{x}| + |y(x) - \tilde{y}|) \text{ as } x \to \tilde{x}$$

implies that

$$y(x) - \tilde{y} = -D_y F(\tilde{x}, \tilde{y})^{-1} \big[D_x F(\tilde{x}, \tilde{y})(x - \tilde{x}) + R(x)\big] \,.$$

Thus, if $D_y F(\tilde{x}, \tilde{y})^{-1} R(x) = o(|x - \tilde{x}|)$, then we obtain the desired differentiability and that

$$Dy(\tilde{x}) = -D_y F(\tilde{x}, \tilde{y})^{-1} D_x F(\tilde{x}, \tilde{y}) \,.$$

The inequality

$$|D_y F(\tilde{x}, \tilde{y})^{-1} R(x)| \leq c(x - \tilde{x})[|x - \tilde{x}| + |y(x) - \tilde{y}|] \leq \tilde{c}\, c(x - \tilde{x})|x - \tilde{x}|$$

which is valid for c with $\lim_{x \to \tilde{x}} c(x - \tilde{x}) = 0$, shows that this is the case and concludes the proof. How do you extend the validity of the formula for the derivative away from the point $(\tilde{x}, \tilde{y})$? $\surd$

11.2 Curves and Surfaces

11.2.1 Motivation and Examples

Motivation. How can we effectively describe "geometric" sets/subsets of $\mathbb{R}^n$? There are two quite general possibilities. One is a *parametric* description, in which case,a set $A \subset \mathbb{R}^n$ is described as the image space $g(U)$ of a function $g : U \overset{o}{\subset} \mathbb{R}^m \to \mathbb{R}^n$ or, more in general, as

$$A = \bigcup_{j=1}^{n} g_j(U_j)\,,\ g_j : U_j \overset{o}{\subset} \mathbb{R}^m \to \mathbb{R}^n\,.$$

If $m = 1$ we obtain curves in n-dimensional space. The graph of a function is a special way of parametrize a set. It occurs when

$$A = \{(x, f(x))\,|\,x \in U \overset{o}{\subset} \mathbb{R}^m\}$$

for some $f : U \to \mathbb{R}^{n-m}$.
A second method of describing sets is *implicit*. In this case sets are described a the zero level set of some function $F : \mathbb{R}^n \to \mathbb{R}^k$

$$A = \{x \in \mathbb{R}^n\,|\,F(x) = 0\}$$

for $k = n - m$. The number m should tell us what the dimension of the *manifold* A is. Why? Precise definitions will given in the rest of the section.

Examples 11.2.1. (a) Consider the circle $\mathbb{S}^1$ of radius 1 in $\mathbb{R}^2$. It can be described as

$$\{(x, y) \in \mathbb{R}^2\,|\,x^2 + y^2 = 1\}\,.$$

Alternatively, it could be described as $\mathbb{S}^1 = g_1\big((0, 2\pi)\big) \cup g_2\big((-\frac{\pi}{2}, \frac{\pi}{2})\big)$ for

$$g_1 : (0, 2\pi) \to \mathbb{R}^2\,,\ \theta \mapsto (\cos(\theta), \sin(\theta))\,,$$
$$g_2 : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}^2\,,\ \theta \mapsto (\cos(\theta), \sin(\theta))\,.$$

Finally we could use the graph patches

$$x = \pm\sqrt{1 - y^2}\,\big[y \in (-1, 1)\big]\,,\ y = \pm\sqrt{1 - x^2}\,\big[x \in (-1, 1)\big]\,.$$

(b) Consider the subset of $\mathbb{R}^2$ determined by

$$\{(x, y) \in \mathbb{R}^2\,|\,y^3 - x^2 = 0\}\,.$$

It can be parametrized by $g : \mathbb{R} \to \mathbb{R}^2$ given by

$$g(t) = (t^3, t^2)\,,$$

or as the graph of the function f given by $f(x) = x^{\frac{2}{3}}$, $x \in \mathbb{R}$. Observe that the set does not seem to be very smooth in spite of the fact that the first

two descriptions only involve smooth functions.
(c) The sphere $\mathbb{S}^2$ of radius 1 has the following representations

$$x^2 + y^2 + z^2 - 1 = 0\,,$$
$$(x, y, z) = \big(\cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)\big)$$
$$\begin{cases} x = \pm\sqrt{1 - y^2 - z^2} \\ y = \pm\sqrt{1 - x^2 - z^2} \\ z = \pm\sqrt{1 - x^2 - y^2} \end{cases}$$

(d) We could also consider $\mathbb{S}^2$ as a subset of $\mathbb{R}^3$ in which case we could use the descriptions

$$\begin{cases} x^2 + y^2 + z^2 - 1 = 0 \\ z = 0 \end{cases} \quad , \qquad \big(\cos(\theta), \sin(\theta), 0\big)\,,$$
$$\begin{cases} (y, z) = (\pm\sqrt{1 - x^2}, 0)\,, \\ (x, z) = (\pm\sqrt{1 - y^2}, 0) \end{cases} \quad .$$

(e) Consider the set $\{(x, y) \in \mathbb{R}^2 \,|\, xy = 0\}$ which is the union of the two lines $\{x = 0\}$ and $\{y = 0\}$. Observe that no single graph or one-to-one parametrization can be given of this set about the origin.

We now turn to precise definitions.

11.2.2 Immersions and Embeddings

Looking at simple fold or cusp singularities, it becomes clear that smooth parametrization clearly exists for them. In this case the smoothness of the parametrization can not be taken as an indicator of the smoothness of a curve. How do we detect “bad points” then?

Example 11.2.2. Consider the surface parametrized by

$$g(t, s) = (t^3, t^2, s)\,,\ t, s \in \mathbb{R}$$

and look at its derivative

$$Dg(t, s) = \begin{bmatrix} 3t^2 & 0 \\ 2t & 0 \\ 0 & 1 \end{bmatrix}$$

Definition 11.2.3. (Immersion)
For a matrix $A \in \mathbb{R}^{m\times n}$ we denote the dimension of its image space $\mathrm{im}(A)$, also called rank of A, by $\mathrm{rank}(A) = r(A)$. Then, for $U \overset{o}{\subset} \mathbb{R}^m$, we call a map

$$g : U \to \mathbb{R}^n\,,\ x \mapsto g(x)$$

an immersion if $r\big(Dg(x)\big) = m$ at every $x \in U$.

Remark 11.2.4. It follows that an immersion is locally one-to-one. It is, however, not necessarily one-to-one globally.

Definition 11.2.5. (Embedding)
An immersion $g : U \to \mathbb{R}^n$ is called *embedding* iff
(i) g is injective (one-to-one) and
(ii) $g^{-1} : g(U) \to \mathbb{R}^m$ is continuous.
Hereby $g(U)$ is to be considered as a metric subspace of $\mathbb{R}^n$ endowed with the induced metric.

Motivation. Why would we consider parametrizations (embeddings) as an interesting object in the study of curves and surfaces? Well, we think of them as a way to introduce (local) coordinates on the curve or surface. Think of the coordinates' domain as a (geographic) map of the curve or surface. Such coordinates should give us an handle on performing some calculus on the curve/surface.
More in general, let U be the parametrization domain of $g : U \to \mathbb{R}^n$ and consider a curve $\gamma : (0, 1) \to U$ through the point x in U, then, clearly, $g \circ \gamma$ is a curve on the surface/manifold $g(U)$ through the point $g(x)$. In particular we can take the simple curves

$$\gamma_j : (0, 1) \to U\,,\ t \to x + te_j \ \text{for}\ j = 1, \dots , m\,.$$

Then $\partial_j g(x) := \frac{d}{dt} g(x + te_j)\big|_{t=0}$ is a vector in $\mathbb{R}^n$ tangent to the manifold $g(U)$ at $g(x)$. In view of the maximal rank condition, if g is an immersion, then

$$\partial_1 g(x), \dots , \partial_m g(x)$$

span the whole tangent $T_{g(x)}g(U)$ space to $g(U)$ at $g(x)$, that is,

$$T_{g(x)}g(U) = \Big\{ \sum_{j=1}^{m} \alpha_j \partial_j g(x) \,\Big|\, \alpha \in \mathbb{R}^m \Big\}\,.$$

Remarks 11.2.6. (a) Let γ be any C^1-curve passing through $x \in U$, then $g \circ \gamma$ is a curve through $g(x) \in g(U)$ and the tangent vector is given by

$$\frac{d}{dt}(g \circ \gamma)\big|_{t=0} = \sum_{j=1}^{m} \frac{d}{dt}\gamma_j(0)\partial_j g(x)\,.$$

(b) In the case of an embedding, any curve through $g(x)$ on $g(U)$ can be viewed as the image of a curve in U since g^{-1} is continuous.

11.2.3 Parametric Description of Manifolds

Definition 11.2.7. (Manifold)
A subset $M_m \subset \mathbb{R}^n$ is called *m-dimensioanl* C^1*-manifold in* $\mathbb{R}^n$ iff

$$\forall\, y \in M_m\ \exists\, V \in \mathcal{U}_{\mathbb{R}^n}(y) \text{ and an embedding } g : U \overset{o}{\subset} \mathbb{R}^m \to \mathbb{R}^n \quad \text{with } g(U) = V \cap M_m\,.$$

Remark 11.2.8. A C^1-manifold is locally parametrized by embeddings, or, in other words, a patchwork of images of embeddings.

Theorem 11.2.9. *Let $g : U \to \mathbb{R}^n$ for $U \overset{o}{\subset} \mathbb{R}^m$. Assume that $g \in \mathrm{C}^1(U, \mathbb{R}^n)$ and that it is an immersion. Then*

$$\forall\, \tilde{x} \in U\ \exists\, U_{\tilde{x}} \in \mathcal{U}(\tilde{x}) \ \textit{s.t.}\ g\big|_{U_{\tilde{x}}} \textit{ is injective and } \quad g(U_{\tilde{x}}) \textit{ is the graph of a } \mathrm{C}^1\textit{-function}\,.$$

Proof. Since $g = (g_1, \ldots, g_n)$ we have

$$Dg(\tilde{x}) = \begin{bmatrix} Dg_1(\tilde{x}) \\ \vdots \\ Dg_n(\tilde{x}) \end{bmatrix}, \ \tilde{x} \in U\,.$$

By assumption Dg has maximal rank and, therefore, we can find m linearly independent rows in it. W.l.o.g we can assume that those are the first m rows $Dg_1(\tilde{x}), \ldots, Dg_m(\tilde{x})$. Considering the system

$$\begin{cases} t_1 = g_1(x_1, \ldots, x_m) \\ t_2 = g_2(x_1, \ldots, x_m) \\ \vdots \\ t_m = g_m(x_1, \ldots, x_m) \end{cases}$$

we can invoke the inverse function theorem 11.1.5 to find neighborhoods $U_{\tilde{x}} \in \mathcal{U}(\tilde{x})$ and $V \in \mathcal{U}(h(\tilde{x}))$ for $h = (g_1, \ldots, g_m)$ such that

$$(x_1, \ldots, x_m) = h^{-1}(t_1, \cdots, t_m) \text{ for } t \in V \text{ and } h^{-1} \in \mathrm{C}^1(V, U_{\tilde{x}})$$

In particular g is injective on $U_{\tilde{x}}$. Next consider

$$\begin{cases} s_1 = g_{m+1}(x_1, \ldots, x_m) \\ \vdots \\ s_{n-m} = g_n(x_1, \ldots, x_m) \end{cases}$$

and rewrite it as $s = \Phi(x)$. Then

$$s = \Phi\big(h^{-1}(t)\big) =: f(t)$$

and $f \in \mathrm{C}^1(V, \mathbb{R}^{n-m})$. Furthermore

$$y = g(x) \text{ for some } x \in U_{\tilde{x}} \iff y = \big(t, f(t)\big) \text{ for some } t \in V$$

which gives the desired graph representation. $\surd$

Examples 11.2.10. (a) Consider the immersion

$$g : \mathbb{R}^1 \to \mathbb{R}^2\,,\ \theta \mapsto \big(\cos(\theta), \sin(\theta)\big) =: (x, y)\,.$$

About any point with $g_2'(\theta) = \cos(\theta) \neq 0$ [$g_1'(\theta) = \sin(\theta) \neq 0$], we can solve for x [y] as a function of y [x]. Choose $(\tilde{x}, \tilde{y}) \in \mathbb{S}^1$ with $\sin(\tilde{\theta}) \neq 0$, then solve for θ as a function of x

$$\theta = \arccos(x)$$

in a small neighborhood about $\tilde{x}$. Finally conclude that

$$y = \sin\big(\arccos(x)\big)\ \big(= \pm\sqrt{1 - x^2}\big)\,.$$

(b) Consider the mapping

$$g : \mathbb{R}^2 \to \mathbb{R}^3\,,\ (\theta, \varphi) \mapsto \big(\cos(\theta)\sin(\varphi), \sin(\theta)\sin(\varphi), \cos(\varphi)\big) = (x, y, z)\,.$$

Then $g(\mathbb{R}^2) = \mathbb{S}^2$. Is g an immersion? Looking at the derivative

$$\begin{bmatrix} -\sin(\theta)\sin(\varphi) & \cos(\theta)\cos(\varphi) \\ \cos(\theta)\sin(\varphi) & \sin(\theta)\cos(\varphi) \\ 0 & -\sin(\theta) \end{bmatrix}$$

we see that its rank is only one if $\sin(\varphi) = 0$, since the first column vanishes. It can be checked that the rank is otherwise 2 and therefore maximal. We get an immersion if we restrict g to where $\sin(\varphi) \neq 0$. Let's take $\varphi \in (0, \pi)$, for instance.
In order to obtain a local graph representation of the sphere, we can try to solve for z as a function of x and y. To do so we can
(ii) determine θ, φ as functions of x, y.
(iii) plug the latter into the expression for z.
Consider therefore the system

$$\begin{cases} x = \cos(\theta)\sin(\varphi) \\ y = \sin(\theta)\cos(\varphi) \end{cases}$$

which has associated Jacobian

$$\det \begin{bmatrix} -\sin(\theta)\sin(\varphi) & \cos(\theta)\cos(\varphi) \\ \cos(\theta)\sin(\varphi) & \sin(\theta)\cos(\varphi) \end{bmatrix} = -\sin(\varphi)\cos(\varphi)\,.$$

Since $\sin(\varphi) \neq 0$ $(0 < \varphi < \pi)$ we only need stay away from the region where $\cos(\varphi) = 0$, which is the equator, in order to be able to make use of the inverse function theorem to solve for *theta* and φ. In fact, in that case

$$\begin{cases} x^2 + y^2 = \sin^2(\varphi) \\ \frac{y}{x} = \tan(\theta) \end{cases}$$

yields

$$\varphi = \arcsin\sqrt{x^2 + y^2} \text{ and } \theta = \arctan(\frac{y}{x})\,.$$

Finally

$$z = \cos(\varphi) = \cos\big(\arcsin\sqrt{x^2 + y^2}\big) = \pm\sqrt{1 - x^2 - y^2}$$

where the sign is determined by which emisphere the point of interest is on. For points lying on the equator we have

$$Dg = \begin{bmatrix} -\sin(\theta) & 0 \\ \cos(\theta) & 0 \\ 0 & -1 \end{bmatrix}$$

and can thus solve for x as a function of y, z by first solving for *theta* and φ in terms of y and z. In this case we obtain

$$\begin{cases} z = \cos(\varphi) \\ y = \sin(\theta)\sin(\varphi) \end{cases} \implies \begin{cases} \varphi = \arccos(z) \\ \theta = \arcsin\big(\frac{y}{\sin\arccos(z)}\big) \end{cases}$$

and finally

$$x = \cos(\theta)\sin(\varphi) = \cos\big(\arcsin(\frac{y}{\sin\arccos(z)})\big) = \pm\sqrt{1-y^2-z^2}\,.$$

Remarks 11.2.11. (a) A C^1-manifold could as well have been defined as being locally the graph of some function. Beware that, just as in the last example considered, the arguments of the function might vary from patch to patch.
(b) A local parametrization $g : U \to V \cap M_m$ of a patch $V \cap M_m$ of a manifold M_m always gives rise to a *local coordinate map* $g^{-1} : V \cap M_m \to U$.
(c) What does the theorem say in terms of coordinate maps? Well, it makes sure that we can always choose coordinates on $M_m \cap V$ such that the coordinate map is just the projection onto a $\mathbb{R}^m$.
(d) Given two overlapping patches $M_n \cap V_1$ and $M_n \cap V_2$ of a manifold M_m with parametrizations g_1 and g_2, we obtain C^1 change of coordinate maps

$$g_2^{-1} \circ g_1 : g_1^{-1}(V_1 \cap V_2 \cap M_m) \to g_2^{-1}(V_1 \cap V_2 \cap M_m)$$
$$g_1^{-1} \circ g_2 : g_2^{-1}(V_1 \cap V_2 \cap M_m) \to g_1^{-1}(V_1 \cap V_2 \cap M_m)\,.$$

(e) Using local coordinates it is possible to transplant calculus from U (which has a linear structure) onto the manifold M_m. For instance, given $f \in C(M_m, \mathbb{R})$, how would we define whether the function is continuously differentiable? A viable definition is

$$f \in C^1(M_m, \mathbb{R}) :\Longleftrightarrow f \circ g \in C^1(U_g, \mathbb{R})\ \forall \text{ coordinate patch } g$$

where U_g is the parametrization domain of g.

11.2.4 Implicit Description of Manifolds

In this section we turn to implicit representations of surfaces/manifolds as the zero-level set of some function. Let $M_m \subset \mathbb{R}^n$ be a C^1-manifold and assume that

$$M_m = \{x \in \mathbb{R}^n \,|\, F(x) = 0\}$$

for some $F \in C^1(\mathbb{R}^n, \mathbb{R}^{n-m})$. Let $\gamma\, C^1\big((0,1), M_m\big)$ a curve in M_m, then it follows that

$$F_j(\gamma(t)) = 0\,\forall\, t \in (0,1)\,,\ j = 1, \dots, n-m\,,$$

which, by differentiation implies

$$0 = \frac{d}{dt}F_j \circ \gamma(t) = \nabla F_j(\gamma(t))\dot\gamma(t)\,\forall\, j = 1, \dots, n-m\,.$$

Geometrically, this means that the vectors $\nabla F_j(\gamma(t))$ and $\dot\gamma(t)$ are orthogonal. Since $\dot\gamma(t)$ is a vector tangent to M_m at $\gamma(t)$ we conclude that

$$\nabla F_j(\gamma(t)) \perp T_{\gamma(t)}M_m \,\forall\, j = 1, \dots, n-m\,.$$

In other words, we can say that the vectors

$$\nabla F_1(\gamma(t)), \dots, \nabla F_{n-m}(\gamma(t))$$

are in the normal space $N_{\gamma(t)}M_m = \big(T_{\gamma(t)}M_m\big)^\perp$ to M_m at $\gamma(t)$. If they were linearly independent, they would span it and thus determine $T_{\gamma(t)}M_m$ itself.

Theorem 11.2.12. *Let $F \in \mathrm{C}^1(\mathbb{R}^n, \mathbb{R}^{n-m})$ with*

$$r(DF(x)) = n - m \,\forall\, x \in \mathrm{L}_c\, F := \{x \in \mathbb{R}^n \,|\, F(x) = c\}\,.$$

Then $\mathrm{L}_c\, F$ is an m-dimensional C^1-manifold.

Proof. Let $x \in \mathrm{L}_c\, F$. Then, by assumption, we can find $n-m$ variables $x_{j_1}, \dots, x_{j_{n-m}}$ such that

$$\partial_{x_{j_k}} F(x)\,,\ k = 1, \dots, n-m\,,$$

are linearly independent. Let $s = (x_{j_k})_{k=1,\dots,n-m}$ and t be the remaining m variables. If we can show that

$$V \cap \mathrm{L}_c\, F = \{(t, s) \,|\, s = f(t)\}$$

for some $V \in \mathcal{U}(x)$, $U \overset{o}{\subset} \mathbb{R}^m$ and $f \in \mathrm{C}^1(U, \mathbb{R}^{n-m})$, we are done. This is a direct consequence of the implicit function theorem since it amounts to solving $F(t, s) = c$ for s as a functions of t which is possible on the very assumption that $D_sF(x)$ be invertible. $\surd$

Example 11.2.13. Consider the set in $\mathbb{R}^3$ defined by

$$x^2 + y^2 + z^2 = c\,.$$

Since

$$DF(x, y, z) = \begin{bmatrix} 2x & 2y & 2z \end{bmatrix} \neq 0 \Longleftrightarrow (x, y, z) \neq 0$$

we see that $\mathrm{L}_c\, F$ is a 2-dimensional C^1-manifold whenever $c \neq 0$. How can you compute the normal space at one of its points?

11.3 Maxima and Minima on Surfaces

Given a manifold M_m a function defined on it, we can ask the question about the existence of extremal points, that is points, of minimum or maximum of it. In Euclidean space we found necessary conditions in terms of the gradient of f. This approach cannot be directly implemented in this case since we don't have a the concept of a gradient for $f \in \mathrm{C}^1(M_m, \mathbb{R})$. What is it then that we can do in order to derive a necessary condition which would provide a computational recipe for searching for extrema?

11.3.1 Lagrange Multipliers

Let $M_m \subset \mathbb{R}^n$ be a C^1-manifold and assume that $f \in \mathrm{C}^1(M_m, \mathbb{R})$. If the manifold is locally described by a parametrization $g : U \to g(U) \subset M_m$, then, clearly, f has a local extremum at $g(x) \in g(U)$ iff $f \circ g$ has one at x. In this case local extrema can be looked for in the way we have learned already by considering the patch maps $f \circ g$ which are defined on open sets of a linear space. But, what if the manifold is described implicitly?
Assume the $M_m = G^{-1}(0)$ for some $G : \mathbb{R}^n \to \mathbb{R}^{n-m}$ and that $f \in \mathrm{C}^1(\mathbb{R}^n, \mathbb{R})$. How can we locate local extrema of $f|_{M_m}$?
Let $k = n - m$ and consider the function

$$H(x, \lambda) := f(x) + \sum_{j=1}^{k} \lambda_j G_j(x)\,,\ x \in \mathbb{R}^n\,,\ \lambda \in \mathbb{R}^k\,.$$

Clearly H coincides with f on the manifold M_m. Its critical points satisfy

$$0 = \nabla H(x, \lambda) = \begin{bmatrix} \nabla f(x) + \sum_{j=1}^{k} \lambda_j \nabla G_j(x) \\ G_1(x) \\ \vdots \\ G_k(x) \end{bmatrix}$$

and are therefore necessarily located on the manifold M_m. The vanishing of the first component says, that at such a point, the gradient of f is a linear combination of vectors normal to the manifold M_m. Its tangential components therefore vanish!

Theorem 11.3.1. *(Lagrange Multipliers)*
Assume that $f \in \mathrm{C}^1(\mathbb{R}^n, \mathbb{R})$ and $G \in \mathrm{C}^1(\mathbb{R}^n, \mathbb{R}^k)$ for some $k \leq n$. Let $\tilde{x} \in \mathbb{R}^n$ with $G(\tilde{x}) = 0$ and such that $r\big(DG(\tilde{x})\big) = k$. If $\tilde{x}$ is a point of local minimum [maximum] for $f|_{G^{-1}(0)}$, that is, if

$$\exists\, U \in \mathcal{U}_{\mathbb{R}^n}(\tilde{x}) \text{ s.t. } f(\tilde{x}) \leq [\geq] f(x)\ \forall\, x \in U \cap G^{-1}(0)\,,$$

then

$$\exists\, \tilde{\lambda} \in \mathbb{R}^k \ s.t. \ H(x, \lambda) = f(x) + \lambda \cdot G(x)$$

has a critical point at $(\tilde{x}, \tilde{\lambda})$.

Warning. The maximal rank condition is important! Consider $G(x, y) = y^3 - x^2$ and $f(x, y) = y$. Then $(0, 0)$ is a point of minimum for f on $G^{-1}(0)$. However, $\nabla f(0, 0) = (0, 1)$ and is clearly not a scalar multiple of $\nabla G(0, 0) = (0, 0)$.

Example 11.3.2. Consider the function $f(x, y) = x^2 - y^2$ and locate points of minimum and maximum within the unit circle $[x^2 + y^2 \leq 1]$. Since $\nabla f(x, y) = 2(x, -y)$ the only candidate point is $(0, 0)$ and therefore in the interior of the circle. By inspecting the function behavior in the vicinity of the origin, we conclude that it is no minimum nor maximum. Extrema in this case can be taken on in the interior or on the boundary. We therefore need to locate extrema on the unit circle itself to complete the task. To that end, consider

$$H(x, y, \lambda) = x^2 - y^2 - \lambda(x^2 + y^2 - 1)$$

the critical points of which are solutions of

$$\begin{cases} 2x + 2\lambda x = 0 \\ -2y + 2\lambda y = 0 \\ x^2 + y^2 - 1 = 0 \end{cases}$$

which has solutions

$$(x, y, \lambda) = (0, \pm 1, 1) \text{ or } (\pm 1, 0, -1)\,.$$

Evaluating f a points nearby we see that $(\pm 1, 0)$ are points of maximum and $(0, \pm 1)$ of minimum. How would you proceed using a parametrization of the circle instead?