Chapter 4

Inverse Function Theorem

北 海 若 曰 ： 井 蛙 不 可 以 語 於 海 者 ， 拘 於 虛 也 ； 夏
蟲 不 可 以 語 於 冰 者 ， 篤 於 時 也 ； 曲 士 不 可 以 語 於
道 者 ， 束 於 教 也 。 今 爾 出 於 崖 涘 ， 觀 於 大 海 ， 乃
知 爾 丑 ， 爾 將 可 與 語 大 理 矣 。 莊 子 秋 水

This chapter is devoted to the proof of the inverse and implicit function theorems. The inverse function theorem is proved in Section 1 by using the contraction mapping principle. Next the implicit function theorem is deduced from the inverse function theorem in Section 2. Section 3 is concerned with various definitions of curves, surfaces and other geometric objects. The relation among these definitions are elucidated by the inverse/implicit function theorems. Finally in Section 4 we prove the Morse Lemma.

4.1 The Inverse Function Theorem

This chapter is concerned with functions between the Euclidean spaces and the inverse and implicit function theorems. We learned these theorems in advanced calculus but the proofs were not emphasized. Now we fill out the gap. Adapting the notations in advanced calculus, a point $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ is sometimes called a vector and we use $|x|$ instead of $\|x\|_2$ to denote its Euclidean norm.

All is about linearization. Recall that a real-valued function on an open interval I is differentiable at some $x_0 \in I$ if there exists some $a \in \mathbb{R}$ such that

$$\lim_{x \to x_0} \left| \frac{f(x) - f(x_0) - a(x - x_0)}{x - x_0} \right| = 0.$$

In fact, the value a is equal to $f'(x_0)$, the derivative of f at x_0. We can rewrite the limit above using the little o notation:

$$f(x_0 + z) - f(x_0) = f'(x_0)z + \circ(z), \quad \text{as } z \to 0.$$

Here $\circ(z)$ denotes a quantity satisfying $\lim_{z\to 0} \circ(z)/|z| = 0$. The same situation carries over to a real-valued function f in some open set in $\mathbb{R}^n$. A function f is called differentiable at p_0 in this open set if there exists a vector $a = (a_1, \cdots, a_n)$ such that

$$f(p_0 + z) - f(p_0) = \sum_{j=1}^{n} a_j z_j + \circ(z) \quad \text{as } z \to 0.$$

Again one can show that the vector a is uniquely given by the gradient vector of f at p_0

$$\nabla f(p_0) = \left(\frac{\partial f}{\partial x_1}(p_0), \cdots, \frac{\partial f}{\partial x_n}(p_0) \right).$$

More generally, a map F from an open set in $\mathbb{R}^n$ to $\mathbb{R}^m$ is called differentiable at a point p_0 in this open set if each component of $F = (f^1, \cdots, f^m)$ is differentiable. We can write the differentiability condition collectively in the following form

$$F(p_0 + z) - F(p_0) = DF(p_0)z + o(z), \tag{4.1}$$

where $DF(p_0)$ is the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by

$$(DF(p_0)z)_i = \sum_{j=1}^{n} a_{ij}(p_0)x_j, \quad i = 1, \cdots, m,$$

where $\big(a_{ij}\big) = \big(\partial f^i/\partial x_j\big)$ is the Jabocian matrix of f. (4.1) shows near p_0, that is, when z is small, the function F is well-approximated by the linear map $DF(p_0)$ up to the constant $F(p_0)$ as long as $DF(p_0)$ is nonsingular. It suggests that the local information of a map at a differentiable point could be retrieved from its a linear map, which is much easier to analyse. This principle, called linearization, is widely used in analysis. The inverse function theorem is a typical result of linearization. It asserts that a map is locally invertible if its linearization is invertible. Therefore, local bijectivity of the map is ensured by the invertibility of its linearization. When $DF(p_0)$ is not invertible, the first term on the right hand side of (4.1) may degenerate in some or even all direction so that $DF(p_0)z$ cannot control the error term $\circ(z)$. In this case the local behavior of F may be different from its linearization.

Theorem 4.1 (**Inverse Function Theorem**). *Let $F : U \to \mathbb{R}^n$ be a C^1-map where U is open in $\mathbb{R}^n$ and $p_0 \in U$. Suppose that $DF(p_0)$ is invertible. There exist open sets V and W containing p_0 and $F(p_0)$ respectively such that the restriction of F on V is a bijection onto W with a C^1-inverse. Moreover, the inverse is C^k when F is $C^k, 1 \le k \le \infty$, in U.*

Example 4.1. The inverse function theorem asserts a local invertibility. Even if the linearization is non-singular everywhere, we cannot assert global invertibility. Let us consider the switching between the cartesian and polar coordinates in the plane:

$$x = r\cos\theta, \quad y = r\sin\theta\ .$$

The function $F: (0,\infty)\times(-\infty,\infty)\to\mathbb{R}^2$ given by $F(r,\theta)=(x,y)$ is a continuously differentiable function whose Jacobian matrix is non-singular except $(0,0)$. However, it is clear that F is not bijective, for instance, all points $(r,\theta+2n\pi), n\in\mathbb{Z}$, have the same image under F.

Example 4.2. An exceptional case is dimension one where a global result is available. Indeed, in MATH2060 we learned that if f is continuously differentiable on (a,b) with non-vanishing f', it is either strictly increasing or decreasing so that its global inverse exists and is again continuously differentiable.

Example 4.3. Consider the map $F: \mathbb{R}^2\to\mathbb{R}^2$ given by $F(x,y)=(x^2,y)$. Its Jacobian matrix is singular at $(0,0)$. In fact, for any point $(a,b), a>0$, $F(\pm\sqrt{a},b)=(a,b)$. We cannot find any open set, no matter how small is, at $(0,0)$ so that F is injective. On the other hand, the map $H(x,y)=(x^3,y)$ is bijective with inverse given by $J(x,y)=(x^{1/3},y)$. However, as the non-degeneracy condition does not hold at $(0,0)$ so it is not differentiable there. In these cases the Jacobian matrix is singular, so the nondegeneracy condition does not hold. We will see that in order the inverse map to be differentiable, the nondegeneracy condition must hold.

A map from some open set in R^n to $\mathbb{R}^m$ is C^k, $1\le k\le\infty$ if all its components belong to C^k. It is called a C^∞-map or a smooth map if its components are C^∞.

The condition that $DF(p_0)$ is invertible, or equivalently the non-vanishing of the determinant of the Jacobian matrix, is called the nondegeneracy condition. Without this condition, the map may or may not be local invertible, see the examples below. Nevertheless, it is necessary for the differentiability of the local inverse. At this point, let us recall the general chain rule.

Let $G: \mathbb{R}^n\to\mathbb{R}^m$ and $F: \mathbb{R}^m\to\mathbb{R}^l$ be C^1 and their composition $H=F\circ G: \mathbb{R}^n\to\mathbb{R}^l$ is also C^1. We compute the first partial derivatives of H in terms of the partial derivatives of F and G. Letting $G=(g_1,\cdots,g_m)$, $F=(f_1,\cdots,f_l)$ and $H=(h_1,\cdots,h_l)$. From

$$h_k(x_1,\cdots,x_n)=f_k(g_1(x),\cdots,g_m(x)),\quad k=1,\cdots,l,$$

we have

$$\frac{\partial h_k}{\partial y_i}=\sum_{i=1}^{n}\frac{\partial f_k}{\partial x_i}\frac{\partial g_i}{\partial x_j}\ .$$

Writing it in matrix form we have

$$DF(G(x))DG(x)=DH(x).$$

For, when the inverse is differentiable, we may apply this chain rule to differentiate the relation $F^{-1}(F(x))=x$ to obtain

$$DF^{-1}(q_0)\ DF(p_0)=I\ ,\quad q_0=F(p_0),$$

where I is the identity map. We conclude that

$$DF^{-1}(q_0) = \big(DF(p_0)\big)^{-1},$$

in other words, the matrix of the derivative of the inverse map is precisely the inverse matric of the derivative of the map. So when the inverse map is C^1, $DF(p_0)$ must be invertible.

Lemma 4.2. *Let L be a linear map from $\mathbb{R}^n$ to itself given by*

$$(Lz)_i = \sum_{j=1}^{n} a_{ij} z_j, \quad i = 1, \cdots n.$$

Then

$$|Lz| \leq \|L\| \ |z|, \quad \forall z \in \mathbb{R}^n,$$

where $\|L\| = \sqrt{\sum_{i,j} a_{ij}^2}$.

Proof. By Cauchy-Schwarz inequality,

$$\begin{aligned} |Lz|^2 &= \sum_i (Lz)_i^2 \\ &= \sum_i \big(\sum_j a_{ij} z_j\big)^2 \\ &\leq \sum_i \big(\sum_j a_{ij}^2\big)\big(\sum_j z_j^2\big) \\ &= \|L\|^2 \ |z|^2 \ . \end{aligned}$$

□

Now we prove Theorem 4.1. We may take $p_0 = F(p_0) = 0$, for otherwise we could look at the new function $\overline{F}(x) = F(x+p_0) - F(p_0)$ instead of $F(x)$, after noting $D\overline{F}(0) = DF(p_0)$. First we would like to show that there is a unique solution for the equation $F(x) = y$ for y near 0. We will use the contraction mapping principle to achieve our goal. After a further restriction on the size of U, we may assume that F is C^1 with $DF(x)$ invertible at all $x \in U$. For a fixed y, define the map in U by

$$T(x) = L^{-1}\left(Lx - F(x) + y\right)$$

where $L = DF(0)$. It is clear that any fixed point of T is a solution to $F(x) = y$. By the lemma,

$$\begin{aligned} |T(x)| &\leqslant \|L^{-1}\| \ |F(x) - Lx - y| \\ &\leqslant \|L^{-1}| \left(|F(x) - Lx| + |y|\right) \\ &\leq \|L^{-1}\| \left(\left| \int_0^1 (DF(tx) - DF(0))dt \ x \right| + |y| \right), \end{aligned}$$

where we have used the formula

$$F(x) - DF(0)x = \int_0^1 \frac{d}{dt}F(tx)dt - DF(0) = \int_0^1 \big(DF(tx) - DF(0)\big)dt\ x,$$

after using the chain rule to get

$$\frac{d}{dt}F(tx) = DF(tx)\cdot x.$$

By the continuity of DF at 0, we can find a small ρ_0 such that

$$\|L^{-1}\|\|DF(x) - DF(0)\| \le \frac{1}{2}, \quad \forall x, \quad |x| \le \rho_0. \tag{4.2}$$

Then for for each y in $B_R(0)$, where R is chosen to satisfy $\|L^{-1}\|R \le \rho_0/2$, we have

$$\begin{aligned}
|T(x)| &\le \|L^{-1}\|\left(\int_0^1 \|(DF(tx) - DF(0))dt\||x| + |y|\right) \\
&\le \frac{1}{2}|x| + \|L^{-1}\||y| \\
&\le \frac{1}{2}\rho_0 + \frac{1}{2}\rho_0 = \rho_0,
\end{aligned}$$

for all $x \in B_{\rho_0}(0)$. We conclude that T maps $\overline{B_{\rho_0}(0)}$ to itself. Moreover, for x_1, x_2 in $B_{\rho_0}(0)$, we have

$$\begin{aligned}
|T(x_2) - T(x_1)| &= \left|L^{-1}\left(F(x_2) - Lx_2 - y\right) - L^{-1}\left(F(x_1) - Lx_1 - y\right)\right| \\
&\leqslant \|L^{-1}\|\ |F(x_2) - F(x_1) - DF(0)(x_2 - x_1)| \\
&\leqslant \|L^{-1}\|\ \left|\int_0^1 DF\left(x_1 + t(x_2 - x_1)\right)(x_2 - x_1)dt - DF(0)(x_2 - x_1)\right|,
\end{aligned}$$

where we have used

$$\begin{aligned}
F(x_2) - F(x_1) &= \int_0^1 \frac{d}{dt}F(x_1 + t(x_2 - x_1))dt \\
&= \int_0^1 DF(x_1 + t(x_2 - x_1))(x_2 - x_1)dt.
\end{aligned}$$

Consequently,

$$|T(x_2) - T(x_1)| \le \frac{1}{2}|x_2 - x_1|.$$

We have shown that $T : \overline{B_{\rho_0}(0)} \to \overline{B_{\rho_0}(0)}$ is a contraction. By the contraction mapping principle, there is a unique fixed point for T, in other words, for each y in the ball $B_R(0)$ there is a unique point x in $\overline{B_{\rho_0}(0)}$ solving $F(x) = y$. Defining $G:\ B_R(0) \to \overline{B_{\rho_0}(0)} \subset X$ by setting $G(y) = x$, G is inverse to F.

Next, we claim that G is continuous. In fact, for $G(y_i) = x_i,\ i = 1, 2$, (not to be mixed up with the x_i above),

$$\begin{aligned}
|G(y_2) - G(y_1)| &= |x_2 - x_1| \\
&= |T(x_2) - T(x_1)| \\
&\le \|L^{-1}\| \left(|F(x_2) - F(x_1) - L(x_2 - x_1)| + |y_2 - y_1|\right) \\
&\le \|L^{-1}\| \left(\left| \int_0^1 \big(DF((1-t)x_1 + tx_2) - DF(0)\big) dt (x_2 - x_1) \right| + |y_2 - y_1| \right) \\
&\le \frac{1}{2}|x_2 - x_1| + \|L^{-1}\| |y_2 - y_1| \\
&= \frac{1}{2}|G(y_2) - G(y_1)| + \|L^{-1}\| |y_2 - y_1|,
\end{aligned}$$

where (4.2) has been used. We deduce

$$|G(y_2) - G(y_1)| \leqslant 2\|L^{-1}\| |y_2 - y_1| \ , \tag{4.3}$$

that's, G is continuous on $B_R(0)$.

Finally, let's show that G is a C^1-map in $B_R(0)$. In fact, for $y_1, y_1 + y$ in $B_R(0)$, using

$$\begin{aligned}
y &= F(G(y_1 + y)) - F(G(y_1)) \\
&= \int_0^1 DF(G(y_1) + t(G(y_1 + y) - G(y_1))dt \ (G(y_1 + y) - G(y_1)),
\end{aligned}$$

we have

$$G(y_1 + y) - G(y_1) = DF^{-1}(G(y_1))y + R,$$

where R is given by

$$DF^{-1}(G(y_1)) \int_0^1 \Big(DF(G(y_1)) - DF(G(y_1) + t(G(y_1 + y) - G(y_1))\Big)(G(y_1 + y) - G(y_1))dt.$$

As G is continuous and F is C^1, we have

$$G(y_1 + y) - G(y_1) - DF^{-1}(G(y_1))y = \circ(1)(G(y_1 + y) - G(y_1))$$

for small y. Using (4.3), we see that

$$G(y_1 + y) - G(y_1) - DF^{-1}(G(y_1))y = \circ(\|y\|) \ ,$$

as $\|y\| \to 0$. We conclude that G is differentiable with derivative equal to $DF^{-1}(G(y_1))$.

After we have proved the differentiability of G, from the formula $DF(G(y))DG(y) = I$ where I is the identity matrix we see that

$$DF^{-1}(y) = (DF(F^{-1}(y)))^{-1}, \quad \forall y \in B_R(0).$$

From linear algebra we know that $DF^{-1}(y)$ can be expressed as a rational function of the entries of the matrix of $DF(F^{-1}(y)$. Consequently, F^{-1} is C^k in y if F is C^k in x for $1 \le k \le \infty$.

The proof of the inverse function theorem is completed by taking $W = B_R(0)$ and $V = F^{-1}(W)$.

Remark 4.1. It is worthwhile to keep tracking and see how ρ_0 and R are determined. Indeed, let

$$M_{DF}(\rho) = \sup_{x \in B_\rho(0)} \|DF(x) - DF(0)\|$$

be the modules of continuity of DF at 0. We have $M_{DF}(\rho) \downarrow 0$ as $\rho \to 0$. From this proof we see that ρ_0 and R can be chosen as

$$M_{DF}(\rho_0) \le \frac{1}{2\|L^{-1}\|}, \quad \text{and } R \le \frac{\rho_0}{2\|L^{-1}\|} \ .$$

Example 4.4. Consider the system of equations

$$\begin{cases} x - y^2 = a, \\ x^2 + y + y^3 = b. \end{cases}$$

We know that $x = y = 0$ is a solution when $(a, b) = (a, b)$. Can we find the range of (a, b) so that this system is solvable? Well, let $F(x, y) = (x - y^2, x^2 + y + y^3)$. We have $F(0, 0) = (0, 0)$ and DF is given by the matrix

$$\begin{pmatrix} 1 & -2y \\ \\ 2x & 1 + 3y^2 \end{pmatrix},$$

which is nonsingular at $(0, 0)$. In fact the inverse matrix of $DF((0, 0))$ is given by the identity matrix, hence $\|L^{-1}\| = 1$ in this case. According to Remark 4.1 a good ρ_0 could be found by solving $M_{DF}(\rho_0) = 1/2$. We have $\|DF((x, y)) - DF((0, 0))\| = 4y^2 + 4x^2 + 9y^2$, which, in terms of the polar coordinates, is equal to $4r^2 + 9\sin^4\theta$. Hence the maximal value is given by $4r^2 + 9r^4$, and so ρ_0 could be chosen to be any point satisfying $4\rho_0^2 + 9\rho_0^4 \le 1/2$. A simple choice is $\rho_0 = \sqrt{1/26}$. Then R is given by $\sqrt{26}/52$. We conclude that whenever a, b satisfy $a^2 + b^2 \le 1/104$, this system is uniquely solvable in the ball $B_{\rho_0}((0, 0))$.

Example 4.5. Determine all points where the function $F(x, y) = (xy^2 - \sin \pi x, y^2 - 25x^2 + 1)$ has a local inverse and find the partial derivatives of the inverse. Well, the Jacobian matrix of F is given by

$$\begin{pmatrix} y^2 - \pi \cos \pi x & 2xy \\ \\ -50x & 2y \end{pmatrix}.$$

Hence, F admits a local inverse at points (x, y) satisfying

$$2y(y^2 - \pi \cos \pi x) + 100x^2 y \neq 0 \ .$$

Derivatives of the inverse function, denoted by $G = (g_1, g_2)$, can be obtained by implicit differentiation of the relation

$$(u, v) = F(G(u, v)) = (g_1 g_2^2 - \sin \pi g_1, g_2^2 - 25g_1^2 + 1),$$

where g_1, g_2 are functions of (u, v). We have

$$\frac{\partial g_1}{\partial u} g_2^2 + 2g_1 g_2 \frac{\partial g_2}{\partial u} - \pi \cos \pi g_1 \ \frac{\partial g_1}{\partial u} = 1,$$

$$2g_2 \frac{\partial g_2}{\partial u} - 50g_1 \frac{\partial g_1}{\partial u} = 0,$$

$$\frac{\partial g_1}{\partial v} g_2^2 + 2g_1 g_2 \frac{\partial g_2}{\partial v} - \pi \cos \pi g_1 \ \frac{\partial g_1}{\partial v} = 0,$$

$$2g_2 \frac{\partial g_2}{\partial v} - 50g_1 \frac{\partial g_1}{\partial v} = 1.$$

The first and the second equations form a linear system for $\partial g_i / \partial u, i = 1, 2$, and the third and the fourth equations form a linear system for $\partial g_i / \partial v, \ i = 1, 2$. By solving it (the solvability is ensured by the invertibility of the Jacobian matrix) we obtain the partial derivatives of the inverse function G. Nevertheless, it is too tedious to carry it out here. An alternative way is to find the inverse matrix of the Jacobian DF. In principle we could obtain all partial derivatives of G by implicit differentiation and solving linear systems.

We end this section by rephrasing the inverse function theorem.

A C^k-map F between open sets V and W is a "C^k-diffeomorphism" if F^{-1} exists and is also C^k. Let $f_1, f_2, \cdots, f_n$ be C^k-functions defined in some open set in $\mathbb{R}^n$ whose Jacobian matrix of the map $F = (f_1, \cdots, f_n)$ is non-singular at some point p_0 in this open set. By Theorem 4.1 F is a C^k-diffeomorphism between some open sets V and W containing p_0 and $F(p_0)$ respectively. To every function Φ defined in W, there corresponds a function defined in V given by $\Psi(x) = \Phi(F(x))$, and the converse situation holds. Thus every C^k-diffeomorphism gives rise to a "local change of coordinates".

4.2 The Implicit Function Theorem

Next we deduce the implicit function theorem from the inverse function theorem.

Theorem 4.3 (Implicit Function Theorem). *Consider C^1-map $F : U \to \mathbb{R}^m$ where U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$. Suppose that $(p_0, q_0) \in U$ satisfies $F(p_0, q_0) = 0$ and*

$D_yF(p_0, q_0)$ is invertible in $\mathbb{R}^m$. There exist an open set $V_1 \times V_2$ in U containing (p_0, q_0) and a C^1-map $\varphi : V_1 \to V_2$, $\varphi(p_0) = q_0$, such that

$$F(x, \varphi(x)) = 0 \ , \qquad \forall x \in V_1 \ .$$

The map φ belongs to C^k when F is $C^k, 1 \le k \le \infty$, in U. Moreover, if ψ is another C^1-map in some open set containing p_0 to V_2 satisfying $F(x, \psi(x)) = 0$ and $\psi(p_0) = q_0$, then ψ coincides with φ in their common set of definition.

The notation $D_yF(p_0, q_0)$ stands for the linear map associated to the Jocabian matrix $(\partial F_i/\partial y_j(p_0, q_0))_{i,j=1,\cdots,m}$ where p_0 is fixed.

Proof. Consider $\Phi : U \to \mathbb{R}^n \times R^m$ given by

$$\Phi(x, y) = (x, F(x, y)).$$

It is evident that $D\Phi(x, y)$ is invertible in $\mathbb{R}^n \times \mathbb{R}^m$ when $D_yF(x, y)$ is invertible in $\mathbb{R}^m$. By the inverse function theorem, there exists a C^1-inverse $\Psi = (\Psi_1, \Psi_2)$ from some open W in $\mathbb{R}^n \times \mathbb{R}^m$ containing $(p_0, 0)$ to an open subset of U. By restricting W further we may assume $\Psi(W)$ is of the form $V_1 \times V_2$. For every $(x, z) \in W$, we have

$$\Phi(\Psi_1(x, z), \Psi_2(x, z)) = (x, z),$$

which, in view of the definition of Φ, yields

$$\Psi_1(x, z) = x, \text{ and } F((\Psi_1(x, z), \Psi_2(x, z)) = z.$$

In other words, $F(x, \Psi_2(x, z)) = z$ holds. In particular, taking $z = 0$ gives

$$F(x, \Psi_2(x, 0)) = 0, \quad \forall x \in V_1 \ ,$$

so the function $\varphi(x) \equiv \Psi_2(x, 0)$ satisfies our requirement.

By restricting V_1 and V_2 further if necessary, we may assume the matrix

$$\int_0^1 D_yF(x, y_1 + t(y_2 - y_1)dt$$

is nonsingular for $(x, y_1), (x, y_2) \in V_1 \times V_2$. Now, suppose ψ is a C^1-map defined near x_0 satisfying $\psi(p_0) = q_0$ and $F(x, \psi(x)) = 0$. We have

$$\begin{aligned} 0 &= F(x, \psi(x)) - F(x, \varphi(x)) \\ &= \int_0^1 D_yF(x, \varphi(x) + t(\psi(x) - \varphi(x))dt(\psi(x) - \varphi(x)), \end{aligned}$$

for all x in the common open set they are defined. This identity forces that ψ coincides with φ in this open set. The proof of the implicit function is completed, once we observe that the regularity of φ follows from the inverse function theorem. □

Example 4.6. Let $F : \mathbb{R}^5 \to \mathbb{R}^2$ be given by $F(x, y, z, u, v) = (xy^2 + xzu + yv^2 - 3, u^3yz + 2xv - u^2v^2 - 2)$. We have $F(1, 1, 1, 1, 1) = (0, 0)$. Show that there are functions $f(x, y, z), g(x, y, z)$ satisfying $f(1, 1, 1) = g(1, 1, 1) = 1$ and $F(x, y, z, f(x, y, z), g(x, y, z)) = (0, 0)$ for (x, y, z) near $(1, 1, 1)$. We compute the "partial" Jacobian matrix of F in (u, v):

$$\begin{pmatrix} xz & 2yv \\ 3u^2yz - 2uv^2 & 2x - 2u^2v \end{pmatrix}.$$

Its determinant at $(1, 1, 1, 1, 1)$ is equal to -2, so we can apply the implicit function theorem to get the desired result. The partial derivatives of f and g can be obtained by implicit differentiations. For instance, to find $\partial f/\partial y$ and $\partial g/\partial y$ we differentiate the relation

$$(xy^2 + xzf + yg^2 - 3, f^3yz + 2xg - f^2g^2 - 2) = (0, 0)$$

to get

$$2xy + xz\frac{\partial f}{\partial y} + g^2 + 2yg\frac{\partial g}{\partial y} = 0,$$

and

$$f^3z + 3f^2yz\frac{\partial f}{\partial y} + 2x\frac{\partial g}{\partial y} - 2fg^2\frac{\partial f}{\partial y} - 2f^2g\frac{\partial g}{\partial y} = 0.$$

By solving this linear system we can express $\partial f/\partial y$ and $\partial g/\partial y$ in terms of x, y, z, f and g. Similarly we can do it for the other partial derivatives.

It is interesting to note that the inverse function theorem can be deduced from the implicit function theorem. Thus they are equivalent. To see this, keeping the notations used in Theorem 4.1. Define a map $\widetilde{F} : U \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\widetilde{F}(x, y) = F(x) - y.$$

Then $\widetilde{F}(p_0, q_0) = 0, q_0 = F(p_0)$, and $D\widetilde{F}(p_0, q_0)$ is invertible. By Theorem 4.3, there exists a C^1-function φ from near q_0 satisfying $\varphi(q_0) = p_0$ and $\widetilde{F}(\varphi(y), y) = F(\varphi(y)) - y = 0$, hence φ is the local inverse of F.

4.3 Curves and Surfaces

A **parametric curve** is a C^1-map γ from an interval I to $\mathbb{R}^2$ which satisfies $|\gamma'| \neq 0$ on I. In this definition a curve is not a geometric object but a map. The condition $|\gamma'| \neq 0$ ensures that the curve does not collapse into a point. A parametric curve becomes a geometric curve only if we look at its image. On the other hand, a **non-parametric curve** is a subset Γ in $\mathbb{R}^2$ satisfies the following condition: For each $p_0 = (x_0, y_0) \in \Gamma$, there exist an open rectangle $R = (x_1, x_2) \times (y_1, y_2)$ containing p_0 such that either for some C^1-function f from (x_1, x_2) to (y_1, y_2) such that $\Gamma \cap R = \{(x, f(x)) : x \in (x_1, x_2)\}$, or for some C^1-function g from (y_1, y_2) to (x_1, x_2) such that $\Gamma \cap R = \{(g(y), y) : y \in (y_1, y_2)\}$. In the following we show the essential equivalence of these two definitions.

Proposition 4.4. *(a). Let* $\gamma : (t_1, t_2) \to \mathbb{R}^2$ *be a parametric curve. For each* $t_0 \in (t_1, t_2)$*, there exists* $(t_1', t_2') \subset (t_1, t_2)$ *containing* t_0 *such that* $\Gamma_1 = \{\gamma(t) \in \mathbb{R}^2 : \ t \in (t_1', t_2')\}$ *is a non-parametric curve.*

(b). Let Γ *be a non-parametric curve in* $\mathbb{R}^2$ *and* p_0 *be a point on it. There exist some open set* G *containing* p_0 *and a parametric curve from some open interval to* G *whose image coincides with* $\Gamma \cap G$*.*

Proof. (a). Let $\gamma = (\gamma_1, \gamma_2)$ be a parametric curve from (t_1, t_2) to $\mathbb{R}^2$ and $p_0 = \gamma(t_0),\ t_0 \in (t_1, t_2)$, a point on its image Γ. From the condition $|\gamma'(t)| \neq 0,\ \forall t \in (t_1, t_2)$, we assume $\gamma_1'(t_0) \neq 0$ first. By the inverse function theorem, we can find a subinterval (t_1', t_2') of (t_1, t_2) containing t_0 and some (x_1, x_2) containing $\gamma_1(t_0)$ such that $\gamma_1 : (t_1', t_2') \to (x_1, x_2)$ is invertible. Denoting the inverse function by τ. We have

$$\Gamma \cap G = \{\gamma(t) : \ t \in (t_1', t_2')\} = \{(x, f(x)) : \ x \in (x_1, x_2)\},$$

where $G = (x_1, x_2) \times (-\infty, \infty)$ and $f(x) = \gamma_2(\tau(x))$ is a C^1-function. We have proved that the image of γ is a non-parametric curve locally. If instead $\gamma_2'(t_0) \neq 0$, a similar argument shows that Γ is a locally a graph over y instead of a graph over x.

(b). Let Γ be a non-parametric curve. According to the definition and because it is always possible to find an open rectangle containing p_0 in G, there are some $R = (x_1, x_2) \times (y_1, y_2)$ containing p_0 and, say, a C^1-function $f : (x_1, x_2) \to (y_1, y_2)$ such that $\Gamma \cap R$ coincides with $\{(x, f(x) : \ x \in (x_1, x_2)\}$. But then the parametric curve $\gamma : (x_1, x_2) \to \mathbb{R}^2$ given by $\gamma(t) = (t, f(t))$ satisfies our requirement. The case when g exists instead of f can be handled similarly. Hence every non-parametric curve is the image of a parametric curve locally. □

Example 4.7. In the statement of Proposition 4.4, we need to restrict to a smaller interval (t_1', t_2'). To explain its necessity image a parametric curve γ_1 defined on (0,1) that, after passing a point $p_0 = \gamma_1(t_1)$ at time t_1, returns to hit it at some instance t_2 transversally. Therefore, near p_0 the image of γ_1 cannot be the graph of a single function no matter it is over the x- or the y-axis. It is possible only if we restrict to a small interval containing t_1 or t_2.

Example 4.8. Even if there is no self-intersection in the image, we still need to restrict (t_1', t_2') for another reason. For instance, consider the parametric curve $\gamma_2(t) = e^t(\cos t, \sin t),\ t \in (0, \infty)$. The image is a spiral starting at the point $(1, 0)$ that coiling around the origin infinitely many times. It never intersects itself and eventually goes to infinity. No matter it is over an interval in the x- or the y-axis, the image consists infinitely many layers of graphs. Only if we restrict to some suitable (t_1', t_2') do we have a non-parametric curve.

Example 4.9. A parametric curve can be defined in any interval. In order to compare it with non-parametric curves, we have restricted to open intervals. A parametric curve $\gamma : [a, b] \to \mathbb{R}^2$ is called a closed curve or a loop if their endpoints coincides with equal derivative, that is, $\gamma(a) = \gamma(b)$ and $\gamma'(a) = \gamma'(b)$. When this happens, we can extend this curve as a periodic function in $(-\infty, \infty)$ with period $b - a$. It follows that Proposition 4.1 applies to closed curves as well.

Example 4.10. While locally a non-parametric curve arises from a parametric curve and vice versa. The global property may be different. Let us use a simple example to illustrate this point. Consider the unit circle $c : [0, 4\pi] \to \mathbb{R}^2$ given by $c(\theta) = (\cos\theta, \sin\theta)$. While the image is the unit circle, as t runs from 0 to 4π, the unit circle is transversed twice. If we calculate the length of the circle using the formula

$$L(c) = \int_0^{4\pi} \sqrt{\dot{c}_1^2 + \dot{c}_2^2}\, dt,$$

we will get $L(c) = 4\pi$ instead of 2π, the "geometric" length of the circle.

Next we consider surfaces in space. A map $\sigma : (s_1, s_2) \times (t_1, t_2) \times \to \mathbb{R}^3$ is called a **parametric surface** if $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ is continuously differentiable and the vectors

$$\left(\frac{\partial\sigma_1}{\partial s}, \frac{\partial\sigma_2}{\partial s}, \frac{\partial\sigma_3}{\partial s}\right) \text{ and } \left(\frac{\partial\sigma_1}{\partial t}, \frac{\partial\sigma_2}{\partial t}, \frac{\partial\sigma_3}{\partial t}\right)$$

are linear independent at every point in $(s_1, s_2) \times (t_1, t_2)$. The linear independence requirement ensures that the image does not collapse into a point, a curve or something of dimension less than 2. A **non-parametric surface** is a subset Σ in $\mathbb{R}^3$ satisfying: For each $p_0 \in \Sigma$, there exists an open rectangular box $B = (x_1, x_2) \times (y_1, y_2) \times (x_3, y_3)$ containing p_0 and either there exists a C^1-function $f : (x_1, x_2) \times (y_1, y_2) \to (z_1, z_2)$ satisfying

$$\Sigma \cap B = \big\{(x, y, f(x, y)) : \ (x, y) \in (x_1, x_2) \times (y_1, y_2)\big\},$$

or there exists a C^1-function $g : (y_1, y_2) \times (z_1, z_2) \to (x_1, x_2)$ satisfying

$$\Sigma \cap B = \big\{(g(y, z), y, z) : \ (y, z) \in (y_1, y_2) \times (z_1, z_2)\big\},$$

or there exists a C^1-function $h : (x_1, x_2) \times (z_1, z_2) \to (y_1, y_2)$ satisfying

$$\Sigma \cap B = \big\{(x, h(x, z), z) : \ (x, z) \in (x_1, x_2) \times (z_1, z_2)\big\}.$$

Parallel to the case of curves, we have

Proposition 4.5. *(a). Let* $\sigma : \ (s_1, s_2) \times (t_1, t_2) \to \mathbb{R}^3$ *be a parametric surface and* Σ *its image. For each* $p_0 = (s_0, t_0)$ *in* Σ*, there exists some open set* $G \subset (s_1, s_2) \times (t_1, t_2)$ *containing* (s_0, t_0) *so that* $\Sigma_1 = \{\sigma(s, t) : \ (s, t) \in G\}$ *is a non-parametric surface.*

(b). Let Σ *be a non-parametric surface in* $\mathbb{R}^3$ *and* $p_0 \in \Sigma$*. There exist some open rectangular box* B *containing* p_0 *such that* $\Sigma \cap B$ *is the image of a parametric surface.*

Proof. (a). By assumption the matrix formed by the two vectors $(\partial\sigma_1/\partial s, \partial\sigma_2/\partial s, \partial\sigma_3/\partial s)$ and $(\partial\sigma_1/\partial t, \partial\sigma_2/\partial t, \partial\sigma_3/\partial t)$ are of rank 2. From linear algebra, we know that the three column vectors

$$\left(\frac{\partial\sigma_1}{\partial s}, \frac{\partial\sigma_1}{\partial t}\right), \quad \left(\frac{\partial\sigma_2}{\partial s}, \frac{\partial\sigma_2}{\partial t}\right), \quad \left(\frac{\partial\sigma_3}{\partial s}, \frac{\partial\sigma_3}{\partial t}\right)$$

span $\mathbb{R}^2$. We can pick two independent vectors among these them. Assuming that it is the first two at the point (s_0, t_0), the matrix

$$\begin{pmatrix} \dfrac{\partial\sigma_1}{\partial s} & \dfrac{\partial\sigma_2}{\partial s} \\ \\ \dfrac{\partial\sigma_1}{\partial t} & \dfrac{\partial\sigma_2}{\partial t} \end{pmatrix}$$

is nonsingular at (s_0, t_0) and we can appeal to the inverse function theorem to conclude that the map $(s,t) \mapsto (\sigma_1(s,t), \sigma_2(s,t))$ from some open G in $(s_1, s_2) \times (t_1, t_2)$ containing (s_0, t_0) onto some open rectangle $R = (x_1, x_2) \times (y_1, y_2)$ has a C^1-inverse Φ. It follows that

$$\Sigma_1 \equiv \{\sigma(s,t) : \ (s,t) \in G\} = \{(x, y, f(x,y)) : \ (x,y) \in R\},$$

where $f(x,y) = \sigma_3(\Phi(x,y)$ is a non-parametric surface in the rectangular box$R \times (-\infty, \infty)$.

(b). Let Σ be a non-parametric surface and p_0 a point on Σ. There exists a rectangular box $B = (x_1, x_2) \times (y_1, y_2) \times (z_1, z_2)$ and a function, say, f from $(x_1, x_2) \times (y_1, y_2)$ to (z_1, z_2) such that

$$\Sigma \cap B = \{(x, y, f(x,y)) : \ (x,y) \in (x_1, x_2) \times (y_1, y_2)\}.$$

It is clear that $\Sigma \cap B$ is the image of the parametric surface $(x,y) \mapsto (x, y, f(x,y))$. The other cases are similar.

□

Example 4.11. Consider the set given by $\sigma : (\theta, z) \to \mathbb{R}^3$ given by

$$\sigma(\theta, z) = (a\cos\theta, a\sin\theta, z),$$

where a is a non-negative number. We have

$$\left(\frac{\partial\sigma_1}{\partial\theta}, \frac{\partial\sigma_2}{\partial\theta}, \frac{\partial\sigma_3}{\partial\theta}\right) = (-a\sin\theta, a\cos\theta, 0)$$

and

$$\left(\frac{\partial\sigma_1}{\partial z}, \frac{\partial\sigma_2}{\partial z}, \frac{\partial\sigma_3}{\partial z}\right) = (0, 0, 1),$$

which are linearly independent for $a > 0$, so Σ defines a non-parametric surface according to Proposition 4.2. However, when $a = 0$, the first vector becomes $(0, 0, 0)$ and Proposition 4.2 cannot apply. In fact, the map σ degenerates into the z-axis.

To generalize, a **parametric hypersurface** is a map σ from $\prod_{j=1}^{n-1}(t_j^1, t_j^2) \to \mathbb{R}^n$ such that the $(n-1)$-vectors in $\mathbb{R}^n$

$$\left(\frac{\partial \sigma_1}{\partial t_j}, \cdots, \frac{\partial \sigma_n}{\partial t_j}\right), \quad j = 1, \cdots, n-1,$$

are linearly independent at every point $(t_1, \cdots, t_{n-1})$. A **non-parametric hypersurface** is a set Σ in $\mathbb{R}^n$ such that for each p_0 in Σ, there exist an open rectangular box $B = \prod_{j=1}^{n}(x_j^1, x_j^2)$ containing p_0 and a function f from $\prod_{j=1, j\neq k}^{n}(x_j^1, x_j^2)$ to (x_k^1, x_k^2) such that $\Sigma \cap B$ coincides with $\{(x_1, \cdots, f(x'), \cdots, x_n)$ where x_k is replaced by $f(x')$ and $x' = (x_1, \cdots, x_n)$ with x_k deleted.

Proposition 4.6. *(a). Let $\sigma : \prod_{j=1}^{n-1}(t_j^1, t_j^2) \to \mathbb{R}^n$ be a parametric hypersurface and Σ its image. For each $p_0 = \sigma(t^0)$ in Σ, there exists some open set $G \subset \prod_{j=1}^{n-1}(t_j^1, t_j^2)$ containing t^0 so that $\Sigma_1 = \{\sigma(t) : \ t \in G\}$ is a non-parametric hypersurface.*

(b). Let Σ be a non-parametric hypersurface in $\mathbb{R}^n$ and $p_0 \in \Sigma$. There exist some open rectangular box B containing p_0 such that $\Sigma \cap B$ is the image of a parametric hypersurface.

The proof of this proposition is similar to that of the previous proposition and is omitted.

Another useful way to describe a hypersurface is to express it as the zero set or locus of some function. For instance, a non-parametric curve in $\mathbb{R}^2$ may be described as the set $\{(x, y) : \ f(x, y) = 0\}$ for some function f and a non-parametric surface in $\mathbb{R}^3$ may be described as the set $\{(x, y, z) : \ g(x, y, z) = 0\}$ for some function g and etc. In general, by the implicit function theorem we have

Proposition 4.7. *Let f be a continuously differentiable function defined on some open set G in $\mathbb{R}^n$ and $\mathcal{Z}$ its zero set, which is assumed to be nonempty. Let $p_0 = (x_1^0, \cdots, x_n^0)$ be a point on $\mathcal{Z}$ such that $\nabla f(p_0) \neq (0, \cdots, 0)$. If $\partial f/\partial x_n(p_0) \neq 0$, say, there exist an open rectangular box B containing $(x_1^0, \cdots, x_{n-1}^0)$, a C^1-function $\varphi : B \to G$ and an open set $G_1 \subset G$ such that*

$$\mathcal{Z} \cap G_1 = \big\{\big(x_1, \cdots, x_{n-1}, \varphi(x_1, \cdots, x_{n-1})\big) : \ (x_1, \cdots, x_{n-1}) \in B\big\}.$$

In particular, $\mathcal{Z} \cap G_1$ is a non-parametric hypersurface.

Example 4.12. In summarizing, we note there are three ways of description of the unit circle centered at the origin. First, it could be described as a parametric curve $\gamma(t) = (\cos t, \sin t), \ t \in [0, 2\pi]$. Next, it is a non-parametric curve described either as the set $\{(x, \pm\sqrt{1-x^2})$ or $\{(\pm\sqrt{1-y^2}, y)$. Finally, it is the zero set of the function $f(x, y) = x^2 + y^2 - 1$.

The last point of view, namely, to regard a hypersurface as the zero set of some function, can be generalized to treat more other geometric objects. A **parametric k-surface** is a continuously differentiable map φ from $R = (a_1, b_1) \times (a_k, b_k)$ to $\mathbb{R}^n$, $1 \leq k \leq$

$n-1$, where the matrix $(\partial\varphi_i/\partial t_j)_{k\times n}$ has rank k everywhere. A parametric k-surface is a parametric curve when $k=1$ and a hypersurface when $k=n-1$. While a non-parametric description of these k-surfaces for $k,\ 1<k<n-1$, is tedious and not so useful, expressing them as the common zero set of several functions is much better. The following result is again a consequence of the implicit function theorem.

Proposition 4.8. *Let $\varphi : R \to \mathbb{R}^n$ be a parametric k-surface and Φ be its image in $\mathbb{R}^n$. For each $p_0 \in \Phi$, there exist an open set G containing p_0 and C^1-functions $f_1, \cdots, f_{n-k}$, defined on G such that*

$$\Phi \cap G = \big\{x \in \mathbb{R}^n :\ f_j(x) = 0,\ j = 1, \cdots, n-k\big\}.$$

4.4 The Morse Lemma

We present a further application of the inverse function theorem. A twice continuously differentiable function or a C^2-function for short is a real-valued function whose second partial derivatives are continuous. Let f be a C^2-function defined in some open set in $\mathbb{R}^n$. We will use $\big(h_{ij}\big)_{i,j=1,\cdots,n}$ to denote its Hessian matrix, that is,

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

By Taylor's expansion theorem, we have

$$f(x) - f(p_0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(p_0)(x_i - p_i) + \sum_{i,j=1}^n h_{ij}(x_0)(x_i-p_i)(x_j-p_j) + \circ(|x-p_0|^2).$$

When p_0 is a critical point of f, the first term on the right hand side vanishes so the second order term becomes important. We call p_0 a **non-degenerate critical point** if the Hessian matrix is non-singular at p_0. From linear algebra we know that every symmetric matrix can be diagonalized by a rotation of axes. Henceforth in a suitable coordinate system $(h_{ij}(p_0))$ can be taken to be a diagonal matrix $(\lambda_j\delta_{ij})$, where λ_j's are the eigenvalues of the Hessian matrix. The number of negative eigenvalues is called the **index** of the critical point. The Morse lemma asserts more can be done. It guarantees a "normal form" of the function near a non-degenerate critical point.

Theorem 4.9 (**Morse Lemma**). *Let f be a smooth function in some open set in $\mathbb{R}^n$ and p_0 be a non-degenerate critical point for f in this open set. There exists a smooth, local change of coordinates $x = \Phi(y), p_0 = \Phi(0)$, such that*

$$\widetilde{f}(y) = f(\Phi(y)) = -y_1^2 - y_1^2 - \cdots - y_m^2 + y_{m+1}^2 + \cdots + y_n^2,$$

where $m, 0 \le m \le n$ is the index of the critical point.

Proof. Replacing f by $f(x+x_0)-f(x_0)$ we may assume $f(0)=0$ and 0 is a non-degenerate critical point of f. Moreover, by a suitable rotation of axes, the Hessian matrix becomes a diagonal one. In particular, $\partial^2 f/\partial x_n^2(0)\neq 0$. By replacing f by $-f$ when necessary, we may assume $\partial^2 f/\partial x_n^2(0)>0$. We will use induction on the dimension n in the following statement:

There is a local change of coordinates such that f assumes the form

$$\delta_1 y_1^2+\delta_2 y_2^2+\cdots+\delta_n y_n^2,$$

where $\delta_i\in\{1,-1\},\ i=1\cdots,n.$

Using the fact that a change of coordinates preserves the index of the critical point, a consequence from linear algebra, the number the negative δ_j's must equal to m.

When $n=1$,

$$\begin{aligned} f(x) &= f(0)+\int_0^1 \frac{df}{dt}(tx)dt \\ &= \int_0^1 f'(tx)dt\ x \\ &\equiv g(x)x. \end{aligned}$$

From $f'(x)=g'(x)x+g(x)$ we know that $g(0)=0$. Repeating the argument above to g, $g(x)=h(x)x$ for some smooth h. Therefore, we have $f(x)=h(x)x^2$ near 0. From $f''(x)=h''(x)x^2+2h'(x)x+2h(x)$, we see that $h(0)=f''(0)/2>0$. By defining $\Phi(x)=\sqrt{h(x)}x$, $\Phi'(0)>0$, so by the inverse function theorem, $y=\Phi(x)$ forms a local change of coordinates in which $\widetilde{f}(\Phi^{-1}(y))\equiv f(x)=y^2$.

Next, assuming the statement holds for dimension $n-1$, we establish it for n. Let $x=(x',z)$ where $x'=(x_1,\cdots,x_{n-1})$. The function $\varphi(x',z)\equiv\partial f/\partial x_n(x',z)$ satisfies $\varphi(0,0)=0$ and $\partial\varphi/\partial x_n(0,0)>0$. By the implicit function theorem there exists an open set $V\times W\subset\mathbb{R}^{n-1}\times\mathbb{R}$ containing $(0,0)$ such that $\varphi(x',h(x'))=0$ holds for some smooth h from V to W. Now consider the function

$$g(x',z)=f(x',z)-f(x',h(x')),\quad (x',z)\in V\times W.$$

We have $g(x',h(x'))=0$ and $\partial g/\partial x_n(x',h(x'))=\varphi(x',h(x'))=0$. As for each fixed x', $\partial^2 g/\partial x_n^2(x',z)=\partial^2 f/\partial x_n^2(x',z)>0$ by shrinking $V\times W$ a little, we may assume $(x',h(x'))$ is the unique minimum for $g(x',z)$ in this open set, so g is non-negative. We next claim that g can be written as the square root of some smooth function j. Indeed, we have

$$\begin{aligned} g(x',z) &= g(x',h(x'))+\int_0^1\frac{dg}{dt}(x',h(x')+t(z-h(x'))dt \\ &= \int_0^1\frac{\partial g}{\partial x_n}(x',h(x')+t(z-h(x'))dt(z-h(x')) \\ &\equiv k(x',z)(z-h(x')), \end{aligned}$$

where k is smooth. By differentiating this relation, we have

$$\frac{\partial g}{\partial x_n}(x', h(x')) = \frac{\partial k}{\partial x_n}(x', z)(z - h(x')) + k(x', h(x')), \tag{4.4}$$

which shows that $k(x', h(x')) = 0$. Arguing as before, we can find a smooth function j so that $k(x', h(x')) = j(x', z)(z - h(x'))$. It follows that

$$g(x', z) = k(x', z)(z - h(x')) = j(x', z)(z - h(x'))^2$$

holds. By the chain rule,

$$\frac{\partial k}{\partial x_n} = \frac{\partial j}{\partial x_n}(z - h(x')) + j(x', z).$$

Since $\partial k/\partial x_n(0,0) = \partial^2 g/\partial x_n(x', z) > 0$, $j(x', z) > 0$ in $V \times W$ and hence $\sqrt{j}$ is smooth. We have succeeded in showing

$$f(x', z) = \left(\sqrt{j}(z - h(x')\right)^2 + f(x', h(x')), \quad \forall V \times W.$$

From $\partial j/\partial x_n(0,0) \neq 0$ we may also assume $T : (x', z) \mapsto (x', \sqrt{j}(z - h(x'))$ is a local change of coordinates. Under this new coordinates,

$$\widetilde{f}(x', u) \equiv f(T^{-1}(x, u)) = u^2 + f(x', h(x')), \quad (x', u) \in V_1 \times W_1,$$

for some $V_1 \times W_1 \subset V \times W$ containing $(0,0)$. Now, one can verify that $0 \in V_1$ is a non-degenerate critical point of the function $f_1(x') \equiv f(x', h(x'))$, so by induction hypothesis, there exists a change of coordinates from x' to some $u_1, \cdots, u_{n-1}$ such that $f_1(x')$ becomes $\sum_{i=1}^{n-1} \delta_i u_i^2$. By composing these two local changes of coordinates, we finally obtain a local change of coordinates $(x', z) \mapsto (u_1, \cdots, u_{n-1}, j)$ so that f becomes $\sum_{i=1}^{n} \delta_i u_i^2$ locally. We have completed the proof of Morse lemma.

□

Comments on Chapter 4. Inverse and implicit function theorems, which reduce complicated structure to simpler ones via linearization, are the most frequently used tool in the study of the local behavior of maps. We learned these theorems and some of its applications in MATH2010 already. In view of this, we basically provide detailed proofs here but leave out many standard applications. You may look up Fitzpatrick, Advance Calculus, to refresh your memory. By the way, the proof in this book does not use the contraction mapping principle. I do know a third proof besides these two.

We discuss the definition of curves, surfaces, etc in some details. Although they are elementary and you may have learn it here or there. I believe it is worthwhile to discuss it in a synthetic way. One step further is the definition of a manifold, which contains all curves, surfaces and k-surfaces. It is the object of study in modern differential geometry. You may google to learn more.

Morse lemma, which provides the "normal form" near a non-degenerate critical point, is the starting point of Morse theory. A smooth function defined in a manifold (take it to be $\mathbb{R}^n$ for simplicity) is called a Morse function if all its critical points are non-degenerate. Let Σ_c be its level set $\{x \in \mathbb{R}^n : \ f(x) = c\}$ where f is a Morse function. This theory is mainly concerned with how the topology of Σ_c changes as c varies. When c_0 is not a critical value, that is $\nabla f(x) \neq 0, \ x \in \Sigma_c$, the topology does not change for all c close to c_0. However, the topology of Σ_c changes when Σ_c contains a critical point. Therefore, the topology of the underlying manifold can be studied via the Morse functions defined on it. A classic on this topic is J. Milnor, Morse Theory. Look up Morse theory in Wikipedia for more.