1 Optimization under Inequality Constraints (June 4)

1.1 Kuhn-Tucker Conditions

Our problem is of the form

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } h_1(x) = \cdots = h_k(x) = 0 \\ &\qquad g_1(x), \ldots, g_l(x) \leq 0. \end{aligned}$$

Definition 1. *Let x_0 satisfy the above constraints. We call the inequality constraint $g_i(x) \leq 0$ active at x_0 if $g_i(x_0) = 0$. Otherwise, it is inactive at x_0.*

Since we are only studying local properties of functions, we will only be concerned with active constraints.

Definition 2. *Suppose there is an index $l' \leq l$ such that $g_1(x_0) =, \ldots, g_{l'}(x_0) = 0$ are active, and $g_{l'+1}(x_0) \leq 0, \ldots, g_l(x_0) \leq 0$ are inactive. We say that x_0 is a regular point of these constraints if the vectors $\nabla h_1(x_0), \ldots, \nabla h_k(x_0), \nabla g_1(x_0), \ldots, \nabla g_{l'}(x_0)$ are linearly independent.*

Theorem 1.1. *(First order necessary conditions for minimizers under inequality constraints) Let $\Omega \subseteq \mathbb{R}^n$ be open and consider C^1 functions $f, h_1, \ldots, h_k, g_1, \ldots, g_l$ on Ω. Suppose x_0 is a local minimizer of f subject to the constraints, and that x_0 is regular as defined above. Then*

(i) There exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $\mu_1, \ldots, \mu_l \in \mathbb{R}^{\geq 0}$ such that

$$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) + \sum \mu_j \nabla g_j(x_0) = 0.$$

(ii) ("Complementary slackness conditions") For all j, $\mu_j g_j(x_0) = 0$, or equivalently, $\sum \mu_j g_j(x_0) = 0$.

These conditions are also known as the *Kuhn-Tucker conditions.*

Suppose the active constraints at x_0 are the first l' constraints. Since each $\mu_j \geq 0$, condition (ii) is equivalent to saying that if $j \geq l' + 1$, then $\mu_j = 0$.

Proof. If x_0 is a local minimizer of f subject to the constraints, then it is certainly a local minimizer of f subject to only the active constraints. That is, x_0 is also a local minimizer of f subject to the equality constraints

$$h_1(x) = \cdots = h_k(x) = g_1(x) = \cdots = g_{l'} = 0.$$

We know how to work with this! Let M be the surface defined by these equality constraints. By the Lagrange multipliers theorem,

$$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) + \sum \mu_j \nabla g_j(x_0) = 0,$$

for some $\lambda_i \in \mathbb{R}, \mu_j \in \mathbb{R}$. (Note that we have not yet shown that the μ_j's are non-negative.)

Note that $g_1(x_0) = \cdots = g_{l'}(x_0) = 0$. Therefore $\mu_{l'+1} = \cdots = \mu_l = 0$, so it follows that

$$\mu_1 g_1(x_0) = 0, \ldots, \mu_{l'} g_{l'}(x_0) = 0,$$

which implies that $\mu_j g_j(x_0) = 0$ for all j. We have proven condition (ii).

We must now verify the non-negativity of the μ_j's. Suppose for the sake of contradiction that some $\mu_j < 0$; WLOG assume $j = 1$. Let

$$\widetilde{M} = \{x \in \Omega : h_i(x) = 0, g_i(x) = 0, j \neq 1\}.$$

Since x_0 is a regular point of M, x_0 is a regular point of $\widetilde{M}$. Therefore

$$T_{x_0}\widetilde{M} = \operatorname{span}(\{\nabla h_1(x_0), \ldots, h_k(x_0), \nabla g_2(x_0), \ldots, \nabla g_l(x_0)\})^{\perp}.$$

The vector $\nabla g_1(x_0)$ does not lie in this span, so there is a $v \in T_{x_0}\widetilde{M}$ such that $\nabla g_1(x_0) \cdot v < 0$. That is, g_1 is strictly decreasing in the direction of v, or in more precise language, $g_1(x_0 + sv) < g_1(x_0)$ for all sufficiently small s, as we have

$$\left.\frac{d}{ds}\right|_{s=0} g_1(x_0 + sv) = \nabla g_1(x_0) \cdot v < 0.$$

Therefore v is a feasible direction for $g_1(x) \leq 0$ at x_0, and also, v is tangential to the other constraints. Since x_0 is a regular point of $\widetilde{M}$, we may find a curve $x(s)$ on $\widetilde{M}$ such that $x(0) = x_0$ and $x'(0) = v$. Also, $s = 0$ is a local minimizer of $f \circ x$, so

$$\left.\frac{d}{ds}\right|_{s=0} f(x(s)) = \nabla f(x_0) \cdot v \geq 0.$$

On the other hand,

$$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) + \mu_1 \nabla g_1(x_0) + \sum_{j=2}^{l'} \mu_j \nabla g_j(x_0) = 0.$$

Taking the dot product of the above equation by v kills the two sums above and gives

$$\nabla f(x_0) \cdot v + \mu_1 \nabla g_1(x_0) \cdot v = 0,$$

implying $\nabla f(x_0) \cdot v < 0$, a contradiction. So every $\mu_j \geq 0$. $\square$