Math 131BH - Week 7
Textbook pages covered: 204-227

- Review of linear transformations and matrices
- Derivatives in several variable calculus
- Total derivatives, partial derivatives, and directional derivatives
- The chain rule in several variable calculus
- Double derivatives; Clairaut's theorem
- Inverse function theorem
- Implicit function theorem

* * * * *

Review of linear transformations

- We shall now switch to a different topic, namely that of differentiation in several variable calculus. More precisely, we shall be dealing with maps $f : \mathbf{R}^n \to \mathbf{R}^m$ from one Euclidean space to another, and trying to understand what the derivative of such a map is.

- Before we do so, however, we need to recall some notions from linear algebra (Math 33A and Math 115A), most importantly that of a linear transformation and a matrix. We shall be rather brief here, since we do not want to review the entirety of Math 115A.

- First recall that elements of $\mathbf{R}^n$ are sometimes referred to as *n-dimensional row vectors*. A typical n-dimensional row vector may take the form $x = (x_1, x_2, \ldots, x_n)$, which we abbreviate as $(x_i)_{1 \leq i \leq n}$; the quantities $x_1, x_2, \ldots, x_n$ are of course real numbers. If $(x_i)_{1 \leq i \leq n}$ and $(y_i)_{1 \leq i \leq n}$ are n-dimensional row vectors, we can define their vector sum by

$$(x_i)_{1 \leq i \leq n} + (y_i)_{1 \leq i \leq n} = (x_i + y_i)_{1 \leq i \leq n},$$

and also if $c \in \mathbf{R}$ is any scalar, we can define the scalar product $c(x_i)_{1\le i\le n}$by

$$c(x_i)_{1\le i\le n} := (cx_i)_{1\le i\le n}.$$

Of course one has similar operations on $\mathbf{R}^m$ as well. However, if $n \neq m$, then we do not define any operation of vector addition between vectors in $\mathbf{R}^n$ and vectors in $\mathbf{R}^m$ (e.g. $(2,3,4)+(5,6)$ is undefined).

- The operations of vector addition and scalar multiplication obey a number of basic axioms, for instance $c(x+y) = cx + cy$; a full list of these axioms can be found in Math 115A (see e.g. my week 1 notes on this). Because of this, we say that $\mathbf{R}^n$ is a *vector space.* (There are many more examples of vector spaces than this, but we will only need this one).

- If $(x_i)_{1\le i\le n} = (x_1, x_2, \ldots, x_n)$ is an n-dimensional row vector, we can define its *transpose* $(x_i)^\dagger_{1\le i\le n} = (x_1, x_2, \ldots, x_n)^\dagger$ by

$$(x_i)^\dagger_{1\le i\le n} = (x_1, x_2, \ldots, x_n)^\dagger := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

  We refer to objects such as $(x_i)^\dagger_{1\le i\le n}$ as *n-dimensional column vectors.* There is no functional difference between a row vector and a column vector (e.g. one can add and scalar multiply column vectors just as well as we can row vectors), however we shall (rather annoyingly) need to transpose our row vectors into column vectors in order to be consistent with the conventions of matrix multiplication, which we will see later.

- We identify n special vectors in $\mathbf{R}^n$, the *standard basis row vectors* $e_1, \ldots, e_n$. For each $1 \le j \le n$, e_j is the vector which has 0 in all entries except for the j^{th} entry, which is equal to 1. For instance, in $\mathbf{R}^3$, we have $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, and $e_3 = (0,0,1)$. Note that if $x = (x_j)_{1\le j\le n}$ is a vector in $\mathbf{R}^n$, then

$$x = x_1e_1 + x_2e_2 + \ldots + x_ne_n = \sum_{j=1}^n x_je_j,$$

or in other words every vector in $\mathbf{R}^n$ is a *linear combination* of the standard basis vectors $e_1, \ldots, e_n$. (The notation $\sum_{j=1}^n x_j e_j$ is unambiguous because the operation of vector addition is both commutative and associative). Of course, just as every row vector is a linear combination of standard basis row vectors, every column vector is a linear combination of standard basis column vectors:

$$x^\dagger = x_1 e_1^\dagger + x_2 e_2^\dagger + \ldots + x_n e_n^\dagger = \sum_{j=1}^n x_j e_j^\dagger.$$

- There are (many) other ways to create a basis for $\mathbf{R}^n$, but this is a topic for Math 115A and will not be discussed here.

- **Definition** A *linear transformation* $T : \mathbf{R}^n \to \mathbf{R}^m$ is any function from one Euclidean space $\mathbf{R}^n$ to another $\mathbf{R}^m$ which obeys the following two axioms:

- (i) (Additivity) For every $x, x' \in \mathbf{R}^n$, we have $T(x + x') = Tx + Tx'$.

- (ii) (Homogeneity) For every $x \in \mathbf{R}^n$ and every $c \in \mathbf{R}$, we have $T(cx) = cTx$.

- **Example** The *dilation operator* $T_1 : \mathbf{R}^3 \to \mathbf{R}^3$ defined by $T_1 x := 5x$ (i.e. it dilates each vector x by a factor of 5) is a linear transformation, since $5(x + x') = 5x + 5x'$ for all $x, x' \in \mathbf{R}^3$ and $5(cx) = c(5x)$ for all $x \in \mathbf{R}^3$ and $x \in \mathbf{R}$.

- **Example** The *rotation operator* $T_2 : \mathbf{R}^2 \to \mathbf{R}^2$ defined by a clockwise rotation by $\pi/2$ radians around the origin (so that $T_2(1,0) = (0,1)$, $T_2(0,1) = (-1,0)$, etc.) is a linear transformation; this can best be seen geometrically rather than analytically.

- **Example** The *projection operator* $T_3 : \mathbf{R}^3 \to \mathbf{R}^2$ defined by $T_3(x,y,z) := (x,y)$ is a linear transformation (why?). The *inclusion operator* $T_4 : \mathbf{R}^2 \to \mathbf{R}^3$ defined by $T_4(x,y) := (x,y,0)$ is also a linear transformation (why?).

- As we shall shortly see, there is a connection between linear transformations and matrices. Recall that an $m \times n$ *matrix* is an object A of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix};$$

  we shall abbreviate this as

$$A = (a_{ij})_{1 \leq i \leq m; 1 \leq j \leq n}.$$

  In particular, n-dimensional row vectors are $1 \times n$ matrices, while n-dimensional column vectors are $n \times 1$ matrices.

- Given an $m \times n$ matrix A and an $n \times p$ matrix B, we can define the *matrix product* AB to be the $m \times p$ matrix defined as

$$(a_{ij})_{1 \leq i \leq m; 1 \leq j \leq n}(b_{jk})_{1 \leq j \leq n; 1 \leq k \leq p} := (\sum_{j=1}^{n} a_{ij} b_{jk})_{1 \leq i \leq m; 1 \leq k \leq p}.$$

  In particular, if $x^\dagger = (x_j)^\dagger_{1 \leq j \leq n}$ is an n-dimensional column vector, and $A = (a_{ij})_{1 \leq i \leq m; 1 \leq j \leq n}$ is an $m \times n$ matrix, then $Ax^\dagger$ is an m-dimensional column vector:

$$Ax^\dagger = (\sum_{j=1}^{n} a_{ij} x_j)^\dagger_{1 \leq i \leq m}.$$

- We now relate matrices to linear transformations. If A is an $m \times n$ matrix, we can define the transformation $L_A : \mathbf{R}^n \to \mathbf{R}^m$ by the formula

$$(L_A x)^\dagger := Ax^\dagger.$$

  Thus, for instance, if A is the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

and $x = (x_1, x_2, x_3)$ is a 3-dimensional row vector, then $L_A x$ is the 2-dimensional row vector defined by

$$(L_A x)^\dagger = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \end{pmatrix}$$

or in other words

$$L_A(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 4x_1 + 5x_2 + 6x_3).$$

More generally, if

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

then we have

$$L_A(x_j)_{1 \leq j \leq n} = (\sum_{j=1}^{n} a_{ij} x_j)_{1 \leq i \leq m}.$$

For any $m \times n$ matrix A, the transformation L_A is automatically linear; one can easily verify that $L_A(x+y) = L_A x + L_A y$ and $L_A(cx) = c(L_A x)$ for any n-dimensional row vectors x, y and any scalar c. (Why?)

- Perhaps surprisingly, the converse is also true, i.e. every linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ is given by a matrix:

- **Lemma 1.** Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Then there exists exactly one $m \times n$ matrix A such that $T = L_A$.

- **Proof.** Suppose $T : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation. Let $e_1, e_2, \dots, e_n$ be the standard basis row vectors of $\mathbf{R}^n$. Then $Te_1, Te_2, \dots, Te_n$ are vectors in $\mathbf{R}^m$. For each $1 \leq j \leq n$, we write Te_j in co-ordinates as

$$Te_j = (a_{1j}, a_{2j}, \dots, a_{mj}) = (a_{ij})_{1 \leq i \leq m},$$

i.e. we define a_{ij} to be the i^{th} component of Te_j. Then for any n-dimensional row vector $x = (x_1, \ldots, x_n)$, we have

$$Tx = T(\sum_{j=1}^{n} x_j e_j),$$

which (since T is linear) is equal to

$$= \sum_{j=1}^{n} T(x_j e_j) = \sum_{j=1}^{n} x_j T e_j$$

$$= \sum_{j=1}^{n} x_j (a_{ij})_{1 \leq i \leq m}$$

$$= \sum_{j=1}^{n} (a_{ij} x_j)_{1 \leq i \leq m}$$

$$= (\sum_{j=1}^{n} a_{ij} x_j)_{1 \leq i \leq m}.$$

But if we let A be the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

then the previous vector is precisely $L_A x$. Thus $Tx = L_A x$ for all n-dimensional vectors x, and thus $T = L_A$.

- Now we show that A is unique, i.e. there does not exist any other matrix

$$B = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \ldots & b_{mn} \end{pmatrix}$$

for which T is equal to L_B. Suppose for contradiction that we could find such a matrix B which was different from A. Then we would have $L_A = L_B$. In particular, we have $L_A e_j = L_B e_j$ for every $1 \leq j \leq n$. But from the definition of L_A we see that

$$L_A e_j = (a_{ij})_{1 \leq i \leq m}$$

and

$$L_B e_j = (b_{ij})_{1 \leq i \leq m}$$

and thus we have $a_{ij} = b_{ij}$ for every $1 \leq i \leq m$ and $1 \leq j \leq m$, thus A and B are equal, contradiction. □

- This Lemma establishes a one-to-one correspondence between linear transformations and matrices, and is one of the fundamental reasons why matrices are so important in linear algebra. (One may ask then why we bother dealing with linear transformations at all, and why we don't just work with matrices all the time. The reason is that sometimes one does not want to work with the standard basis $e_1, \ldots, e_n$, but instead wants to use some other basis. In that case, the correspondence between linear transformations and matrices changes, and so it is still important to keep the notions of linear transformation and matrix distinct. See Math 115A for more information).

- If $T = L_A$, then A is sometimes called the *matrix representation of* T, and is sometimes denoted $A = [T]$. We shall avoid this notation here, however.

- If $T : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, and $S : \mathbf{R}^p \to \mathbf{R}^n$ is a linear transformation, then the composition $TS : \mathbf{R}^p \to \mathbf{R}^m$ of the two transforms, defined by $TS(x) := T(S(x))$, is also a linear transformation (why? Expand $TS(x+y)$ and $TS(cx)$ carefully, using plenty of parentheses). The next lemma shows that the operation of composing linear transformations is connected to that of matrix multiplication.

- **Lemma 2.** Let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix. Then $L_A L_B = L_{AB}$.

- **Proof.** See Week 7 homework. □

* * * * *

Derivatives in several variable calculus

- Now that we've reviewed some linear algebra, we turn now to our main topic, which is that of understanding differentiation of functions of the form $f : \mathbf{R}^n \to \mathbf{R}^m$, i.e. functions from one Euclidean space to another. For instance, one might want to differentiate the function $f : \mathbf{R}^3 \to \mathbf{R}^4$ defined by

$$f(x, y, z) = (xy, yz, xz, xyz).$$

- In single variable calculus, when one wants to differentiate a function $f : E \to \mathbf{R}$ at a point x_0, where E is a subset of $\mathbf{R}$ that contains x_0, this is given by

$$f'(x_0) := \lim_{x \to x_0; x \in E \backslash \{x_0\}} \frac{f(x) - f(x_0)}{x - x_0}.$$

  One could try to mimic this definition in the several variable case $f : E \to \mathbf{R}^m$, where E is now a subset of $\mathbf{R}^n$, however we encounter a difficulty in this case: the quantity $f(x) - f(x_0)$ will live in $\mathbf{R}^m$, and $x - x_0$ lives in $\mathbf{R}^n$, and we do not know how to divide an m-dimensional vector by an n-dimensional vector.

- To get around this problem, we will rewrite the concept of derivative in a way which does not involve division of vectors.

- Conversely, if f behaves like a linear function near x_0, then f is differentiable at x_0:

- **Lemma 3.** Let E be a subset of $\mathbf{R}$, $f : E \to \mathbf{R}$ be a function, $x_0 \in E$, and $L \in \mathbf{R}$. Then the following two statements are equivalent.

- (a) f is differentiable at x_0, and $f'(x_0) = L$.

- (b) We have $\lim_{x \to x_0; x \in E - \{x_0\}} \frac{|f(x) - (f(x_0) + L(x - x_0))|}{|x - x_0|} = 0$.

- **Proof.** See Week 7 homework. □

- In light of the above lemma, we see that the derivative $f'(x_0)$ can be interpreted as the number L for which $|f(x) - (f(x_0) + L(x - x_0))|$ is small, in the sense that it tends to zero as x tends to x_0, even if we divide out by the very small number $|x - x_0|$. More informally, the derivative is the quantity L such that we have the approximation $f(x) - f(x_0) \approx L(x - x_0)$.

- This does not seem too different from the usual notion of differentiation, but the point is that we are no longer explicitly dividing by $x - x_0$. (We are still dividing by $|x - x_0|$, but this will turn out to be OK). When we move to the several variable case $f : E \to \mathbf{R}^m$, where $E \subseteq \mathbf{R}^n$, we shall still want the derivative to be some quantity L such that $f(x) - f(x_0) \approx L(x - x_0)$. However, since $f(x) - f(x_0)$ is now an m-dimensional vector and $x - x_0$ is an n-dimensional vector, we no longer want L to be a scalar; we want it to be a linear transformation. More precisely:

- **Definition.** Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, $x_0 \in E$ be a point, and let $L : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. We say that f is *differentiable at x_0 with derivative L* if we have

$$\lim_{x \to x_0; x \in E - \{x_0\}} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Here $\|x\|$ is the length of x (as measured in the l^2 metric):

$$\|(x_1, x_2, \ldots, x_n)\| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}.$$

- **Example.** Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the map $f(x, y) := (x^2, y^2)$, let x_0 be the point $x_0 := (1, 2)$, and let $L : \mathbf{R}^2 \to \mathbf{R}^2$ be the map $L(x, y) := (2x, 4y)$. We claim that f is differentiable at x_0 with derivative L. To see this, we compute

$$\lim_{(x,y) \to (1,2) : (x,y) \neq (1,2)} \frac{\|f(x, y) - (f(1, 2) + L((x, y) - (1, 2)))\|}{\|(x, y) - (1, 2)\|}.$$

Making the change of variables $(x, y) = (1, 2) + (a, b)$, this becomes

$$\lim_{(a,b) \to (0,0) : (a,b) \neq (0,0)} \frac{\|f(1 + a, 2 + b) - (f(1, 2) + L(a, b))\|}{\|(a, b)\|}.$$

Substituting the formula for f and for L, this becomes

$$\lim_{(a,b)\to(0,0):(a,b)\neq(0,0)} \frac{\|((1+a)^2,(2+b)^2)-(1,4)-(2a,4b))\|}{\|(a,b)\|},$$

which simplifies to

$$\lim_{(a,b)\to(0,0):(a,b)\neq(0,0)} \frac{\|(a^2,b^2)\|}{\|(a,b)\|}.$$

We use the squeeze test. The expression $\frac{\|(a^2,b^2)\|}{\|(a,b)\|}$ is clearly non-negative. On the other hand, we have by the triangle inequality

$$\|(a^2,b^2)\| \leq \|(a^2,0)\| + \|(0,b^2)\| = a^2+b^2$$

and hence

$$\frac{\|(a^2,b^2)\|}{\|(a,b)\|} \leq \sqrt{a^2+b^2}.$$

Since $\sqrt{a^2+b^2} \to 0$ as $(a,b) \to 0$, we thus see from the squeeze test that the above limit exists and is equal to 0. Thus f is differentiable at x_0 with derivative L.

- As you can see, verifying that a function is differentiable from first principles can be somewhat tedious. Later on we shall find better ways to verify differentiability, and to compute derivatives.

- We have to check a basic fact, which is a function can have at most one derivative at any *interior* point of its domain:

- **Lemma 4.** Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, $x_0 \in E$ be an *interior point* of E, and let $L_1 : \mathbf{R}^n \to \mathbf{R}^m$ and $L_2 : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformations. Suppose that f is differentiable at x_0 with derivative L_1, and also differentiable at x_0 with derivative L_2. Then $L_1 = L_2$.

- **Proof.** See Week 7 homework. $\square$

- Because of Lemma 4, we can now talk about *the* derivative of f at interior points x_0, and we will denote this derivative by $f'(x_0)$. Thus $f'(x_0)$ is the unique linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ such that

$$\lim_{x \to x_0; x \neq x_0} \frac{\|f(x) - (f(x_0) + f'(x_0)(x - x_0))\|}{\|x - x_0\|} = 0.$$

  Informally, this means that derivative $f'(x_0)$ is the linear transformation such that we have

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0)$$

  or equivalently

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

  (this is known as *Newton's approximation*).

- Another consequence of Lemma 4 is that if you know that $f(x) = g(x)$ for all $x \in E$, and f, g are differentiable at x_0, then you also know that $f'(x_0) = g'(x_0)$ at every *interior* point of E. However, this is not necessarily true if x_0 is a boundary point of E; for instance, if E is just a single point $E = \{x_0\}$, merely knowing that $f(x_0) = g(x_0)$ does not imply that $f'(x_0) = g'(x_0)$. We will not deal with these boundary issues here, and only compute derivatives on the interior of the domain.

- We will sometimes refer to f' as the *total derivative* of f, to distinguish this concept from that of partial and directional derivatives below.

* * * * *

Connection with partial and directional derivatives

- We now connect the notion of differentiability with that of partial and directional derivatives, which you should have already seen in Math 32A.

- **Definition** Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, let x_0 be an interior point of E, and let v be a vector in $\mathbf{R}^n$. If the limit

$$\lim_{t \to 0; t \neq 0, x_0 + tv \in E} \frac{f(x_0 + tv) - f(x_0)}{t}$$

exists, we say that f is *differentiable in the direction v at x_0*, and we denote the above limit by $D_v f(x_0)$:

$$D_v f(x_0) := \lim_{t \to 0; t \neq 0} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

Equivalently, we have

$$D_v f(x_0) := \frac{d}{dt} f(x_0 + tv)|_{t=0}.$$

- Note that this time we are dividing by a scalar t, rather than a vector, so this definition makes sense, and $D_v f(x_0)$ will be a vector in $\mathbf{R}^m$. It is sometimes possible to also define directional derivatives on the boundary of E, if the vector v is pointing in an "inward" direction (this generalizes the notion of left derivatives and right derivatives from single variable calculus); but we will not pursue these matters here.

- **Example.** We use the function $f : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $f(x, y) := (x^2, y^2)$ from before, and let $x_0 := (1, 2)$ and $v := (3, 4)$. Then

$$D_v f(x_0) = \lim_{t \to 0; t \neq 0} \frac{f(1 + 3t, 2 + 4t) - f(1, 2)}{t}$$

$$= \lim_{t \to 0; t \neq 0} \frac{(1 + 6t + 9t^2, 4 + 16t + 16t^2) - (1, 4)}{t}$$

$$= \lim_{t \to 0; t \neq 0} (6 + 9t, 16 + 16t) = (6, 16).$$

- Directional derivatives are connected with total derivatives as follows:

- **Lemma 5.** Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, x_0 be an interior point of E, and let v be a vector in $\mathbf{R}^n$. If f is differentiable at x_0, then f is also differentiable in the direction v at x_0, and

$$D_v f(x_0) = f'(x_0) v.$$

- **Proof.** See Week 7 homework. □

- Thus total differentiability implies directional differentiability. However, the converse is not true; see homework.

- A special case of directional derivative is when the direction v is equal to one of the basis vectors e_j. In this case we sometimes write $\frac{\partial f}{\partial x_j}(x_0)$ or $\frac{\partial}{\partial x_j} f(x_0)$ for $D_{e_j} f(x_0)$, and refer to $\frac{\partial f}{\partial x_j}(x_0)$ as the *partial derivative of f with respect to x_j*. Thus

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t\to 0; t\neq 0, x_0+tv\in E} \frac{f(x_0+te_j)-f(x_0)}{t} = \frac{d}{dt} f(x_0+te_j)|_{t=0}.$$

Informally, the partial derivative can be obtained by holding all the variables other than x_j fixed, and then applying the single-variable calculus derivative in the x_j variable. Note that if f takes values in $\mathbf{R}^m$, then so will $\frac{\partial f}{\partial x_j}$. Indeed, if we write f in components as $f = (f_1, \ldots, f_m)$, it is easy to see (why?) that

$$\frac{\partial f}{\partial x_j}(x_0) = (\frac{\partial f_1}{\partial x_j}(x_0), \ldots, \frac{\partial f_m}{\partial x_j}(x_0)),$$

i.e. to differentiate a vector-valued function one just has to differentiate each of the components separately.

- We sometimes replace the variables x_j in $\frac{\partial f}{\partial x_j}$ with other symbols. For instance, if we are dealing with the function $f(x,y) = (x^2, y^2)$, then we might refer to $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ instead of $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$. (In this case, $\frac{\partial f}{\partial x}(x,y) = (2x, 0)$ and $\frac{\partial f}{\partial y}(x,y) = (0, 2y)$). One should caution however that one should only relabel the variables if it is absolutely clear which symbol refers to the first variable, which symbol refers to the second variable, etc.; otherwise one may become unintentionally confused. For instance, in the above example, the expression $\frac{\partial f}{\partial x}(x,x)$ is just $(2x, 0)$, however one may mistakenly compute

$$\frac{\partial f}{\partial x}(x,x) = \frac{\partial}{\partial x}(x^2, x^2) = (2x, 2x);$$

the problem here is that the symbol x is being used for more than just the first variable of f. (On the other hand, it is true that $\frac{d}{dx} f(x,x)$ is equal to $(2x, 2x)$; thus the operation of total differentiation $\frac{d}{dx}$ is not the same as that of partial differentiation $\frac{\partial}{\partial x}$).

- From Lemma 5, we know that if a function is differentiable at a point x_0, then all the partial derivatives $\frac{\partial f}{\partial x_j}$ exist at x_0, and that

$$\frac{\partial f}{\partial x_j}(x_0) = f'(x_0)e_j.$$

Also, if $v = (v_1, \ldots, v_n) = \sum_j v_j e_j$, then we have

$$D_v f(x_0) = f'(x_0) \sum_j v_j e_j = \sum_j v_j f'(x_0) e_j$$

(since $f'(x_0)$ is linear) and thus

$$D_v f(x_0) = \sum_j v_j \frac{\partial f}{\partial x_j}(x_0).$$

Thus one can write directional derivatives in terms of partial derivatives, *provided that* the function is actually differentiable at that point.

- As mentioned before, however, just because the partial derivatives exist at a point x_0, we cannot conclude that the function is differentiable there. However, if we know that the partial derivatives not only exist, but are continuous, then we can in fact conclude differentiability:

- **Theorem 6.** Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, F be a subset of E, and x_0 be an interior point of F. If all the partial derivatives $\frac{\partial f}{\partial x_j}$ exist on F and are continuous at x_0, then f is differentiable at x_0, and the linear transformation $f'(x_0) : \mathbf{R}^n \to \mathbf{R}^m$ is defined by

$$f'(x_0)(v_j)_{1 \leq j \leq n} = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x_0).$$

- **Proof.** Let $L : \mathbf{R}^n \to \mathbf{R}^m$ be the linear transformation

$$L(v_j)_{1 \leq j \leq m} := \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x_0).$$

We have to prove that

$$\lim_{x \to x_0; x \in E - \{x_0\}} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Let $\varepsilon > 0$. It will suffice to find a radius $\delta > 0$ such that

$$\frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} \leq \varepsilon$$

for all $x \in B(x_0, \delta)\backslash\{x_0\}$. Equivalently, we wish to show that

$$\|f(x) - f(x_0) - L(x - x_0)\| \leq \varepsilon\|x - x_0\|$$

for all $x \in B(x_0, \delta)\backslash\{x_0\}$.

Because x_0 is an interior point of F, there exists a ball $B(x_0, r)$ which is contained inside F. Because each partial derivative $\frac{\partial f}{\partial x_j}$ is continuous on F, there thus exists an $0 < \delta_j < r$ such that $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon/nm$ for every $x \in B(x_0, \delta_j)$. If we take $\delta = \min(\delta_1, \ldots, \delta_n)$, then we thus have $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon/nm$ for every $x \in B(x_0, \delta)$ and every $1 \leq j \leq n$.

Let $x \in B(x_0, \delta)$. We write $x = x_0 + v_1 e_1 + v_2 e_2 + \ldots + v_n e_n$ for some scalars $v_1, \ldots, v_n$. Note that

$$\|x - x_0\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

and in particular we have $|v_j| \leq \|x - x_0\|$ for all $1 \leq j \leq n$. Our task is to show that

$$\|f(x_0 + v_1 e_1 + \ldots + v_n e_n) - f(x_0) - \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon\|x - x_0\|.$$

Write f in components as $f = (f_1, f_2, \ldots, f_m)$ (so each f_i is a function from E to $\mathbf{R}$). From the mean value theorem in the x_1 variable, we see that

$$f_i(x_0 + v_1 e_1) - f_i(x_0) = \frac{\partial f_i}{\partial x_1}(x_0 + t_i e_1) v_1$$

for some t_i between 0 and v_1. But we have

$$|\frac{\partial f_i}{\partial x_j}(x_0 + t_i e_1) - \frac{\partial f_i}{\partial x_j}(x_0)| \leq \|\frac{\partial f}{\partial x_j}(x_0 + t_i e_1) - \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon/nm$$

and hence

$$|f_i(x_0 + v_1 e_1) - f_i(x_0) - \frac{\partial f_i}{\partial x_1}(x_0) v_1| \leq \varepsilon |v_1|/nm.$$

Summing this over all $1 \leq i \leq m$ (and noting that $\|(y_1, \ldots, y_m)\| \leq |y_1| + \ldots + |y_m|$ from the triangle inequality) we obtain

$$\|f(x_0 + v_1 e_1) - f(x_0) - \frac{\partial f}{\partial x_1}(x_0) v_1\| \leq \varepsilon |v_1|/n;$$

since $|v_1| \leq \|x - x_0\|$, we thus have

$$\|f(x_0 + v_1 e_1) - f(x_0) - \frac{\partial f}{\partial x_1}(x_0) v_1\| \leq \varepsilon \|x - x_0\|/n.$$

A similar argument gives

$$\|f(x_0 + v_1 e_1 + v_2 e_2) - f(x_0 + v_1 e_1) - \frac{\partial f}{\partial x_2}(x_0) v_2\| \leq \varepsilon \|x - x_0\|/n.$$

and so forth up to

$$\|f(x_0 + v_1 e_1 + \ldots + v_n e_n) - f(x_0 + v_1 e_1 + \ldots + v_{n-1} e_{n-1}) - \frac{\partial f}{\partial x_n}(x_0) v_n\| \leq \varepsilon \|x - x_0\|/n.$$

If we sum these n inequalities and use the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$, we obtain a telescoping series which simplifies to

$$\|f(x_0 + v_1 e_1 + \ldots + v_n e_n) - f(x_0) - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x_0) v_j\| \leq \varepsilon \|x - x_0\|$$

as desired. □

- From Theorem 6 and Lemma 5 we see that if the partial derivatives of a function $f : E \to \mathbf{R}^m$ exist and are continuous on some set F, then all the directional derivatives also exist at every interior point x_0 of F, and we have the formula

$$D_{(v_1, \ldots, v_n)} f(x_0) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

In particular, if $f : E \to \mathbf{R}$ is a real-valued function, and we define the *gradient* $\nabla f(x_0)$ of f at x_0 to be the n-dimensional row vector $\nabla f(x_0) := (\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0))$, then we have the familiar formula

$$D_v f(x_0) = v \cdot \nabla f(x_0)$$

whenever x_0 is in the interior of the region where the gradient exists and is continuous.

- More generally, if $f : E \to \mathbf{R}^m$ is a function taking values in $\mathbf{R}^m$, with $f = (f_1, \ldots, f_m)$, and x_0 is in the interior of the region where the partial derivatives of f exist and are continuous, then we have from Theorem 6 that

$$f'(x_0)(v_j)_{1 \leq j \leq n} = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x_0)$$
$$= (\sum_{j=1}^n v_j \frac{\partial f_i}{\partial x_j}(x_0))_{i=1}^m,$$

which we can rewrite as

$$L_{Df(x_0)}(v_j)_{1 \leq j \leq n}$$

where $Df(x_0)$ is the $m \times n$ matrix

$$Df(x_0) = (\frac{\partial f_i}{\partial x_j}(x_0))_{1 \leq i \leq m; 1 \leq j \leq n} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \ldots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \ldots & \frac{\partial f_2}{\partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \ldots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}.$$

Thus we have

$$(D_v f(x_0))^\dagger = (f'(x_0)v)^\dagger = Df(x_0)v^\dagger.$$

- The matrix $Df(x_0)$ is sometimes also called the *derivative* or *differential* of f at x_0, although we will try to avoid this notation in order to separate the matrix $Df(x_0)$ from the linear transformation $f'(x_0)$. One can also write Df as

$$Df(x_0) = (\frac{\partial f}{\partial x_1}(x_0)^\dagger, \frac{\partial f}{\partial x_2}(x_0)^\dagger, \ldots, \frac{\partial f}{\partial x_n}(x_0)^\dagger),$$

i.e. each of the columns of $Df(x_0)$ is one of the partial derivatives of f, expressed as a column vector. Or one could write

$$Df(x_0) = \begin{pmatrix} \nabla f_1(x_0) \\ \nabla f_2(x_0) \\ \vdots \\ \nabla f_m(x_0) \end{pmatrix}$$

i.e. the rows of $Df(x_0)$ are the gradient of various components of f. In particular, if f is scalar-valued (i.e. $m = 1$), then Df is the same as ∇f.

- **Example.** Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the function $f(x, y) = (x^2 + xy, y^2)$. Then $\frac{\partial f}{\partial x} = (2x + y, 0)$ and $\frac{\partial f}{\partial y} = (x, 2y)$. Since these partial derivatives are continuous on $\mathbf{R}^2$, we see that f is differentiable on all of $\mathbf{R}^2$, and

$$Df(x, y) = \begin{pmatrix} 2x + y & x \\ 0 & 2y \end{pmatrix}.$$

Thus for instance, the directional derivative in the direction (v, w) is

$$D_{(v,w)}f(x, y) = ((2x + y)v + xw, 2yw).$$

* * * * *

The several variable calculus chain rule

- We are now ready to state the several variable calculus chain rule. Recall that if $f : X \to Y$ and $g : Y \to Z$ are two functions, then the composition $g \circ f : X \to Z$ is defined by $g \circ f(x) := g(f(x))$ for all $x \in X$.

- **Several variable calculus chain rule.** Let E be a subset of $\mathbf{R}^n$, and let F be a subset of $\mathbf{R}^m$. Let $f : E \to F$ be a function, and let $g : F \to \mathbf{R}^p$ be another function. Let x_0 be a point in the interior of E. Suppose that f is differentiable at x_0, and that $f(x_0)$ is in the interior of F. Suppose also that g is differentiable at $f(x_0)$. Then $g \circ f : E \to \mathbf{R}^p$ is also differentiable at x_0, and we have the formula

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

- **Proof.** See Week 7 homework. □

- Intuitively, one can think of the several variable chain rule as follows. Let x be close to x_0. Then Newton's approximation asserts that

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0)$$

and in particular $f(x)$ is close to $f(x_0)$. Since g is differentiable at $f(x_0)$, we see from Newton's approximation again that

$$g(f(x)) - g(f(x_0)) \approx g'(f(x_0))(f(x) - f(x_0)).$$

Combining the two, we obtain

$$g \circ f(x) - g \circ f(x_0) \approx g'(f(x_0))f'(x_0)(x - x_0)$$

which then should give $(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$. This argument however is rather imprecise; to make it more precise one needs to manipulate limits rigorously (see homework).

- As a corollary of the chain rule and Lemma 2 (and Lemma 1), we see that

$$D(g \circ f)(x_0) = Dg(f(x_0))Df(x_0);$$

i.e. we can write the chain rule in terms of matrices and matrix multiplication, instead of in terms of linear transformations and composition.

- **Example** Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$ be differentiable functions. We form the combined function $h : \mathbf{R}^n \to \mathbf{R}^2$ by defining $h(x) := (f(x), g(x))$. Now let $k : \mathbf{R}^2 \to \mathbf{R}$ be the multiplication function $k(a, b) := ab$. Note that

$$Dh(x_0) = \begin{pmatrix} \nabla f(x_0) \\ \nabla g(x_0) \end{pmatrix}$$

while

$$Dk(a, b) = (b, a)$$

(why?). By the chain rule, we thus see that

$$D(k \circ h)(x_0) = (g(x_0), f(x_0)) \begin{pmatrix} \nabla f(x_0) \\ \nabla g(x_0) \end{pmatrix} = g(x_0)\nabla f(x_0) + f(x_0)\nabla g(x_0).$$

But $k \circ h = fg$ (why?), and $D(fg) = \nabla(fg)$. We have thus proven the *product rule*

$$\nabla(fg) = g\nabla f + f\nabla g.$$

- A similar argument gives the sum rule $\nabla(f+g) = \nabla f + \nabla g$, or the difference rule $\nabla(f-g) = \nabla f - \nabla g$. In the homework you will be asked to similarly derive a several variable calculus quotient rule from the chain rule. As you can see, the several variable chain rule is quite powerful, and can be used to deduce many other rules of differentiation.

- We do record one useful application of the chain rule. Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Observe that T is continuously differentiable at every point, and in fact $T'(x) = T$ for every x. (This equation may look a little strange, but perhaps it is easier to swallow if you view it in the form $\frac{d}{dx}(Tx) = T$). Thus, for any differentiable function $f : E \to \mathbf{R}^n$, we see that $Tf : E \to \mathbf{R}^m$ is also differentiable, and hence by the chain rule

$$(Tf)'(x_0) = T(f'(x_0)).$$

  This is a generalization of the single-variable calculus rule $(cf)' = c(f')$ for constant scalars c.

- Another special case of the chain rule which is quite useful is the following: if $f : \mathbf{R}^n \to \mathbf{R}^m$ is some differentiable function, and $x_j : \mathbf{R} \to \mathbf{R}$ are differentiable functions for each $j = 1, \ldots n$, then

$$\frac{d}{dt} f(x_1(t), x_2(t), \ldots, x_n(t)) = \sum_{j=1}^{n} x_j'(t) \frac{\partial f}{\partial x_j}(x_1(t), x_2(t), \ldots, x_n(t)).$$

  (Why is this a special case of the chain rule?).

* * * * *

Double derivatives and Clairaut's theorem

- We now investigate what happens if one differentiates a function twice.

- **Definition** Let E be an open subset of $\mathbf{R}^n$, and let $f : E \to \mathbf{R}^m$ be a function. We say that f is *continuously differentiable* if the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ exist and are continuous on E. We say that f is *twice continuously differentiable* if it is continuously differentiable, and the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are themselves continuously differentiable.

- Continuously differentiable functions are sometimes called C^1 functions; twice continuously differentiable functions are sometimes called C^2 functions. One can also define C^3, C^4, etc. but we shall not do so here.

- **Example** Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the function $f(x, y) = (x^2 + xy, y^2)$. Then f is continuously differentiable because the partial derivatives $\frac{\partial f}{\partial x}(x, y) = (2x+y, 0)$ and $\frac{\partial f}{\partial y}(x, y) = (x, 2y)$ exist and are continuous on all of $\mathbf{R}^2$. It is also twice continuously differentiable, because the double partial derivatives $\frac{\partial}{\partial x}\frac{\partial f}{\partial x}(x, y) = (2, 0)$, $\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(x, y) = (1, 0)$, $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x, y) = (1, 0)$, $\frac{\partial}{\partial y}\frac{\partial f}{\partial y}(x, y) = (0, 2)$ all exist and are continuous.

- Observe in this example that the double derivatives $\frac{\partial}{\partial y}\frac{\partial f}{\partial x}$ and $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}$ are the same. This is a general fact:

- **Clairaut's theorem.** Let E be an open subset of $\mathbf{R}^n$, and let $f : E \to \mathbf{R}$ be a twice continuously differentiable function on E. Then we have $\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x_0) = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x_0)$ for all $1 \leq i, j \leq n$.

- **Proof.** The claim is trivial if $i = j$, so we shall assume that $i \neq j$. We shall prove the theorem for $x_0 = 0$; the general case is similar. (Actually, once one proves Clairaut's theorem for $x_0 = 0$, one can immediately obtain it for general x_0 by applying the theorem with $f(x)$ replaced by $f(x - x_0)$).

- Let a be the number $a := \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(0)$, and a' denote the quantity $a' := \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(0)$. Our task is to show that $a' = a$.

- Let $\varepsilon > 0$. Because the double derivatives of f are continuous, we can find a $\delta > 0$ such that

$$|\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x) - a| \leq \varepsilon$$

and

$$|\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) - a'| \leq \varepsilon$$

whenever $|x| \leq 2\delta$.

- Now we consider the quantity

$$X := f(\delta e_i + \delta e_j) - f(\delta e_i) - f(\delta e_j) - f(0).$$

From the fundamental theorem of calculus in the e_i variable, we have

$$f(\delta e_i + \delta e_j) - f(\delta e_j) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j)\ dx_i$$

and

$$f(\delta e_i) - f(0) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i)\ dx_i$$

and hence

$$X = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i)\ dx_i.$$

But by the mean value theorem, for each x_i we have

$$\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) = \delta\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x_i e_i + x_j e_j)$$

for some $0 \leq x_j \leq \delta$. By our construction of δ, we thus have

$$|\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a| \leq \varepsilon\delta.$$

Integrating this from 0 to δ, we thus obtain

$$|X - \delta^2 a| \leq \varepsilon\delta^2.$$

- We can run the same argument with the role of i and j reversed (note that X is symmetric in i and j), to obtain

$$|X - \delta^2 a'| \leq \varepsilon\delta^2.$$

From the triangle inequality we thus obtain

$$|\delta^2 a - \delta^2 a'| \leq 2\varepsilon\delta^2,$$

and thus

$$|a - a'| \leq 2\varepsilon.$$

But this is true for all $\varepsilon > 0$, and a and a' do not depend on ε, and so we must have $a = a'$, as desired. □

- One should caution that Clairaut's theorem fails if we do not assume the double derivatives to be continuous; see homework.

* * * * *

Digression: the contraction mapping theorem

- Before we turn to the next topic - namely, the inverse function theorem - we need to develop a useful fact from the theory of complete metric spaces, namely the contraction mapping theorem.

- **Definition** Let (X, d) be a metric space, and let $f : X \to X$ be a map. We say that f is a *contraction* if we have $d(f(x), f(y)) \leq d(x, y)$ for all $x, y \in X$. We say that f is a *strict contraction* if there exists a constant $0 < c < 1$ such that $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$.

- **Example** The map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) := x+1$ is a contraction but not a strict contraction. The map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) := x/2$ is a strict contraction. The map $f : [0, 1] \to [0, 1]$ defined by $f(x) := x - x^2$ is a contraction but not a strict contraction.

- One can easily verify that contractions are continuous (why?). Also, every strict contraction is of course also a contraction, but not conversely.

- **Definition** Let $f : X \to X$ be a map, and $x \in X$. We say that x is a *fixed point* of f if $f(x) = x$.

- Contractions do not necessarily have any fixed points; for instance, the map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x + 1$ does not. However, it turns out that *strict* contractions always do, at least when X is complete:

- **Contraction mapping theorem.** Let (X, d) be a metric space, and let $f : X \to X$ be a strict contraction. Then f can have at most one fixed point. Moreover, if we also assume that X is non-empty and complete, then f has *exactly* one fixed point.

- **Proof.** See Week 7 homework. □

- The contraction mapping theorem is one example of a *fixed point theorem* - a theorem which guarantees, assuming certain conditions, that a map will have a fixed point. There are a number of other fixed point theorems which are also useful. One amusing one is the so-called *hairy ball theorem*, which (among other things) states that any continuous map $f : S^2 \to S^2$ from the sphere $S^2 := \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 = 1\}$ to itself, must contain either a fixed point, or an anti-fixed point (a point $x \in S^2$ such that $f(x) = -x$). A proof of this theorem can be found in Math 121.

- We shall give one consequence of the contraction mapping theorem which is important for our application to the inverse function theorem. Basically, this says that any map f on a ball which is a "small" perturbation of the identity map, remains one-to-one and cannot create any internal holes in the ball.

- **Lemma 7.** Let $B(0, r)$ be a ball in $\mathbf{R}^n$ centered at the origin, and let $g : B(0, r) \to \mathbf{R}^n$ be a map such that $g(0) = 0$ and

$$\|g(x) - g(y)\| \leq \frac{1}{2}\|x - y\|$$

for all $x, y \in B(0, r)$ (here $\|x\|$ denotes the length of x in $\mathbf{R}^n$). Then the function $f : B(0, r) \to \mathbf{R}^n$ defined by $f(x) := x + g(x)$ is one-to-one, and furthermore the image $f(B(0, r))$ of this map contains the ball $B(0, r/2)$.

- **Proof.** We first show that f is one-to-one. Suppose for contradiction that we had two different points $x, y \in B(0, r)$ such that $f(x) = f(y)$. But then we would have $x + g(x) = y + g(y)$, and hence

$$\|g(x) - g(y)\| = \|x - y\|.$$

The only way this can be consistent with our hypothesis $\|g(x)-g(y)\| \leq \frac{1}{2}\|x-y\|$ is if $\|x-y\|=0$, i.e. if $x=y$, contradiction. Thus f is one-to-one.

- Now we show that $f(B(0,r))$ contains $B(0,r/2)$. Let y be any point in $B(0,r/2)$; our objective is to find a point $x \in B(0,r)$ such that $f(x)=y$, or in other words that $x=y-g(x)$. So the problem is now to find a fixed point of the map $x \mapsto y-g(x)$.

- Let $F: B(0,r) \to B(0,r)$ denote the function $F(x) := y - g(x)$. Observe that if $x \in B(0,r)$, then

$$\|F(x)\| \leq \|y\|+\|g(x)\| \leq \frac{r}{2}+\|g(x)-g(0)\| \leq \frac{r}{2}+\frac{1}{2}\|x-0\| < \frac{r}{2}+\frac{r}{2} = r,$$

so F does indeed map $B(0,r)$ to itself. Also, for any x, x' in $B(0,r)$ we have

$$\|F(x)-F(x')\| = \|g(x')-g(x)\| \leq \frac{1}{2}\|x'-x\|$$

so F is a strict contraction. By the contraction mapping theorem, F has a fixed point, i.e. there exists an x such that $x = y - g(x)$. But this means that $f(x)=y$, as desired. □

* * * * *

The inverse function theorem in several variable calculus

- We recall the inverse function theorem in single variable calculus, which asserts that if a function $f : \mathbf{R} \to \mathbf{R}$ is invertible, differentiable, and $f'(x_0)$ is non-zero, then f^{-1} is differentiable at $f(x_0)$, and

$$(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}.$$

- In fact, one can say something even when f is not invertible, as long as we know that f is *continuously* differentiable. If $f'(x_0)$ is non-zero, then $f'(x_0)$ must be either strictly positive or strictly negative, which implies (since we are assuming f' to be continuous) that $f'(x)$ is either strictly positive for x near x_0, or strictly negative for x near x_0. In particular, f must be either strictly increasing near x_0, or strictly decreasing near

x_0. In either case, f will become invertible if we restrict the domain and range of f to be sufficiently close to x_0 and to $f(x_0)$ respectively. (The technical terminology for this is that f is *locally invertible near* x_0).

- The requirement that f be continuously differentiable is important. For instance, if one lets $f : \mathbf{R} \to \mathbf{R}$ be the function $f(x) := x + x^2 \sin(1/x^4)$ for $x \neq 0$ and $f(0) := 0$, then one can show that f is differentiable and $f'(0) = 1$, but f is not increasing on any open set containing 0 (this is easiest to see by sketching the graph of f; alternatively, one can show that the derivative of f can turn negative arbitrarily close to 0).

- It turns out that a similar theorem is true for functions $f : \mathbf{R}^n \to \mathbf{R}^n$ from one Euclidean space to the same space. However, the condition that $f'(x_0)$ is non-zero must be replaced with a slightly different one, namely that $f'(x_0)$ is invertible. We first remark that the inverse of a linear transformation is also linear:

- **Lemma 8** Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be a linear transformation which is also invertible. Then the inverse transformation $T^{-1} : \mathbf{R}^n \to \mathbf{R}^n$ is also linear.

- **Proof.** See Week 7 homework. □

- **Inverse function theorem.** Let E be an open subset of $\mathbf{R}^n$, and let $T : E \to \mathbf{R}^n$ be a function which is continuously differentiable on E. Suppose $x_0 \in E$ is such that the linear transformation $f'(x_0) : \mathbf{R}^n \to \mathbf{R}^n$ is invertible. Then there exists an open set U in E containing x_0, and an open set V in $\mathbf{R}^n$ containing $f(x_0)$, such that f is a bijection from U to V. In particular, there is an inverse map $f^{-1} : V \to U$. Furthermore, this inverse map is differentiable at $f(x_0)$, and

$$(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}.$$

- **Proof.** (Optional) We first observe that once we know the inverse map f^{-1} is differentiable, the formula $(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}$ is automatic. This comes from starting with the identity

$$I = f^{-1} \circ f$$

on U, where $I : \mathbf{R}^n \to \mathbf{R}^n$ is the identity map $Ix := x$, and then differentiating both sides using the chain rule at x_0 to obtain

$$I'(x_0) = (f^{-1})'(f(x_0))f'(x_0).$$

Since $I'(x_0) = I$, we thus have $(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}$ as desired.

- (Remark: This argument shows that if $f(x_0)$ is *not* invertible, then there is no way that an inverse f^{-1} can exist and be differentiable at $f(x_0)$).

- Next, we observe that it suffices to prove the theorem under the additional assumption $f(x_0) = 0$. The general case then follows from the special case by replacing f by a new function $\tilde{f}(x) := f(x) - f(x_0)$ and then applying the special case to $\tilde{f}$ (note that V will have to shift by $f(x_0)$). Note that $f^{-1}(y) = \tilde{f}^{-1}(y + f(x_0))$ - why?. Henceforth we will always assume $f(x_0) = 0$.

- In a similar manner, one can make the assumption $x_0 = 0$. The general case then follows from this case by replacing f by a new function $\tilde{f}(x) := f(x + x_0)$ and applying the special case to $\tilde{f}$ (note that E and U will have to shift by x_0). Note that $f^{-1}(y) = \tilde{f}^{-1}(y) + x_0$ - why? Henceforth we will always assume $x_0 = 0$. Thus we now have that $f(0) = 0$ and that $f'(0)$ is invertible.

- Finally, one can assume that $f'(0) = I$ where $I : \mathbf{R}^n \to \mathbf{R}^n$ is the identity transformation $Ix = x$). The general case then follows from this case by replacing f with a new function $\tilde{f} : E \to \mathbf{R}^n$ defined by $\tilde{f}(x) := f'(0)^{-1}f(x)$, and applying the special case to this case. Note from Lemma 8 that $f'(0)^{-1}$ is a linear transformation. In particular, we note that $\tilde{f}(0) = 0$ and that

$$\tilde{f}'(0) = f'(0)^{-1}f'(0) = I,$$

so by the special case of the inverse function theorem we know that there exists an open set U' containing 0, and an open set V' containing 0, such that $\tilde{f}$ is a bijection from U' to V', and that $\tilde{f}^{-1} : V' \to U'$ is differentiable at 0 with derivative I. But we have $f(x) = f'(0)\tilde{f}(x)$, and hence f is a bijection from U' to $f'(0)(V')$ (note that $f'(0)$ is also a bijection). Since $f'(0)$ and its inverse are both continuous, $f'(0)(V')$ is open, and it certainly contains 0. Now consider the inverse function $f^{-1} : f'(0)(V') \to U'$. Since $f(x) = f'(0)\tilde{f}(x)$, we see that $f^{-1}(y) = \tilde{f}^{-1}(f'(0)^{-1}y)$ for all $y \in f'(0)(V')$ (why? use the fact that $\tilde{f}$ is a bijection from U' to V'). In particular we see that f^{-1} is differentiable at 0.

- So all we have to do now is prove the inverse function theorem in the special case, when $x_0 = 0$, $f(x_0) = 0$, and $f'(x_0) = I$. Let $g : E \to \mathbf{R}^n$ denote the function $f(x) - x$. Then $g(0) = 0$ and $g'(0) = 0$. In particular
$$\frac{\partial g}{\partial x_j}(0) = 0$$
for $j = 1, \ldots, n$. Since g is continuously differentiable, there thus exists a ball $B(0, r)$ in E such that
$$\|\frac{\partial g}{\partial x_j}(x)\| \leq \frac{1}{2n^2}$$
for all $x \in B(0, r)$. (There is nothing particularly special about $\frac{1}{2n^2}$, we just need a nice small number here). In particular, for any $x \in B(0, r)$ and $v = (v_1, \ldots, v_n)$ we have
$$\|D_v g(x)\| = \|\sum_{j=1}^{n} v_j \frac{\partial g}{\partial x_j}(x)\|$$
$$\leq \sum_{j=1}^{n} |v_j| \|\frac{\partial g}{\partial x_j}(x)\|$$
$$\leq \sum_{j=1}^{n} \|v\| \frac{1}{2n^2} \leq \frac{1}{2n}\|v\|.$$
But now for any $x, y \in B(0, r)$, we have by the fundamental theorem of calculus
$$g(y) - g(x) = \int_0^1 \frac{d}{dt} g(x + t(y - x))\ dt$$
$$= \int_0^1 D_{y-x} g(x + t(y - x))\ dt.$$

By the previous remark, the vectors $D_{y-x}g(x+t(y-x))$ have a magnitude of at most $\frac{1}{2n}\|y-x\|$. Thus every component of these vectors has magnitude at most $\frac{1}{2n}\|y-x\|$. Thus every component of $g(y)-g(x)$ has magnitude at most $\frac{1}{2n}\|y-x\|$, and hence $g(y)-g(x)$ itself has magnitude at most $\frac{1}{2}\|y-x\|$ (actually, it will be substantially less than this, but this bound will be enough for our purposes). In other words, g is a contraction. By Lemma 7, the map $f = g+I$ is thus one-to-one on $B(0,r)$, and the image $f(B(0,r))$ contains $B(0,r/2)$. In particular we have an inverse map $f^{-1} : B(0,r/2) \to B(0,r)$ defined on $B(0,r/2)$.

- Applying the contraction bound with $y = 0$ we obtain in particular that
$$\|g(x)\| \leq \frac{1}{2}\|x\|$$
for all $x \in B(0,r)$, and so by the triangle inequality
$$\frac{1}{2}\|x\| \leq \|f(x)\| \leq \frac{3}{2}\|x\|$$
for all $x \in B(0,r)$.

- Now we set $V := B(0,r/2)$ and $U := f^{-1}(B(0,r))$. Then by construction f is a bijection from U to V. V is clearly open, and $U = f^{-1}(V)$ is also open since f is continuous. (Notice that if a set is open relative to $B(0,r)$, then it is open in $\mathbf{R}^n$ as well). Now we want to show that $f^{-1} : V \to U$ is differentiable at 0 with derivative $I^{-1} = I$. In other words, we wish to show that
$$\lim_{x\to 0; x\in V\setminus\{0\}} \frac{\|f^{-1}(x) - f^{-1}(0) - I(x-0)\|}{\|x\|} = 0.$$
Since $f(0) = 0$, we have $f^{-1}(0) = 0$, and the above simplifies to
$$\lim_{x\to 0; x\in V\setminus\{0\}} \frac{\|f^{-1}(x) - x\|}{\|x\|} = 0.$$
Let $(x_n)_{n=1}^\infty$ be any sequence in $V\backslash 0$ that converges to 0. By Proposition 1(b) of Week 3 notes, it suffices to show that
$$\lim_{n\to\infty} \frac{\|f^{-1}(x_n) - x_n\|}{\|x_n\|} = 0.$$

Write $y_n := f^{-1}(x_n)$. Then $y_n \in B(0,r)$ and $x_n = f(y_n)$. In particular we have

$$\frac{1}{2}\|y_n\| \leq \|x_n\| \leq \frac{3}{2}\|y_n\|$$

and so since $\|x_n\|$ goes to 0, $\|y_n\|$ goes to zero also, and their ratio remains bounded. It will thus suffice to show that

$$\lim_{n\to\infty} \frac{\|y_n - f(y_n)\|}{\|y_n\|} = 0.$$

But since y_n is going to 0, and f is differentiable at 0, we have

$$\lim_{n\to\infty} \frac{\|f(y_n) - f(0) - f'(0)(y_n - 0)\|}{\|y_n\|} = 0$$

as desired (since $f(0) = 0$ and $f'(0) = I$). □

- The inverse function theorem gives a useful criterion for when a function is (locally) invertible at a point x_0 - all we need is for its derivative $f'(x_0)$ to be invertible (and then we even get further information, for instance we can compute the derivative of f^{-1} at $f(x_0)$). Of course, this begs the question of how one can tell whether the linear transformation $f'(x_0)$ is invertible or not. Recall that we have $f'(x_0) = L_{Df(x_0)}$, so by Lemmas 1 and 2 we see that the linear transformation $f'(x_0)$ is invertible if and only if the matrix $Df(x_0)$ is. There are many ways to check whether a matrix such as $Df(x_0)$ is invertible; for instance, one can use determinants, or alternatively Gaussian elimination methods. We will not pursue this matter here, but see math 115A for more details.

- If $f'(x_0)$ exists but is non-invertible, then the inverse function theorem does not apply. In such a situation it is not possible for f^{-1} to exist and be differentiable at x_0; this was remarked in the above proof. But it is still possible for f to be invertible. For instance, the single-variable function $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^3$ is invertible despite $f'(0)$ not being invertible.

* * * * *

The implicit function theorem

- Recall that a function $f : \mathbf{R} \to \mathbf{R}$ gives rise to a *graph*

$$\{(x, f(x)) : x \in \mathbf{R}\}$$

which is a subset of $\mathbf{R}^2$, usually looking like a curve. However, not all curves are graphs, they must obey the vertical line test, that for every x there is exactly one y such that (x, y) is in the curve. For instance, the circle $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$ is not a graph, although if one restricts to a semicircle such as $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1, y > 0\}$ then one again obtains a graph. Thus while the entire circle is not a graph, certain local portions of it are. (The portions of the circle near $(1, 0)$ and $(0, 1)$ are not graphs over the variable x, but they are graphs over the variable y).

- Similarly, any function $f : \mathbf{R}^n \to \mathbf{R}$ gives rise to a graph $\{(x, f(x)) : x \in \mathbf{R}^n\}$ in $\mathbf{R}^{n+1}$, which in general looks like some sort of n-dimensional surface in $\mathbf{R}^{n+1}$ (the technical term for this is a *hypersurface*). Conversely, one may ask which hypersurfaces are actually graphs of some function, and whether that function is continuous or differentiable.

- If the hypersurface is given geometrically, then one can again invoke the vertical line test to work out whether it is a graph or not. But what if the hypersurface is given algebraically, for instance the surface $\{(x, y, z) \in \mathbf{R}^3 : xy + yz + zx = -1\}$? Or more generally, a hypersurface of the form $\{x \in \mathbf{R}^n : f(x) = 0\}$, where $f : \mathbf{R}^n \to \mathbf{R}$ is some function? In this case, it is still possible to say whether the hypersurface is a graph, locally at least, by means of the *implicit function theorem.*

- **Implicit function theorem** Let E be an open subset of $\mathbf{R}^n$, let $f : E \to \mathbf{R}$ be a continuously differentiable function, and let $y = (y_1, \ldots, y_n)$ be a point in E such that $f(y) = 0$ and $\frac{\partial f}{\partial x_n}(y) \neq 0$. Then there exists an open subset U of $\mathbf{R}^{n-1}$ containing $(y_1, \ldots, y_{n-1})$, an open subset V of E containing y, and a function $g : U \to \mathbf{R}$ such that $g(y_1, \ldots, y_{n-1}) = y_n$, and

$$\{(x_1, \ldots, x_n) \in V : f(x_1, \ldots, x_n) = 0\}$$

$$= \{(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) : (x_1, \ldots, x_{n-1}) \in U\}.$$

In other words, the set $\{x \in V : f(x) = 0\}$ is a graph of a function over U. Moreover, g is differentiable at $(y_1, \ldots, y_{n-1})$, and we have

$$\frac{\partial g}{\partial x_j}(y_1, \ldots, y_{n-1}) = -\frac{\partial f}{\partial x_j}(y) / \frac{\partial f}{\partial x_n}(y)$$

for all $1 \leq j \leq n-1$.

- **Remark.** This last equation is sometimes derived using *implicit differentiation.* Basically, the point is that if you know that

$$f(x_1, \ldots, x_n) = 0$$

then (as long as $\frac{\partial f}{\partial x_n} \neq 0$) the variable x_n is "implicitly" defined in terms of the other $n-1$ variables, and one can differentiate the above identity in, say, the x_j direction using the chain rule to obtain

$$\frac{\partial f}{\partial x_j} + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial x_j} = 0$$

which is the above formula in disguise (we are using g to represent the implicit function defining x_n in terms of $x_1, \ldots, x_n$). Thus, the implicit function theorem allows one to define a dependence implicitly, by means of a constraint rather than by a direct formula of the form $x_n = g(x_1, \ldots, x_{n-1})$.

- **Proof.** This theorem looks somewhat fearsome, but actually it is a fairly quick consequence of the inverse function theorem. Let $F : E \to \mathbf{R}^n$ be the function

$$F(x_1, \ldots, x_n) := (x_1, \ldots, x_{n-1}, f(x_1, \ldots, x_n)).$$

This function is continuously differentiable. Also note that

$$F(y) = (y_1, \ldots, y_{n-1}, 0)$$

and

$$DF(y) = (\frac{\partial f}{\partial x_1}(y)^\dagger, \frac{\partial f}{\partial x_2}(y)^\dagger, \ldots, \frac{\partial f}{\partial x_n}(y)^\dagger)$$

$$= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \frac{\partial f}{\partial x_1}(y) & \frac{\partial f}{\partial x_2}(y) & \dots & \frac{\partial f}{\partial x_{n-1}}(y) & \frac{\partial f}{\partial x_n}(y) \end{pmatrix}.$$

Since $\frac{\partial f}{\partial x_n}(y)$ is assumed by hypothesis to be non-zero, this matrix is invertible; this can be seen either by computing the determinant, or using row reduction, or by computing the inverse explicitly, which is

$$DF(y)^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ -\frac{\partial f}{\partial x_1}(y)/a & \frac{\partial f}{\partial x_2}(y)/a & \dots & \frac{\partial f}{\partial x_{n-1}}(y)/a & 1/a \end{pmatrix},$$

where we have written $a = \frac{\partial f}{\partial x_n}(y)$ for short. Thus the inverse function theorem applies, and we can find an open set V in E containing y, and an open set W in $\mathbf{R}^n$ containing $F(y) = (y_1, \dots, y_{n-1}, 0)$, such that F is a bijection from V to W, and that F^{-1} is differentiable at $(y_1, \dots, y_{n-1}, 0)$.

- Let us write F^{-1} in co-ordinates as

$$F^{-1}(x) = (h_1(x), h_2(x), \dots, h_n(x))$$

where $x \in W$. Since $F(F^{-1}(x)) = x$, we thus have $h_j(x_1, \dots, x_n) = x_j$ for all $1 \leq j \leq n-1$ and $x \in W$, and $f(x_1, \dots, x_{n-1}, h_n(x_1, \dots, x_n)) = x_n$. Also, h_n is differentiable at $(y_1, \dots, y_{n-1}, 0)$ since F^{-1} is.

- Now we set $U := \{(x_1, \dots, x_{n-1}) \in \mathbf{R}^{n-1} : (x_1, \dots, x_{n-1}, 0) \in W\}$. It is easy to check that U is open and contains $(y_1, \dots, y_{n-1}, 0)$. Now we define $g : U \to \mathbf{R}$ by $g(x_1, \dots, x_{n-1}) := h_n(x_1, \dots, x_{n-1}, 0)$. Then g is differentiable at $(y_1, \dots, y_{n-1})$. Now we prove that

$$\{(x_1, \dots, x_n) \in V : f(x_1, \dots, x_n) = 0\}$$

$$= \{(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) : (x_1, \dots, x_{n-1}) \in U\}.$$

First suppose that $(x_1, \ldots, x_n) \in V$ and $f(x_1, \ldots, x_n) = 0$. Then we have $F(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1}, 0)$, which lies in W. Thus $(x_1, \ldots, x_{n-1})$ lies in W. Applying F^{-1}, we see that $(x_1, \ldots, x_n) = F^{-1}(x_1, \ldots, x_{n-1}, 0)$. In particular $x_n = h_n(x_1, \ldots, x_{n-1}, 0)$, and hence $x_n = g(x_1, \ldots, x_{n-1})$. Thus every element of the left-hand set lies in the right-hand set. The reverse inclusion comes by reversing all the above steps and is left to the reader.

- Finally, we show the formula for the partial derivatives of g. From the above set identity we have

$$f(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) = 0$$

for all $(x_1, \ldots, x_{n-1}) \in U$. Since g is differentiable at $(y_1, \ldots, y_{n-1})$, and f is differentiable at $(y_1, \ldots, y_{n-1}, g(y_1, \ldots, y_{n-1})) = y$, we may use the chain rule, differentiating in x_j, to obtain

$$\frac{\partial f}{\partial x_j}(y) + \frac{\partial f}{\partial x_n}(y)\frac{\partial g}{\partial x_j}(y_1, \ldots, y_{n-1})$$

and the claim follows by simple algebra. □

- **Example** Consider the surface $S := \{(x, y, z) \in \mathbf{R}^3 : xy + yz + zx = -1\}$, which we rewrite as $\{(x, y, z) \in \mathbf{R}^3 : f(x, y, z) = 0\}$, where $f : \mathbf{R}^3 \to \mathbf{R}$ is the function $f(x, y, z) := xy + yz + zx + 1$. Clearly f is continuously differentiable, and $\frac{\partial f}{\partial z} = y + x$. Thus for any (x_0, y_0, z_0) in S with $y_0 + x_0 \neq 0$, one can write this surface (near (x_0, y_0, z_0)) as a graph of the form $\{(x, y, g(x, y)) : (x, y) \in U\}$ for some open set U containing (x_0, y_0), and some function g which is differentiable at (x_0, y_0). Indeed one can implicitly differentiate to obtain that

$$\frac{\partial g}{\partial x}(x_0, y_0) = -\frac{y_0 + z_0}{y_0 + x_0} \text{ and } \frac{\partial g}{\partial y}(x_0, y_0) = -\frac{x_0 + z_0}{y_0 + x_0}.$$

- In the implicit function theorem, if the derivative $\frac{\partial f}{\partial x_n}$ equals zero at some point, then it is unlikely that the set $\{x \in \mathbf{R}^n : f(x) = 0\}$ can be written as a graph of the x_n variable in terms of the other $n - 1$ variables near that point. However, if some other derivative $\frac{\partial f}{\partial x_j}$ is zero, then it would be possible to write the x_j variable in terms of the other $n-1$ variables, by a variant of the implicit function theorem. Thus as long as the gradient ∇f is not entirely zero, one can write this set $\{x \in \mathbf{R}^n : f(x) = 0\}$ as a graph of *some* variable x_j in terms of the other $n-1$ variables. (The circle $\{(x,y) \in \mathbf{R}^2 : x^2 + y^2 - 1 = 0\}$ is a good example of this; it is not a graph of y in terms of x, or x in terms of y, but near every point it is one of the two. And this is because the gradient of $x^2 + y^2 - 1$ is never zero on the circle). However, if ∇f does vanish at some point x_0, then we say that f has a *critical point* at x_0 and the behavior there is much more complicated. For instance, the set $\{(x,y) \in \mathbf{R}^2 : x^2 - y^2 = 0\}$ has a critical point at $(0,0)$ and there the set does not look like a graph of any sort (it is the union of two lines).

- Sets which look like graphs of continuous functions at every point have a name, they are called *manifolds*. Thus $\{x \in \mathbf{R}^n : f(x) = 0\}$ will be a manifold if it contains no critical points of f. The theory of manifolds is very important in modern geometry (Especially differential geometry and algebraic geometry), but we will not discuss it here as it is a graduate level topic.