Math 4220/6220. Lecture 0,
Review and summary of background information

Introduction: The most fundamental concepts used in this course are those of continuity and differentiability (hence linearity), and integration.

Continuity

Continuity is fundamentally the idea of approximation, since a continuous function is one for which f(x) approximates f(a) as well as desired whenever x approximates a well enough. The precise version of this is couched in terms of "neighborhoods" of a point. In that language we say f is continuous at a, if whenever a neighborhood V of f(a) is specified, there exists a corresponding neighborhood U of a, such that every point x lying in U has f(x) lying in V.

Then the intuitive statement "if x is close enough to a, then f(x) is as close as desired to f(a)", becomes the statement: "for every neighborhood V of f(a), there exists a neighborhood U of a, such that if x is in U, then f(x) is in V".

Neighborhoods in turn are often defined in terms of distances, for example an "r - neighborhood" of a, consists of all points x having distance less than r from a. In the language of distances, continuity of f at a becomes: "if a distance $r > 0$ is given, there is a corresponding distance $s > 0$, such that if $\mathrm{dist}(x,a) < s$, (and f is defined at x) then $\mathrm{dist}(f(x),f(a)) < r$".

More generally we say f(x) has limit L as x approaches a, if for every nbhd V of L, there is a nbhd U of a such that for every point of U except possibly a, we have f(x) in V. Notice that the value f(a) plays no role in the definition of the limit of f at a. Then f is continuous at a iff f(x) has limit equal to f(a) as x approaches a.

Differentiability

Differentiability is the approximation of non - linear functions by linear ones. Thus making use of differentiability requires one to know how to calculate the linear function which approximates a given differentiable one, to know the properties of the approximating linear function, and how to translate these into analogous properties of the original non linear function. Hence a prerequisite for understanding differentiability, is understanding linear functions and the linear spaces on which they are defined. I.e. linear algebra is a a prerequisite for understanding differential calculus.

Linearity

Linear spaces capture the idea of flatness, and allow the concept of dimension. A line with a specified point of origin is a good model of a one dimensional linear space. A Euclidean plane with an origin is a good model of a two dimensional linear space. Every point in a linear space is thought of as equivalent to the arrow drawn to it from the specified origin. This makes it possible to add points in a linear space by adding their position vectors via the parallelogram law, and to "scale" points by real numbers or "scalars", by stretching the arrows by this scale factor, (reversing the direction if the scalar is negative).

We often call the points of a linear space "vectors" and the space itself a "vector space". A linear function, or linear map, is a function from one linear space to another which commutes with these operations, i.e. f is linear if $f(v+w) = f(v)+f(w)$ and $f(cv) = cf(v)$, for all scalars c, and all vectors v,w.

The standard model of a finite dimensional linear space is R^n. A fundamental example of an infinite dimensional linear space is the space of all infinitely differentiable functions on R.

Linear Dimension

This is an algebraic version of the geometric idea of dimension. A line is one - dimensional. This means: given any point on the line except the origin, the resulting non zero vector can be scaled to give any other vector on the line. Thus a linear space L is one dimensional if it contains a non zero vector v such that given any other vector x, there is a real number c such that x = cv. We say then "v spans the one – dimensional space L".

A plane S has the two dimensional property that if we pick two distinct points both different from the origin, and not collinear with the origin, then every point of the plane is the vector sum of multiples of the two corresponding vectors. Thus a linear space S is two dimensional if it contains two non zero vectors v,w, such that w is not a multiple of v, but every vector in S has form cv+dw for some real numbers c,d. We say the set {v,w} spans the 2 dimensional space S.

In general a set of vectors $\{v_j\}$ spans a space S if every vector in S has form $\sum c_j v_j$ where the sum is finite. The space is finite dimensional if the set $\{v_j\}$ can be taken to be finite. A space has dimension r if it can be spanned by a set of r vectors but not by any set of fewer than r vectors. If S, T are finite dimensional linear spaces of the same dimension, and T contains S, then S = T.

Linear maps

Unlike continuous maps, linear maps cannot raise dimension, and bijective linear maps preserve dimension exactly. More precisely, if f:S-->T is a surjective linear map, then dim(T) ≤ dim(S), whereas if f:S-->T is an injective linear map, then dim(T) ≥ dim(S). Still more precisely, if ker(f) = $f^{-1}(0)$, and im(f) = {f(v): v is in S} (a subset of T), then ker(f) and im(f) are both linear spaces, and dim(ker(f)) + dim(im(f)) = dimS. This is the most fundamental and important property of dimension. This is often stated as follows. The "rank" of a linear map f:S-->T is the dimension of im(f), and the "nullity" of f is the dimension of ker(f). Then for f:S-->T, we have rank(f) + nullity(f) = dim(S).

It follows that f is injective if and only if ker(f) = {0}, and surjective if dimT = dim(im(f)) < ∞. A linear map f:S-->T with a linear inverse is called an <u>isomorphism</u>. A linear map is an isomorphism if and only if it is bijective. If dimS = dimT < ∞, a linear map f:S-->T is bijective if and only if f is injective, if and only if f is surjective. A simple and important example of a linear map is projection $R^n \times R^m$-->R^n, taking (v,w)-->v. This map is trivially surjective with kernel $\{0\} \times R^m \approx R^m$.

The theory of dimension gives a strong criterion for proving the existence of solutions of linear equations f(x) = w in finite dimensional spaces. Assume dimS = dimT < ∞, f:S-->T is linear, and f(x) = 0 only if x = 0. Then for every w in T, the equation f(x) = w has a unique solution.

More generally, if S,T are finite dimensional, f:S-->T is linear, and dim(ker(f)) = dim(S) - dim(T) = r, then every equation f(x) = w has an r dimensional set of solutions. We describe the set of solutions more precisely below.

Differentiation D:f-->f' is a linear map from the space of infinitely differentiable functions on R to itself. The mean value theorem implies the kernel of D is the one dimensional space of constant functions, and the fundamental theorem of calculus implies D is surjective.

More generally, for every constant c, the differential operator (D-c) is surjective with kernel the one dimensional space of multiples of e^{ct}, hence a composition of n such operators has n dimensional kernel. One can deduce that a linear combination $\sum c_j D^j$ $0 \le j \le n$, $c_n \neq 0$, with constant coefficients c_j, of compositions of D with maximum order n, has n dimensional kernel.

Geometry of linear maps.

If f:S-->T is a linear surjection of finite dimensional spaces, then $\ker(f) = f^{-1}(0)$ is a linear space of dimension r = dim(T)-dim(S), and for every w in T, the set $f^{-1}(w)$ is similar to a linear space of dimension r, except it has no specified origin. I.e. if v is any solution of f(v) = w, then the correspondence x-->x+v, is a bijection from $f^{-1}(0)$ to $f^{-1}(w)$. Hence the choice of v as "origin" in $f^{-1}(w)$ allows us to define a unique structure of linear space making $f^{-1}(w)$ isomorphic to $f^{-1}(0)$. Thus $f^{-1}(w)$ is a translate of an r - dimensional linear space.

In this way, f "fibers" or "partitions" the space S into the disjoint union of the "affine" linear sets" $f^{-1}(w)$. There is one fiber $f^{-1}(w)$ for each w in T, all of which are translates of the linear space $\ker(f) = f^{-1}(0)$. If f:S-->T is surjective and linear, and dimT = dimS - 1, then the fibers of f are all one dimensional, so f fibers S into a family of parallel lines, one line over each point of T. If f:S-->T is surjective (and linear), but dimT = dimS - r with r > 0, then f fibers S into a family of parallel affine linear sets $f^{-1}(w)$ each of dimension r.

The matrix of a linear map R^n-->R^m

If S, T are linear spaces of dimension n and m, and $\{v_1,....,v_n\}$, $\{w_1,....,w_m\}$ are sets of vectors spanning S,T respectively, then for every v in S, and every w in T, the scalar coefficients a_i, b_j in the expressions $v = \sum a_i v_i$, and $w = \sum b_j w_j$, are unique. Then given these minimal spanning sets, a linear map f:S-->T determines and is determined by the "m by n matrix" [å$_{ij}$] of scalars where: $f(v_j) = \sum S_i\ c_{ij} w_i$, for all j = 1,....,n. If $S = T = R^n$, we may take $v_i = w_i = (0,....,0,1,0,....,0) = e_i$ = the "ith unit vector", where the 1 occurs in the ith place.

If S is a linear space of dimension n and $\{v_1,....,v_n\}$ is a minimal spanning set, we call $\{v_1,....,v_n\}$ a basis for S. Then there is a unique isomorphism S-->R^n that takes v_i to e_i, where the set of unit vectors $\{e_1,....,e_n\}$ is called the "standard" basis of R^n. Conversely under any isomorphism S-->R^n, the vectors corresponding to the set $\{e_1,....,e_n\}$ form a basis for S. Thus a basis for an n dimensional linear space S is equivalent to an isomorphism of S with R^n. Since every linear space has a basis, after choosing one, a finite dimensional vector space can be regarded as essentially equal to some R^n.

In the context of the previous sentence, every linear map can be regarded as a map f:R^n-->R^m. The matrix of such a map, with respect to the standard bases, is the m by n matrix whose jth column is the coordinate vector $f(e_j)$ in R^m.

If f:S-->T is any linear surjection of finite dimensional spaces, a careful choice of bases for S,T can greatly simplify the matrix of the corresponding map R^n-->R^m. In fact there are bases for S,T such that under the corresponding isomorphisms, f is equivalent to a projection $R^{(n-m)}xR^m$-->R^m. Thus up to linear isomorphism, every linear surjection is equivalent to the simplest example, a projection.

This illustrates the geometry of a linear surjection as in the previous subsection. I.e. a projection $f:R^n x R^m \to R^m$ fibers the domain space $R^n x R^m$ into the family of disjoint parallel affine spaces $f^{-1}(v) = R^n x\{v\}$, with the affine space $R^n x\{v\}$ lying over the vector v. Since every linear surjection is equivalent to a projection, every linear surjection fibers its domain into a family of disjoint affine spaces linearly isomorphic to this family. We will see that the implicit function theorem gives an analogous local statement for differentiable functions.

The determinant of a linear map $R^n \to R^n$.

For each linear map $f:R^n \to R^n$ there is an important associated number $\det(f) = \det(c_{ij})$ = the sum of the products $\sum_{s(i)} c_{is(i)}$, where s ranges over all permutations of the integers (1,2,3....,n). det(f) is the oriented volume of the parallelepiped (i.e. block) spanned by the image of the ordered set of unit vectors $f(e_1),....,f(e_n)$. Then f is invertible iff $\det(f) \pm 0$. The intuition is that this block has non zero n dimensional volume iff the vectors $f(e_1),....,f(e_n)$ span R^n, iff f is surjective, iff f is invertible.

Derivatives: Approximating non linear functions by linear ones.

Ordinary Euclidean space R^n is a linear space in which an absolute value is defined, say by the Euclidean "norm", $|v| = (x_1^2+....+x_n^2)^{1/2}$, where $v = (x_1,....,x_n)$, hence also a distance is defined by $\mathrm{dist}(v,w) = |v-w|$. The set of points x such that $|x-a| < r$, is called the open ball of radius r centered at a. An "open set" is any union of open balls, and an open neighborhood of the point a is an open set containing a. If $f:R^n \to R^m$ is any map, then f(x) has limit L as x approaches a, iff the real valued function $|f(x)-L|$ has limit 0 as x approaches a.

In a linear space with such an absolute value or norm we can define differentiability as follows. A function h is "tangent to zero" at a, if h(a) = 0 and the quotient $|h(x)|/|x-a|$ has limit zero as x approaches a. I.e. if "rise" over "run" approaches zero in all directions. In particular then h(x) approaches zero as x approaches a. Two functions f,g are tangent at a, if the difference f-g is tangent to zero at a.

A function f defined on a nbhd of a, is <u>differentiable</u> at a if there is a linear function L such that L(v) is tangent to f(v+a)-f(a) at 0. Then L = f'(a) is unique and is called the derivative of f at a. I.e. f has derivative L = f'(a) at a, iff the quotient $|(f(x)-f(a)-L(x-a))|/|x-a|$ has limit zero as x approaches a. If f is itself linear, then f'(a)(v) = f(v), for all a. I.e. then $a \to f'(a)$ is a constant (linear map valued) function, with value f everywhere.

Chain Rule

The most important property of derivatives is the chain rule for the derivative of a composite function. If f is differentiable at a and g is differentiable at f(a), then gof is differentiable at a and (gof)'(a) = g'(f(a))of'(a). I.e. the derivative of the composition, is the composition (as linear functions) of the derivatives. Since the derivative of the identity map is the identity map, this says roughly "the derivative is a functor", i.e. it preserves compositions and identity maps.

As a corollary, if a differentiable function has a differentiable inverse, the derivative of the inverse function is the inverse linear function of the derivative. I.e. If f^{-1} exists and is differentiable, then $(f^{-1})'(f(a)) = (f'(a))^{-1}$. In particular, since a linear function can be invertible only if the domain and range have the same dimension, the same holds for a differentiable function. E.g. a differentiable function $f:R^2 \to R$ cannot have a differentiable inverse, because R^2 and R have different linear dimensions. (Continuously invertible, non differentiable, functions also preserve dimension, but this is much harder to prove in general. It is easy in low dimensions however. Can you prove there is no continuously invertible function $f:R^2 \to R$?)

Calculating derivatives

The usual definition of the derivative of a one variable function from R to R, agrees with the definition above, in the sense that if f'(a) is the usual derivative, i.e. if f'(a) is the $\lim_{h\to 0} (f(a+h)-f(a))/h$), then f(a+h)-f(a) is tangent at zero to the linear function f'(a).h of the variable h. I.e. the usual derivative is the number occurring in the 1 by 1 matrix of the derivative thought of as a linear function. There is an analogous way to compute the matrix of the derivative in general.

A function $f:R^n \to R^m$ is made up of m component functions $g_1,\ldots,g_m$, and if in the ith component function g_i, we hold all but the jth variable constant, and define the real valued function h(t) of one variable by $h(t) = g_i(a_1,\ldots,a_j+t,\ldots,a_n)$, we call $h'(0) = \partial g_i/\partial x_j(a)$, the jth partial derivative of g_i at a. If f is differentiable at a, then all partials of f exist at a, and the matrix of the derivative L = f'(a) of f at a is the "Jacobian" matrix of partials $[\partial g_i/\partial x_j(a)]$.

It is useful to have a criterion for existence of a derivative that does not appeal to the definition. It is this: if all the partials of f exist not only at a but in a nbhd of a, and these partials are all continuous at a, then f is differentiable at a, and the derivative is given by the matrix of partials. We can thus check the invertibility of f'(a), by computing the determinant of this Jacobian matrix.

Inverse function and implicit function theorems

The "inverse function theorem" is a criterion for f to have a local differentiable inverse as follows: If f is differentiable on a neighborhood of a, and if the derivative f'(x) is a continuous function of x in that nbhd, (i.e. if the entries in the matrix of f'(x) are continuous functions of x), and if f'(a) is invertible, then f is differentiably invertible when restricted to some nbhd U of a. I.e. then f maps some open nbhd U of a bijectively onto an open nbhd V = f(U) of f(a), and f^{-1} is defined and differentiable on V, and $f^{-1}(V) = U$.

More generally, the implicit function theorem characterizes differentiable functions which are locally equivalent to projection maps, as follows. If f is differentiable on a neighborhood of a in R^n with values in R^m, and if the derivative f'(x) is a continuous function of x, and if f'(a) is surjective, then on some nbhd U of a, f is differentiably isomorphic to a projection

.

I.e. if $f:R^n \to R^m$ is continuously differentiable near a with surjective derivative at a, then there are open sets U in R^n, W in R^{n-m}, V in R^m, with U a nbhd of a, V a nbhd of f(a), and a differentiable isomorphism h:U-->WxV, such that the composition $f\circ h^{-1}$:WxV-->V, is the projection map (x,y)-->y. Then the parallel flat sets Wx{y} which fiber the rectangle WxV, are carried by h^{-1} into "parallel" curved sets which fiber the nbhd U of a. The fiber passing through a, suitably restricted, is the graph of a differentiable function, hence the name of the theorem.

I.e. one can take a smaller nbhd of a within U, of form XxY, with X in W, and the map XxY-->WxV of form (x,y)-->(x,f(x,y)). Then the flat set Xx{f(a)} pulls back by h^{-1} to some subset G of XxY in which every point is determined by its "X-coordinate". I.e. given x in X, there is a unique point of form (x, f(a)), hence a unique point h^{-1}(x,f(a)) in the set G = h^{-1}(Xx{f(a)}). Since on G, the Y coordinate of every point is determined by the X coordinate, and every x coordinate in X occurs, G is the graph of a function X-->Y. This function is differentiable since it is a composite of differentiable functions: i.e. (projection) o (h^{-1}) o (id,f(a)). We are more interested in the simpler geometric interpretation than in the "implicit function" interpretation.

Compactness

In proving many results, we will often need the important ideas of connectedness and compactness from point set topology. In Euclidean space recall that an open set is a union of open balls. Compactness is a replacement for finiteness as follows: a set Z is called compact if whenever Z is "covered by" a collection of open sets (i.e. Z is contained in the union of those open sets), then a finite number of those same open sets already cover Z. A set is called "closed" if it is the complement of an open set.

A subset of R^n is compact if and only if it is closed and contained in some finite open ball, i.e. if and only if it is closed and "bounded". It follows that the product of two compact sets of Euclidean space is compact.

If f is a continuous function, and Z a compact subset of its domain, then f(Z) is also compact. Hence a real valued continuous function defined on a compact set Z assumes a global maximum there, namely the least upper bound of its values on Z. Likewise it assumes a global minimum on Z.

If Z is a compact subset of R^n then any open cover $\{U_j\}$ of Z has a "Lebesgue number". I.e. given any collection of open sets $\{U_j\}$ covering Z, there is a positive number r > 0, such that every open ball of radius r centered at any point of Z is wholly contained in some open set U_j of the given cover. This number is the minimum of the continuous function assigning to each point p of Z the least upper bound of its distances from the outside of all the sets U_j, i.e. the least upper bound of all r > 0 such that the open ball of radius r about p is contained in some set U_j. This function is positive valued since the sets U_j cover Z, hence it has a positive minimum.

A sequence contained in a compact set Z has a subsequence converging to a point of Z. In R^n this property implies in turn that Z is closed and bounded hence compact.

Connectedness

This is one of the most intuitive concepts in topology. Ask anyone, mathematician or not, which set is connected, the interval [0,1], or the two point set {0,1}, and they will always get it correct. Fortunately it is also one of the most important and powerful concepts. A set Z is connected if whenever Z is contained in the union of two open sets A,B, then either some point of Z is in both A and B, or Z is entirely contained in one of the sets A or B. I.e. you cannot separate a connected Z into two non empty disjoint parts AinZ and BinZ. Either AinZ and BinZ meet, or one of them is empty.

The empty set is connected. Any one point set is connected. The only connected subsets of R are the intervals, either finite or infinite, open or closed, half open or half closed. The image of a connected set under any continuous map is again connected. Thus an integer valued continuous function on an interval is constant. If f is a continuous real valued function defined on an interval of R, the set of values of f is also an interval. In calculus this is called the intermediate value theorem.

If $f:S^1 \to R^2$ is a continuous injection from the circle to the plane, then $R^2 - f(S^1)$ is a disjoint union of exactly two non empty connected open sets, the inside and the outside of the closed loop $f(S^1)$. This, the "Jordan curve theorem", is famously hard to prove, but we will prove it easily when f is continuously differentiable.