Lecture 6
Differential Calculus on Inner-product Spaces

In this section, we will use without proof standard facts that you should have seen in your multi-variable calculus classes.

6.1 Review of Basic Topology.

V will denote a finite dimensional inner-product space, and $e_1, \ldots, e_n$ some orthonormal basis for V. We will use this basis to identify $\mathbf{R}^n$ with V, via the map $(x_1, \ldots, x_n) \mapsto x_1e_1 + \cdots + x_ne_n$, and we recall that this preserves inner-products and norms, and hence distances between points. In particular, this means that a sequence $v^k = v_1^k e_1 + \ldots + v_n^k e_n$ of vectors in V converges to a vector $v = v_1e_1 + \ldots + v_ne_n$ if an only if each of the n component sequences v_i^k of real numbers converges to the corresponding real number v_i, and also, that the sequence v^k is Cauchy if and only if each component sequence of real numbers is Cauchy. This allows us to reduce questions about convergence of sequences in V to more standard questions about convergence in $\mathbf{R}$. For example, since $\mathbf{R}$ is complete (i.e., every Cauchy sequence of real numbers is convergent) it follows that V is also complete.

Recall that a subset S of $\mathbf{R}^n$ is called *compact* if any sequence of points in S has a subsequence that converges to a point of S, and the Bolzano-Weierstrass Theorem says that S is compact if and only if it is closed and bounded. (Bounded means that $\{\|s\| \mid s \in S\}$ is a bounded set of real numbers, and closed means that any limit of a sequence of points of S is itself in S.) Because of the distance preserving identification of V with $\mathbf{R}^n$ it follows that for subsets of V too, compact is the same as closed and bounded.

▷ **6.1—Exercise 1.** Recall that a set $O \subseteq V$ is called *open* if whenever $p \in O$ there is an $\epsilon > 0$ such that $\|x - p\| < \epsilon$ implies that $x \in O$. Show that O is open in V if and only if its complement is closed.

If $S \subseteq V$ and $f : S \to W$ is a map of S into some other inner-product space, then f is called *continuous* if whenever a sequence s_k in S converges to a point s of S, it follows that $f(s_k)$ converges to $f(s)$. (More succinctly, $\lim_{k\to\infty} f(s_k) = f(\lim_{k\to\infty} s_k)$, so one sometimes says a map is continuous if it "commutes with taking limits".)

▷ **6.1—Exercise 2.** Show that if $f : S \to W$ is continuous, then if A is an open (resp. closed) subset of W, then $f^{-1}(A)$ is open (resp. closed) in S. Deduce from this that $S(V)$, the unit sphere of V, is a closed and hence compact subset of V.

▷ **6.1—Exercise 3.** Show that any continuous real-valued function on a compact subset S of V must be bounded above and in fact there is a point s of S where f assumes its maximum value. (Hint # 1: If it were not bounded above, there would be a sequence s_n such that $f(s_n) > n$. Hint #2: Choose a sequence s_n so that $f(s_n)$ converges to the least upper bound of the values of f.)

6.2 Differentiable maps Between Inner-Product Spaces.

In this section, V and W are inner-product spaces.

The key concept in differential calculus is approximating a non-linear map $f : V \to W$ near some point p, by a linear map, T, called its *differential* at p. To be more precise, if v is close to zero, we should have $f(p+v) = f(p) + Tv + R_p(v)$ where the error term $R_p(v)$ should be "small" in a sense we shall make precise below.

We start with the one-dimensional case. A map $\sigma : (a,b) \to V$ of an interval (a,b) into V is called a *curve* in V, and it is said to be *differentiable* at the point $t_0 \in (a,b)$ if the limit $\sigma'(t_0) := \lim_{h \to 0} \frac{\sigma(t+h)-\sigma(t_0)}{h}$ exists. Note that this limit $\sigma'(t_0)$ is a vector in V, called the *derivative* (or the tangent vector or velocity vector) of σ at t_0. In this case, the differential of σ at t_0 is the linear map $D\sigma_{t_0} : \mathbf{R} \to V$ defined by $D\sigma_{t_0}(t) = t\sigma'(t_0)$, so that the error term is $R_p(t) = \sigma(t_0+t) - \sigma(t_0) - t\sigma'(t_0)$. Thus, not only is $R_p(t)$ small when t is small, but even when we divide it by t the result, $\frac{\sigma(t_0+t)-\sigma(t_0)}{t} - \sigma'(t_0)$, is still small and in fact it approaches zero as $t \to 0$. That is, we have $\sigma(t_0+t) = \sigma(t_0) + D\sigma_{t_0}(t) + |t|\rho(t)$ where $|\rho(t)| \to 0$ as $t \to 0$. We use this to define the notion of differentiability more generally by analogy.

6.2.1 Definition. . Let O be open in V, $F : O \to W$ a map, and $p \in O$. We say that F is *differentiable* at p if there exists a linear map $T : V \to W$ such that for v near 0 in V, $F(p+v) = F(p) + T(v) + \|v\|\,\rho(v)$ where $\rho(v) := \frac{F(p+v)-F(p)-T(v)}{\|v\|} \to 0$ as $\|v\| \to 0$. We call T the *differential of F at p* and denote it by DF_p. If $F : O \to W$ is differentiable at each point of O and if $DF : O \to L(V,W)$ is continuous we say that F is *continuously differentiable (or C^1) in O.*

6.2.2 Definition. Assuming $F : O \to W$ is as above and is differentiable at $p \in O$, then for $v \in V$, we call $DF_p(v)$ the *directional derivative of F at p in the direction v.*

▷ **6.2—Exercise 1.** Show that DF_p is well-defined, i.e., if $S : V \to W$ is a second linear map satisfying the same property as T, then necessarily $S = T$. (Hint: By subtraction one finds that for all v close to zero $\frac{\|(S-T)(v)\|}{\|v\|} \to 0$ as $\|v\| \to 0$. If one now replaces v by tv and uses the linearity of S and T, one finds that for fixed v, $\frac{\|(S-T)(v)\|}{\|v\|} \to 0$ as $t \to 0$.)

Chain Rule. *Let U, V, W be inner-product spaces, Ω an open set of U and O and open set of V. Suppose that $G : \Omega \to V$ is differentiable at $\omega \in \Omega$ and that $F : O \to W$ is differentiable at $p = G(\omega) \in O$. Then $F \circ G$ is differentiable at ω and $D(F \circ G)_\omega = DF_p \circ DG_\omega$.*

▷ **6.2—Exercise 2.** Prove the Chain Rule.

▷ **6.2—Exercise 3.** Show that if $p, v \in V$ then the map $\sigma : \mathbf{R} \to V$, $\sigma(t) = p + tv$ is differentiable at all $t_0 \in \mathbf{R}$ and that $\sigma'(t_0) = v$ for all $t_0 \in \mathbf{R}$. More generally, if $w_0 \in W$ and $T \in L(V,W)$ show that $F : V \to W$ defined by $F(v) := w_0 + Tv$ is differentiable at all v_0 in V and that $DF_{v_0} = T$. (So a linear map is its own differential at every point.)

Using the Chain Rule, we have a nice geometric interpretation of the directional derivative.

▷ **6.2—Exercise 4.** Let $F : O \to W$ be differentiable at p, let $v \in V$. Let $\sigma : \mathbf{R} \to V$ be any curve in V that is differentiable at t_0 with $\sigma(t_0) = p$ and $\sigma'(t_0) = v$. Then the curve in W, $F \circ \sigma : \mathbf{R} \to W$ is also differentiable at t_0 and its tangent vector at t_0 is $DF_p(v)$, the directional derivative of F at p in the direction v.

6.3 But Where Are All the Partial Derivatives?

Let's try to tie this up with what you learned in multi-variable calculus. As above, let us assume that O is open in V and that $F : O \to W$ is differentiable at $p \in O$. Let $e_1, \ldots, e_n$ be an orthonormal basis for V and $\epsilon_1, \ldots, \epsilon_m$ be an orthonormal basis for W, and let $p = p_1 e_1 + \cdots + p_n e_n$.

If $x = x_1 e_1 + \cdots + x_n e_n$ is in O, then its image $F(x) \in W$ will have an expansion in the basis ϵ_i, $F(x) = F_1(x)\epsilon_1 + \cdots + F_m(x)\epsilon_m$. If as usual we identify x with its n-tuple of components $(x_1, \ldots, x_n)$, then we have m functions of n variables, $F_1(x_1, \ldots, x_n), \ldots, F_m(x_1, \ldots, x_n)$ that describe the mapping F relative to the two bases e_i and ϵ_j.

▷ **6.3—Exercise 1.** Show that the partial derivatives of the F_i at $(p_1, \ldots, p_n)$ all exist, and in fact, show that the $\frac{\partial F_i(p_1,\ldots,p_n)}{\partial x_j}$ are the components of the directional derivative of F at p in the direction e_j relative to the basis ϵ_i.

▷ **6.3—Exercise 2.** The $n \times m$ matrix $\frac{\partial F_i(p_1,\ldots,p_n)}{\partial x_j}$ is called the *Jacobian matrix* of F at p (relative to the two bases e_j and ϵ_i). Show that it is the matrix of DF_p relative to these two bases, so that if $v = v_1 e_1 + \cdots + v_n e_n$ then the i-th component of the directional derivative of F at p in the direction v is $\sum_{j=1}^{n} F_{ij} v_j$.

6.3.1 Remark. It is clear from these exercises that differentials and Jacobian matrices are logically equivalent. So which should one use? For general theoretical discussions it is usually more natural and intuitive (and easier) to reason directly about differentials of maps. However, when it comes to a question concerning a particular map F that requires computing some of its properties, then one often must work with its partial derivatives to carry out the necessary computations.

The following is a standard advanced calculus result that provides a simple test for when a map $F : O \to W$ such as above is C^1.

Theorem. *A necessary and sufficient condition for a map $F : O \to W$ as above to be C^1 is that all its partial derivatives $\frac{\partial F_i(x_1,\ldots,x_n)}{\partial x_j}$ are continuous functions on O.*

▷ **6.3—Exercise 3.** Consider the map $F : \mathbf{R}^2 \to \mathbf{R}$ defined by $F(x, y) = \frac{2xy^2}{x^2+y^2}$ for $(x, y) \neq (0, 0)$ and $F(0, 0) = 0$. Show that the partial derivatives of F exist everywhere and are continuous except at the origin. Show also that F is actually linear on each straight line through the origin, but nevertheless F is **not** differentiable at the origin. (Hint: In polar coordinates, $F = r \sin(2\theta) \cos(\theta)$.)

6.4 The Gradient of a Real-Valued Function

Let's specialize to the case $W = \mathbf{R}$, i.e., we consider a differentiable **real-valued** function $f : V \to \mathbf{R}$ (perhaps only defined on an open subset O of V). In this case it is customary to denote the differential of f at a point p by df_p rather than Df_p. Notice that df_p is in the dual space $V^* = L(V, R)$ of V. We recall the "meaning" of df_p, namely $df_p(v)$ is the directional derivative of f in the direction v, i.e., the rate of change of f along any path through p in the direction v. So if $\sigma(t)$ is a smooth curve in V with $\sigma(0) = p$ and with tangent vector $\sigma'(0) = v$ at p, then $df_p(v) = (\frac{d}{dt})_{t=0} f(\sigma(t))$.

Next recall the self-duality principle for inner-product spaces: any ℓ of V^* can be expressed in the form $\ell(v) = \langle v, \omega \rangle$ for a unique $\omega \in V$. So in particular, for each $p \in O$, there is a unique vector ∇f_p in V such that

$$df_p(v) = \langle v, \nabla f_p \rangle \,,$$

and the vector ∇f_p defined by this identity is called the gradient of f at p.

A set of the form $f^{-1}(a) = \{v \in V \mid f(v) = a\}$ (where $a \in \mathbf{R}$) is called a *level set* of f and more precisely the a-level of f.

▷ **6.4—Exercise 1.** Show that if the image of a differentiable curve $\sigma : (a, b) \to V$ is in a level set of f then $\nabla f_{\sigma(t)}$ is orthogonal to $\sigma'(t)$. (Hint: The derivative of $f(\sigma(t))$ is zero.)

A point p where df (or ∇f) vanishes is called a *critical point* of f. So for example any local maximum or local minimum of f is a critical point.

▷ **6.4—Exercise 2.** If p is not a critical point of f then show that $\frac{\nabla f_p}{\|\nabla f_p\|}$ is the unit vector in the direction in which f is increasing most rapidly. (Hint: Use the Schwartz inequality.)

▷ **6.4—Exercise 3.** Suppose that $\sigma : (a, b) \to V$ and $\gamma : (a, b) \to V$ are differentiable curves. Show that $\frac{d}{dt} \langle \sigma(t), \gamma(t) \rangle = \langle \sigma'(t), \gamma(t) \rangle + \langle \sigma(t), \gamma'(t) \rangle$, and in particular $\frac{d}{dt} \|\sigma(t)\|^2 = 2 \langle \sigma(t), \sigma'(t) \rangle$. Deduce that if $\sigma : (a, b) \to V$ has its image in the unit sphere $S(V)$ then $\sigma(t)$ and $\sigma'(t)$ are orthogonal.

▷ **6.4—Exercise 4.** For $p \in S(V)$, define $T_pS(V)$, the tangent space to $S(V)$ at p, to be all $v \in V$ of the form $\sigma'(t_0)$ where $\sigma(t)$ is a differentiable curve in $S(V)$ having $\sigma(t_0) = p$. By the previous exercise, p is orthogonal to everything in $T_pS(V)$. Show that conversely any v orthogonal to p is in $T_pS(V)$, so the tangent space to $S(V)$ at p is exactly the orthogonal complement of p. (Hint: Define $\sigma(t) := \cos(\|v\| \, t)p + \sin(\|v\| \, t) \frac{v}{\|v\|}$. Check that $\sigma(0) = p$, $\sigma'(t) = v$, and that $\sigma(t) \in S(V)$ for all t.

▷ **6.4—Exercise 5.** Let T be a self-adjoint operator on V and define $f : V \to \mathbf{R}$ by $f(v) := \frac{1}{2} \langle Tv, v \rangle$. Show that f is differentiable and that $\nabla f_v = Tv$.

Proof of Spectral Lemma. We must find a p in $S(V)$ that is an eigenvector of the self-adjoint operator $T : V \to V$. By Exercise 4 we must show that ∇f_p is a multiple of p. But by Exercise 4, the scalar multiples of p are just those vectors orthogonal to $T_pS(V)$, so it will suffice to find p witht ∇f_p orthogonal to $T_pS(V)$. If we choose p a point of $S(V)$ where f assumes its maximum value on $S(V)$, that is automatic.