Midterm Exam Answers

Part I: Successive Approximation

If X is any set and $f : X \to X$ a mapping of X to itself, then for each positive integer n we define a mapping $f^{(n)} X \to X$ by composing f with itself n times. That is, $f^{(1)}(x) = f(x)$, $f^{(2)}(x) = f(f(x))$, $f^{(3)}(x) = f(f(f(x)))$, etc. To be more formal, we define the sequence $f^{(n)}$ inductively by: $f^{(1)} := f$ and $f^{(n+1)} := f \circ f^{(n)}$.

▷ **Problem 1.** Show that $f^{(n)} \circ f^{(k)} = f^{(n+k)}$.

Answer. We fix k and proceed by induction on n. For $n = 1$ the result is true by definition. Assuming the result for n result for $n+1$ follows by associativity of composition as follows:
$f^{(n+1)} \circ f^{(k)} = (f \circ f^{(n)}) \circ f^{(k)} = f \circ (f^{(n)} \circ f^{(k)}) = f \circ f^{(n+k)} = f^{((n+1)+k)}$ ∎

▷ **Problem 2.** Let X be a metric space and suppose that f satisfies a Lipschitz condition with constant K. (Recall this means that $\rho(f(x_1), f(x_2)) \leq K\rho(x_1, x_2)$ for all $x_1, x_2 \in X$.) Show that $f^{(n)}$ satisfies a Lipschitz condition with constant K^n.

Answer. This is an absolutely trivial induction!

In what follows, we suppose that X is a metric space and that $f : X \to X$ is a contraction mapping, i.e., we assume that f satisfies a Lipschitz condition with constant $K < 1$. (We refer to K as a contraction constant for f.) We recall that in an earlier assignment you proved the so-called Fundamental Inequality For Contraction Mappings, namely, for all $x_1, x_2 \in X$,

$$\rho(x_1, x_2) \leq \frac{1}{1-K}\Big(\rho(x_1, f(x_1)) + \rho(x_2, f(x_2))\Big).$$

▷ **Problem 3.** Show that a contraction mapping $f : X \to X$ can have at **most** one fixed point, i.e., there is at most one point $x \in X$ such that $f(x) = x$.

Answer. If x_1 and x_2 are both fixed points, then $\rho(x_1, f(x_1)) = 0$ and $\rho(x_2, f(x_2)) = 0$ so the Fundamental Inequality gives $\rho(x_1, x_2) \leq 0$, implying $x_1 = x_2$.

▷ **Problem 4.** Show that if $f : X \to X$ is a contraction mapping with contraction constant K and if x is **any** point of X then

$$\rho(f^{(n)}(x), f^{(m)}(x)) \leq \left(\frac{K^n + K^m}{1-K}\right)\rho(x, f(x)),$$

and deduce that $f^{(n)}(x)$ is a Cauchy sequence.

Answer. Since $f \circ f^{(n)} = f^{(n+1)} = f^{(n)} \circ f$ (by Problem 1), $\rho(f^{(n)}(x), f(f^{(n)}(x))) = \rho(f^{(n)}(x), f^{(n)}(f(x))) \leq K^n \rho(x, f(x))$ by Problem 2. If we substitute $f^{(n)}(x)$ for x_1 and $f^{(m)}(x)$ for x_2 in the Fundamental Inequality, we now get the stated inequality. Then, since $K < 1$ implies that both K^n and K^m tend to zero as n and m approach infinity, $\lim_{m,n\to\infty} \rho(f^{(n)}(x), f^{(m)}(x)) = 0$, i.e., the sequence $\{f^{(n)}(x)\}$ is Cauchy. ∎

▷ **Problem 5.** Now prove:

Banach Contraction Principle. *If X is a complete metric space and $f : X \to X$ is a contraction mapping, then f has a unique fixed point p and if x is any point of X then the sequence $\{f^{(n)}(x)\}$ converges to p.*

Answer. Pick any $x \in X$. By the preceding problem, the sequence $\{f^{(n)}(x)\}$ is Cauchy and since X is complete, it converges to a point p of X. Since a Lipschitz map is continuous, it follows that $\{f(f^{(n)}(x))\} = \{f^{(n+1)}(x)\}$ converges to $f(p)$. But $\{f^{(n+1)}(x)\}$ is a subsequence of $\{f^{(n)}(x)\}$ so it converges to p, and hence $f(p) = p$. So p is a fixed point of f, and by Problem 3, it is the unique fixed point. ∎

Important Caveat! In applications, X is frequently a closed subset of a Banach space V (hence complete) and f is some mapping from X into V for which one can prove that f satisfies a Lipschitz condition with constant $K < 1$. **But that is not enough!** One must also prove that f maps X into itself in order to apply the Contraction Principle.

▷ **Problem 6.** Show that if $f : X \to X$ is a contraction mapping and p is the unique fixed point of f, then for any x in X, $\rho(f^{(n)}(x), p) \leq \left(\frac{K^n}{1-K}\right) \rho(x, f(x))$

Answer. In the inequality of Problem 4, take the limit as m tends to infinity.

Remark. The sequence $\{f^{(n)}(x)\}$ is usually referred to as *the sequence of iterates of x under f*, and the process of locating the fixed point p of a contraction mapping f by taking the limit of a sequence of iterates of f goes by the name *"the method of successive approximations"*. To make this into a rigorous algorithm, we must have a "stopping rule". That is, since we cannot keep iterating f forever, we must know when to stop. One rather rough approach is to keep on iterating until successive iterates are "close enough", but a better method is provided by the previous problem. Suppose we decide to be satisfied with the approximation $f^{(n)}(x)$ if we can be sure that $\rho(f^{(n)}(x), p) \leq \epsilon$ where ϵ is some "tolerance" given in advance. We first compute $f(x)$, then $\rho(f(x), x)$, and then solve $\left(\frac{K^n}{1-K}\right) \rho(x, f(x)) = \epsilon$ for n and iterate $n - 1$ more times to get our acceptable approximation $f^{(n)}(x)$ to p.

▷ **Problem 7.** Solve $\left(\frac{K^n}{1-K}\right) \rho(x, f(x)) = \epsilon$ for n in terms of ϵ, K, and $\rho(x, f(x))$.

Answer. Take the log of both sides and solve for n. This gives

$$n = \frac{\log(\epsilon) - \log(\rho(x, f(x)) + \log(1-K)}{\log(K)}.$$

Since we want an integer, we take the smallest integer that exceeds this value,

▷ **Problem 8.** Carry out the following (third) Matlab Project. (Make a printout of your version of the M-File and submit it with the exam, but also send an electronic version to Izi and me as an email attachment.)

Third Matlab Project.

Write an Matlab M-file that implements the Successive Approximations algorithm. Name it SuccessiveApprox.m, and use it to define a Matlab function SuccessiveApprox($f, K, x,$ eps). Assume that $f : \mathbf{R}^n \to \mathbf{R}^n$ is known to be a contraction mapping with contraction constant K, that $x \in \mathbf{R}^n$, and you want to compute iterates of x until you are within eps of the fixed point p of f. Use a subfunction to compute the number of times n you need to iterate f starting from x to get within eps of p, and then use a loop and feval to iterate applying f to x the appropriate number of times.

Answer.

```
function      p = SuccessiveApprox(f,K,x,eps)
if (K>=1)
   error('Error:  K must be less than 1');
end
p = feval(f,x);
dif = x-p;
rho = sqrt(dif*dif'); % the distance from x to f(x)
n = getN(rho,eps,K);
for i=[1:n-1]
   p = feval(f,p);
end
function n = getN(distance,eps,K)
n = ceil((log(eps) - log(distance) + log(1-K))/log(K));
```

Part II: Inverse Function Theorems

I don't think you will need any convincing that "solving equations" is an essential mathematical activity. For us, solving an equation will mean that we have normed spaces V and W, a map f from a subset X of V into W, and given $y \in W$ we would like to find (some or all) $x \in X$ such that $f(x) = y$. In practice, what frequently happens is that we start with an x_0 and y_0 satisfying $f(x_0) = y_0$, and given y close to y_0 we would like to find the x close to x_0 that satisfy $f(x) = y$. A so-called "inverse function theorem" is a theorem to the effect that, under certain assumptions on f, for each y in some neighborhood U of y_0, there is a **unique** $x = g(y)$ near x_0 that solves $f(x) = y$. In this case g is called the local inverse for f near x_0. Perhaps the mother of all inverse function theorems is the Lipschitz Inverse Function Theorem, which we state below after reviewing some standard notation.

Notation. In the following, I denotes the identity mapping of the space V, so $F_0 = I - f$ means the mapping $F_0(x) = x - f(x)$ and more generally for any y in V, $F_y := I - f + y$ means the map $F_y(x) := x - f(x) + y$. Also we denote the closed ball of radius ϵ in V centered at v_0 by $\bar{B}_\epsilon(v_0, V) := \{v \in V \mid \|v - v_0\| \leq \epsilon\}$.

Lipschitz Inverse Function Theorem. *Let V be a Banach space and $f : O \to V$ a map of a neighborhood of the origin into V such that $f(0) = 0$, and suppose also that $F_0 := I - f$ satisfies a Lipschitz condition with Lipschitz constant $K < 1$. Then:*

1) f satisfies a Lipschitz condition with constant $1 + K$.

2) For $r > 0$ small enough that, $\bar{B}_r(0, V) \subseteq O$, the map $F_y := I - f + y$ is a contraction mapping of $\bar{B}_r(0, V)$, provided y in $\bar{B}_{(1-K)r}(0, V)$.

3) For each y in $\bar{B}_{(1-K)r}(0, V)$, there is a unique $x = g(y)$ in $\bar{B}_r(0, V)$ such that $f(x) = y$.

4) This inverse map $g : \bar{B}_{(1-K)r}(0, V) \to V$ satisfies a Lipschitz condition with Lipschitz constant $\frac{1}{1-K}$.

Some general hints for the following problems, The next problems will lead you through the proof of the Lipschitz Inverse Function Theorem, and the following hints may be of some help. But first try to do the problems without looking at the hints. (You quite possibly will think of some or all of them on your own anyway.)

a) Note that (since the "y"s cancel) $F_y(v_1) - F_y(v_2) = F_0(v_1) - F_0(v_2)$, so any Lipschitz condition satisfied by one of the F_y is also satisfied by all the others.

b) If you write out what it means to say that F_0 satisfies a Lipschitz condition with constant K (and rearrange terms a bit) you will find $\|(f(v_1) - f(v_2) - (v_1 - v_2)\| \leq K \|v_1 - v_2\|$.

c) What does it mean for x to be a fixed point of F_y?

d) You are probably used to thinking of the triangle inequality in the form $\|x + y\| \leq \|x\| + \|y\|$, but if you replace x by $x - y$ you end up with $\|x\| \leq \|x - y\| + \|y\|$, and quite often it is this second form of the triangle inequality that is easier to apply.

▷ **Problem 9.** Prove conclusion 1) of the Lipschitz Inverse Function Theorem.

Answer. Since $f = I - F_0$, $f(x) - f(y) = (x - y) - (F_0(x) - F_0(y))$, so $\|f(x) - f(y)\| \leq \|(x - y)\| + \|F_0(x) - F_0(y)\| \leq \|(x - y)\| + K \|(x - y)\| = (1 + K) \|(x - y)\|$. ∎

▷ **Problem 10.** Prove conclusion 2) of the Lipschitz Inverse Function Theorem. (Hint: The only slightly tricky part is showing that F_y maps $\bar{B}_r(0, V)$ to itself provided $\|y\| \leq (1 - K)r$, i.e., that for such y, if $\|x\| \leq r$, then also $\|F_y(x)\| \leq r$. See the "Important Caveat!" immediately after the statement of the Banach Contraction Principle.)

Answer. $\|F_y(x)\| = \|F_0(x) + y\| \leq \|F_0(x) - F_0(0)\| + \|y\| \leq K\,\|x\|$, so if $\|x\| \leq r$ and $\|y\| \leq (1-K)r$ then $\|F_y(x)\| \leq r$, i.e., if y in $\bar{B}_{(1-K)r}(0,V)$ then F_y maps $\bar{B}_r(0,V)$ into itself, and by hint a) it satisfies a Lipschitz condition with constant K.

▷ **Problem 11.** Prove conclusion 3) of the Lipschitz Inverse Function Theorem. (Hint: You may want to look at general hint c) for this.)

Answer. By the preceding problem, if y in $\bar{B}_{(1-K)r}(0,V)$, then F_y has a unique fixed point $x = g(y)$ in $\bar{B}_r(0,V)$. But $x = F_y(x) := x - f(x) + y$ is equivalent to $f(x) = y$. ∎

▷ **Problem 12.** Prove conclusion 4) of the Lipschitz Inverse Function Theorem.

Answer. Let y_1 and y_2 be in $\bar{B}_{(1-K)r}(0,V)$ and let $x_i = g(y_i)$, so that $f(x_i) = y_i$. Now by definition of F_0, $x_1 - x_2 = (f(x_1) - f(x_2)) - (F_0(x_1) - F_0(x_2))$ so $\|x_1 - x_2\| \leq \|y_1 - y_2\| + \|F_0(x_1) - F_0(x_2)\| \leq \|y_1 - y_2\| + K\,\|x_1 - x_2\|$. It follows that $\|g(y_1) - g(y_2)\| = \|x_1 - x_2\| \leq \frac{1}{1-K}\,\|y_1 - y_2\|$. ∎

Remark. The principle application of the Lipschitz Inverse Function Theorem is as a lemma to prove the (far more important) Differentiable Inverse Function Theorem. We consider this next.

More Notation. If V and W are normed spaces, then we denote by $L(V,W)$ the space of all continuous linear maps $T : V \to W$. (If V and W finite dimensional, then **every** linear map $T : V \to W$ is automatically continuous, so this is consistent with our earlier use of $L(V,W)$.) We saw that there was a natural choice of norm for $L(V,W)$, namely $|||\,T\,||| := \sup_{\|v\|=1} \|Tv\|$, or equivalently, $|||\,T\,||| := \sup_{\|v\|\neq 0} \frac{\|Tv\|}{\|v\|}$. We also saw that $|||\,T\,|||$ was the smallest Lipschitz constant for T. If O is open in V and $F : O \to W$ is differentiable at every point of O, then we have a map $DF : O \to L(V,W)$ called the differential of F, namely $p \mapsto DF_p$, and we say that F is C^1 (or continuously differentiable) in O if this mapping is continuous. We recall also that if O is convex and if $|||\,DF_p\,||| \leq K$ for all $p \in O$ then we showed that K was a Lipschitz constant for F.

▷ **Problem 13.** Assume that $F : O \to W$ is C^1, $p_0 \in O$ and $K > |||\,DF_{p_0}\,|||$. Show that there is a neighborhood U of p_0 such that K is a Lipschitz constant for f restricted to U.

Answer. Since F is C^1, $|||\,DF_{p_0}\,|||$ is continuous, so the set where it is less than K is open. In particular, since $|||\,DF_{p_0}\,||| < K$, there is an $\epsilon > 0$ such that $|||\,DF_p\,||| < K$ in $B_\epsilon(p_0,V)$. Since open balls in normed spaces are convex, it follows that K is a Lipschitz constant for F in $B_\epsilon(p_0,V)$.

Differentiable Inverse Function Theorem (1st Case). *Let V be a Banach space and $f : O \to V$ a C^1 map of a neighborhood of the origin into V such that $f(0) = 0$ and $Df_0 = I$, the identity map of V. If $\epsilon > 0$ is sufficiently small, then there is an $r > 0$ and a unique "inverse" map $g : B_r(0,V) \to B_\epsilon(0,V)$ such that $f(g(v)) = v$ for all v in $B_r(0,V)$. Moreover g is differentiable at the origin and $Dg_0 = I$.*

▷ **Problem 14.** Prove the 1st case of the Differentiable Inverse Function Theorem. Hint: First show that if ϵ is sufficiently small then $I - f$ satisfies a Lipschitz condition with constant $\frac{1}{2}$ in $B_\epsilon(0, V)$ and apply the Lipschitz Inverse Function Theorem.

Answer. Since the differential of $F_0 = I - f$ at 0 is 0, it follows from Problem 13 that any positive constant K is a Lipschitz constant for F in $B_\epsilon(0, V)$ if ϵ is sufficiently small, so the existence of the inverse g follows from the Lipschitz inverse function theorem, and also that g is Lipschitz with constant $\frac{1}{1-K}$. Since $DF_0 = I$, we have $f(x) = x + \|x\|\,\rho(x)$ where $\lim_{x\to 0}\rho(x) = 0$. Then $x = f(g(x)) = g(x) + \|g(x)\|\,\rho(g(x))$, so $g(x) = x + \|x\|\,\rho'(x)$, where $\rho'(x) = -\frac{\|g(x)\|}{\|x\|}\rho(g(x))$, and since g is Lipschitz and $g(0) = 0$ if follows easily that $\lim_{x\to 0}\rho'(x) = 0$, which proves that $Dg_0 = I$.

Differentiable Inverse Function Theorem (2nd Case). *Let V and W be Banach spaces and $F : O \to W$ a C^1 map of a neighborhood of the origin of V into W such that $f(0) = 0$ and DF_0 has a continuous linear inverse. If $\epsilon > 0$ is sufficiently small, then there is an $r > 0$ and a unique "inverse" map $G : B_r(0, W) \to B_\epsilon(0, V)$ such that $F(G(w)) = w$ for all w in $B_r(0, W)$. Moreover G is differentiable at the origin of W and $DG_0 = (DF_0)^{-1}$.*

▷ **Problem 15.** Prove the 2nd case of the Differentiable Inverse Function Theorem. Hint: Prove this by reducing it to the 1st case of the Differentiable Inverse Function Theorem. Namely, define $f : O \to V$ by $f := (DF_0)^{-1} \circ F$, and carry on from there.

Answer. If we define f as in the hint, then since a continuous linear map is its own differential at every point, it follows from the Chain Rule that f is differentiable at the origin and its differential is the identity. Thus f is locally inveritible, and if g is its local inverse, then $g \circ (DF_0)^{-1}$ is a local inverse for F. ∎

And finally, the following general case of the Differentiable Inverse Function Theorem follows from the "2nd Case" simply by replacing $F(v)$ by $F(v + v_0) - F(v_0)$!

Differentiable Inverse Function Theorem. *Let V and W be Banach spaces, $v_0 \in V$, and $F : O \to W$ a C^1 map of a neighborhood of v_0 in V into W such that DF_{v_0} has a continuous linear inverse. If $\epsilon > 0$ is sufficiently small, then there is an $r > 0$ and a unique "inverse" map $G : B_r(F(v_0), W) \to B_\epsilon(x_0, V)$ such that $F(G(w)) = w$ for all w in $B_\epsilon(0, W)$. Moreover G is also C^1, and in fact, if $v = G(w)$ then $DG_w = (DF_v)^{-1}$.*