15.084J Recitation Handout 6

Sixth Week in a Nutshell:

- Penalty/Barrier Methods
- Quiz Review

Penalty Methods

How should we solve constrained optimization problems, given what we know of the unconstrained case?
Idea: Add a penalty to the infeasible region, and then solve as if you were unconstrained!
What should the penalty look like?

- It should clearly be zero in the feasible region – you don't want to affect what is going on in feasible cases.
- It should start at zero at the edge, so that being "a little bit" infeasible is only "a little bit" bad.
- It should be continuous and continuously differentiable. We like continuous and differentiable functions.
- Ideally it should have derivative/gradient zero at the boundary, since that make the objective plus the penalty continuously differentiable, assuming the objective is.

For any given (continuous, differentiable, convex) constraint function $g_i(x)$, $c_i(\max(g_i(x), 0))^2$ (with any positive constant c_i is good:

- It clearly satisfies "zero in feasible region, positive elsewhere"
- It is continuous, convex, differentiable (via the chain rule), and continuously differentiable if g_i is.
- Under weak assumptions on g_i, it has gradient zero at the border.

How to use this?

Solve the unconstrained problem min $f(x) + cg(x)$

If you get a solution in the interior of the original feasible region, clearly you win. Otherwise, increase c. As c approaches infinity, all infeasible points have values approaching infinity, so your minimum had better approach something feasible.

Convergence

Why is it good?

At each iteration, you have a point at least as good as the constrained maximum (since the constrained max is feasible and has the same objective in the new program).

At each iteration, you have a point with a higher objective than the previous iteration.

Thus, you've got an increasing sequence, bounded above, so it goes to a limit; that limit can't be infeasible, since any infeasible point would have an objective trending to infinity, so it converges to a point in the feasible region. Thus, it converges to the constrained optimum.

Also, at each step your optimum x^k satisfies the unconstrained necessary conditions: $\nabla f(x^k) + \sum \nabla(c_i(\max 0, g_i(x^k)^2)) = 0$. For any satisfied constraint, the gradient for that constraint is zero, and for any unsatisfied constraint, by the chain rule, it is $2g_i(x^k)\nabla g_i(x^k)$. Now, consider what this does in the limit: for any constraint not tight in the limit, after some point in the sequence, g_i is zero, so $2c_i g_i(x^k)$ is zero, so the term drops from the limit... For any tight constraint, $2c_i g_i(x)$ is nonnegative (since all three parts are nonnegative). Thus, this term tends to some non-negative limit u_i. Thus, in the limit, you are approaching a situation where $\nabla f(\bar{x}) + \sum_{\text{tight}} u_i \nabla g_i(\bar{x}) = 0$ and $u_i \geq 0$, which is exactly the KKT conditions!

Quiz

Thursday, in class. Closed book. 1.5 hours long.

Things to know:

- Unconstrained optimality: necessary, sufficient.
- Be careful of SPSD versus SPD.
- Matrix stuff: eigenvectors, eigenvalues, factoring lemmas.
- Unconstrained algorithms: newton, steepest descent.
- Rate of convergence, line search
- Constrained optimality: FJ versus KKT
- When is KKT necessary (slater, linearity, convexity)?
- Projection Methods: Why? When do we use them?
- Penalty Methods: good penalty functions, when do we use them