In this note we will consider some basic results related to the differentiation of functions of several variables.

Let $U \subseteq \mathbb{R}^n$ be an open set, and let f be a map $U \to \mathbb{R}^m$. The map f has the components $f_1, \dots, f_m : U \to \mathbb{R}$. We say that f is C^1 if each component is continuously differentiable, i.e. the first order partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous on U. More generally, f is said to be C^r if it has continuous partial derivatives of all orders $\leq r$, and it is called C^∞ or *smooth* if it is C^r for all r. In this course we will mainly be concerned with smooth maps, but the notion of C^r for $r \in \mathbb{N}$ is sometimes useful.

The first order partial derivatives are conveniently arranged in the *Jacobian matrix*

$$Jf(p) = \left(\frac{\partial f_i}{\partial x_j}(p) \right)_{i=1,\dots,m, j=1,\dots,n},$$

which is the $m \times n$ matrix that has the partial derivatives of f_i in its i-th row.

Definition The *differential* of f at $p \in U$ is the linear map $df_p : \mathbb{R}^n \to \mathbb{R}^m$ which in the canonical bases is represented by the Jacobi matrix $Jf(p)$.

Example. If $f : \mathbb{R}^3 \to \mathbb{R}^2$ is given by $f(x, y, z) = (x^2 + y^2 + z^2, z)$ then

$$Jf(p) = \begin{pmatrix} 2p_1 & 2p_2 & 2p_3 \\ 0 & 0 & 1 \end{pmatrix}$$

and df_p is the linear map $x = (x_1, x_2, x_3) \mapsto (2p \cdot x, x_3)$.

The notation is particularly useful when it comes to composition of maps. The rule for differentiation of a composed map is as follows.

Theorem 1. The chain rule. *Let $f : U \to \mathbb{R}^m$ and $g : V \to \mathbb{R}^l$ be smooth maps, where $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open sets with $f(U) \subseteq V$. Then $g \circ f : U \to \mathbb{R}^l$ is smooth and satisfies*

$$d(g \circ f)_p = dg_{f(p)} \circ df_p. \tag{1}$$

Proof. The equality (1) between linear maps is equivalent with the matrix equation

$$J(g \circ f)(p) = Jg(f(p)) \cdot Jf(p). \tag{2}$$

This is a standard result from multivariable calculus.

However, in many calculus books the statement concerns maps that are C^1, not C^∞ as here. It thus remains to be seen that if f and g are C^∞, then so is $g \circ f$. In fact, it can be proved by induction on $r \in \mathbb{N}$ that if f and g are C^r, then so is $g \circ f$. This statement is clearly true for $r = 0$. For the general case, it is sufficient to prove that the entries of $J(g \circ f)$ are C^{r-1}. In the rule (2) these entries are expressed by the matrix product of $Jg \circ f$ and Jf. By the induction hypothesis, the entries of the composed map $Jg \circ f$ are C^{r-1}, and since the entries of Jf are also C^{r-1}, we are done. $\square$

Example. Let $m = 1$, $k = 2$ and $n = 3$. Write the coordinates of $f(u, v) \in \mathbb{R}^3$ as $(x(u, v), y(u, v), z(u, v))$. Then the composed function is $g(x(u, v), y(u, v), z(u, v))$. According to the theorem this is a differentiable function of u and v with

$$\frac{\partial g}{\partial u} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial u}$$

and similarly for the differentiation w.r.t. v.

Recall from Note 1 the *implicit function theorem* for functions $f\colon \mathbb{R}^n \to \mathbb{R}$ (see also the remark in Note 1). It is convenient to write elements in $\mathbb{R}^n$ as (x, y) where $x = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$ and $y \in \mathbb{R}$.

Theorem 2. *Let $f\colon U \to \mathbb{R}$ be a smooth function, where $U \subseteq \mathbb{R}^n$ is open. Let*

$$\mathcal{C} = \{(x, y) \in U \mid f(x, y) = 0\}$$

and let $p = (x_0, y_0) \in \mathcal{C}$. Assume that $\partial f/\partial y \neq 0$ at p.

Then there exist open neighborhoods $I \subseteq \mathbb{R}^{n-1}$ and $J \subseteq \mathbb{R}$ around x_0 and y_0, respectively, such that $W = I \times J \subseteq U$, and a smooth map $h\colon I \to J$ such that

$$\mathcal{C} \cap W = \{(x, h(x)) \mid x \in I\},$$

that is, in a neighborhood of p the set $\mathcal{C}$ is the graph of h.

Differentiating the expression $f(x, h(x)) = 0$ by means of the chain rule we obtain $\frac{\partial f}{\partial x_j}(x, h(x)) + \frac{\partial f}{\partial y}(x, h(x))\frac{\partial h}{\partial x_j}(x) = 0$, and hence

$$\frac{\partial h}{\partial x_j}(x) = -\frac{\frac{\partial f}{\partial x_j}(x, h(x))}{\frac{\partial f}{\partial y}(x, h(x))} \tag{3}$$

(this expression is also derived in Note 1).

We will now generalize Theorem 2 to functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$, that is, to the case where the set $\mathcal{C}$ consists of the simultaneous solutions to m equations. It is then convenient to write elements in $\mathbb{R}^n$ as (x, y) where $x = (x_1, \dots, x_{n-m}) \in \mathbb{R}^{n-m}$ and $y = (y_1, \dots, y_m) \in \mathbb{R}^m$.

Theorem 3. The implicit function theorem. *Let $f\colon U \to \mathbb{R}^m$ be a smooth function, where $U \subseteq \mathbb{R}^n$ is open. Let*

$$\mathcal{C} = \{(x, y) \in U \mid f(x, y) = 0\}$$

and let $p = (x_0, y_0) \in \mathcal{C}$. Assume that the determinant of the $m \times m$ matrix

$$A = \frac{\partial f_i}{\partial y_j}(p),$$

consisting of the last m columns of the jacobian $Jf(p)$, is non-zero.

Then there exist open neighborhoods $I \subseteq \mathbb{R}^{n-m}$ and $J \subseteq \mathbb{R}^m$ around x_0 and y_0, respectively, such that $W = I \times J \subseteq U$, and a smooth map $h\colon I \to J$ such that

$$\mathcal{C} \cap W = \{(x, h(x)) \mid x \in I\},$$

that is, in a neighborhood of P the set $\mathcal{C}$ is the graph of $h\colon \mathbb{R}^{n-m} \to \mathbb{R}^m$.

Proof. The theorem is proved by induction from the special case $m = 1$ already obtained in Theorem 2. Thus, we assume that the conclusion of the theorem is valid for functions into $\mathbb{R}^{m-1}$.

Since $\det A$ is non-zero, A is invertible. We want to replace $f\colon \mathbb{R}^n \to \mathbb{R}^m$ by the function $A^{-1} \circ f\colon \mathbb{R}^n \to \mathbb{R}^m$, obtained by multiplying all image vectors $f(x, y) \in \mathbb{R}^m$ with the constant matrix A^{-1}. Since multiplication by A^{-1} is a bijection, the solution sets for the equations $f(x, y) = 0$ and $A^{-1}f(x, y) = 0$ are identical, and it follows from the chain rule (2) that

$$J(A^{-1} \circ f)(p) = A^{-1} \cdot Jf(p),$$

from which we see that the last m columns of $J(A^{-1} \circ f)(p)$ comprise a unit matrix δ_{kj}. The effect of the replacement is thus that we may assume that A itself is a unit matrix. This we assume from now on, that is, $\partial f_k/\partial y_j = \delta_{kj}$.

In particular, for the function f_m whose derivatives are in the last row of Jf, we have that $\partial f_m/\partial y_j(p) = 0$ for $j < m$ and $\partial f_m/\partial y_m(p) = 1$. It follows from Theorem 2, applied with the last variable y_m as y, that there exists a neighborhood of p in which the set of solutions to $f_m(x, y) = 0$ is the graph of a smooth function $h\colon \mathbb{R}^{n-1} \to \mathbb{R}$, that is, $f_m(x, y) = 0$ if and only if

$$y_m = h(x, y'), \quad y' = (y_1, \ldots, y_{m-1}).$$

Moreover, by (3) the derivatives of h at p are given by

$$\frac{\partial h}{\partial x_i}(p) = -\frac{\partial f_m}{\partial x_i}(p), \quad \text{and} \quad \frac{\partial h}{\partial y_j}(p) = -\frac{\partial f_m}{\partial y_j}(p) = 0$$

for $i = 1, \ldots, n - m$ and $j = 1, \ldots, m - 1$.

Let the function $F\colon \mathbb{R}^{n-1} \to \mathbb{R}^{m-1}$ be defined (on a neighborhood of (x_0, y'_0)) by

$$F_k(x, y') = f_k(x, y', h(x, y')) \tag{4}$$

for $k = 1, \ldots, m - 1$, where as before $y' = (y_1, \ldots, y_{m-1})$. The partial derivatives of F_k are obtained by applying the chain rule to (4):

$$\frac{\partial F_k}{\partial y_j} = \frac{\partial f_k}{\partial y_j} + \frac{\partial f_k}{\partial y_m}\frac{\partial h}{\partial y_j}$$

and at p we thus have $\partial F_k/\partial y_j = \partial f_k/\partial y_j = \delta_{kj}$. The determinant of this matrix being non-zero, we can apply our induction hypothesis to F, and we obtain the existence of a function $g\colon \mathbb{R}^{n-m} \to \mathbb{R}^{m-1}$ such that the solution set for the equation $F(x, y') = 0$ is the graph of g in a neighborhood of (x_0, y'_0), that is, $F(x, y') = 0$ if and only if $y' = g(x)$.

We now see that (in a neighborhood of p) with $y = (y', y_m)$

$$f(x, y) = 0$$

if and only if

$$f_k(x, y) = 0, \quad k = 1, \ldots, m$$

if and only if

$$f_k(x, y) = 0, \quad k = 1, \ldots, m-1 \quad \text{and} \quad y_m = h(x, y')$$

if and only if

$$F(x, y') = 0 \quad \text{and} \quad y_m = h(x, y')$$

if and only if

$$y' = g(x) \quad \text{and} \quad y_m = h(x, y')$$

if and only if

$$y = (g(x), h(x, g(x))).$$

The function $x \mapsto (g(x), h(x, g(x)))$ is thus seen to be the desired function whose graph is $\mathcal{C}$ in a neighborhood of p. □

Definition Let $U, V \subseteq \mathbb{R}^m$ be open sets. A map $f\colon U \to V$ is called a *diffeomorphism* if

1. f is smooth,
2. $f\colon U \to V$ is bijective, and
3. $f^{-1}\colon V \to U$ is also smooth.

Proposition. *If $f\colon U \to V$ is a diffeomorphism, then its differential $df_q\colon \mathbb{R}^m \to \mathbb{R}^m$ is bijective at each $q \in U$ (equivalently,* $\det Jf(q) \neq 0$*). Moreover, the differential of f^{-1} at $z = f(q)$ is given by*

$$d(f^{-1})_z = (df_q)^{-1}.$$

Proof. Follows immediately from the chain rule by differentiation of the expression $f \circ f^{-1} = I$. □

The following fundamental result from multivariable calculus plays a very prominent role in differential geometry.

Theorem 4. The inverse function theorem. *Let $F\colon U \to \mathbb{R}^m$ be smooth, where $U \subseteq \mathbb{R}^m$ is open. Suppose that at a given point $q \in U$ the differential $dF_q\colon \mathbb{R}^m \to \mathbb{R}^m$ is bijective. Then there exists an open set $W \subseteq U$ containing q and an open set $V \subseteq \mathbb{R}^m$ containing $F(q)$ such that $V = F(W)$ and F is a diffeomorphism of W onto V.*

Proof. We shall apply the implicit function theorem with $n = 2m$ to the map $f\colon \mathbb{R}^m \times U \to \mathbb{R}^m$ given by $f(x, y) = -x + F(y)$ where $x \in \mathbb{R}^m$, $y \in U$. Notice that $f(x, y) = 0$ if and only if $F(y) = x$, hence if we can exhibit the solution set to this equation as the graph $y = h(x)$ of a function h, then h will be inverse to F.

Let $y_0 \in \mathbb{R}^m$ denote the given point q, and let $x_0 = F(y_0)$. The matrix A of Theorem 3 is exactly $JF(q)$, hence it has a non-vanishing determinant. Thus, according to the theorem there exist open neighborhoods I and J of x_0 and y_0, respectively, and a smooth map $h\colon I \to J$ such that $f(x, y) = 0$ if and only if $y = h(x)$, for all $(x, y) \in I \times J$. Let $W = J \cap F^{-1}(I)$, then W is open (since F is continuous). It is now seen, as remarked above, that $F\colon W \to I$ and $h\colon I \to W$ are the inverse maps of each other. Hence F is a diffeomorphism of W onto $V = I$. $\square$

Remark There is a fundamental difference between the inverse function theorem for functions of one variable and those of several variables. The theorem we have proved is *local*, as it only asserts the existence of an inverse to f in a neighborhood of $f(p)$. Even if the condition, that $df_p\colon \mathbb{R}^n \to \mathbb{R}^n$ is bijective, is satisfied everywhere in U, a global inverse of f need not exist, as seen in the example below. This contrasts the situation for $n = 1$: If $f'(x) \neq 0$ on an interval, then f is monotone on that interval, hence bijective.

Example Let $U \subseteq \mathbb{R}^2$ denote the right half plane

$$U = \{(r, \theta) \in \mathbb{R}^2 \mid r > 0\}$$

and let $f\colon U \to \mathbb{R}^2$ be given by

$$f(r, \theta) = (r\cos\theta, r\sin\theta)$$

that is, $f(r, \theta)$ is the point having polar coordinates (r, θ). The Jacobian of f,

$$Jf(r, \theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

is regular for all $(r, \theta) \in U$, hence the inverse function theorem implies that f is locally invertible. Since $f(r, \theta + 2\pi) = f(r, \theta)$, f is not globally injective. However, it follows from the corollary below, that the restriction of f to $\{(r, \theta) \mid r > 0,\ \theta \in I\}$, where I is any open interval of length $\leq 2\pi$, has a differentiable inverse.

Corollary. *Let $F\colon U \to \mathbb{R}^m$ be injective and smooth, where $U \subseteq \mathbb{R}^m$ is open. Suppose that $dF_q\colon \mathbb{R}^m \to \mathbb{R}^m$ is bijective for each $q \in U$. Then $F(U)$ is open and F is a diffeomorphism of U onto $F(U)$.*

Proof. Since F is injective, it has an inverse map $F^{-1}\colon F(U) \to U$. Let $F(q) \in F(U)$ be given, then according to Thm. 4 there exists an open neighborhood $V = F(W)$ of $F(q)$ in $F(U)$, and the restriction of F^{-1} to that neigborhood is smooth. It follows that F^{-1} is smooth. $\square$