Handout 1
BACKGROUND MATERIAL FROM MULTIVARIABLE CALCULUS
(CORRECTED 1/16/13)

Differential geometry is based on the theory of multivariable calculus. Most of this theory is covered (at least in the 2-dimensional case) in Math 326, and can be found in Chapters 6–9 of the standard Math 326 textbook, *Advanced Calculus*, 3rd Edition (1983), by Taylor and Mann. Chapter 12 of that book is not typically covered in 326, but it includes a good treatment of this theory in arbitrary dimensions, and is useful as a reference.

1. Scalar-Valued and Vector-Valued Functions

The main subject of this note is functions (or, synonymously, maps or mappings) between Euclidean spaces. Because we want to be able to take partial derivatives of our functions in all directions, we will generally insist that they be defined on *open* subsets of their domains.

Thus suppose m and n are positive integers, U is an open subset of $\mathbb{R}^m$, and $f\colon U \to \mathbb{R}^n$ is a function. When $n = 1$, we call f a ***scalar-valued function***, and when $n > 1$, it is a ***vector-valued function***. Here are some specific interpretations of such functions:

- When $m = n = 1$, $f\colon \mathbb{R} \to \mathbb{R}$ is just an ordinary real-valued function of one real variable.
- When $m = 1$, $n > 1$, and $U \subseteq \mathbb{R}$ is an interval, $f\colon U \to \mathbb{R}^n$ can be interpreted as a ***parametric curve*** in $\mathbb{R}^n$. For each $t \in U$, we think of the point $f(t) \in \mathbb{R}^n$ as representing the position of a moving particle at time t.
- When $m > 1$ and $n = 1$, a function $f\colon U \to \mathbb{R}$ is often called a ***scalar field*** on U. In applications, it can represent any quantity that assigns a numerical value to each point in a region of the plane or space, such as temperature in a room or varying density of an object.
- When $m = n > 1$ (usually $m = 2$ or 3), a vector-valued function $f\colon U \to \mathbb{R}^n$ is sometimes called a ***vector field on U***. It can represent a quantity that has both magnitude and direction at each point, such as a magnetic field or the velocity field of a moving fluid.
- In general when $m > 1$ and $n > 1$, a function $f\colon U \to \mathbb{R}^n$ can be interpreted as a ***transformation***, that is, a function that takes points of U and maps them to points of $\mathbb{R}^n$. For example, the polar coordinate map $f(r, \theta) = (r\cos\theta, r\sin\theta)$ takes the subset $\{(r, \theta) : r > 0\} \subseteq \mathbb{R}^2$ and maps it onto $\mathbb{R}^2 \setminus \{0\}$ (wrapping around infinitely many times).

In general, given an open subset $U \subseteq \mathbb{R}^m$ and a function $f\colon U \to \mathbb{R}^n$ as above, at each point $x = (x_1, \ldots, x_m) \in U$, we can write the value $f(x) \in \mathbb{R}^n$ in components as

$$f(x) = (f_1(x), \ldots, f_n(x)).$$

This determines n scalar-valued functions $f_1, \ldots, f_n\colon U \to \mathbb{R}$, called the ***coordinate functions*** or ***component functions of f***, and f is in turn determined by them. Given a nonnegative integer k, we say that f is a ***function of class C^k*** if each component function of f has continuous partial derivatives of all orders less than or equal to k. (A function of class C^0 is just a continuous function.) If f is of class C^k for every k, we say it is ***of class C^∞***. Functions of class C^∞ are also called ***smooth functions***.

A map $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is called a ***linear map*** if it satisfies the following identity for all $x, y \in \mathbb{R}^m$ and all $a, b \in \mathbb{R}$:

$$f(ax + by) = af(x) + bf(y).$$

It is shown in linear algebra courses that linear maps from $\mathbb{R}^m$ to $\mathbb{R}^n$ are exactly those maps that can be expressed in the form of matrix multiplication. More precisely, a map $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is linear

if and only if there is a $n \times m$ matrix $A = (A_{ij})$ of real numbers, such that $f(x)$ is obtained by expressing both x and $f(x)$ as column matrices (of dimensions $m \times 1$ and $n \times 1$, respectively), and carrying out the following matrix multiplication:

$$f(x) = \begin{pmatrix} A_{11} & \dots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} A_{11}x_1 + \cdots + A_{1m}x_m \\ \vdots \\ A_{n1}x_1 + \cdots + A_{nm}x_m \end{pmatrix} \in \mathbb{R}^n.$$

Thus the matrix A completely determines the function f. Conversely, A can be obtained from f in the following way. For each $i = 1, \dots, m$, let $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ be the vector in $\mathbb{R}^m$ with a 1 in the ith place and zeros elsewhere. In column form, they can be written

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ \dots, \ \mathbf{e}_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

These vectors are called the ***standard basis vectors*** for $\mathbb{R}^m$. If a linear map f is known, its matrix A can be computed by observing that the columns of A are the images under f of the standard basis vectors: $f(\mathbf{e}_1), \ \dots, \ f(\mathbf{e}_m)$. It follows from this computation that if f is a linear map, its ith component function is of the form $f_i(x) = A_{i1}x_1 + \cdots + A_{im}x_m$, which is infinitely differentiable; thus every linear map is smooth.

If f and g are linear maps represented by the matrices A and B, respectively, the composite map $f \circ g$ is represented by the product matrix AB. The identity map $\mathrm{id}\colon \mathbb{R}^n \to \mathbb{R}^n$ is easily seen to be linear. It is represented by the $n \times n$ ***identity matrix*** I_n, which is the matrix that has 1's on the main diagonal and 0's elsewhere:

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

It follows that $I_n A = A = A I_m$ for any $m \times n$ matrix A.

The following theorem is proved in linear algebra courses:

Theorem 1.1. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear map, and let A be its matrix. The following statements are equivalent:*

(a) *f has a linear inverse map $f^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$.*
(b) *f is bijective.*
(c) *f is injective or surjective.*
(d) $\det A \neq 0$,
(e) *The columns of A are linearly independent.*
(f) *There is an $n \times n$ matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$.*

If any (and hence all) of the conditions in the preceding theorem are true, then we say f is ***invertible*** or is an ***isomorphism***, and its matrix A is said to be ***invertible*** or ***nonsingular***. Because all linear maps are smooth, every invertible linear map is a diffeomorphism.

There is a slightly more general type of map that is closely related to linear maps. A map $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is called an ***affine map*** if there exist a linear map $f_0\colon \mathbb{R}^m \to \mathbb{R}^n$ and a vector $b \in \mathbb{R}^n$ such that $f(x) = f_0(x) + b$ for all $x \in \mathbb{R}^m$. (In other words, f is equal to a linear map plus a constant.) Note that the functions $f\colon \mathbb{R} \to \mathbb{R}$ of the form $f(x) = mx + b$, which are referred to in

high-school algebra as "linear functions," are actually *affine* functions, not linear ones in the sense we have defined here (unless $b = 0$).

Fair warning: different authors use different notations and terminology for matrices. Abate & Tovena use vertical bars instead of parentheses to denote a matrix, while many other books use vertical bars to denote the *determinant* of a matrix. To avoid confusion, I will not use vertical bars at all; I'll use parentheses to denote a matrix, and the notation "det" to denote a determinant, so the determinant of the matrix A will be denoted $\det A$.

2. The Differential of a Map

Henceforth, we assume that $U \subseteq \mathbb{R}^m$ is open and $f\colon U \to \mathbb{R}^n$ is smooth. (All of these results are valid under weaker differentiability assumptions, but assuming f is smooth frees us from having to count how many derivatives we need to take in any given argument.) For any point $p \in U$, the ***Jacobian matrix of f at p*** is the matrix

$$\operatorname{Jac} f(p) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(p) & \cdots & \dfrac{\partial f_1}{\partial x_m}(p) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n}{\partial x_1}(p) & \cdots & \dfrac{\partial f_n}{\partial x_m}(p) \end{pmatrix}.$$

(Each row of this matrix consists of the partial derivatives of one of the component functions of f.) In the special case $m = n$, where $\operatorname{Jac} f(p)$ is a square matrix, its determinant, denoted by $\det \operatorname{Jac} f(p)$, is called the ***Jacobian determinant of f at p***. (Another warning about terminology: the term "Jacobian" by itself sometimes refers to the Jacobian matrix, and sometimes to the Jacobian determinant. To avoid confusion, we will always say explicitly which one we mean.)

The principal significance of the Jacobian matrix is that it defines a linear map that approximates the nearby behavior of the function f. Here's how that works. Given an open subset $U \subseteq \mathbb{R}^m$ and a smooth map $f\colon U \to \mathbb{R}^n$, for each $p \in U$ we define the ***differential of f at p*** to be the linear map $df_p\colon \mathbb{R}^m \to \mathbb{R}^n$ whose matrix is $\operatorname{Jac} f(p)$. The differential of f at p represents a "best linear approximation" to the difference between the value of f at p and its value at a nearby point $p + v$, in the sense that will be made precise later (see Theorem 1.4 below).

The differential unifies most of the different notions of derivatives that occur in multivariable calculus. In three special cases, it has familiar interpretations:

- When both m and n are equal to 1, f is just a real-valued function of one real variable, and the differential df_{t_0} is the linear map whose 1×1 matrix has the ordinary derivative $f'(t_0)$ as its sole entry.
- In the case $m > 1$ and $n = 1$, $f\colon U \to \mathbb{R}$ is a scalar field on U, and df_p is represented by the row matrix with components $(\partial f/\partial x_1(p), \ldots, \partial f/\partial x_m(p))$. The vector with these components is called the ***gradient of f at p***, and is denoted by $\nabla f(p) = (\partial f/\partial x_1(p), \ldots, \partial f/\partial x_m(p))$.
- When $m = 1$ and $n > 1$, and $U = (a, b)$ is an open interval in $\mathbb{R}$, f is called a ***parametrized curve*** in $\mathbb{R}^n$, and we often denote the variable in the domain (the "parameter") by t and think of it as "time." In this case, for each $t_0 \in (a, b)$, df_{t_0} is represented by the column matrix with components $(f_1'(t_0), \ldots, f_m'(t_0))$. This vector is called the ***velocity of f at time t_0*** and is often denoted by $f'(t_0)$ or $\dot{f}(t_0)$.

The other two types of derivatives that you might have seen are the ***divergence*** and ***curl*** of a vector field $f\colon \mathbb{R}^3 \to \mathbb{R}^3$. The divergence is actually the ***trace*** of the differential of f (the sum of the entries on the main diagonal), and we will make use of it later in the course. The curl is

a vector field built out of a more complicated combination of partial derivatives; we will not have occasion to use it in this course.

3. The Chain Rule

You have seen various versions of the chain rule for computing derivatives of composite functions. They all have a very concise and unified formulation in terms of differentials.

Theorem 1.2 (The Chain Rule)**.** *Suppose $V \subseteq \mathbb{R}^m$ and $U \subseteq \mathbb{R}^n$ are open subsets, and $f\colon V \to U$ and $g\colon U \to \mathbb{R}^k$ are smooth functions. Then the composite function $g \circ f\colon V \to \mathbb{R}^k$ is smooth, and for each $p \in V$ its differential is given by*

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

The composition on the right-hand side represents composition of linear maps; in terms of matrices, it can be interpreted as matrix multiplication. Thus if we let $(x_1, \dots, x_m)$ denote the coordinates in V and $(u_1, \dots, u_n)$ those in U, the component in the ith row and jth column of $d(g \circ f)_p$ is

$$\frac{\partial}{\partial x_j}(g_i \circ f)(p) = \sum_{l=1}^{n} \frac{\partial g_i}{\partial u_l}(f(p)) \frac{\partial f_l}{\partial x_j}(p).$$

Here are the situations in which we will most frequently use the chain rule.

- $m = n = k = 1$: This is the classic chain rule of one-variable calculus:
$$\frac{d}{dx}(g \circ f)(p) = \frac{dg}{du}(f(p)) \frac{df}{dx}(p).$$
- $m = n = 1$, $k > 1$: In this case, g is a parametrized curve in $\mathbb{R}^k$, and f is a "reparametrization" of g. The formula above applies to each component of $g \circ f$:
$$\frac{d}{dx}(g_i \circ f)(p) = \frac{dg_i}{du}(f(p)) \frac{df}{dx}(p), \qquad i = 1, \dots, k.$$
- $m = 1$, $n > 1$, $k = 1$: Here f is a parametrized curve in $U \subseteq \mathbb{R}^n$, and g is a scalar field on U, so $g \circ f(t)$ represents the function g evaluated along the curve f at time t. In this case, the chain rule reduces to the following formula for its derivative:
$$\frac{d}{dt}(g \circ f)(t_0) = \sum_{l=1}^{n} \frac{\partial g}{\partial u_l}(f(t_0)) \frac{df_l}{dt}(t_0) = \nabla g(f(p)) \cdot f'(t_0).$$
- $m > 1$, $n = k = 1$: Now f is a scalar field on $V \subseteq \mathbb{R}^m$, and g is an ordinary real-valued function of one real variable. The chain rule reads
$$\frac{\partial}{\partial x_j}(g \circ f)(p) = g'(f(p)) \frac{\partial f}{\partial x_j}(p), \qquad j = 1, \dots, m,$$
or
$$\nabla(g \circ f)(p) = g'(f(p)) \nabla f(p),$$
- $m = 1$, $n > 1$, $k > 1$: In this case, we think of f as a parametrized curve in $U \subseteq \mathbb{R}^n$, and g is a transformation from U to some subset of $\mathbb{R}^k$ (where k might be equal to n or not). In this case, we can interpret the column matrix df_{t_0} as the velocity of f at time t_0, and write it in the form $f'(t_0)$. Likewise, the composition $g \circ f$ is also a parametrized curve in $\mathbb{R}^n$, and the chain rule gives a formula for its velocity:
$$(g \circ f)'(t_0) = dg_{f(t_0)}(f'(t_0)).$$
In other words, the velocity of the composite curve is the differential of g applied to the velocity of f.

- If m and n are abitrary and $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is a linear map, then the differential of f at any point $p \in \mathbb{R}^m$ is just the map f itself:
$$df_p = f.$$
(You will be asked to prove this for homework.)

Using the chain rule, we can prove the promised theorem that shows how the differential provides a good approximation to nearby values of the function. First we recall the following standard lemma about the dot product in $\mathbb{R}^n$.

Lemma 1.3 (Cauchy–Schwartz Inequality)**.** *For all $v, w \in \mathbb{R}^n$,*
$$|v \cdot w| \le \|v\|\,\|w\|.$$

Proof: See Exercise 3.1 on the *Metric Spaces* handout from Math 441. □

The following theorem says that, in the special case of a real-valued function $f\colon U \to \mathbb{R}$, $df_p(v)$ is a good approximation to the difference $\Delta f(v) = f(p+v) - f(p)$ in the sense that by taking v sufficiently small, we can make the difference between $\Delta f(v)$ and $df_p(v)$ much smaller than v.

Theorem 1.4. *Suppose $U \subseteq \mathbb{R}^m$ is an open set, $f\colon U \to \mathbb{R}$ is a smooth function, and fix $p \in U$. For any $v \in \mathbb{R}^m$ small enough that $p + v \in U$, define $\Delta f(v) = f(p+v) - f(p)$. Then*
$$\lim_{v \to 0} \frac{|\Delta f(v) - df_p(v)|}{\|v\|} = 0.$$

Proof: Because U is open, we can choose $r > 0$ such that the open ball $B(p, r)$ is contained in U. For any vector $v \in \mathbb{R}^m$ with $\|v\| < r$, define a function $g_v\colon [0,1] \to \mathbb{R}$ by $g_v(t) = f(p + tv)$. The mean-value theorem applied to g_v guarantees that there exists some real number $\tau(v) \in (0,1)$ (depending on v, of course), such that $g_v(1) - g_v(0) = g_v'(\tau(v))$. By the chain rule, for each $t \in (0,1)$,
$$g_v'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p + tv) v_i,$$
and therefore
$$\Delta f(v) = g_v(1) - g_v(0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p + \tau(v)v) v_i,$$
which yields
$$\begin{aligned} \Delta f(v) - df_p(v) &= g_v(1) - g_v(0) - df_p(v) \\ &= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(p + \tau(v)v) - \frac{\partial f}{\partial x_i}(p) \right) v_i \\ &= \big(\nabla f(p + \tau(v)v) - \nabla f_p(v) \big) \cdot v. \end{aligned}$$

Let $\varepsilon > 0$ be given. Because the component functions of the gradient vector field $\nabla f\colon U \to \mathbb{R}^m$ are the partial derivatives of f, each of which is continuous at p, we can choose $\delta > 0$ such that $\|x - p\| < \delta$ and $x \in U$ implies $\|\nabla f(x) - \nabla f(p)| < \varepsilon$. If $\|v\| < \delta$, the fact that $0 < \tau(v) < 1$ implies $\|(p + \tau(v)v) - p\| = \tau(v)\|v\| < \|v\|$, and therefore by the Cauchy–Schwartz inequality,
$$|\Delta f(v) - df_p(v)| \le \|\nabla f(p + \tau(v)v) - \nabla f(p)\|\ \|v\| < \varepsilon \|v\|.$$
From this it follows that if $v \neq 0$ and $\|v\| < \delta$, we have
$$\frac{|\Delta f(v) - df_p(v)|}{\|v\|} < \varepsilon,$$

which is what we needed to prove. □

4. The Inverse and Implicit Function Theorems

There are two extremely important theorems that allow us to draw deep conclusions about the behavior of a smooth function from information about the behavior of its differential: the inverse function theorem and the implicit function theorem. Proofs of both of these theorems can be found in Chapter 12 of *Advanced Calculus* by Taylor and Mann.

If $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ are open subseets and k is a positive integer, a bijective C^k map $f\colon U \to V$ with C^k inverse $f^{-1}\colon V \to U$ is called a ***C^k diffeomorphism***. Since C^k maps are continuous, every diffeomorphism is in particular a homeomorphism. If we don't specify the integer k, we will generally assume by default that our diffeomorphisms are smooth (i.e., of class C^k for every k).

The following theorem is an easy consequence of the chain rule.

Theorem 1.5. *Suppose $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ are open subsets, and $f\colon U \to V$ is a diffeomorphism. Then for each $p \in U$, the linear map df_p is invertible, with $d(f^{-1})_{f(p)}$ as its inverse.*

Proof: Using the fact that $f^{-1} \circ f = \mathrm{id}_U$, the chain rule implies the following for each $p \in U$:

$$d(f^{-1})_{f(p)} \circ df_p = d(\mathrm{id})_p = \mathrm{id}\,.$$

Similarly, $f \circ f^{-1} = \mathrm{id}_V$ implies

$$df_p \circ d(f^{-1})_{f(p)} = d(\mathrm{id})_{f(p)} = \mathrm{id}\,.$$

Thus df_p is an invertible linear map. □

Corollary 1.6 (Smooth Invariance of Dimension)**.** *If $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ are open subsets and $f\colon U \to V$ is a diffeomorphism, then $m = n$.*

Proof: The preceding theorem shows that df_p is an invertible linear map. Because invertible linear maps take bases to bases, it follows that $m = n$. □

The converse of Theorem 1.5 is not true—it is not too hard to come up with an example of a smooth map whose differential is invertible at each point, but which is not bijective. However, the following theorem (whose proof is much more involved than that of Theorem 1.5) provides a sort of local converse.

Theorem 1.7 (Inverse Function Theorem)**.** *Suppose n is a positive integer, $U \subseteq \mathbb{R}^n$ is an open subset, and $f\colon U \to \mathbb{R}^n$ is a smooth map. If $p \in U$ is a point such that df_p is invertible, then there exist an open subset $U_0 \subseteq U$ containing p and an open subset $V_0 \subseteq \mathbb{R}^n$ containing $f(p)$ such that $f(U_0) = V_0$ and $f|_{U_0}\colon U_0 \to V_0$ is a diffeomorphism.*

The next theorem gives a theoretical answer to the question "when is it possible to solve an equation for one variable in terms of all the others?"

Theorem 1.8 (Implicit Function Theorem)**.** *Suppose $U \subseteq \mathbb{R}^{n+1}$ is an open subset and $f\colon U \to \mathbb{R}$ is a smooth function. Suppose also that $p = (p_1, \ldots, p_{n+1})$ is a point in U such that $\partial f/\partial x_{n+1}(p) \neq 0$, and set $c = f(p)$. Then there exist open sets $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}$ such that $p \in V \times W \subseteq U$, and a smooth function $g\colon V \to W$, such that for all $x = (x_1, \ldots, x_{n+1}) \in V \times W \subseteq U$, we have $f(x) = c$ if and only if $x_{n+1} = g(x_1, \ldots, x_n)$.*

In other words, the conclusion says that provided the partial derivative of f with respect to x_{n+1} is nonzero at p, we can solve the equation $f(x_1, \ldots, x_{n+1}) = c$ for x_{n+1} in terms of the other variables for x sufficiently close to p.

Another way to phrase the implicit function theorem is in terms of graphs. If V and W are any two sets and $g\colon V \to W$ is a function, the ***graph of g*** is the following subset of $V \times W$:

$$\Gamma_g = \{(x, y) \in V \times W : x \in V \text{ and } y = g(x)\}.$$

Corollary 1.9. *Under the same hypotheses as in the implicit function theorem, there exist open sets $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}$ with $p \in V \times W \subseteq U$, such that $f^{-1}(c) \cap (V \times W)$ is the graph of a smooth function $g\colon V \to W$.*

In the statement of the implicit function theorem and its corollary, we have used the last variable as the one to be solved for, just for convenience of notation. But the same argument works for any other variable: if $\partial f/\partial x_i \neq 0$ at some point, then near that point it is possible to solve smoothly for x_i as a function of the other variables.