2.098/6.255/15.093 - Recitation 10

Michael Frankovich and Shubham Gupta

December 4, 2009

1 Karush Kuhn Tucker Necessary Conditions

$$\begin{array}{llrcll} \text{P:} & \min & f(x) & & & \\ & \text{s.t.} & g_j(x) & \leq & 0, & j = 1, \ldots, p \\ & & h_i(x) & = & 0, & i = 1, \ldots, m \end{array}$$

Theorem. (KKT Necessary Conditions for Optimality)
If $\hat{x}$ is local minimum of P and the following *Constraint Qualification Condition (CQC)* holds:

- The vectors $\nabla g_j(\hat{x}),\ j \in \mathcal{I}(\hat{x})$ and $\nabla h_i(\hat{x}),\ i = 1, \ldots, m$, are linearly independent, where $\mathcal{I}(\hat{x}) = \{j\ : g_j(\hat{x}) = 0\}$ is the set of indices corresponding to active constraints at $\hat{x}$.

Then, there exist vectors (u, v) s.t.:

1. $\nabla f(\hat{x}) + \sum_{j=1}^{p} u_j \nabla g_j(\hat{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\hat{x}) = 0$
2. $u_j \geq 0,\ j = 1, \ldots, p$
3. $u_j g_j(\hat{x}) = 0,\ j = 1, \ldots, p$ (or equivalently $g_j(\hat{x}) < 0 \Rightarrow u_j = 0,\ j = 1, \ldots, p$)

Theorem. (KKT + Slater)
If $\hat{x}$ is local minimum of P and the following *Slater Condition* holds:

- There exists some feasible solution $\bar{x}$ such that $g_j(\bar{x}) < 0,\ \forall j \in \mathcal{I}(\hat{x})$, where $\mathcal{I}(\hat{x}) = \{j\ : g_j(\hat{x}) = 0\}$ is the set of indices corresponding to active constraints at $\hat{x}$.

Then, there exist vectors (u, v) s.t.:

1. $\nabla f(\hat{x}) + \sum_{j=1}^{p} u_j \nabla g_j(\hat{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\hat{x}) = 0$
2. $u_j \geq 0,\ j = 1, \ldots, p$

3. $u_j g_j(\hat{x}) = 0,\ j = 1, \ldots, p$ (or equivalently $g_j(\hat{x}) < 0 \Rightarrow u_j = 0,\ j = 1, \ldots, p$)

Example.
Solve

$$\begin{array}{lll} \min & x_1^2 + x_2^2 + x_3^2 & \\ \text{s.t.} & x_1 + x_2 + x_3 & \leq & -18 \end{array}$$

Solution.
For some x to be a local minimum, condition (1) requires that $\exists u$ s.t. $2x_i + u = 0,\ i = 1, 2, 3$.

Now, the constraint can either be active or inactive:

- If it is inactive, then $u < 0$ by condition (3). This would imply $x_1 = x_2 = x_3 = 0$, but $x = (0, 0, 0)^\top$ is infeasible, so this cannot be true of a local minimum.
- If it is active, then $x_1 + x_2 + x_3 = -18$ and $2x_i + u = 0,\ i = 1, 2, 3$. This is a system of four linear equations in four unknowns. We solve and obtain $u = 12,\ x = (-6, -6, -6)^\top$. Since $u = 12 \geq 0$, there exists a u as desired. To check the regularity requirement, we simply confirm that $\nabla x = (1, 1, 1)^\top \neq 0$. Also, we could have checked that the Slater condition is satisfied (eg use $\bar{x} = (-10, -10, -10)^\top$).

Hence $(-6, -6, -6)^\top$ is the only candidate for a local minimum. Now, the question is: is it a local minimum? (Note that since this is the unique candidate, this is the same as asking if the function has a local minimum over the set.)

Observe that the objective function is convex. Why? Because it is a positive combination of convex functions. Now, is the feasible set convex? Answer: yes, since it is of the form $\{x \in \mathbb{R}^n : f(x) \leq 0\}$, where f is a convex function.

So we may apply a stronger version of the KKT condtions, the KKT *sufficient* conditions, which imply that any x which satisfies the KKT necessary conditions and also meets these two convexity requirements is in fact a global minimum.

So the point $x = (-6, -6, -6)^\top$ is the unique global minimum (unique since it was the only candidate).

Example.[1]
Solve

$$\begin{array}{llll} \min & -\log\,(x_1 + 1) - x_2 & & \\ \text{s.t.} & g(x) \triangleq 2x_1 + x_2 - 3 & \leq & 0 \\ & x_1, x_2 & \geq & 0 \end{array}$$

[1]Thanks to Andy Sun.

Solution.
Firstly, observe that this is a convex optimization problem, since $f(x)$ is convex (a positive combination of the convex functions $-x_2$ and $-\log\ (x_1+1)$), and the constraint functions $g(x)$, $-x_1$ and $-x_2$ are convex (again, this is required for the feasible space to be convex, since we have $\leq$ constraints). Alternatively, in this case we can see that the feasible space is a polyhedron, which we know to be convex.

Now, in order to use KKT, we need to assume which inequalities are active. Let's start by assuming that at a local minimum x, only $g(x) \leq 0$ is active. This leads to the system: $\begin{bmatrix} \frac{-1}{x_1+1} \\ -1 \end{bmatrix} + u \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 0$, which gives $u = 1$ and $x_1 = -0.5$, which is not feasible, so our assumption cannot be correct.

Now try assuming $g(x) \leq 0$ and $-x_1 \leq 0$ are active, giving the system $\begin{bmatrix} \frac{-1}{x_1+1} \\ -1 \end{bmatrix} + u_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + u_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = 0$, which gives $u_1 = 1, u_2 = 1$ and $x_2 = 3$ (recall we assumed $x_1 = 0$).

Now since it's a convex optimization problem, and the Slater condition is trivially satisfied, this is a global minimum by the KKT sufficient conditions.

Example.
The following example shows that the KKT theorem may not hold if the regularity condition is violated: Consider

$$\begin{array}{rrcl} \min & x_1 + x_2 & & \\ \text{s.t.} & (x_1-1)^2 + x_2^2 - 1 & = & 0 \\ & (x_1-3)^2 + x_2^2 - 9 & = & 0 \end{array}$$

The feasible region is the intersection of two circles, one centered at the point $(1,0)$ with radius 1, the other at the point $(3,0)$ with radius 3. The intersection occurs at the origin, which is the optimal solution by inspection.

We have $\nabla f(\hat{x}) = (1,1)^\top$, $\nabla h_1(\hat{x}) = (2x_1 - 2, 2x_2)^\top = (-2,0)^\top$, and $\nabla h_1(\hat{x}) = (2x_1 - 6, 2x_2)^\top = (-6,0)^\top$. So condition (1) cannot hold.

2 Conjugate Gradient Method[2]

Consider minimizing the quadratic function $f(x) = \frac{1}{2}x^\top Qx + c^\top x$.

1. $d_1, d_2, \ldots, d_m$ are **Q-conjugate** if

$$d_i^\top Q d_j = 0,\ \forall i \neq j$$

[2]Thanks to Allison Chang for notes

2. Let x_0 be our initial point.

3. Direction $d_1 = -\nabla f(x_0)$.

4. Direction $d_{k+1} = -\nabla f(x_{k+1}) + \lambda_k d_k$, where $\lambda_k = \frac{\nabla f(x_{k+1})^\top d_k}{d_k^\top Q d_k}$ in the quadratic case (and $\lambda_k = \frac{||\nabla f(x_{k+1})||^2}{||\nabla f(x_k)||^2}$ in the general case). It turns out that with each d_k constructed in this way, $d_1, d_2, \dots, d_k$ are **Q**-conjugate.

5. By Expanding Subspace Theorem, x_{k+1} minimizes $f(x)$ over the affine subspace $S = x_0 + \text{span}\{d_1, d_2 \dots, d_k\}$.

6. Hence finite convergence (n steps).

3 Barrier Methods

A barrier function $G_(x)$, is a continuous function with the propertiy that it approaches ∞ as one of the $g_j(x)$ approaches 0 from below.

Examples:

$$-\sum_{j=1}^{p} \log[-g_j(x)] \quad \text{and} \quad -\sum_{j=1}^{p} \frac{1}{g_j(x)}$$

Consider the primal/dual pair of linear optimization problems

$$\begin{array}{llrcl} \text{P:} & \min & c^\top x & & \\ & \text{s.t.} & Ax & = & b \\ & \text{s.t.} & x & \geq & 0 \end{array} \qquad \begin{array}{llrcl} \text{D:} & \max & b^\top p & & \\ & \text{s.t.} & A^\top p + s & = & c \\ & \text{s.t.} & s & \geq & 0 \end{array}$$

To solve P, we define the following barrier problem:

$$\begin{array}{lll} \text{BP:} & \min & B_\mu(x) \triangleq \; c^\top x - \mu \sum_{j=1}^n \log x_j \\ & \text{s.t.} & Ax = b \end{array}$$

Assume that for all $\mu > 0$, BP has an optimal solution $x(\mu)$. This optimum will be unique. Why?

As μ varies, the $x(\mu)$ form what is called the *central path*.

Theorem. $\lim_{\mu \to 0} x(\mu)$ exists and $x^* = \lim_{\mu \to 0} x(\mu)$ is an optimal solution to P.

Then the barrier problem from the dual problem is

$$\begin{array}{lll} \text{BD:} & \max & b^\top p + \mu \sum_{j=1}^n \log s_j \\ & \text{s.t.} & A^\top p + s = c \end{array}$$

Theorem. Let $\mu > 0$. Then $x(\mu), s(\mu), p(\mu)$ are optimal solutions to BP and BD if and only if the following hold:

$$\begin{aligned} Ax(\mu) &= b \\ x(\mu) &\geq 0 \\ A^\top p(\mu) + s(\mu) &= c \\ s(\mu) &\geq 0 \\ x_j(\mu)s_j(\mu) &= \mu, \ \forall j \end{aligned}$$

To solve BP using the *Primal path following algorithm*, we:

1. Start with a feasible interior point solution $x_0 > 0$
2. Step in the Newton direction $d(\mu) = \big(I - X^2A^\top(AX^2A^\top)^{-1}A\big)\big(Xe - \frac{1}{\mu}X^2c\big)$
3. Decrement μ
4. Iterate until convergence is obtained (complementary slackness above is ϵ-satisfied)

Note if we were to fix μ and carry out several Newton steps, then x would converge to $x(\mu)$. By taking a single step in the *Newton direction* we can guarantee that x stays "close to" $x(\mu)$, i.e. the *central path.* Hence following the iterative Primal path following algorithm we will converge to an optimal solution by this result and the first theorem above.

15.093J / 6.255J Optimization Methods
Fall 2009