MA 8, WEEK 7: OPTIMIZATION, CRITICAL POINTS AND EXTREMA, CONVEX FUNCTIONS

1. Optimization

1.1. Critical points and local extrema.

Definition 1. Let f be a once continuously differentiable function on an interval I. We say that a point $a \in I$ is a *critical point* of f if $f'(a) = 0$

Critical points are important because they are the *only* points that can be local extrema. So let's first define what we mean by local extrema precisely.

Definition 2. We say that a is a local minimum (resp. local maximum) of f if there exists $\delta > 0$ such that for all $x \in (a-\delta, a+\delta)$, we have $f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$).

If a is either a local minimum or a local maximum, we say that a is a local extremum.

Proposition 3. *Let f be a once continuously differentiable function on an interval I. Suppose a is a local extremum of f. Then $f'(a) = 0$.*

Proof. We proceed by contradiction. We will only prove that when a is a local minimum of f, then $f'(a) = 0$. The case when a is a local maximum will follow by the symmetry of argument.

Suppose a is a local minimum of f and suppose that $f'(a) \neq 0$. Then, either $f'(a) > 0$ or $f'(a) < 0$. We can assume without loss of generality that $f'(a) > 0$.

Since f is once continuously differentiable, f' is a continuous function. In particular, we can find $\delta_1 > 0$ such that $f'(x) > 0$ for all $x \in (a-\delta_1, a+\delta_1)$.

Now, since a is a local minimum, we can find $\delta_2 > 0$ such that $f(x) \geq f(a)$ for all $x \in (a-\delta_2; a+\delta_2)$. Now, let $\delta = \min(\delta_1, \delta_2)$.

Since f is continuous and has a local minimum at a, we can find $b_1, b_2 \in (a-\delta, a+\delta)$ such that $f(b_1) = f(b_2)$. By the Mean Value Theorem, we can find $c \in (b_1, b_2) \subset (a-\delta, a+\delta)$ such that

$$f'(c) = \frac{f(b_2) - f(b-1)}{b_2 - b_1} = 0$$

However, we chose δ such that whenever $x \in (a-\delta, a+\delta)$, we have $f'(x) > 0$. Hence, existence of such c is a contradiction and thus, we conclude that $f'(a) = 0$. □

So we see that local extrema must be critical points of f. But is the converse also true? Unfortunately, the converse is not true in general. For example, consider $f(x) = x^3$. Then, $f'(0) = 0$ and thus, 0 is a critical point of f but we also know that x^3 is a monotone increasing function. Therefore, 0 cannot be a local extremum.

What is true, however, is that with a few extra conditions on f, we can check if a point a is a local extremum of f.

1.2. First and Second Derivative Tests.

Proposition 4 (First Derivative Test). *Let f be once continuously differentiable function on an interval I. Let $a \in I$ be a critical point of f.*

- *Suppose there exists $\delta > 0$ such that $f'(x) < 0$ for all $x \in (a-\delta, a)$ and $f'(x) > 0$ for all $x \in (a, a+\delta)$. Then, a is a local minimum of f*
- *Suppose there exists $\delta > 0$ such that $f'(x) > 0$ for all $x \in (a-\delta, a)$ and $f'(x) < 0$ for all $x \in (a, a+\delta)$. Then, a is a local maximum of f*

If f is twice continuously differentiable, this extra condition looks particularly nice.

Proposition 5 (Second Derivative Test). *Suppose f is a twice continuously differentiable function on I. Let $a \in I$ be a critical point of f.*

- *Suppose $f''(a) > 0$. Then, a is a local minimum of f.*
- *Suppose $f''(a) < 0$. Then, a is a local maximum of f.*

Example 6. Let's end this section with an optimization problem.

Suppose Caltech Bookstore sells 200 iPad Mini$^{\mathrm{TM}}$ with Retina Display$^{\mathrm{TM}}$ (hereafter referred to as just iPad) a week at \$350 each. A market survey indicates that for each \$10 rebate offered to buyers, the number of iPads sold per week will increase by 20 units.

Write the price and the revenue as functions of number of units sold per week. How large should the rebate be if Caltech Bookstore wanted to maximize revenue?

Solution. Let x be the number of iPads sold in a week. Then, the increase in sales by offering a rebate is $x - 200$. The market survey says that for each increment of 20 units sold per week, the rebate offered is \$10. Therefore, we can write the price function, $p(x)$ as

$$P(x) = 350 - \frac{10}{20}(x - 200) = 450 - \frac{1}{2}x$$

Now, revenue is simply the price multiplied by sales i.e.

$$R(x) = xP(x) = 450x - \frac{1}{2}x^2$$

To maximize $R(x)$ we need to find the critical points but it isn't hard at all.

$$R'(x) = 450 - x \quad \Rightarrow R'(450) = 0$$

Hence, $x = 450$ is the only critical point of $R(x)$. Noting that $R''(x) = -1 < 0$ for all x, we conclude by the second derivative test that $x = 450$ is indeed the local maximum of $R(x)$.

Since there are no more local extrema other than $x = 450$, it is in fact the global maximum. Therefore, to maximize revenue, Caltech Bookstore must offer a rebate which would cause an increase in sales of $450 - 200 = 250$. This corresponds to a rebate of $\frac{\$20}{10} \cdot 250 = \$125$! In other words, Caltech Bookstore should be selling iPad Mini's at $\$350 - \$125 = \$225$ after rebate! $\square$

2. Convex Functions

You might have seen convex and concave functions before from highschool math classes (some teachers say "concave up" and "concave down" for convex and concave). An example of a convex function is $f(x) = x^2$ and an example of a concave function is $g(x) = -x^2$. Intuitively, we think that convex functions can be minimized and concave functions can be maximized. Using the second derivative test, this would mean that convex functions should have nonnegative second derivative and concave functions should have nonpositive second derivative. Our intuition serves us well this time and we will see in a bit that $f(x)$ is convex (resp. concave) if and only if $f''(x) \geq 0$ (resp. $f''(x) \leq 0$).

But first, let's give a formal definition of convex/concave functions.

Definition 1. Let f be a function. f is called *convex* if for all $\theta \in [0,1]$ and for all x, y we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

f is called *concave* if for all $\theta \in [0,1]$ and for all x, y we have

$$f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta)f(y)$$

As we said in the beginning, we have the following proposition.

Proposition 2. *f is convex (resp. concave) if and only if $f''(x) \geq 0$ (resp. $f''(x) \leq 0$) for all x*

In this week's problem set, you will need to prove this!

Theorem 3 (Jensen's Inequality)**.** *Let f be a convex function and let $a_1, \ldots, a_n > 0$. Then, for all $x_1, \ldots, x_n$ we have*

$$f\left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{j=1}^n a_j}\right) \leq \frac{\sum_{i=1}^n a_i f(x_i)}{\sum_{j=1}^n a_j}$$

If g is a concave function, then we have

$$g\left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{j=1}^n a_j}\right) \geq \frac{\sum_{i=1}^n a_i g(x_i)}{\sum_{j=1}^n a_j}$$

You should think of Jensen's inequality as an extension of the definition of convexity/concavity of a function to more than 2 points. In short, it says that if you have a convex function, the function's value at the weighted average of x_i's is at most the weighted average of $f(x_i)$'s. And for concave functions, Jensen's inequality says that the function's value at the weighted average of x_i's is at least the weighted average of $f(x_i)$'s.

Using Jensen's inequality, we can prove the arithmetic mean-geometric mean (a.k.a. AM-GM) inequality. So let's first state AM-GM inequality!

Theorem 4 (AM-GM Inequality)**.** *Let $a, b > 0$. Then,*

$$\frac{a+b}{2} \geq \sqrt{ab}$$

Proof. We could prove this using Jensen's inequality but that would be *very* inefficient. Instead, we will give the shortest possible proof using basic arithmetic.

We first have $0 \leq (a-b)^2 = a^2 - 2ab + b^2$. Therefore, $4ab \leq a^2 + 2ab + b^2 = (a+b)^2$. Finally, putting square roots on both sides we get $2\sqrt{ab} \leq a + b$ and this gives the desired inequality. $\square$

That was easy! And we didn't need to use Jensen's inequality. Let's try to prove the generalized AM-GM inequality now.

Theorem 5 (Generalized AM-GM Inequality)**.** *Let $x_1, \ldots, x_n > 0$. Then,*

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}$$

Proof. This time, we will use Jensen's inequality to attack this problem. Our choice of weapon (i.e. convex/concave function)? Let's take $f(x) = \log x$.

First, we claim that $\log x$ is a concave function that is strictly increasing on its domain of definition, which is $(0, +\infty)$. First, we have $(\log x)' = \frac{1}{x} > 0$ for all $x \in (0 < +\infty)$ so $\log x$ is monotone increasing. Differentiating $\log x$ another time, we get

$$(\log x)'' = ((\log x)')' = \left(\frac{1}{x}\right)' = -\frac{1}{x^2} < 0 \text{ for all } x \in (0, +\infty)$$

By Proposition **??** we see that $\log x$ is a concave function.

Now, let $x_1, \ldots, x_n > 0$. Then, by Jensen's inequality for concave functions (where we let $a_1 = \cdots a_n = 1$), we get

$$\begin{aligned} \log\left(\frac{x_1 + \cdots + x_n}{n}\right) \quad \geq \quad \frac{\log x_1 + \cdots + \log x_n}{n} \quad &= \quad \tfrac{1}{n}\log(x_1 \cdots x_n) \\ &= \quad \log\left((x_1 \cdots x_n)^{\frac{1}{n}}\right) \end{aligned}$$

But we showed that $\log x$ is monotone increasing on $(0, +\infty)$. Therefore, above inequality implies

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}$$

which is precisely the generalized AM-GM inequality. □

3. More on Convex Functions

We briefly introduced concave and convex functions last time. However, in most applications of mathematics, convex functions are much more important and are the standard way to approach applications, so it's worth it to become more familiar with this notion. The only exception that I know of is economics, since concavity there is a manifestation of the law of diminishing marginal returns. Note that the theories of of concave and convex functions are more or less equivalent, because of *a function f is concave if and only if $-f$ is convex.*

Let's briefly recall the definition of a convex function.

Definition 1. We say that a function $f : (a, b) \to \mathbf{R}$ is **convex** if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

whenever $a < x < b, a < y < b$, and $0 < \lambda < 1$. We say f is **strictly convex** if the inequality above is strict.

The difference, intuitively, is that convex functions can have linear parts, while strictly convex functions do not.

Remark 2. This definition is equivalent to the definition given on Tuesday, because if $\lambda = 0$ or 1, then the inequality is always satisfied.

These two definitions are also equivalent to the one given in class (that we'll go over later) and it's a good exercise to work out that the two notions are equivalent.

If a function is twice differentiable, the easiest way to characterize convex functions is via the following criterion that you have to prove on your homework.

Proposition 3. *A function f is convex if and only if $f''(x) \geq 0$ for all x.*

From this, we have some obvious examples of convex functions, like

- $f(x) = x^n$ where n is even.
- $f(x) = e^x$, arguably the most important example.

We call such functions *convex* because the region "above them" on a graph forms a **convex set**. Recall that a subset S of Euclidean space (that, is $\mathbf{R}^n$ for some n) is **convex** if for every pair of points in the set, the straight line segment connecting those two points is also contained in S.

Examples of convex and non-convex sets.

Thus, we see in the case of x^2 that we clearly get a convex set.

Why is $f(x) = x^3$ not convex?

Actually you have to be careful. Just like differentiability, convexity is relative term. Namely, $f(x)$ is convex on $x \geq 0$ and concave on $x \leq 0$. You can see this by looking at the second derivative.

The **inflection point**, that is, places where $f''(x) = 0$ and the sign of $f''(x)$ changes are where the function changes from convex to concave or vice versa.

Something like $\sin(x)$ is not a convex, and one way to see this is to look at the following criterion for distinguishing convex functions, which was given in class.

Proposition 4. *A function f is convex if and only if the line segment between any two points on the graph of a function lie above the graph.*

*This line segment connecting these two points is called the **chord**. If the line is extended infinitely in both directions, it is called a **secant line**. Technically, Prof. Katz should have given his definition in class in terms of chords rather than secant lines, but you can easily see what he meant.*

One reason for the importance of convex functions is their relation to optimization problems. For instance, a strictly convex function on an open set (that is, an open interval) has at most one minimum, so any local minimum must be a global minimum. Indeed, there's a whole branch of math called **convex optimization** which deals with minimizing convex functions over convex sets. A surprising number of problems can be phrased in this way.

There are also lots of industry software that allows you to solve such problems quickly.

Remark 5. This is most relevant for functions $f : \mathbf{R}^n \to \mathbf{R}$, but here in order for a function to be **convex** not only does the inequality have to be satisfied, but the set on which you are defining your function must be a convex set.

It is also important because convexity is a good way to understand exponential growth, which underlies, of course, much of the "growth" phenomenon in the real world. As we saw above e^x is a convex function.

Roughly speaking, the distinction between exponential growth and convex growth is the exponential growth means "increasing at a rate proportional to the current value" whereas convex growth means "increasing at an increasing rate (but not necessarily proportionally to current value." We'll see a precise formulation of this result below.

Here are some basic operations we can do with convex functions.

- If f is convex and $c > 0$ is a constant, then the function cf is convex.
- If f and g are convex, then $f + g$ is convex.

- (This is on your homework, in the language of concave functions) If $f(u)$ is convex and increasing and $u(x)$ is convex, then $(f \circ u)(x)$ is convex.

We'll begin with some easy properties about convex functions.

Proposition 6. *If $f(x)$ is convex on $[a, b]$ and is not constant, then it cannot have a local maximum on (a, b).*

Proof. Suppose that $x_0 \in (a, b)$ is a local maximum. Then there exists points $x_1, x_2 \in (a, b)$ such that $x_1 < x_0 < x_2$ and so

$$x_0 = \theta_1 x_1 + \theta_2 x_2$$

for some $\theta_1, \theta_2 > 0$ and $\theta_1 + \theta_2 = 1$. Moreover, $f(x_1) \leq f(x_0)$ and $f(x_2) \leq f(x_0)$, with at least one of the two inequalities being strict. Multiply the first inequality by θ_1 and the second inequality by θ_2 and then add them, to obtain

$$f(x_0) > \theta_1 f(x_1) + \theta_2 f(x_2)$$

which contradicts the convexity of f. □

Proposition 7. *Suppose that f is a function that is convex on $x \in [a, b]$. Fix a points $x_1, x_2 \in [a, b]$ with $x_1 < x_2$. Then the defining equality*

$$f(\theta_1 x_1 + \theta_2 x_2) \leq \theta_1 f(x_1) + \theta_2 f(x_2)$$

for the above x_1 and x_2 and for all $x \in (x_1, x_2)$ is either always an equality or is always strict.

Proof. Consider the function

$$g(x) = f(x) - \ell(x) - f(x) + (-\ell(x))$$

for $x \in [x_1, x_2]$, where $\ell(x)$ is chord, that is,

$$\ell(x) = \frac{x_2 - x}{x_2 - x_1} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2).$$

Then g is the sum of two convex functions and so must be convex. Furthermore, $g(x_1) = g(x_2) = 0$. By the previous proposition, $g(x)$ is either a constant or it cannot have a local minimum for $x \in (x_1, x_2)$, which corresponds to our two results. □

Lemma 8. *If f is convex in (a, b) and if $a < s < t < u < b$, then*

$$\frac{f(t) - f(s)}{t - s} \leq \frac{f(u) - f(s)}{u - s} \leq \frac{f(u) - f(t)}{u - t}.$$

Intuitively, this says that convex functions increase at an increasing rate.

Proof. Let $\lambda = \frac{t-s}{u-s}$, noting that $0 < \lambda < 1$. Then $1 - \lambda = \frac{u-t}{u-s}$ and so

$$t = \lambda u + (1 - \lambda)s.$$

Therefore,

$$f(t) = f(\lambda u + (1 - \lambda)s) \leq \lambda f(u) + (1 - \lambda) f(s)$$

by convexity. Plugging in our value for λ, we obtain

$$f(t) \leq \frac{t - s}{u - s} f(u) + \frac{u - t}{u - s} f(s),$$

that is,

$$(u - s) f(t) \leq (t - s) f(u) + (u - t) f(s). \tag{$*$}$$

From here, we just need to do a little algebra:

$$(u-s)f(t)-(u-s)f(s) \le (t-s)f(u)+(u-t)f(s)-(u-s)f(s)$$
$$(u-s)(f(t)-f(s)) \le (t-s)(f(u)-f(s))$$

and so we obtain our first inequality

$$\frac{f(t)-f(s)}{t-s} \le \frac{f(u)-f(s)}{u-s}.$$

Now, let's go back to $(*)$ and multiply everything by -1, obtaining

$$-(t-s)f(u)-(u-t)f(s) \le -(u-s)f(t).$$

Once again, we do some algebra:

$$(u-s)f(u)-(t-s)f(u)-(u-t)f(s) \le (u-s)f(u)-(u-s)f(t)$$
$$(u-t)(f(u)-f(s)) \le (u-s)(f(u)-f(t))$$

and so

$$\frac{f(u)-f(s)}{u-s} \le \frac{f(u)-f(t)}{u-t}$$

as desired. □

Proposition 9. *Every convex function $f : (a,b) \to \boldsymbol{R}$ is continuous.*

Proof. Consider real numbers $a < s < x < t < y < u < b$. By the lemma above, we have

$$m := \frac{f(t)-f(s)}{t-s} \le \frac{f(t)-f(x)}{t-x} \le \frac{f(y)-f(t)}{y-t} \le \frac{f(u)-f(t)}{u-t} =: M.$$

Since $m \le \frac{f(t)-f(x)}{t-x} \le M$ and $m \le \frac{f(y)-f(t)}{y-t} \le M$, we have

$$(t-x)m \le f(t)-f(x) \le (t-x)M \text{ and } (y-t)m \le f(y)-f(t) \le (y-t)M.$$

Therefore, we have

$$\lim_{x\to t^-}(t-x)m \le \lim_{x\to t^-}[f(t)-f(x)] \le \lim_{x\to t^-}(t-x)M$$
$$\lim_{y\to t^+}(y-t)m \le \lim_{y\to t^+}[f(y)-f(t)] \le \lim_{y\to t^+}(y-t)M.$$

Thus,

$$0 \le \lim_{x\to t^-}[f(t)-f(x)] \le 0 \text{ and } 0 \le \lim_{y\to t^+}[f(y)-f(t)] \le 0$$

and so

$$\lim_{x\to t^-} f(x) = f(t) = \lim_{y\to t^+} f(y).$$

Hence, $\lim_{x\to t} f(x) = f(t)$. Since t was arbitrary in (a,b), we conclude that f is continuous on (a,b). □

Here's a more intuitive proof that uses the "geometric" definition.

Proof. Let $\epsilon > 0$. Let a be a point of the interval. Pick some $x_0 > a$. Notice that no matter what value $f(x_0)$ may have, the line segment between $(a, f(a))$ and $(x_0, f(x_0))$ eventually lies below the horizontal line at height $f(a)+\epsilon$. Since the graph of f must lie below this line on the interval (a, x_0), this shows that

$$f(x) < f(a)+\epsilon$$

for all $x > a$ sufficiently close to a. A similar argument works for all $x < a$ sufficiently close to a.

It remains to show that $f(x) > f(a) - \epsilon$ for all x sufficiently close to a. If $f(x) \geq f(a)$, there is nothing to prove, so suppose that $f(x_0) < f(a)$ for some x_0 with $x_0 > a$, say. Then we must have $f(y) > g(a)$ for all $y < a$ by convexity, so all $y < a$ must satisfy $f(y) > f(a) - \epsilon$. Moreover, if we pick some $y_0 < a$, then the line segment between $(y_0, f(y_0))$ and $(a, f(a))$ lies above the horizontal line at height $f(a) - \epsilon$ in some interval to the right of a. Since the graph of f must lie *above* this line to the right of a, it follows that $f(x) > f(a) - \epsilon$ for all $x > a$ sufficiently close to a.

$\square$

Example 10. Here's an example of a discontinuous convex function from a *closed* interval to $\mathbf{R}$:

With some work, you can show that these are the only discontinuities that can arise.

Proposition 11. *Suppose that f is a continuous real function defined in (a, b) such that*

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$$

for all $x, y \in (a, b)$*. Prove that* f *is convex.*

Remark 12. Note that the converse is an immediate corollary of the definition of convex function.

Proof. Let $0 < t < 1$. For any $\epsilon > 0$, there is a number k of the form $m/2^n$ which is so close to t that

$$|f(kx + (1-k)y) - f(tx + (1-t)y)| < \epsilon$$
$$|[kf(x) + (1-k)f(y)] - [tf(x) + (1-t)f(y)]| < \epsilon.$$

Then

$$\begin{aligned} f(tx + (1-t)y) &< f(kx + (1-k)y) + \epsilon \\ &< kf(x) + (1-k)f(y) + \epsilon \\ &< tf(x) + (1 - tf(y)) + 2\epsilon. \end{aligned}$$

Thus $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$. The following diagram shows that if strict inequality holds for even one t, then it holds for all t (by applying the weak inequality to x and $tx + (1-t)y$ or to $tx + (1-t)y$ and y). But we have strict inequality for t of the form $m/2^n$, so we must have strict inequality for all t.

$\square$