UNIVERSITY OF CALIFORNIA, BERKELEY

Lecture 16 - Optimality Conditions

Oscar Ortega

July 16, 2021

1 REVIEW KKT CONDITIONS

Given an optimization problem of the following form,

$$\min f_0(x)$$

$$\text{s.t } f_i(x) \leq 0$$

$$h_i(x) = 0$$

The **KKT Conditions** are as follows:

1. $f_i(x) \leq 0, i = 1, ..., n$
2. $h_i(x) = 0, i = 1, ..., m$
3. $\lambda_i \geq 0, i = 1, ..., n$
4. $\lambda_i f_i(x) = 0$
5. $\nabla \mathscr{L}(x^*, \nu^*, \lambda^*)_x = 0$

Also recall the following:

1. If $\tilde{x}, \lambda, \tilde{\nu}$ are optimal pts. and strong duality holds → KKT conditions will be satisfied
2. If $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT **and problem is convex** → then $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ are optimal pts.

2 OPTIMALITY CONDITIONS

Theorom:
Consider once again the following Optimization Problem:

$$\min_{x\in\mathscr{X}} f_0(x)$$

and assume $f_0(x)$ is convex and differentiable, and the set $\mathscr{X}$ is convex. then,

$$x \in \mathscr{X} \text{ is optimal } \rightarrow \nabla f_0(x)^T(y-x) \geq 0, \forall y \in \mathscr{X}$$

Proof:→ Recall that for any $x, y \in \mathbf{dom} f_0$ the first order conditions tell us the following,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y-x)$$

this along with the fact that the statement holds for all $x, y \in \mathbf{dom} f$ tells us that $f_0(y) \geq f_0(x)$, which implies optimality.
Proof: ← If $\nabla f_0(x) = 0$ the statement holds true trivially. Consider the case where the $\nabla f_0(x) \neq 0$. For the sake of contradiction, let $\exists y_0, \nabla f_0(x^*)(y_0 - x^*) < 0$ Let $g(t) = f_0(ty_0 + (1-t)x^*)$

$$\nabla g(0) = \nabla f_0(x^*)(y_0 - x)$$

$$g(t_0) = f_0(t_0 y_0 + (1-t_0)x^*) \leq f_0(x^*) + f_0(x^*)(y_0 - x) \leq f_0(x^*)$$

By using the first order approximation of the function. This contradicts the optimality of x^*.

3 GRADIENT DESCENT AND ARMIJO'S CONDITION

Consider now the unconstrained optimization problem:

$$p^* = \min f_0(x)_{x\in\mathbf{R}^n}$$

Recall Taylor's theorom:

$$f_0(x+\Delta) = f(x) + \nabla f_0(x)^T(\Delta)$$

And reparametrize as follows:

$$f_0(x+sv)_{s\geq 0} = f(x) + \nabla f_0(x)^T sv$$

$$f_0(x+sv)_{s\geq 0} = f(x) + \langle \nabla f_0(x), sv\rangle$$

Notice the following:

- $\langle \nabla f_0(x), sv\rangle \geq 0 \rightarrow f(x+sv) \geq f(x)$
- $\langle \nabla f_0(x), sv\rangle \geq 0 \rightarrow f(x+sv) < f(x)$

By Cauchy-Schwarz:

$$\langle \nabla f_0(x), v\rangle \leq \|f_0(x)\|_2 \|v\|_2$$

So, if we want to minimize this convex function, this means we want to find the direction of steepest descent. this is true when we set v to the following:

$$v = -\frac{\nabla f_0(x)}{\|\nabla f_0(x)\|_2}$$

This gives rise to our gradient descent update rule.

$$x^{k+1} = x^k - s\nabla f_0(x)$$

how do we choose the learning rate s?

3.1 General Descent Algorithms

In general, a descent algorithm will consist of the following.
Given a starting point x:

- Determine a descent direction ∇x
- Line Search, choose a step size $t > 0$
- Update. $x^{k+1} = x^k + t\nabla x$
- go to top, unless a stopping criterion is satisfied.

3.2 Exact Line Search

Choose $t = \text{argmin}_{s\geq 0} f(x + s\delta x)$ Repeat until convergence.
This is used when the cost of computing the objective function is lower than computation of the gradient.

3.3 Armijo Condition

Consider $\alpha \in (0,1)$ The Armijo condition for convergence is the following:

$$f_0(x_k + sv_k) \leq f_0(x_k) + s\alpha f_0(x_k)^T v_k$$

Consider, $\phi(s) = f_0(x_k + sv_k)$, $\delta_k = f_0(x)^T v_k$

$$\begin{aligned}\phi(s) &\approx f_0(x_k) + s\delta_k \\ &= \phi(0) + s\delta_k \\ \bar{l}(s) &= \phi(0) + s\alpha\delta_k\end{aligned}$$

- Note that $\bar{l}(s) \geq l(s)$

- this implies the Armijo condition is that we want it to be the case that the approximation $\bar{l}(s)$ also gives decrease.

Convergence Theorom: Given a differentiable function with a stepsize that satisfies the Armijo Condition for a fixed alpha. Then the gradient will converge to a point where

$$\lim_{k\to\infty} \| f_0(x_k) \| = 0$$

.