MATH 3334-SUMMARY

VERN I. PAULSEN

Abstract.

1. The Space $\mathbb{R}^n$

First we introduce the basic concepts that are needed to study limits and continuity in $\mathbb{R}^n$.

Recall that given $\mathbf{x} = (\mathbf{x_1}, ..., \mathbf{x_n}) \in \mathbb{R}^\mathbf{n}$ we set $\|\mathbf{x}\| = \sqrt{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i^2}}$ and we call this the *norm of* $\mathbf{x}$. Given two vectors, $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$ the distance between them is, $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

The *open ball centered at* $\mathbf{x}$ **of radius r** is the set $B(\mathbf{x}; \mathbf{r}) = \{\mathbf{y} \in \mathbb{R}^\mathbf{n} : \|\mathbf{x} - \mathbf{y}\| < \mathbf{r}\}$.

A subset $\mathcal{O} \subseteq \mathbb{R}^n$ is called *open* provided that for every $\mathbf{x} \in \mathcal{O}$, there exists an $r > 0$ such that $B(\mathbf{x}; \mathbf{r}) \subseteq \mathcal{O}$.

Problem 1. *Show that every open ball is in fact open*

Problem 2. *Let $\mathcal{O} \subseteq \mathbb{R}^2$ be defined by $\mathcal{O} = \{\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}) : \mathbf{x_1} > \mathbf{0}, \mathbf{x_2} > \mathbf{0}\}$. Show that $\mathcal{O}$ is an open set.*

Given a set $E \subseteq \mathbb{R}^n$ a point $\mathbf{y} \in \mathbb{R}^\mathbf{n}$ is called a *limit point of E* provided that for every $r > 0$ the set $E \cap B(\mathbf{y}; \mathbf{r})$ has an element *other than* $\mathbf{y}$. The collection of all limit points of a set E is denoted E'.

As the name suggests limit points are exactly the points where it makes sense to talk about taking limits.

A set is *closed* if it contains all of its limit points, that is , if $E' \subseteq E$.

Problem 3. *Find the limit points of the set $\mathcal{O}$ defined above.*

Problem 4. *Show that for any $\mathbf{x}$ and any $r > 0$, the set $\{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| \leq \mathbf{r}\}$ is closed. This set is called the* closed ball centered at **x of radius r.**

Problem 5. *Show that if E is a set with only finitely many points, then E' is the empty set. Conclude that every set with finitely many points is closed.*

The following fact is very useful and some other texts use it as the definition of closed.

Theorem 6. *Let $\mathcal{O} \subseteq \mathbb{R}^n$ and let E denote the complement of $\mathcal{O}$. Then $\mathcal{O}$ is open if and only if E is closed.*

2000 *Mathematics Subject Classification.* Primary 46L15; Secondary 47L25.

2. Limits of Functions

Let $E \subseteq \mathbb{R}^n$ and let $f : E \to \mathbb{R}$ be a function. If $\mathbf{y} \in \mathbf{E}'$ is a limit point of E, then we wish to define what we mean by *the limit of f as x approaches y.* First for $\mathbf{x}$ to approach $\mathbf{y}$ we would like for the distance between them to go to 0 and for the limiting value we would like that the numbers $f(\mathbf{x})$ approach the limiting value, say L. Following is the precise $\epsilon - \delta$ definition.

Definition 7. *Let $E \subseteq \mathbb{R}^n$, let $f : E \to \mathbb{R}$ be a function and let $\mathbf{y} \in \mathbf{E}'$ be a limit point of E, then we write,* $\lim_{\mathbf{x} \to \mathbf{y}} f(\mathbf{x}) = \mathbf{L}$ *provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $\mathbf{x} \in \mathbf{E}$ and $0 < \|\mathbf{x} - \mathbf{y}\| < \delta$, then $|f(\mathbf{x}) - \mathbf{L}| < \epsilon$.*

The inequality, $0 < \|\mathbf{x} - \mathbf{y}\|$ is just another way to say that in limits we require $\mathbf{x} \neq \mathbf{y}$.

Example 8. *Let and let $f : E \to \mathbb{R}$ be defined by $f((x_1, x_2)) = \frac{x_1 x_2}{x_1^2 + x_2^2}$, then* $\mathbf{y} = (\mathbf{0}, \mathbf{0})$ *is a limit point of E but* $\lim_{\mathbf{x} \to \mathbf{0}} f(\mathbf{x})$ *fails to exist.*

To see this, note that $f((t, 0)) = 0$, while, $f((s, s)) = 1/2$ since we can take points of the form $(t, 0)$ and (s, s) arbitrarily close to $(0, 0)$, no matter what we tried to choose for L, we could find points within δ of $\mathbf{y}$ so that $f(\mathbf{x})$ is not close to L.

Example 9. *This shows that the limit of the function depends on the domain. Fix $a > 0$ and let $E = \{(at, t) : t \neq 0\}$ and let $f((x_1, x_2)) = \frac{x_1 x_2}{x_1^2 + x_2^2}$ be defined as before, but with a different domain. Now since for any point in $E, \mathbf{x} = (\mathbf{at}, \mathbf{t})$ we have $f(\mathbf{x}) = \frac{\mathbf{att}}{(\mathbf{at})^2 + \mathbf{t}^2} = \frac{\mathbf{a}}{\mathbf{1} + \mathbf{a}^2}$, and so now,* $\lim_{\mathbf{x} \to (\mathbf{0}, \mathbf{0})} f(\mathbf{x}) = \frac{\mathbf{a}}{\mathbf{1} + \mathbf{a}^2}$ *and just by changing the domain the limit exists!*

The key point to the above example is that to decide if limits exist it is important that we pay attention to the domains!

Example 10. *Let $E = \mathbb{R}^2 / \{(0, 0)\}$ and let $f : E \to \mathbb{R}^2$ be defined by* $f((x_1, x_2)) = \frac{x_1^4}{x_1^2 + x_2^2}$. *Show that* $\lim_{\mathbf{x} \to (\mathbf{0}, \mathbf{0})} f(\mathbf{x}) = \mathbf{0}$.

To show a limit directly, requires some inequalities. Here we see that $x_1^2 \leq x_1^2 + x_2^2 = \|\mathbf{x}\|^{\mathbf{2}}$ and hence, $x_1^4 \leq (x_1^2 + x_2^2)^2 = \|\mathbf{x}\|^{\mathbf{4}}$. Dividing this last inequality by $\|\mathbf{x}\|^{\mathbf{2}}$ we see that, $0 \leq f((x_1, x_2)) \leq \|\mathbf{x}\|^{\mathbf{2}}$. Thus, if we are given any $\epsilon > 0$ and we pick $\delta = \sqrt{\epsilon}$, then when $\|\mathbf{x} - (\mathbf{0}, \mathbf{0})\| < \delta$, we have that

$$|f(\mathbf{x}) - \mathbf{0}| = \mathbf{f}(\mathbf{x}) \leq \|\mathbf{x}\|^{\mathbf{2}} < \delta^{\mathbf{2}} = \epsilon.$$

These two examples, show that for quotients of polynomials in two or more variables, deciding if limits exists, is not as simple as just dividing and cancelling like we did in one variable.

Problem 11. *Let E be as in the last example and let $f(x_1, x_2)) = \frac{x_1^3}{x_1^2 + x_2^2}$. Show that* $\lim_{\mathbf{x} \to (\mathbf{0}, \mathbf{0})} f(\mathbf{x}) = \mathbf{0}$.

3. Continuity of Functions from $\mathbb{R}^n$ to $\mathbb{R}$

Let $E \subseteq \mathbb{R}^n$ and let $f : E \to \mathbb{R}$ be a function. A function is "continuous" if whenever tow points, $\mathbf{x}, \mathbf{y}$ are "close", then the numbers, $f(\mathbf{x}), \mathbf{f}(\mathbf{y})$ are "close". There are two equivalent ways to make this precise.

Definition 12. *Let $E \subseteq \mathbb{R}^n$, let $f : E \to \mathbb{R}$ and let $\mathbf{x} \in \mathbf{E}$. Then f is* **continuous at y** *provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that if $\mathbf{x} \in \mathbf{E}$ with $\|\mathbf{x} - \mathbf{y}\| < \delta$, then $|f(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \epsilon$. The function f is called* **continuous** *or* **continuous on E** *provided that it is continuous at every point in E.*

The following theorem gives an equivalent way to define continuity, that is often used instead.

Theorem 13. *Let $E \subseteq \mathbb{R}^n$ and let $f : E \to \mathbb{R}$. Then f is continuous on E if and only if for every limit point $\mathbf{y}$ of E that is also in E, we have that* $\lim_{\mathbf{x} \to \mathbf{y}} f(\mathbf{x}) = \mathbf{f}(\mathbf{y})$.

We will not be too concerned about proving whether or not functions are continuous, but it is important to be able to recognize functions that are continuous.

The following result is useful.

Proposition 14. *Let $\mathbf{x} = (\mathbf{x_1}, ..., \mathbf{x_n})$. Then any function that is a polynomial or continuous function of the quantities, $x_1, ..., x_n$ is continuous on all of $\mathbb{R}^n$. Functions that are quotients of such functions are continuous at any point where the denominator is not 0.*

So by the proposition, the function $f((x_1, x_2)) = x_1^3 x_2 + cos(x_2)$ is continuous on all of $\mathbb{R}^2$, while the function $g((x_1, x_2)) = \frac{x_1 x_2}{x_1^2 + x_2^2}$ would be continuous everywhere except possibly at (0,0). In the case of g, discussing continuity at (0,0) doesn't even make sense, because it has no value defined there.

A slightly more subtle example, is the function, $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f((x_1, x_2)) = \begin{cases} \frac{x_1^4}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0,0) \\ 0 & (x_1, x_2) = (0,0) \end{cases}.$$

Since, $\lim_{\mathbf{x} \to (\mathbf{0},\mathbf{0})} f(\mathbf{x}) = \mathbf{0}$, we have that f is continuous on all of $\mathbb{R}^2$.

Like for limits, determining whether or not a quotient, where both the numerator and denominator vanish, can be given a "value" at that point that will make things continuous is tricky.

4. Limits and Continuity for Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let $E \subseteq \mathbb{R}^n$ and suppose that we are given m functions $f_i : E \to \mathbb{R}, i = 1, ..., m$, then we can build a function $F : E \to \mathbb{R}^m$ by setting $F(\mathbf{x}) = (\mathbf{f_1}, (\mathbf{x}), ..., \mathbf{f_m}(\mathbf{x}))$. In this case the functions $f_1, ..., f_m$ are called the *component functions of F*. We will often refer to such a function as a *vector-valued function.*

In fact, it is quite easy to see, that any time we have a function, $F : E \to \mathbb{R}^m$, it must be of this form.

The definitions of limits and continuity for vector-valued functions is the same as for scalar-valued functions, except that the limits are vectors and instead of the absolute value of the difference, we use the norm of the difference, since this is the quantity that measures the distance between two image vectors.

For the record, we give the definitions below.

Definition 15. *Let $E \subseteq \mathbb{R}^n$, let $F : E \to \mathbb{R}^m$ be a function and let $\mathbf{x} \in \mathbf{E}'$ be a limit point of E, then we write, $\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{y}) = \mathbf{L}$, where $\mathbf{L} = (\mathbf{L_1}, ..., \mathbf{L_m})$, provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $\mathbf{y} \in \mathbf{E}$ and $0 < \|\mathbf{x} - \mathbf{y}\| < \delta$, then $\|F(\mathbf{y}) - \mathbf{L}\| < \epsilon$.*

Definition 16. *Let $E \subseteq \mathbb{R}^n$, let $F : E \to \mathbb{R}^m$ be a function and let $\mathbf{x} \in \mathbf{E}$. Then F is* **continuous at x** *provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that if $\mathbf{y} \in \mathbf{E}$ with $\|\mathbf{x} - \mathbf{y}\| < \delta$, then $\|F(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| < \epsilon$. The function F is called* **continuous** *or* **continuous on E** *provided that it is continuous at every point in E.*

For our purposes, the following characterizations of limits and continuity of vector-valued functions is useful.

Theorem 17. *Let $E \subseteq \mathbb{R}^n$, let $F : E \to \mathbb{R}^m$ be a function with component functions, $f_1, ..., f_m$ and let $\mathbf{x} \in \mathbf{E}'$ be a limit point of E, then, $\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{y}) = \mathbf{L}$, where $\mathbf{L} = (\mathbf{L_1}, ..., \mathbf{L_m})$, if and only if $\lim_{\mathbf{y}\to\mathbf{x}} f_i(\mathbf{y}) = \mathbf{L_i}$ for $i = 1, ..., m$.*

Theorem 18. *Let $E \subseteq \mathbb{R}^n$, let $F : E \to \mathbb{R}^m$ be a function with component functions, $f_1, ..., f_m$ and let $\mathbf{x} \in \mathbf{E}$, then F is continuous at $\mathbf{x}$ if and only if each of the component functions, $f_1, ..., f_m$ is continuous at $\mathbf{x}$. Consequently, F is continuous on E if and only if $f_1, ...f_m$ are all continuous on E.*

Problem 19. *Let $\mathbf{v} = (\mathbf{v_1}, ..., \mathbf{v_n})$ and $\mathbf{w} = (\mathbf{w_1}, ..., \mathbf{w_n})$ be two points in $\mathbb{R}^n$. Show that if $|v_i - w_i| < r$ for all i, then $\|\mathbf{v} - \mathbf{w}\| < \mathbf{r}\sqrt{\mathbf{n}}$.*

Problem 20. *Give $\epsilon - \delta$ proofs of the last two theorems, using the above problem.*

5. Partial Derivatives and Directional Derivatives

Let $U \subseteq \mathbb{R}^n$ be an open set, let $f : U \to \mathbb{R}$ be a function and let $\mathbf{x} \in \mathbf{U}$. If

$$\lim_{t\to 0} \frac{f(x_1, ..., x_{i-1}, x_i + t, x_{i+1}, ..., x_n) - f(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n)}{t}$$

exists, then we say that the *i-th partial derivative of f exists at* $\mathbf{x}$ and we denote this value by

$$\frac{\partial f}{\partial x_i}(\mathbf{x}).$$

More generally, if $\mathbf{u} \in \mathbb{R}^{\mathbf{n}}$ is a unit vector, then whenever

$$\lim_{t\to 0} \frac{f(\mathbf{x}+\mathbf{tu})-\mathbf{f}(\mathbf{x})}{t}$$

exists, we say that the *partial derivative of f in the direction* $\mathbf{u}$ *exists* and we denote this value by $D_{\mathbf{u}}f(\mathbf{x})$.

Note that if we let $\mathbf{e_i}$ denote the vector whose i-th coordinate is equal to 1 and whose remaining entries are 0, then

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \mathbf{D_{e_i}f}(\mathbf{x}).$$

Thus, partial derivatives are just directionalderivatives inthe directions of the coordinate axes.

More generally, if $F : U \to \mathbb{R}^m$, has component functions, $f_1, ..., f_m$, then when all the corresponding partial derivatives and directional derivatives exist, we set

$$\frac{\partial F}{\partial x_i} = (\frac{\partial f_1}{\partial x_i}, ..., \frac{\partial f_m}{\partial x_i})$$

and

$$D_{\mathbf{u}}F(\mathbf{x}) = (\mathbf{D_u f_1}(\mathbf{x}), ..., \mathbf{D_u f_m}(\mathbf{x})$$

and we call these corresponding vectors the *partial derivative* and *directional derivative* of F at $\mathbf{x}$.

Problem 21. *Get out your Calculus III text find 10 problems involving partial and directional derivatives and do them.*

Problem 22. *Show that the directional derivative of F exists at* $\mathbf{x}$ *in the direction* $\mathbf{u}$ *and is equal to the vector* $\mathbf{v}$ *if and only if*

$$\lim t \to 0 \| \frac{F(\mathbf{x}+\mathbf{tu})-\mathbf{F}(\mathbf{x})}{t} - \mathbf{v}\| = \mathbf{0}.$$

The following example shows why just the existence of partial and directional derivatives turns out to be inadequate for many results in analysis.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f((x_1, x_2)) = \begin{cases} \frac{x_1 x_2^2}{x_1^2 + x_2^4} & (x_1, x_2) \neq (0,0) \\ 0 & (x_1, x_2) = (0,0) \end{cases}.$$

If this function was continuous, then $\lim_{t\to 0} f((t^2, t)) = f(0,0) = 0$, but for $t \neq 0, f((t^2,t)) = \frac{(t^2)(t)^2}{(t^2)^2+(t)^4} = 1/2$, so this function is NOT continuous at (0,0).

However, we will now show that $D_{\mathbf{u}}f((0,0))$ exists for every direction $\mathbf{u}$. To see this we set $\mathbf{u} = (\mathbf{a}, \mathbf{b})$ where $a^2 + b^2 = 1$ and compute,

$$\lim_{t\to 0} \frac{f((0,0)+t\mathbf{u}) - \mathbf{f}((\mathbf{0},\mathbf{0}))}{t} = \lim t \to 0 \frac{f(t\mathbf{u})}{t} = \lim t \to 0 \frac{(ta)(tb)^2}{t[(ta)^2 + (tb)^4]} = \lim_{t\to 0} \frac{ab^2}{a^2 + t^2 b^4} = \frac{b^2}{a},$$

as long as $a \neq 0$, and when $a = 0$, the limit is easily seen to be 0. Hence, the directional derivative exists in every direction.

Thus, we see that the existence of directional derivatives at a point, in every direction isn't even enough to guarantee that a function is continuous at that point!

For this reason the right notion of differentiability in higher dimensions is a slightly stronger notion.

6. The Total Derivative and Linear Approximation

First let's look at the one dimensional derivative from a different viewpoint. So suppose that $f : \mathbb{R} \to \mathbb{R}$ has a derivative at the point x with $f'(x) = a$. Then we have that, $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} - a = 0$. Putting this expression over a common denominator, this is equivalent to $\lim_{h\to 0} \frac{f(x+h)-f(x)-ah}{h} = 0$. Thus, not only does $|f(x+h) - (f(x) + ah)| \to 0$ as $|h| \to 0$, but it does it so rapidly, that even after dividing by $|h|$, the ratio still tends to 0. This tells us that for h sufficiently small $f(x) + ah$ is a very good approximation to $f(x+h)$. In fact this is what we called the *tangent line approximation to f(x+h)* in calculus.

We wish to use this to motivate a definition of derivative in multi-dimensions. So let $F : \mathbb{R}^n \to \mathbb{R}^m$ so that $\mathbf{x}, \mathbf{h} \in \mathbb{R}^{\mathbf{n}}$ are now vectors and $F(\mathbf{x}), \mathbf{F}(\mathbf{x}+\mathbf{h}) \in \mathbb{R}^{\mathbf{m}}$ are vectors. If we wish to approximate $F(\mathbf{x}+\mathbf{h})$ by a quantity of the form $F(\mathbf{x}) + \mathbf{Ah}$, then we need $A : \mathbb{R}^n \to \mathbb{R}^m$. If we also want to do this in a "linear" fashion, then we need A to be an $m \times n$ matrix. This is all meant to motivate the following definition.

Definition 23. *Let $U \subseteq \mathbb{R}^n$ be an open set, let $F : U \to \mathbb{R}^m$ be a function and let $\mathbf{x} \in \mathbf{U}$. We say that F is* differentiable at $\mathbf{x}$ *provided that there is an $m \times n$ matrix A such that*

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x}+\mathbf{h}) - \mathbf{F}(\mathbf{x}) - \mathbf{Ah}\|}{\|\mathbf{h}\|} = 0.$$

In this case, we set $F'(\mathbf{x}) = \mathbf{A}$ and we call A the derivative of F at $\mathbf{x}$.

Some books refer to A as the *total derivative of F at* $\mathbf{x}$. However, since it is the only concept of the derivative at a point that is used, we find the extra word "total" redundant.

The following result relates the derivative of F to partial derivatives.

Proposition 24. *Let $U \subseteq \mathbb{R}^n$ be an open set, let $F : U \to \mathbb{R}^m$, with component functions $f_1, ..., f_m$, and let $\mathbf{x} \in \mathbf{U}$. If F is differentiable at $\mathbf{x}$, then $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ exists for all i and j is equal to the (i,j)-th entry of the matrix $F'(\mathbf{x})$.*

Thus, when the derivative exists, we have that, in matrix notation,

$$F'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{bmatrix}.$$

One immediate advantage of this definition of differentiability is the following.

Proposition 25. *Let $U \subseteq \mathbb{R}^n$, let $F : U \to \mathbb{R}^m$ and let $\mathbf{x} \in \mathbf{U}$. If F is differentiable at $\mathbf{x}$, then F is continuous at $\mathbf{x}$.*

By the two results above, we see that a necessary condition for F to be differentiable at $\mathbf{x}$ is that all of the partial derivatives exist at $\mathbf{x}$. However, by the example of the last section, we see that a function can have all of its partial derivatives exist at a point and still not be continuous at that point, which by the last result means that it can't be differentiable at that point.

Thus, while existence of all the partial derivatives are necessary for a function to be differentiable, their existence is not sufficient. This makes it very hard in general to tell exactly when a function is differentiable. For this reason the following concept is very important.

Definition 26. *Let $U \subseteq \mathbf{R^n}$ be an open set and let $F : U \to \mathbb{R}^m$ have component functions, $f_i : U \to \mathbb{R}$. If for all i and j, $\frac{\partial f_i}{\partial x_j}$ exists at every point in U and is a continuous function on U, then we say that F is* continuously differentiable on U, *or is of class $\mathcal{C}^1$ on U. We let $\mathcal{C}^1(U;\mathbb{R}^m)$ denote the set of all such functions.*

Theorem 27. *Let $U \subseteq \mathbb{R}^n$ be an open set and let $F : U \to \mathbb{R}^m$. If F is continuously differentiable on U, then F is differentiable at every point of U.*

By analogy with the one variable case, we refer to the quantity,

$$F(\mathbf{x}) + \mathbf{F}'(\mathbf{x})\mathbf{h}$$

as the *tangent approximation to* $F(\mathbf{x}+\mathbf{h})$ or as the *linear approximation to* $F(\mathbf{x}+\mathbf{h})$.

Note that for the purposes of matrix multiplication it is useful to think of vectors as *column vectors,* even though for the purposes of these notes it is more convenient to write vectors as rows. For this reason to emphasize this distinction we will often write a vector as $\mathbf{h} = (\mathbf{h_1}, ..., \mathbf{h_n})^{\mathbf{t}}$ with the superscript indicating the *transpose*, which converts a row vector into a column.

Example 28. *Let $F : \mathbb{R}^2 \to \mathbb{R}^3$ be the function whose components are given by $f_1(x_1, x_2) = x_1 cos(x_2) + sin(x_1 x_2) + 3, f_2(x_1, x_2) = x_1 + 2x_2 + 1, f_3(x_1, x_2) = x_1^3 + x_2^2 + 2x_1 + 5$. Assuming that $F'(\mathbf{0})$, exists compute this matrix and use it to find the tangent approximation to $F((.1, .3))$.*

We have that $\frac{\partial f_1}{\partial x_1} = cos(x_2) + x_2 cos(x_1 x_2)$, $\frac{\partial f_1}{\partial x_2} = -x_1 sin(x_2) + x_1 cos(x_1 x_2)$, $\frac{\partial f_2}{\partial x_1} = 1$, $\frac{\partial f_2}{\partial x_2} = 2$, $\frac{\partial f_3}{\partial x_1} = 3x_1^2 + 2$, $\frac{\partial f_3}{\partial x_2} = 2x_2$. Evaluating these at $\mathbf{0}$ we obtain that,

$$F'(\mathbf{0}) = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 2 & 0 \end{bmatrix}.$$

Hence the tangent approximation is

$$F((.1,.3)) \approxeq F(\mathbf{0}) + \mathbf{f}'(\mathbf{0})\begin{bmatrix}.1\\.3\end{bmatrix} = \begin{bmatrix}3\\1\\5\end{bmatrix} + \begin{bmatrix}.1\\.7\\.2\end{bmatrix} = \begin{bmatrix}3.1\\1.7\\5.2\end{bmatrix}.$$

Problem 29. *Let $F : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $f_1((x_1, x_2)) = 2x_1 + x_1x_2, f_2((x_1, x_2)) = 3x_1 + 2x_2 + x_1x_2, f_3((x_1, x_2)) = x_1 + 3x_2 + x_1^3$. Assuming that F is differentiable at $\mathbf{0}$, compute $F'(\mathbf{0})$, the tangent approximation to $F(\mathbf{0} + \mathbf{h})$ for $\mathbf{h} = (\mathbf{h_1}, \mathbf{h_2})^{\mathbf{t}}$ and the error, $F(\mathbf{0} + \mathbf{h}) - \mathbf{F}(\mathbf{0}) - \mathbf{F}'(\mathbf{0})\mathbf{h}$.*

Problem 30. *Show that if $F : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$ then the directional derivative of f exists in every direction $\mathbf{u}$ and that $D_{\mathbf{u}}F(\mathbf{x})$ is the product of the matrix, $F'(\mathbf{x})$ times the vector $\mathbf{u}$.*

The formulation of the derivative as a matrix is also good for understanding the chain rule.

Theorem 31 (Chain Rule)**.** *Let $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$, let $F : U \to \mathbb{R}^m$ and let $G : V \to \mathbb{R}^p$ with $F(U) \subseteq V$ so that the function, $G \circ F : U \to \mathbb{R}^p$ is defined. If F is differentiable at $\mathbf{x}$ and G is differentiable at $F(\mathbf{x})$, then $G \circ F$ is differentiable at $\mathbf{x}$ and $(G \circ F)'(\mathbf{x})$ is equal to the matrix product, $G'(F(\mathbf{x}))\mathbf{F}'(\mathbf{x})$.*

Problem 32. *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$ have component functions, $f_1(x_1, x_2) = x_1^2 + 3x_2, f_2(x_1, x_2) = 5x_1 + x_2^3, g_1(y_1, y_2) = y_1^3 + 4y_2, g_2(y_1, y_2) = 2y_1 + 3y_2$. Compute, $(G \circ F)'((1, 2))$, two ways. First, by finding the function $G \circ F$ explicitly and differentiating it and second, by using the chain rule.*

7. Newton's Method and Roots of Equations

As one application we take a look at a multi-variable version of Newton's method. Suppose that we have a function, $F : \mathbb{R}^n \to \mathbb{R}^n$ and we wish to find a solution to the equation, $F(\mathbf{x}) = \mathbf{0}$.

As in the one-variable version of Newton's method. Given an initial guess $\mathbf{x_0}$ to a solution, for $\mathbf{x_1}$ sufficiently close to $\mathbf{x_0}$, we have that $F(\mathbf{x_1}) \approxeq \mathbf{F}(\mathbf{x_0}) + \mathbf{F}'(\mathbf{x_0})(\mathbf{x_1} - \mathbf{x_0})$. Thus, if we want $F(\mathbf{x_1}) = \mathbf{0}$, then assuming that $\approxeq$ is really "=", leads to, $\mathbf{0} = \mathbf{F}(\mathbf{x_0}) + \mathbf{F}'(\mathbf{x_0})(\mathbf{x_1} - \mathbf{x_0})$. Solving this last equation for $\mathbf{x_1}$, leads to,

$$\mathbf{x_1} = \mathbf{x_0} - \mathbf{F}'(\mathbf{x_0})^{-\mathbf{1}}\mathbf{F}(\mathbf{x_0}),$$

where $F'(\mathbf{x_0})^{-\mathbf{1}}$ denotes the inverse of the matrix $F'(\mathbf{x_0})$(which we need to assume exists).

This formula leads to *Newton's algorithm for finding the root of an equation:*

Given an initial guess for a solution, $\mathbf{x_0}$, inductively define,

$$\mathbf{x_{n+1}} = \mathbf{x_n} - \mathbf{F}'(\mathbf{x_n})^{-\mathbf{1}}\mathbf{F}(\mathbf{x_n}),$$

assuming that the inverses of these matrices exist. We call this sequence of vectors(when it exists) the *sequence of Newton approximants starting from* $\mathbf{x_0}$.

The following theorem tells us that under certain assumptions the sequence of vectors given by Newton's algorithm converges to an actual root of the equation.

Theorem 33 (Newton's Algorithm)**.** *Let* $F : \mathbb{R}^n \to \mathbb{R}^n$ *and assume that* $\mathbf{z} \in \mathbb{R}^{\mathbf{n}}$ *is a point that satisfies,* $F(\mathbf{z}) = \mathbf{0}$. *If* F *is continuously differentiable in a neighborhood of* $\mathbf{z}$ *and* $F'(\mathbf{z})$ *is invertible, then there is* $r > 0$, *such that for any* $\mathbf{x_0}$ *with* $\|\mathbf{z} - \mathbf{x_0}\| < \mathbf{r}$, *the sequence of Newton's approximants starting from* $\mathbf{x_0}$ *will converge to* $\mathbf{z}$, *that is,* $\lim_{n\to+\infty} \|\mathbf{z} - \mathbf{x_n}\| = \mathbf{0}$.

Problem 34. *Let* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *have component functions,* $f_1(x, y) = x^2 + y^2 - 4, f_2(x, y) = 2x - 3y$. *Sketch the graphs of the sets satisfying* $f_1(x, y) = 0$ *and of* $f_2(x, y) = 0$, *to show that there are exactly two points in* $\mathbb{R}^2$ *with* $F(\mathbf{z}) = \mathbf{0}$. *Using an initial guess for the root, of* $\mathbf{x_0} = (\mathbf{1}, \mathbf{1})$, *compute* $\mathbf{x_1}, \mathbf{x_2}$. *Based on this(flimsy) evidence, which root does it appear that the sequence of Newton approximants starting from (1,1) will converge to?*

8. The Inverse Function Theorem

We first discuss inverse functions in several variables with an explicit example. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f_1(x_1, x_2) = x_1^2 + x_2^2 = y_1, f_2(x_1, x_2) = x_1^2 - x_2^2 = y_2$. An inverse of this funcion would be a function that expressed the x_i's as functions of the y_i's. Note that on the whole of $\mathbb{R}^2$ this would be impossible since $F((x_1, x_2)) = F((\pm x_1, \pm x_2))$. Thus, every point in the range of F has four pre-images. However, if we set $U = \{(x_1, x_2) : x_1 > 0, x_2 > 0\}$ and $V = \{(y_1, y_2) : y_1 > y_2, y_1 > 0\}$, then we will show that F maps U one-to-one and onto V and hence there is a unique function $G : V \to U$ such that $G(F((x_1, x_2))) = (x_1, x_2)$.

To find G, note that if (y_1, y_2) is in the range of F, then $y_1 > 0$ and $y_1 + y_2 = 2x_1^2 > 0$, so that $x_1 = \sqrt{\frac{y_1+y_2}{2}} = g_1(y_1, y_2)$. Also, $y_1 - y_2 = 2x_2^2 > 0$, so that $y_1 > y_2$ and $x_2 = \sqrt{\frac{y_1-y_2}{2}} = g_2(y_1, y_2)$.

Now it is easily checked that for any $(x_1, x_2) \in U, F((x_1, x_2)) \in V$ and $G(F((x_1, x_2))) = (x_1, x_2)$, while for any $(y_1, y_2) \in V, G((y_1, y_2)) \in U$ and $F(G((y_1, y_2))) = (y_1, y_2)$. These two equations combined show that V is the image of U under F, that U is the image of V under G, that F maps U one-to-one onto V and that G maps V one-to-one onto U.

Finally, if $H((x_1, x_2)) = (x_1, x_2)$ is the identity function, then it is easily seen that $H'((x_1, x_2)) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$, the 2×2 identity matrix. Thus, applying the chain rule to the function $G \circ F = H$ we see that

$$G'(F((x_1, x_2))F'((x_1, x_2)) = H'((x_1, x_2)) = I_2,$$

and so we have that,

$$G'(F((x_1,x_2)) = [F'((x_1,x_2))]^{-1}.$$

That is, if G is the inverse mapping of F, then the derivative of G is the matrix inverse of the derivative of F.

In general, we aren't so lucky as to be able to compute inverse functions so explicitly. This is why the following theorem that tells us that they exist and that their derivatives obey this rule is so important.

Theorem 35 (Inverse Function Theorem). *Let $U \subseteq \mathbb{R}^n$ be open and let $F : U\mathbb{R}^n$ be continuously differentiable on U. If $\mathbf{x_0} \in \mathbf{U}$ and the matrix $F'(\mathbf{x_0})$ is invertible, then there is an open set U_0 with $\mathbf{x_0} \in \mathbf{U_0}$, an open set V_0 with $F(\mathbf{x_0}) \in \mathbf{V_0}$ and a continuously differentiable function, $G : V_0 \to U_0$ such that $G(F(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbf{U_0}$ and $F(G(\mathbf{y})) = \mathbf{y}$ for every $\mathbf{y} \in \mathbf{V_0}$. Moreover, for every $\mathbf{x} \in \mathbf{U_0}, \mathbf{F}'(\mathbf{x})$ is invertible, for every $\mathbf{y} \in \mathbf{V_0}, \mathbf{G}'(\mathbf{y})$ is invertible and $G'(F(\mathbf{x})) = [\mathbf{F}'(\mathbf{x})]^{-\mathbf{1}}$.*

We show one of the practical applications of this theorem. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$, have component functions, $f_1(x_1,x_2) = x_1x_2^2, f_2(x_1,x_2) = x_1^2 + x_2$. Note that $F((1,-1)) = (1,0)$. Find a "good" approximate solution to the pair of equations, $x_1x_2^2 = 1.1, x_1^2 + x_2 = .2$ that is near to the point (1,-1).

To solve this note that if we let G denote the inverse of the function F, then $G((1,0)) = (1,-1)$, since $(1.1,.2)$ is near to $(1,0)$, namely, $(1.1,.2) = (1,0) = (.1,.2)$, the tangent approximation is

$$G((1.1,.2)) \approxeq G((1,0)) + G'((1,0)) \begin{bmatrix} .1 \\ .2 \end{bmatrix}.$$

By the inverse function theorem, since $F((1,-1)) = (1,0)$, we have that

$$G'((1,0)) = [F'((1,-1))]^{-1} = \begin{bmatrix} (-1)^2 & 2(1) \\ 2(1)(-1) & 1 \end{bmatrix}^{-1} = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}.$$

Thus, the approximation is given by

$$G((1.1,.2) \approxeq \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \frac{1}{5}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} .1 \\ .2 \end{bmatrix} = \begin{bmatrix} .94 \\ -.92 \end{bmatrix}$$

Thus, $F((.94,-.92)) \approxeq (1.1,.2)$.

Problem 36. *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ have component functions $f_1(x_1,x_2) = x_1^2 + x_2^2, f_2(x_1,x_2) = x_1 + 2x_2$. We have that $F((1,1)) = (2,3)$. Use the inverse function theorem to give an approximate solution to $f_1(x_1,x_2) = 2.2, f_2(x_1,x_2) = 3.3$.*

9. The Implicit Function Theorem

Recall that an equation like, $x^2 + y^2 - 1 = 0$, implicitly defines y as a function of x and that we can compute $\frac{dy}{dx}$ by implicit differentiation, obtaining, $2x + 2y\frac{dy}{dx} = 0$.

In a similar fashion, the set of equations,

$$f_1(x_1,x_2,y_1,y_2) = 3x_2 - y_1^2 - y_2^2 - x_1^2 = 0,$$
$$f_2(x_1,x_2,y_1,y_2) = 2x_2^2 + 4y_1^2 + y_2^2 + 6x_1^2 - 13 = 0,$$

implicitly define y_1, y_2 as functions of x_1, x_2, so we would like to be able to compute the four partial derivatives, $\frac{\partial y_i}{\partial x_j}$.

As one practical application of having such partial derivatives, note that $(x_1,x_2,y_1,y_2) = (-1,1,1,-1)$ satisfies the above equations. If we know the partial derivatives, then we will be able to give the tangent approximation to the values of y_1, y_2 that satisfy these equations when $x_1 = -1.1, x_2 = 1.3$.

The implicit function theorem answers when the y's can be given as functions of the x's in such a case and tells how to compute the derivative.

To explain it, we first focus on the above example. Computing partial derivatives, with respect to x_1, yields

$$0 - 2y_1\frac{\partial y_1}{\partial x_1} - 2y_2\frac{\partial y_2}{\partial x_1} - 2x_1 = 0,$$

$$8y_1\frac{\partial y_1}{\partial x_1} + 2y_2\frac{\partial y_2}{\partial x_1} + 12x_1 = 0.$$

While partial derivatives with respect to x_2 yield,

$$3 - 2y_1\frac{\partial y_1}{\partial x_2} - 2y_2\frac{\partial y_2}{\partial x_2} = 0,$$

$$4x_2 + 8y_1\frac{\partial y_1}{\partial x_2} + 2y_2\frac{\partial y_2}{\partial x_2} = 0.$$

Writing this system of four equations in four unknowns, in matrix notation yields:

$$\begin{pmatrix} -2y_1 & -2y_2 \\ 8y_1 & 2y_2 \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} + \begin{pmatrix} -2x_1 & 3 \\ 12x_1 & 4x_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus, the solution is given by,

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} = -\begin{pmatrix} -2y_1 & -2y_2 \\ 8y_1 & 2y_2 \end{pmatrix}^{-1} \begin{pmatrix} -2x_1 & 3 \\ 12x_1 & 4x_2 \end{pmatrix}$$

and we see that for a solution to exist we need the matrix to be invertible.

At the point $(-1,1,1,-1)$, we have that

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} = -\begin{pmatrix} -2 & 2 \\ 8 & -2 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 3 \\ -12 & 4 \end{pmatrix} = \begin{pmatrix} -5/3 & 7/6 \\ -2/3 & 8/3 \end{pmatrix}.$$

Thus, to find the tangent approximation for $x_1 = -1.1, x_2 = 1.3$, we have that $\Delta x_1 = -1.1 - (-1) = -.1, \Delta x_2 = 1.3 - 1 = .3$, and so

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \approxeq \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} -5/3 & 7/6 \\ -2/3 & 8/3 \end{pmatrix} \begin{pmatrix} -.1 \\ .3 \end{pmatrix} = \begin{pmatrix} 31/60 \\ 26/30 \end{pmatrix}.$$

To put this problem into the format of the implicit function theorem, note that the pair of equations can be thought of as the component functions of a function $F : \mathbb{R}^4 \to \mathbb{R}^2$. Only in this case, we wish to regard the domain as $\mathbb{R}^2 \times \mathbb{R}^2$, writing a point in the domain as $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}), \mathbf{y} = (\mathbf{y_1}, \mathbf{y_2})$. Thus, the equation becomes, $F(\mathbf{x}, \mathbf{y}) = \mathbf{0}$. Now assuming that $y_1 = h_1(x_1, x_2), y_2 = h_2(x_1, x_2)$, we have that

$$f_1(x_1, x_2, h_1(x_1, x_2), h_2(x_1, x_2)) = 3x_2 - h_1(x_1, x_2)^2 - h_2(x_1, x_2)^2 - x_1^2 = 0,$$

$$f_2(x_1, x_2, h_1(x_1, x_2), h_2(x_1, x_2)) = 2x_2^2 + 4h_1(x_1, x_2)^2 + h_2(x_1, x_2)^2 + 6x_1^2 - 13 = 0,$$

or $F(\mathbf{x}, \mathbf{H}(\mathbf{x})) = \mathbf{0}$. Hence, the derivative of this composite function must be 0.

Separating the partial deriviatives of F into their parts depending on $\mathbf{x}$ and $\mathbf{y}$, we can write the 2×4 matrix as, $F' = (F_x, F_y)$, where F_x and F_y are 2×2 matrices.

A careful application of the chain rule, yields

$$F_x + F_y H' = 0,$$

and so the solution is, $H' = -F_y^{-1} F_x$.

This hopefully clarifies the statement of the implicit function theorem.

Theorem 37 (Implicit Function Theorem)**.** *Write a vector in $\mathbb{R}^n \times \mathbb{R}^m$ as $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^\mathbf{n}, \mathbf{y} \in \mathbb{R}^\mathbf{m}$, let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be continuously differentiable and let $F(\mathbf{x_0}, \mathbf{y_0}) = \mathbf{0}$. If $F_{\mathbf{y_0}}$ is invertible, then there is a neighborhood U of $\mathbf{x_0}$ and a continuously differentiable function, $H : U \to \mathbb{R}^m$ with $H(\mathbf{x_0}) = \mathbf{y_0}$ and satisfying, $F(\mathbf{x}, \mathbf{H}(\mathbf{x})) = \mathbf{0}$, for all $\mathbf{x} \in \mathbf{U}$. Moreover, in this case,*

$$H'(\mathbf{x_0}) = -\mathbf{F}'_\mathbf{y}(\mathbf{y_0})^{-\mathbf{1}} \mathbf{F}_\mathbf{x}(\mathbf{x_0}).$$

10. The Jacobian and Local Max-Min

A **critical point** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a point where either $f'(\mathbf{x}) = \mathbf{0}$ or the derivative fails to exist. In this section we discuss the multi-variable analogue of the second derivative test for classifying local maxima and minima. We will need some concepts and results from linear algebra.

An $n \times n$ matrix $P = (p_{i,j})$ is called **positive definite**, provided that for every non-zero vector, $x = (x_1, ..., x_n)$, we have that

$$\sum_{i,j=1}^{n} x_i x_j p_{i,j} > 0.$$

The following result is useful for determining whether or not a matrix is positive definite. Given an $n \times n$ matrix $P = (p_{i,j})_{i,j=1}^n$, the matrices, $P_k = (p_{i,j})_{i,j=1}^k, k = 1, 2, ..., n$ are called the **principal submatrices of P.**

Proposition 38. *Let $P = (p_{i,j})$ be an $n \times n$ matrix. Then the following are equivalent:*

- *P is positive definite,*
- *$P = P^t$ and every eigenvalue of P is strictly positive,*
- *$P = P^t$ and $det(P_k) > 0, k = 1, 2, ..., n$.*

Theorem 39. *Let $f : \mathbb{R}^n \to \mathbb{R}$ have continuous 2nd order partial derivatives. If $f'(\mathbf{x_0}) = \mathbf{0}$ and $(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x_0})$ is positive definite, then $\mathbf{x_0}$ is a local minimum of f. If the negative of this matrix is positive definite, then $\mathbf{x_0}$ is a local maximum.*

The matrix $(\frac{\partial^2 f}{\partial x_i \partial x_j})$ is called the **Jacobian of f.** When the Jacobian matrix is neither positive definite nor the negative of a positive definite, then we say that the 2nd derivative test fails. generally, this means that the point $\mathbf{x_0}$ is a **saddle point**, that is the graph of the function is cupped up in some directions and cupped down in other directions. These directions are determined, respectively, by the eigenvectors corresponding to strictly positive and strictly negative eigenvalues of the Jacobian matrix.

Problem 40. *Find and classify all the critical points of*

$$f(x,y) = -x^4 - y^4 + 6x^3 - 2x^2y^2 - 16x^2 + 28x + 6xy^2 - 8y^2 + 4y + 2x^2y - 8xy + 2y^3 - 12.$$

Problem 41. *Let $A = (a_{i,j})$ be an $n \times n$ matrix. Show that $-A$ is positive definite if and only if $A = A^t$ and $det(A_k)$ is strictly positive for k even and strictly negative for k odd.*

Consequently, to test for a local maximum, we need

Department of Mathematics, University of Houston, Houston, Texas 77204-3476, U.S.A.
E-mail address: vern@math.uh.edu