MA 1C RECITATION 04/17/13

1. Warmup: Another way to think about the gradient

Example 1. Find a differentiable function whose maximal directional derivative at $(3, 5)$ is equal to 1 in the direction $(1, 0)$.

This kind of problem, which you also have on your homework, is a preview of an important concept which we will study later.

Recall that the directional derivative in the direction $(1,0)$ is just $\nabla f(3,5)\cdot(1,0)$. Therefore, we need to find a function f such that

$$\frac{\partial f}{\partial x}(3,5) = 1. \tag{1}$$

Now we need to find something that satisfies this "maximal" condition. Now, for $(1, 0)$ to be the direction of maximal increase, $(1, 0)$ must be the vector that maximizes $\nabla f(3, 5) \cdot \mathbf{v}$. However, recall the following property of the dot product:

$$|\mathbf{a} \cdot \mathbf{b}| = ||\mathbf{a}||||\mathbf{b}|| \cos(\theta),$$

where θ is the angle between the vectors. Since $\cos(\theta)$ is maximized when $\cos(\theta) = 1$, we note that the dot product is maximized when $\theta = 0$, that is, when $\mathbf{a}$ and $\mathbf{b}$ lie on the same line. Thus, if we find a function such that

$$\nabla f(3,5) = (1,0) \tag{2}$$

we are done.

There are many functions that satisfy criteria (1) and (2). For instance, $f(x, y) = (x^3/3 - 5x) + (y - 5)^2$ is one possible solution.

2. Critical Points

We now return to the scalar field ($f : \mathbf{R}^n \to \mathbf{R}$) case. Last time we saw that if a point is a local extremum of a scalar function f, then the gradient of f is zero at that point.

However, the converse statement is not true. Consider $f(x, y) = x^2 - y^2$ at the point $(0, 0)$. Since we are always interested in finding extrema, even if we have to search for them among a given set of points, we give these points a name. A point where the gradient of f is zero is called a **critical point** (or **stationary point**) of f.

Why are these sometimes called "stationary points"? I think the definition makes the most sense if you think about the problem physically. Say f measures the temperature in a certain $\mathbf{R}^n$ space. Recalling our reasoning on the last problem from last week, we saw that the gradient points in the *direction of greatest change*, in other words, it will always point to the locally "hottest point." Suppose you take the gradient of a scalar field at point p, then move a tiny bit in the direction the gradient is pointing at p, then evaluate the gradient and repeat the process. If the temperature function is bounded, then by following this process, you will

Date: April 17, 2013.

eventually will arrive at the "hottest point," and since you are at the hottest point, no direction will point to a hotter point, and so the gradient will be zero, and you will remain stationary.

This thinking is slightly misleading in general, since it applies to "coldest points" as well, but it gives you an idea behind the terminology. Actually, for functions $f(x, y)$ in two variables, the stationary points correspond to peaks, pits, and saddle points. In Math 2a, you will study how points converges to these stationary points, which is a fascinating subject in itself.

To study critical points, we need to look at higher derivatives, much like how we studied maxima in the one-variable case by looking at the second-derivatives at that point. To do this, we introduce the following important object.

Definition 2.1. The **Hessian** matrix of a scalar field $f : \mathbf{R}^n \to \mathbf{R}$ is defined to be

$$H(\mathbf{x}) = [D_{ij} f(\mathbf{x})]_{i,j=1}^n$$

Remark 2.2. Note that

$$H(f)(x) = J(\nabla f)(x).$$

In particular, the Hessian matrix is symmetric, so there exists a basis of $\mathbf{R}^n$ consisting of the eigenvectors of H.

Observe that if $\mathbf{x}$ is an eigenvector of H, then the sign of the derivative in the direction of $\mathbf{x}$ is $\mathbf{x}H(\mathbf{a})\mathbf{x}^T$. Therefore, 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, we have three cases.

- (H is *negative definite*) If all the eigenvalues of H are negative, then f has a maximum.
- (H is *positive definite*) If all the eigenvalues of H are positive, then f has a minimum.
- If there are eigenvalues of both signs, then f has a saddle point.

Remark 2.3. Note that this is just a generalization of the second-derivative criteria for extrema in the one-variable case, i.e. if the function is concave down, then we have a local maximum; if concave up, then we have a local minimum. We just have an additional ambiguity here with the saddle point since we are in the multiple-variable setting.

Example 2. Find and classify the critical points of $f(x, y) = x^2 + y^2$.

Solution. We have $\nabla f = (2x, 2y)$. Therefore, the only critical point of f is $(0, 0)$. Computing the Hessian, we find that

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

Therefore, the eigenvalues of H_f are 2 and 2, so H_f is positive definite, implying that $(0, 0)$ is a local minimum. □

Critical points are subtle objects. They may not always behave like you'd imagine.

For instance, $f(x, y, z) = \sin(x^2 + y^2 + z^2)$ is an example that shows that critical points don't need to be isolated, since every point in the 3-sphere is a critical point of f. (Work it out!)

A function may also have no critical points whatsoever. Consider the (slightly modified) Gaussian function $f(x,y) = \int_x^y e^{-t^2}\,dt$. Then $\nabla f(x,y) = (-e^{-x^2}, e^{-y^2})$, which is never zero.

Warning 2.4. To find extrema for a function restricted to a region, you must check the critical points *and the boundary points*, just like in the single-variable case. This is because if a function is defined on a region, you may have global extrema that are not critical points.

2.1. **How to find extrema of multivariable functions.**

(a) Find and classify all critical points in the defined region and classify them as local/relative extrema or saddle points by looking at the Hessian at that point.
(b) Find the extrema on the boundary by evaluating the function at the boundary points, and see if any of these are global extrema (larger or smaller than your local extrema found above). Since the gradient may not be zero here, you don't have to worry about finding relative extrema or saddle points.
(c) Neatly and clearly describe all the local extrema, the saddles, and the global extrema on the interior of the region defined, as well as the global extrema that may lie on the boundary.

3. Lagrange Multipliers

Lagrange multipliers are a method to optimize a function $f(\overline{x})$ under some constraint $g(\overline{x}) = c$ for some constant c. It is a useful technique that is almost always faster doing things the "long way," that is, by finding the critical points, seeing which ones fit the constraint, and seeing which one is the maximum or minimum.

The reason why Lagrange multipliers work is a pretty clever application of the gradient and it's easy and satisfying to see how *geometric* these results are.

I'll state a commonly used version of the Lagrange multiplier theorem. More general versions are available, but I find that this relatively simple version helps illustrate most of the features of Lagrange multipliers.

Theorem 3.1. *Suppose that the constraint equation $g(\overline{x}) = c$ is nonsingular, that is, the function g is differentiable and $\nabla g \neq 0$ at all points in the set $\{g = c\}$. If $\overline{x}$ is a maximizing (or minimizing) input, then $\overline{x}$ also satisfies the equation*

$$(\nabla f)(\overline{x}) = \lambda(\nabla g)(\overline{x})$$

for some scalar λ.

In other words, for, say, the three-dimensional case, we can solve the system of equations

$$\begin{aligned} g(x,y,z) &= c \\ \frac{\partial f}{\partial x} &= \lambda \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} &= \lambda \frac{\partial g}{\partial y} \\ \frac{\partial f}{\partial z} &= \lambda \frac{\partial g}{\partial z} \end{aligned}$$

to find a list of f-inputs to check ("checkpoints"). Then we can plug these into f to get a set of f-outputs, and one of these checkpoints will be the maximum (or minimum) if it exists.

3.1. Algebra Tips for Lagrange Multipliers.

- **Solving for λ and equating the results** is usually the fastest way to start, but sometimes a clever trick will work faster.
- **Be careful when you divide by 0!** If you solve this (or any system of equations), where you have to divide by an expression, it is *critical* that you consider a *separate* case where that expression is zero. For instance, the equation $AB = AC$ implies that $B = C$ *or* $A = 0$ (or both). You will miss check points if you don't do this!
- **Finding "too many" checkpoints** is not a problem, as long as they all satisfy the constraint $g = c$. When we check the outputs, we can quickly rule out the points that are not extrema. However, you must not miss any potential checkpoints!

3.2. Presentation Tips.

- **Make a list of (or draw a box around) all of your "checkpoints"** once you complete your problem. In other words, make it absolutely clear all the f-inputs that you checked, along with their outputs. Lagrange multiplier calculation can get messy quickly, so don't lose points unnecessarily!
- **Clearly indicate the cases considered** (e.g. by underlining them) to make sure that the grader knows which cases you considered.

Example 3. What is maximum and minimum of $f(x, y) = x^2 + y^3$ given the constraint $g(x, y) = x^4 + y^6 = 2$.

Solution. Check that $g(x, y) = 2$ is a nonsingular constraint. By applying the method of Lagrange multipliers, we have the system of equations

$$\begin{aligned} x^4 + y^6 &= 2 \qquad (*) \\ 2x &= \lambda 4x^3 \\ 3y^2 &= \lambda 6y^5 \end{aligned}$$

We want to divide by x and by y to solve for λ, so we need to consider the cases where x or y is zero.

Case $x = 0$: In this case, $(*)$ tells us that $y^6 = 2$ and $y = \pm\sqrt[6]{2}$, so we have the checkpoints $(0, \pm\sqrt[6]{2})$.

Case $y = 0$: Here, $(*)$ says $x^4 = 2$, so $x = \pm\sqrt[4]{2}$ and so we have checkpoints $(\pm\sqrt[4]{2}, 0)$.

Case $x \neq 0$ and $y \neq 0$: Since x and y are both nonzero, we can divide by x and y to see that $\lambda = 1/(2x^2) = 1/(2y^3)$, so $x^2 = y^3$. Then $(*)$ says that $x^4 + x^4 = 2$, so $x = \pm 1$ and $y^3 = x^2 = 1$, so we have checkpoints $(\pm 1, 1)$.

Now that we have all the checkpoints, let's check their outputs under f:

$$\begin{aligned} f(0, \sqrt[6]{2}) &= \sqrt{2} \\ f(0, -\sqrt[6]{2}) &= -\sqrt{2} \qquad \text{(min)} \\ f(\pm\sqrt[4]{2}, 0) &= \sqrt{2} \\ f(\pm 1, 1) &= 2 \qquad \text{(max)}. \end{aligned}$$

□

Remark 3.2. Observe that if we didn't consider the cases of $x = 0$ and $y = 0$ separately, we would not have found the minimum.

We could also have solved the third case differently. When $x^2 = 1$, we could have found y by solving $x^4 + y^6 = 2$, so we'd get "extra" checkpoints $(\pm 1, -1)$. These don't satisfy the full Lagrange system of equations because $x^2 \neq y^3$, but since they do satisfy the constraint equation, the theorem says that the checking process will show that they are not extrema. *Moral:* Having extra checkpoints is OK, as long as they satisfy the original constraint and that you check them at the end.

We could also have solved the example differently. For instance, if we divided by $x - 1$, then we need to check the case where $x = 1$. This also applies to more general situations. For example, if we divide by $y - z^3$, then we need to separately check the case when $y = z^3$.