On KKT optimality conditions

MinSeok Song

Theorem 1. *Assume that functions $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable. We are trying to solve the optimization problem*

$$\begin{aligned} &\textit{Minimize } f_0(x) \\ &\textit{subject to } f_i(x) \leqslant 0, \quad i = 1, 2, \ldots, m, \\ &\qquad\qquad h_j(x) = 0, \quad j = 1, 2, \ldots, p. \end{aligned}$$

x^* is a solution of this problem and the strong duality holds if and only if x^* satisfies

$$\begin{aligned} &f_i(x^*) \leqslant 0, \quad i = 1, \ldots, m \\ &h_i(x^*) = 0, \quad i = 1, \ldots, p \\ &\lambda_i^*(x) \geqslant 0, i = 1, \ldots, m \\ &\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m \\ &\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0 \\ &\textit{for some } \lambda^* \geq 0 \end{aligned}$$

Proof. ($\Leftarrow$) The key here is to notice that $\lambda \geqslant 0$ because this makes L, a duality function, convex. It follows that x^* gives a minimum solution of L by the last condition. Using the fourth and the second condition, we can see that $f_0(x^*) = g(\lambda^*, \nu^*)$

($\Rightarrow$) This holds by strong duality. □

- $\lambda^T f_i(x) = 0$ is called complemtary slackness. The term complementary comes from the idea of either or that, and slackness means "leftover." This condition is hard to verify in practice.
- Cavet: if the strong duality condition does not hold, then KKT condition might not hold. Further, of course, by no means do we have existence and uniqueness.
- Some of variants exist, one of which includes when f is not convex, given below.

Lemma 2. *Consider now the equality constrained problem, adding the condition*

$$h_i(x) = 0, i = 1, 2, \dots, p$$

from the previous setup. Assume all functions are continuously differentiable. Let x^* be the global minimum of a problem. Assume that the gradients of $h_i(x), i = 1, 2, \dots, p$ are linearly independent at x^*. There exist $\nu_1, \nu_2, \dots, \nu_p \in \mathbb{R}$ (the Lagrange multipliers) such that
$\nabla f(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0.$

Remark 1.

Note that complementary slackness and linearity of h_i are compensated by independence, with a simpler setup here - in a way, independence is a pretty strong condition here.

Proof. It follows by observing that $\nabla f(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$. By independence, we also have a uniqueness. □

Theorem 3. *Assume further that we have continuously differentiable functions $g_i(x) \leqslant 0, i = 1, 2, \dots, m$.*
Let us assume that the vector set composed of $\nabla h_i(x)$ and the gradients of all active inequality constraints form a linear independent set. At x^*, the KKT condition holds.

Remark 2.

Gradient is the steepest direction, and independence signifies the well-behavedness of the constraint, in the sense that whenever x^* is in the 'edge' of the inequality constraint, this direction should give a new information. The theorem says that this well-behavedness is sufficient condition for KKT.

Proof.

1. It is enough to show for the case g_i's are all active at x^*, because $\mu_i = 0$ whenever $g_i \neq 0$ since $\mu_i \geqslant 0$ (otherwise we do not have minimum at x^*).
2. Since $g_i \leqslant 0$, we can also assume that inequality is equality, and use the previous theorem to show the existence of Lagrangian constants that satisfy $\nabla L = \nabla_x f(x^*) + \sum_{i=1}^{m} \mu_i \nabla_x g_i(x^*) + \sum_{i=1}^{p} \nu_i \nabla_x h_i(x^*) = 0$.
3. Now it suffices to prove $\mu_i \geqslant 0$ for all i's. Assume, to the contrary, that $\mu_i < 0$ for some i. First note that $\sum_{i=1}^{m} \mu_i \nabla_x g_i(x^*) + \sum_{i=1}^{p} \nu_i \nabla_x h_i(x^*) = -\nabla_x f(x^*) = 0$ since x^* achieves minimum. Use constant rank theorem on $F(t, x) = (g_1(x), \dots t + g_i(x), \dots, g_m(x), h_1(x) \dots, h_p(x))$ (we may need additional condition, such as independence of gradients). We can derive the contradiction by having $\nabla_x f(x^*) = -\sum_{i=1}^{m} \mu_i \nabla_x g_i(x^*) - \sum_{i=1}^{p} \nu_i \nabla_x h_i(x^*) \neq 0$.

□

Remark 3.

Dual is not necessarily symmetric in mathematics in general. For example the dual of L^1 is L^∞ but the dual of L^∞ is not L^1.

Under certain condition, it is; for example if f is convex and closed, (epigraph is a closed set) then $f^{**} = f$.

Another relevant concept for convex analysis (in particular, for function) is Legendre transform.