MA1C ANALYTIC RECITATION 5/3/12
MIDTERM REVIEW

ALDEN WALKER

1. Interesting Problem

Suppose there is a rubber band with one end tied to a post in the ground, and the other end tied to the tail of a kangaroo. The two ends start 1 meter apart. There is a flea sitting on the post, which jumps 1 cm onto the stretched rubber band. The kangaroo then jumps 1 m, stretching the rubber band accordingly. The flea then jumps 1 cm toward the kangaroo, and so on. Does the flea catch the kangaroo?

One way to think about the problem is to consider the ratio between the distance from the post to the flea and the post to the kangaroo. The kangaroo jumping does not change this ratio, so the only thing we have to figure out is how much the flea jump changes the ratio. The ratio starts at $\frac{1\text{cm}}{100\text{cm}} = 1/100$. After one jump by the kangaroo and one by the flea, we have the ratio $\frac{2\text{cm}+1\text{cm}}{200\text{cm}}$, and so on. In general, if we have the ratio $\frac{x}{y}$, then the kangaroo jump gives us

$$\frac{x\frac{y+100}{y}}{y+100}$$

(it's the same, but written to keep the correct number of centimeters on the top and bottom) and the flea jump gives

$$\frac{x\frac{y+100}{y}+1}{y+100} = \frac{x}{y} + \frac{1}{y+100}$$

That is, if the kangaroo is at y, then after the two jumps we add $\frac{1}{y+100}$ to the ratio. Therefore after n jumps we've added $\sum_{i=1}^{n} \frac{1}{100i+100} = \frac{1}{100}\sum_{i=1}^{n}\frac{1}{i+1}$ to the ratio. However, a comparison with the series $\sum \frac{1}{2i}$ shows that this sum diverges. That is, there is an n such that $\sum_{i=1}^{n}\frac{1}{i+1} > 99$, which makes the ratio larger than $1/100 + (99/100) = 1$, i.e. the flea has caught the kangaroo.

2. Open Sets

- A set A in $\mathbb{R}^n$ is **open** if for every point $x \in A$ there exists a ball about $\mathbf{x}$ entirely contained in A.
- A set is **closed** if its complement (in $\mathbb{R}^n$) is open.
- The above definition of closed is equivalent to: a set A is closed if it contains all of its limit points, where a limit point is the limit of a sequence contained in A.
- The **interior** of a set A is the largest open set contained in A.

Let's denote the ball of radius r about $\mathbf{x}$ by $B_r(\mathbf{x})$. That is, $B_r(\mathbf{x}) = \{\mathbf{a} \in \mathbb{R}^n \mid \|\mathbf{a} - \mathbf{x}| < r\}$.

2.1. **How To Prove Something Is Open.** Given $\mathbf{x} \in A$, you somehow produce an r such that you can show that $B_r(\mathbf{x}) \subseteq A$. Your method works for every $\mathbf{x}$ in A, perhaps by taking cases on what $\mathbf{x}$ is, so you have shown that every point has a ball about it, and thus A is open.

2.2. **Example.** An arbitrary union $\bigcup_{i\in I} A_i$ of open sets is open. To prove this, pick some point x in the union. It must be contained in some A_i, which is open. Therefore, there is some ball B so $x \in B \subset A_i$, so $B \in \bigcup A_i$, and this $\bigcup A_i$ is open.

3. Continuity

3.1. **Definition.**

- A function $f : \mathbb{R}^n \to \mathbb{R}^m$ (we might also consider functions whose domain is a subset of $\mathbb{R}^n$) is **continuous** at a point $\mathbf{a} \in \mathbb{R}^n$ if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$. A function which is continuous at all points is just called continuous.

Notice that this definition, translated into a verison that is useful for proving things, is: for all $\epsilon > 0$ there exists $\delta > 0$ such that $\|\mathbf{x}-\mathbf{a}\| < \delta \Rightarrow \|f(\mathbf{x}) - f(\mathbf{a})\| < \epsilon$. Just like with one-dimensional continuity, it is often very tedious and/or annoyingly difficult to prove that a function is continuous straight from the definition.

3.2. **How to Prove Something is Continuous.** Use pages 248–250! They outline lots of facts that you can put together to get many nice facts. For instance, is $x^2y^2/(x+1)$ continuous when $x \neq -1$? Yes! Because Example 5 on page 249 says so! Here is a basic outline of what to do in general when you are asked where a function is continuous (note this question implies you need to prove that it is continuous where you say AND that it is not continuous elsewhere!).

- Figure out for yourself where the function is continuous
- Use pages 248–250 to prove the vast majority of points are continuous, leaving a handful of points for you to deal with by hand.
- Deal with the leftover points: for each one, either prove the function is continuous there (with a limiting argument involving δ and ϵ), or show it is not continuous there (usually by exhibiting a sequence $x_n \to z$ such that $\lim_{n\to\infty} f(x_n) \neq f(z)$).

3.3. **Facts About Continuous Functions.**

- The preimage of an open set under a continuous function is open. This is actually the definition of a continuous function in more general settings. This fact is true because if you have some point x in the preimage $f^{-1}(S)$ of an open set S, then there is some ball $B_\epsilon(f(x))$ about $f(x)$, as the image is open. By continuity, there is some δ such that $\|x - a\| < \delta \Rightarrow \|f(x) - f(a)\| < \epsilon$. Therefore, if $y \in B_\delta(x)$, then $f(y) \in B_\epsilon(f(x))$, so $f(y) \in S$, so $y \in f^{-1}(S)$. That is, $B_\delta(x)$ is an open ball about x inside S.
- The image of an open set is not necessarily open! The function x^2 maps $(-1,1)$ to $[0,1)$.

4. Derivatives in General

The definition is:

$$\mathbf{f}'(\mathbf{a};\mathbf{y}) = \lim_{h\to 0} \frac{\mathbf{f}(\mathbf{a}+h\mathbf{y}) - \mathbf{f}(\mathbf{a})}{h}$$

etc.

The definition of the total derivative of f at $\mathbf{a}$ is a linear transformation $T_{\mathbf{a}}$ such that

$$f(\mathbf{a}+\mathbf{v}) = f(\mathbf{a}) + T_{\mathbf{a}}(\mathbf{v}) + \|\mathbf{v}\|E(\mathbf{a},\mathbf{v})$$

For all $\mathbf{v}$ in some ball about $\mathbf{a}$, and where $E(\mathbf{a},\mathbf{v}) \to 0$ as $\|\mathbf{v}\| \to 0$. Basically what this means is that if you take $(f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a}))/\|\mathbf{v}\|$ as $\|\mathbf{v}\| \to 0$, then it has a limit, and that limit is given by $T_{\mathbf{a}}$.

It is convenient to split up into component functions, so $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_m(\mathbf{x}))$. Now instead of the gradient, the total derivative (assuming it exists) expressed in terms of the standard basis is the Jacobian:

$$D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

This may seem complicated, but think about this: $\mathbf{f}$ eats things in $\mathbb{R}^n$ and spits out things in $\mathbb{R}^m$. If you give $D\mathbf{f}(\mathbf{a})$ a direction in $\mathbb{R}^n$, it will tell you how the image of $\mathbf{a}$ under $\mathbf{f}$ is changing in that direction.

4.1. **How to Tell if a Function is Differentiable.** We have all these things that are useful if a function is differentiable, but no good way to tell if a function is differentiable. Here is a good condition (Theorem 8.7): If all the partials of f exist in some ball about $\mathbf{a}$, and the partials are continuous at $\mathbf{a}$, then f is differentiable at $\mathbf{a}$.

4.1.1. *Example.* Let $f : \mathbb{C} \to \mathbb{C}$ be defined $f(z) = z^2$. Note that f can be thought of as a function $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x + iy) = (x + iy)(x + iy) = x^2 + 2ixy - y^2$, or rather $f(x, y) = (x^2 - y^2, 2xy)$. We can see that f has continuous partials, so it is differentiable everywhere.

Interesting note you don't have to know: It is possible to define complex differentiation (actually, it looks the same: $f'(z) = \lim_{h\to 0} \frac{f(z+h)-f(z)}{h}$, where $h \in \mathbb{C}$), but you obviously won't ever need to know anything about this for this course. Notice that because in that limit h is complex, being complex differentiable is a very strong condition. I mention this because it turns out that a function is complex differentiable if and only if $\frac{\partial f_1}{\partial x} = \frac{\partial f_2}{\partial y}$ and $\frac{\partial f_1}{\partial y} = -\frac{\partial f_2}{\partial x}$. Those are called the Cauchy-Riemann equations.

4.2. **Gradient.** A nice special case of the total derivative is the gradient. Let f be a function $f : \mathbb{R}^n \to \mathbb{R}$. The gradient of a f at a point $\mathbf{a}$, which is denoted $\nabla f(\mathbf{a})$ (often just ∇f to denote the function from $\mathbb{R}^n \to \mathbb{R}^n$), is defined to be

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{a})\right)$$

Notice that when f is differentiable, the gradient of f *is* the total derivative of f, written in the standard basis.

4.2.1. *The Best Way to Compute Directional Derivatives.* If f is differentiable, then we noted above that $f'(\mathbf{a}; \mathbf{y}) = T_{\mathbf{a}}(\mathbf{y})$, and we also noted that $T_{\mathbf{a}}$, expressed in the standard basis, is $\nabla f(\mathbf{a})$. Therefore, we see that $f'(\mathbf{a}; \mathbf{y}) = \nabla f(\mathbf{a}) \cdot \mathbf{y}$.

5. The Chain Rule in General

The chain rule in general is about as nice as it could possibly be: suppose that $\mathbf{h} = \mathbf{f} \circ \mathbf{g}$ and $\mathbf{g}$ is differentiable at $\mathbf{a}$ and $\mathbf{f}$ is differentiable at $\mathbf{g}(\mathbf{a})$. Then

$$D\mathbf{h}(\mathbf{a}) = D\mathbf{f}(\mathbf{g}(\mathbf{a}))D\mathbf{g}(\mathbf{a})$$

Where the multiplication is just matrix multiplication. Looking again at the chain rule from before, we see that that was a special case of this.

6. Critical Points

If you have a scalar valued function, then at a local extreme point, the gradient is zero. A glance at $x^2 - y^2$ at $(0, 0)$ shows that the converse in not true. A place where the gradient is zero is called a stationary point of the function. To explore the nature of the stationary point, we must look at higher order derivatives (just like the 2nd derivative test in one dimension).

The Hessian matrix of f is defined to be $H(\mathbf{x}) = [D_{ij} f(\mathbf{x})]_{i,j=1}^n$, and it is important because if f has continuous second partial derivatives in a ball about $\mathbf{a}$, then for $\mathbf{y}$ such that $\mathbf{a} + \mathbf{y} \in B(\mathbf{a})$ we have

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}\mathbf{y}H(\mathbf{a})\mathbf{y}^T + \|\mathbf{y}\|^2 E_2(\mathbf{a}, \mathbf{y})$$

The last term is just an error term—don't worry about it. The point is that if you rearrange that you see that the sign of the derivative in the direction $\mathbf{y}$ is $\frac{1}{2}\mathbf{y}H(\mathbf{a})\mathbf{y}^T$. The reason this is useful is that the Hessian matrix, under the above assumptions, is symmetric (I would say it is self-adjoint, because the following theorem applies to that more general case), which means that there is a basis of $\mathbb{R}^n$ consisting of eigenvectors of H. If $\mathbf{y}$ is an eigenvector, then the sign of $\frac{1}{2}\mathbf{y}H(\mathbf{a})\mathbf{y}^T$ is the sign of the eigenvalue of $\mathbf{y}$. This means that the sign of the derivative of f in the various directions at $\mathbf{a}$ corresponds with the signs of the eigenvalues of H. In particular,

- If all the eigenvalues of H are negative, then f has a maximum.
- If all the eigenvalues of H are positive, then f has a minimum.
- If there are eigenvalues of both signs, f has a saddle.

6.1. **Example.** Let's look at the critical points of $f(x,y) = x^2+y^2$. First, we find the gradient $\nabla f = (2x, 2y)$. Clearly, the only place where this is 0 is at $(0,0)$. To see what kind of stationary point this is, we compute the Hessian:

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x \partial x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y \partial y} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

The eigenvalues of which are clearly $(2,2)$. Therefore this matrix is positive definite, and $(0,0)$ is therefore a local minimum. Let's show this is a global minimum. Take a disk of radius 1. If $\|v\| \geq 1$, then $v_1^2+v_2^2 \geq 1 > 0$, so the minimum of f must be achieved inside the disk of radius 1, so it must be achieved at a critical point, so it must be achieved at $(0,0)$.

6.2. **Example.** The same calculation with the function $f(x,y) = x^2-y^2$ yields the Hessian $H_f = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, showing that the stationary point $(0,0)$ is a saddle for f.

6.3. **Example.** Here's a handy way of writing a quadratic function. You already know this, but you may not know you know it. Given any pureply quadratic function (only degree 2 terms), we can write it as

$$f(x,y) = (x,y) \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} (x,y)^T$$

Notice that the Hessian matrix of this function *is* the matrix (times two). And the gradient of f is $(2a_{11}x + 2(a_{12}+a_{21})y, 2a_{22}y + 2(a_{12}+a_{21})x)$.

Also note that if the matrix has only positive eigenvalues, then the function f is always positive!

7. Lagrange Multipliers

It is common that you would want to optimize a function given some constraint. A nice way to do this is to use the method of Lagrange Multipliers: if you have a function $f : \mathbb{R}^n \to \mathbb{R}$ and constraints $g_1, \cdots g_m : \mathbb{R}^n \to \mathbb{R}$ (i.e. $g(\mathbf{x}) = 0, \cdots, g(\mathbf{x}) = 0$), then where f has a relative extremum subject to the constraints we have:

$$\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_m \nabla g_m$$

This may seem arbitrary, but there is good intuition here: suppose that you are at some point in the domain of f. To increase the function, you want to follow the gradient. If the gradient points outside of the region to which you are constrained (the feasible region), then you can imagine being held in the feasible region and having the gradient be a force vector pulling you in that direction. Then even though the gradient points outside the feasible region, you will be pulled in the component of the gradient inside the feasible region, and you will continue to move until the gradient points in a direction perpendicular to the feasible region, at which point you are at a local maximum.

This method does not work if the ∇g_i are not independent. The book has an example in three dimensions on p. 317, but a simple example is to consider finding extrema of $f(x,y) = x^2+y^2$ subject to the constraint $g(x,y) = x^2 - (y-1)^3 = 0$. Obviously, there is a local (actually global) minimum at $(0,1)$, but $\nabla g(0,1) = 0$, so there clearly is no λ such that $\nabla f(0,1) = (0,2) = \lambda \cdot 0$.

7.1. **Example.** Minimize $f(x,y,z) = xyz$ subject to the constraints $0 = g_1(x,y,z) = x+y+z$ and $0 = g_2(x,y,z) = x^2+y^2-z$.

Ok, let's use Lagrange multipliers. Incidentally, what does the feasible region look like? It's the intersection of a parabola and a plane, appropriately arranged, so it's a circle. In particular, it's compact, so in order to find a global minimum, we only need to consider the Lagrange critical points.

We have:

$$\begin{aligned} \nabla f(x,y,z) &= (yz, xz, xy) \\ \nabla g_1(x,y,z) &= (1,1,1) \\ \nabla g_2(x,y,z) &= (2x, 2y, -1) \end{aligned}$$

We need:

$$(yz, xz, xy) = \lambda_1(1,1,1) + \lambda_2(2x, 2y, -1)$$

So we just solve these:

$$yz = \lambda_1 + 2x\lambda_2$$
$$xz = \lambda_1 + 2y\lambda_2$$
$$xy = \lambda_1 - \lambda_2$$

Plugging in for $\lambda_1 = xy + \lambda_2$:

$$yz = xy + (2x+1)\lambda_2$$
$$xz = xy + (2y+1)\lambda_2$$

So $(x-y)z = 2(y-x)\lambda_2$

Therefore, either $x = y$, or $z = -2\lambda_2$. If $x = y$, then combining g_1 and g_2, we have $0 = 2x + z$ and $0 = 2x^2 - z$, so $0 = 2x + 2x^2 = 2x(1+x)$, so $x = 0$ or $x = -1$. If $x = y = 0$, then $z = 0$. If $x = y = -1$, then by g_1, we have $z = 2$, and this works in g_2 also.

If $z = -2\lambda_2$, then we have equations: So we just solve these:

$$yz = \lambda_1 - xz$$
$$xz = \lambda_1 - yz$$
$$xy = \lambda_1 - \lambda_2$$

So adding the first two, $2(x+y)z = \lambda_1$, so

$$yz = 2(x+y)z - xz$$
$$xz = 2(x+y)z - yz$$
$$xy = 2(x+y)z + (1/2)z$$

So

$$0 = (x+y)z$$
$$xy = 2xz + 2yz + (1/2)z$$

So $z = 0$ or $x = -y$. If $z = 0$, then x or y is zero, which by g_1 means all are zero. We already have that point. If $x = -y$, then $z = 0$ again.

Therefore, our points are $(0,0,0)$ and $(-1,-1,-2)$. We evaluate the function and find function values of 0 and -2, respectively, so -2 is the absolute maximum constrained to this set.