Chapter -1: Differential Calculus and Regular Surfaces

Peter Perry

August 2008

Contents

1 Introduction

These notes review some basic ideas of differential calculus and define the notion of a regular submanifold of $\mathbb{R}^n$. Sources include chapters two and 5.1 of Michael Spivak's (strongly recommended) book, *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus* and chapter 3 of Manfredo do Carmo's book *Differential Forms and Applications*. These notes draw freely on both of these texts (some proofs are taken almost verbatim from the references), and the reader is strongly urged to go to the originals for a more thoroughgoing treatment of these ideas! I've titled these notes Chapter -1 because they logically precede Chapter 0 of do Carmo's book *Riemannian Geometry* which is the primary text for this course.

If f is a mapping from an open subset U of $\mathbb{R}^n$ to $\mathbb{R}^m$, we'll write $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$. We'll also write

$$f(x) = (f_1(x), f_2(x), \cdots, f_m(x))$$

where f_i, $1 \leq i \leq m$, are the *component functions* of f and $f_i : U \subset \mathbb{R}^n \to \mathbb{R}$. We'll denote by $\partial f_i/\partial x_j$, or sometimes $D_j f_i$, the partial derivative of the ith component function with respect to the jth independent variable:

$$(D_j f_i)(x) = \lim_{h\to 0} \frac{f_i(x+he_j) - f(x)}{h}.$$

We'll denote by I the identity matrix (acting on $\mathbb{R}^k$ where k is understood in context) and by Id the identity function acting from $\mathbb{R}^n$ to itself. The notation $\langle v, w\rangle$ denotes the inner (dot) product of vectors v and w belonging to $\mathbb{R}^n$. If $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, then (a, b) denotes the corresponding element of $\mathbb{R}^n \times \mathbb{R}^m$. Finally, if $A : \mathbb{R}^m \to \mathbb{R}^p$ and $B : \mathbb{R}^n \to \mathbb{R}^m$, $A\cdot B$ denotes the composition of the linear maps, while if $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g : V \subset \mathbb{R}^m \to \mathbb{R}^p$ with $f(U) \subset V$, $g \circ f : U \to \mathbb{R}^p$ is the composition of g with f.

2 The Derivative as a Linear Map

A function $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is *differentiable* at $a \in M$ if there is a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ with the property that

$$\lim_{h\to 0} \frac{\|f(a+h) - f(a) - L(h)\|}{\|h\|} = 0.$$

The linear mapping L is called the *differential of* f at a and is denoted df_a or $f'(a)$. If such a linear map L exists, it is unique. The affine approximation to f at a is given by the mapping

$$\ell(x) = f(a) + L(x-a).$$

Let's examine several cases of this definition.

Example 1 *(one-variable calculus): Suppose I is an open interval on the real line, $a \in I$, and $f : I \to \mathbb{R}$ is a mapping. If there is a number b so that*

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - bh)|}{|h|} = 0$$

then b is called the derivative of f at a, denoted $f'(a)$. The map $h \mapsto bh$ is a linear map from $\mathbb{R}^1$ to itself. The affine function $\ell(x) = f(a) + b(x-a)$ gives the best affine approximation to f near $x = a$. The graph of this function is the line tangent to the graph of f at $(a, f(a))$

Example 2 *(tangent lines to parametric curves): Suppose I is an open interval, $a \in I$, and $\gamma : I \to \mathbb{R}^n$ is a mapping. If there is a vector $v \in \mathbb{R}^n$ so that*

$$\lim_{h\to 0} \frac{\|\gamma(a+h) - \gamma(a) - hv\|}{|h|} = 0$$

then v is called the tangent vector to γ at a and is also denoted $\gamma'(a)$. Note that the mapping $h \mapsto hv$ is a linear mapping from $\mathbb{R}^1$ to $\mathbb{R}^n$. The graph of the affine function $\ell(t) = \gamma(a) + v(t-a)$ is the tangent line to γ at a.

Example 3 *(tangent planes to the graph of a function): Suppose that* $U \subset \mathbb{R}^2$ *is open, that* $(x_0, y_0) \in U$*, and that* $f : U \to \mathbb{R}$*. If there is a vector* $v = (v_1, v_2) \in \mathbb{R}^2$ *with the property that*

$$\lim_{h \to 0} \frac{|f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) - (v_1 h_1 + v_2 h_2)|}{\|h\|}$$

(here we've written $h = (h_1, h_2)$*), then* (v_1, v_2) *is the differential (or gradient) of* f *and* (x_0, y_0)*, and is denoted* $(\nabla f)(x_0, y_0)$*. Note that* $(v_1 h_1 + v_2 h_2) = \langle v, h \rangle$*. The map* $h \mapsto \langle v, h \rangle$ *is a linear map from* $\mathbb{R}^2$ *to* $\mathbb{R}^1$*. The affine function* $\ell(x, y) = f(x_0, y_0) + v_1 (x - x_0) + v_2(y - y_0)$ *gives the best affine approximation to* f *near* (x_0, y_0)*. The graph of this affine function is the plane tangent to the graph of* f *at* $(x_0, y_0, f(x_0, y_0))$*.*

Example 4 *(the Jacobian matrix) Suppose that* $U \subset \mathbb{R}^n$ *is open, that* $a \in U$*, and that* $f : U \to \mathbb{R}^m$*. If there is an* $m \times n$ *matrix* A *with the property that*

$$\lim_{h \to 0} \frac{\|f(a + h) - f(x) - Ah\|_{\mathbb{R}^m}}{\|h\|_{\mathbb{R}^n}} = 0$$

then f *is differentiable at* a*, and the matrix* A *is called the Jacobian matrix of* f *at* a*, denoted* $df(a)$*. Note that the mapping* $h \mapsto Ah$ *is a linear mapping from* $\mathbb{R}^n$ *to* $\mathbb{R}^m$*. The affine mapping* $\ell(x) = f(a) + A(x - a)$ *is the best linear approximation to* f *near* a*.*

We can think of a function $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ as $f(x) = (f_1(x), \ldots, f_m(x))$ for component functions $f_1, \ldots, f_m$. Its Jacobian matrix is a matrix whose rows are the gradients of the component functions. The Jacobian matrix $df(a)$ is the $m \times n$ matrix given by $\dfrac{\partial f}{\partial x_1}$

$$df(a) = \begin{bmatrix} (\nabla f_1)(a) \\ \vdots \\ (\nabla f_m)(a) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \frac{\partial f_n}{\partial x_2}(a) & \cdots & \frac{\partial f_n}{\partial x_m}(a) \end{bmatrix}$$

Multiplying $df(a)$ by a vector h gives a vector whose entries are $\langle (\nabla f_k)(a), h \rangle$, the linear approximation to the change in f_k due to a displacement h.

The Jacobian matrix is the matrix of the linear mapping A with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

An easy consequence of the derivative is that differentiability implies continuity.

Exercise 5 *Prove that if* $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $a \in U$*, then* f *is continuous at* a*.*

Exercise 6 *Prove that if $f : U \subset \mathbb{R}^n \to \mathbb{R}$ and f is differentiable at a, then*

$$\frac{\partial f}{\partial x_i}(a) = \langle (\nabla f)(a), e_i \rangle$$

where e_i is the ith basis vector in the usual basis of $\mathbb{R}^n$. To prove this, use the definition of $\nabla f(a)$ as the differential together with the definition of the partial derivative as

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}.$$

Exercise 7 *Prove the statement above that the differentiable $df(a)$ of a map from $U \subset \mathbb{R}^n$ to $\mathbb{R}^m$ has rows consisting ot the gradients of the component functions f_i.*

Exercise 8 *Show that if $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$, then $\operatorname{tr}(df(a))$ is the divergence of f, defined as*

$$(\operatorname{div} f)(x) = \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(x)$$

Exercise 9 *Show that if $f : U \subset \mathbb{R}^3 \to \mathbb{R}^3$, then the antisymmetric part of $(df(a))$ takes the form*

$$\begin{pmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & v_1 \\ v_2 & -v_1 & 0 \end{pmatrix}$$

where $v = (v_1, v_2, v_3)$ is $\nabla \times f$ (the "curl" of f).

In what follows we will denote by $f'(a)$ or $df(a)$ the linear mapping T that occurs in the definition of the derivative of f at a.

If $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ has continuous partial derivatives up to order k in U, we will write $f \in C^k(U)$, and if such an f has continuous partial derivatives of all orders in U, we will write $f \in C^\infty(U)$ and say that f is a smooth function. In this course we will deal almost exclusively with smooth functions. In the statements of theorems we will however impose the minimal smoothness hypotheses needed to make them work.

3 The "Big Theorems" of Differential Calculus

The "big theorems" of differential calculus all concern the connection between the local properties of a differentiable function and the properties of its linearization, the derivative. These are the *chain rule*, the *inverse function theorem*, and the *implicit function theorem*.

3.1 The Chain Rule

Theorem 10 *(Chain Rule) Suppose that $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g : V \subset \mathbb{R}^m \to \mathbb{R}^p$. Suppose that $f(U) \subset V$ and that $a \in U$. Finally, suppose that f is differentiable at a and that g is differentiable at $f(a)$. Then $g \circ f$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$$

where the $\cdot$ denotes composition of linear maps.

Inotherwords, *the derivative of a composition is the composition of derivatives.* Of course, if f and g are smooth functions, then the composition is also a smooth function.

Exercise 11 *Use the chain rule to prove the following formula. Suppose that $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$, that $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\gamma(0)$, and that $\gamma(0) \in U$. Prove that*

$$\left.\frac{d}{dt}\right|_{t=0} (f \circ \gamma)(t) = \langle (\nabla f)(\gamma(0)), \gamma'(0) \rangle$$

Show that if $\gamma(t) = x_0 + te_i$ where e_i is the ith basis vector in the standard basis, then

$$\left.\frac{d}{dt}\right|_{t=0} (f \circ \gamma)(t) = \frac{\partial f}{\partial x_i}(x_0).$$

3.2 The Inverse Function Theorem

Theorem 12 *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ and suppose that $f'(a)$ is nonsingular. There is a neighborhood V of a so that the map $f : V \to \mathbb{R}^n$ is invertible on $f(V)$. Moreover, the inverse f^{-1} is differentiable and*

$$\left(f^{-1}\right)'(a) = \left[f'(f(a))\right]^{-1} \tag{1}$$

Inotherwords, *a differentiable mapping with invertible derivative is locally invertible.* We won't give a proof but refer the reader to any standard text on multivariate calculus. The idea is to establish the existence of an inverse by using the linear approximation to f and a contraction mapping argument.

If the function f is smooth, it has a smooth inverse.

Exercise 13 *Assuming that f^{-1} is differentiable, use the Chain Rule to prove (1).*

3.3 The Implicit Function Theorem

The implicit function theorem allows us to assert that the solution set of an equation of the form $f(x) = 0$ for certain nonlinear, differentiable maps admits a local parameterization. As such it plays a fundamental role in differential geometry.

Theorem 14 *(Implicit Function Theorem) Suppose that $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open, and $f : U \times V \to \mathbb{R}^m$ is differentiable at (a, b). Suppose further that the $m \times m$ submatrix*

$$C = \{D_{n+j} f_i(a, b)\}_{1 \le i,j \le m}$$

is invertible. Then there is a neighborhood W of a and a function $g : W \to \mathbb{R}^m$ with $g(a) = b$ so that $f(x, g(x)) = f(a, b)$ for all $x \in W$.

Inotherwords, the set of (x, y) near (a, b) with $f(x, y) = f(a, b)$ is a "parameterized curve" in $\mathbb{R}^{n+m}$ parameterized locally by n parameters. The proof actually shows more, namely that the "parameterized curve" $\{(x, f(x)) : x \in W\}$ is the *only* "curve" (actually, n-dimensional submanifold) passing through (a, b) with this property.

Proof. Without loss of generality we can assume that $f(a, b) = 0$ (replace f by $f - f(a, b)$). Define $F : U \times V \subset \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ by

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

The Jacobian matrix of F at (a, b) takes the form

$$dF(a, b) = \begin{bmatrix} I & 0 \\ * & M \end{bmatrix}$$

and so has full rank by hypothesis. It follows from the inverse function theorem that there is a neighborhood $U_1 \times V_1$ of (a, b) and a neighborhood $U_2 \times V_2$ of $(a, 0)$ together with a differentiable mapping $\overline{h} : U_2 \times V_2 \to U_1 \times V_1$ with the property that $F \circ \overline{h} = Id$. The mapping $\overline{h}$ necessarily takes the form

$$\overline{h}(x, y) = (x, k(x, y))$$

for a differentiable mapping $k : U_2 \times V_2 \to V_1$. We then compute

$$\left(F \circ \overline{h}\right)(x, y) = (x, f(x, k(x, y))) = (x, y)$$

so setting $y = 0$ we have

$$f(x, k(x, 0)) = (x, 0).$$

We can take $g(x) = \overline{h}(x, 0)$ and obtain the desired inverse function. ∎

Remark 15 *The proof of the Implicit Function Theorem actually shows a bit more. The mapping $\overline{h}$ constructed there has the property that*

$$(f \circ \overline{h})(x, y) = y$$

for all $(x, y) \in U_1 \times V_1$. We'll use this remark to prove a stronger version of the Implicit Function Theorem in what follows.

To understand this theorem it is useful to consider the following linear model. Suppose that $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is a linear map with matrix $A = [B \;\; C]$ where

B is $m \times n$ and C is $m \times m$. If C is invertible then we can parameterize $\ker A$ with n parameters. For $x = (y, z) \in \mathbb{R}^n \times \mathbb{R}^m$ with $x \in \ker A$, we have

$$By + Cz = 0$$

or

$$z = -C^{-1}By$$

which describes the kernel in terms of n parameters. Notice that in this situation, the matrix A has maximal rank (namely m).

Thus, we can summarize the implicit function theorem as follows: *if the kernel of the differential can be described by n parameters, then the zero set of the differentiable mapping can be described locally by n parameters.*

This paraphrase suggests a stronger version of the implicit function theorem which is in fact true. We'll change notation a bit from the previous version because we now assume that the differential of the mapping F has rank k.

Theorem 16 *Suppose that $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ is continuously differentiable, that $p \leq n$. and $a \in U$. If $f(a) = 0$ and the $p \times n$ matrix*

$$\{D_j f_i(a)\}_{1 \leq j \leq n, 1 \leq i \leq p}$$

has rank p, then there is an open set $A \subset U$ containing a and a differentiable function $h : A \to \mathbb{R}^n$ with differentiable inverse with the property that

$$(f \circ h)(x_1, \cdots, x_n) = (x_{n-p+1}, \cdots, x_n).$$

In particular, $f = 0$ on the image of the set $A \cap \{x : x_{n-p+1} = \ldots = x_n = 0\}$

Proof. We can consider f as a function $f : \mathbb{R}^{n-p} \times \mathbb{R}^p \to \mathbb{R}^p$. Since the Jacobian has rank p, there is a set of indices $1 \leq j_1 < \cdots < j_p \leq n$ so that the $p \times p$ matrix formed from columns $j_1, \cdots, j_p$ has full rank (i.e., is invertible). Let $\pi : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map (permutation) that maps x_{j_k} to x_{n-p+k} for $1 \leq k \leq p$. The composition $f \circ \pi$ then satisfies the hypothesis of the Implicit Function Theorem, so we may define

$$h = \pi \circ \overline{h}$$

where $\overline{h}$ is the function constructed in that proof for the function $f \circ \pi$. Now use Remark 15. ■

Exercise 17 *Let S^n be the zero set of the function $F(x) = |x|^2 - 1$ where $F : \mathbb{R}^{n+1} \to \mathbb{R}$. Use the implicit function theorem to show that there is a neighborhood $U \subset \mathbb{R}^n$ containng 0 and a differentiable map $g : U \to \mathbb{R}$ with $g(0) = 1$ and $F(x, g(x)) = 0$. Also, compute the map g directly. Note that if we define $f(x) = (x, g(x))$ we obtain a* local parameterization *of the sphere S^n near its north pole* $(0, \ldots, 0, 1)$.

4 Regular Submanifolds of $\mathbb{R}^n$

The sphere S^n is one of the simplest examples of a *regular submanifold* of $\mathbb{R}^n$ – it is implicitly defined and, by Exercise 17, it can be locally parameterized by n parameters near any point $p \in S^n$. More generally:

Definition 18 *A subset M of $\mathbb{R}^n$ is a* regular submanifold *of $\mathbb{R}^n$ of dimension k if for each $p \in M$, there exists a neighborhood V of p in $\mathbb{R}^n$, an open subset U_α of $\mathbb{R}^k$, and a map f_α from U_α onto $V \cap M$ such that:*
(i) f_α is a differentiable homeomorphism,
(ii) the differential $df_\alpha(q)$ is injective for each $q \in U_\alpha$.

In other words, M can be parameterized locally by k parameters. Familiar examples include smooth curves in $\mathbb{R}^n$ ($k = 1$), hypersurfaces in $\mathbb{R}^n$ ($k = n-1$), and Euclidean space itself (a “cheat” where $k = n$). As the example of the sphere clearly shows, there need not be a *global* parameterization of the submanifold and indeed there generally isn't.

The pair (f_α, U_α) is called a *local parameterization* or a *coordinate chart* for M. The map f_α is called a *coordinate map.*

Our primary interest is in *smooth submanifolds* of $\mathbb{R}^n$, where we require the maps f_α to be not simply differentiable but also smooth, i.e., have derivatives of all orders. That is a smooth submanifold of $\mathbb{R}^n$ is a regular submanifold of $\mathbb{R}^n$ where the coordinate maps are smooth.

In order to take the notion of smooth manifold a step further it is very, very useful to realize that smooth submanifolds of $\mathbb{R}^n$ are an example of a larger class of objects that can also be defined intrinsically.

Definition 19 *A k-*dimensional differentiable manifold *is a set M together with a family of injective maps $f_\alpha : U_\alpha \to M$ of open sets U_α in $\mathbb{R}^k$ into M such that:*
(1) $\cup_\alpha f_\alpha(U_\alpha) = M$, and
(2) For each pair (α, β) with $W := f_\alpha(U_\alpha) \cap f_\beta(U_\beta) \neq \emptyset$, the sets $f_\alpha^{-1}(W)$ and $f_\beta^{-1}(W)$ are open sets in $\mathbb{R}^n$ and the maps $f_\beta^{-1} \circ f_\alpha$ and $f_\alpha^{-1} \circ f_\beta$ are differentiable.
(3) The family $\{(U_\alpha, f_\alpha)\}$ is maximal relative to (1) and (2).

Remark 20 *The pair (U_α, f_α) with $p \in M$ is called a* coordinate chart *(or* parameterization *or* coordinate system*) for M near p, and $f_\alpha(U_\alpha)$ is called a* coordinate neighborhood *of p. A family $\{(U_\alpha, f_\alpha)\}$ satisfying (1) and (2) is caled a* differentiable structure *on M. The maps $f_\alpha^{-1} \circ f_\beta$ are called* transition maps*.*

Remark 21 *Condition (3) is a technical condition and can be satisfied by taking the union of all families that satisfy (1) and (2).*

Before proceeding much further it would be nice to see that Definition 19 really is a generalization of Definition 18.

Theorem 22 *Suppose that M is a k-dimensional smooth submanifold of $\mathbb{R}^n$. Then M is also a k-dimensional differentiable manifold.*

Proof. We need to show that the transition maps are smooth. Consider a pair of coordinate charts (U_α, f_α) and (U_β, f_β) with $W = f_\alpha(U_\alpha) \cap f_\beta(U_\beta)$ nonempty. Since f_α and f_β are homeomorphisms the sets $f_\alpha(U_\alpha)$ and $f_\beta(U_\beta)$ are open and so W and its inverse images under f_α^{-1} and f_β^{-1} are open subsets of $\mathbb{R}^k$. To show that

$$f_\beta^{-1} \circ f_\alpha : f_\alpha^{-1}(W) \to f_\beta^{-1}(W)$$

is a smooth map we will consider f_β^{-1}. Pick $a \in U_\beta$. The differential df_β is injective and so its Jacobian matrix has maximal rank k, so that there are indices $1 \le i_1 < \cdots < i_k \le n$ with the property that the $k \times k$ matrix

$$\{D_j f_{i_m}(a) : 1 \le j, m \le k\}$$

is nonsingular. Now write $(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$, denote by $j_1, \ldots, j_{n-k}$ the indices from $\{1, \ldots, n\}$ that don't belong to the set $\{i_1, \ldots, i_k\}$, and define a new function

$$F_\beta : U_\beta \times \mathbb{R}^{n-k} \to \mathbb{R}^n$$

by

$$F_\beta(x) = f_\beta(x) + (0, \ldots, x_{i_1}, 0, \cdots, x_{i_k})$$

(that is, the second right-hand term has zeros in the $j_1, j_2, \cdots, j_k$ slots but $x_{i_1}, \cdots, x_{i_{n-k}}$ in the remaining ones), then a bit of thought shows that:
(i) F_β satisfies the hypothesis of the inverse function theorem and so has a smooth inverse F_β^{-1} in a neighborhood of $f_\beta(x)$ in $\mathbb{R}^n$, and
(ii) If $z \in M$ then $F_\beta^{-1}(z) = f_\beta^{-1}(z)$.
Since $f_\alpha(U_\alpha) \subset M$ we can then write

$$f_\beta^{-1} \circ f_\alpha = F_\beta^{-1} \circ f_\alpha$$

which exhibits $f_\beta^{-1} \circ f_\alpha$ as a composition of smooth maps. ∎

Next, we'll turn to implementing ideas of differential calculus of smooth submanifolds of $\mathbb{R}^n$.