15.084J Recitation Handout 1

First Week in a Nutshell:

- The Basic Problem
- Basic Definitions
- Weirstrass Theorems
- Necessary and Sufficient Conditions for Optimality

The Basic Problem:

- $min f(x)$ subject to $x \in S, g(x) \leq 0, h(x) = 0$
- Variations can be put into this form; maximize rather than minimize, other direction inequalities, etc.
- If there is no $g(x)$ or $h(x)$, problem is 'unconstrained'
- This is very general, and applies to a lot of problems

Definitions: The exact definitions are in the lecture notes; I'm going to instead give the intuitive definitions, to help build up your understanding.

- Feasible Point: one that satisfies $x \in S, g(x) \leq 0, h(h) = 0$
- Local Minimum: intuitively, no feasible point in a small neighborhood has lower objective value
- Global Minimum: no feasible point has a lower objective value
- Strict (local or global) Minimum: no feasible point (in a small neighborhood for local) has the same or lower objective value
- Maxima have the same definitions as minima, but with lower replaced by higher
- Gradient: If $f(x)$ is differentiable, the gradient $\nabla f(x)$ corresponds to the first derivative for single variable functions; it gives the rate of change of the function in any direction: $\nabla f(\bar{x})^t d$ is the rate of change in the direction d, at the point $\bar{x}$.
- Hessian: If $f(x)$ is twice differentiable, the hessian corresponds to the second derivative for single variable functions; it gives the rate of change of the gradient in any direction.
- Symmetric Positive (Semi-)Definite Matrix: M is SPSD if $x^t M x > (\geq) 0$ for all x. We often just assume symmetric.
- Negative (Semi-)Definite Matrix: As above, but $< (\leq)$
- Convex Set: For any two points in the set, the line between them is entirely within the set
- Closed Set: The set contains its boundaries
- Compact Set: A set that is closed and bounded
- Convex Function: A function that always curves upward – it is entirely a bove a linear approximation at any point

- Descent Direction: A direction d at a point x where moving small distances from x in direction d lowers the objective value

Wierstrass Theorem

If a set F is compact, and x_i is a sequence within F, then there exists a convergent subsequence converging to a point in F.

Intuitively, if you have infinitely many points in F, then they have to be clustering somewhere (F is bounded, so they can't be marching off to infinity). That somewhere must be in F, because it is closed. Thus, some of the x_i have to be converging to a point in F.

If a function $f(x)$ is continuous over a compact set F, then there exists a point in F that minimizes $f(x)$ over F.

Intuitively, pick a point $x_1 \in F$. If $f(x_1)$ minimizes $f(x)$ over F, we are done. Otherwise, since F is continuous, there must be another point $x_2 \in F$ with $f(x_2) = \frac{1}{2}(f(x_1) + inf_{x\in F} f(x))$. Repeat. You now have an infinite sequence of points, which converges to something in F by the above, and by continuity of f, their objective values converge to $inf_{x\in F} f(x)$.

This tells us that optimizing over a compact set, using a continuous objective function, there exists an optimum, so we will never be trapped in a situation where no optimal value exists.

Necessary Conditions for Optimality (in the Unconstrained Case)

If $\bar{x}$ is a local minimum, then clearly there are no descent directions at $\bar{x}$, because if there were, you could clearly do better. Since the gradient gives the rate of change of the objective, it must therefore be zero, otherwise moving opposite the direction of the gradient gives you a descent direction.

Thus, if f is differentiable, $\nabla f(\bar{x}) = 0$ is necessary for $\bar{x}$ to be a local minimum. If f isn't differentiable, life is more complicated, but you clearly still need to have no descent directions. If there are constraints, those will also complicate matters.

If the above is satisfied, there is still the possibility that you have a maximum, or what in one dimension would be a point of inflection... Thus, just as in one dimension we look for second derivative non-negative, we here look for the Hessian to be positive semidefinite. If it isn't, then there is some direction in which the gradient is zero but turning negative, so going in that direction will make the objective go down.

Thus, if f is differentiable, $H(\bar{x})$ must be SPSD in order for $\bar{x}$ to be a local minimum. Or looking at the maximization problem, SNSD for a local maximum.

Sufficient Conditions for Optimality (in the Unconstrained Case)

In order to be a strict local minimum, it has to be the case that moving a little bit in any direction increases the objective value. We argued above that the gradient must be zero, so to first order approximation, moving in any direction does nothing. But to a second order approximation, the objective will change as you move ϵ in a direction d by $\frac{\epsilon^2}{2} d^t H(\bar{x}) d$. Thus, if $H(\bar{x})$ is positive definite, you are increasing the objective in any direction.

Thus, if at $\bar{x}$ the gradient is zero and the hessian is positive definite, then $\bar{x}$ is a local minimum.

Alternately, we know that if $f(x)$ is convex, then $H(x)$ is everywhere positive definite, so any point where the gradient is zero is a local minimum, and further, if you had two local minimums, you could consider what the gradient does along the line between them and get a contradiction, so if $f(x)$ is convex, it has at most one local minimum, which is also a global minimum.