Karush-Kuhn-Tucker (KKT) and Fritz John Optimality Conditions

March 17, 2015

Massachusetts Institute of Technology
Lecture Outline

• Inequality Constrained Problems
• Karush-Kuhn-Tucker Necessary Conditions
• Fritz John Optimality Conditions
• Constraint Qualifications

Inequality constrained problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h(x) = 0, \quad g(x) \leq 0,
\end{align*}
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R}, \ h : \mathbb{R}^n \rightarrow \mathbb{R}^m, \ g : \mathbb{R}^n \rightarrow \mathbb{R}^r \) are continuously differentiable. Here

\[
\begin{align*}
h = (h_1, \ldots, h_m), \quad g = (g_1, \ldots, g_r).
\end{align*}
\]
Basic Results

Karush-Kuhn-Tucker Necessary Conditions: Let \( x^* \) be a local minimum and a regular point. Then there exist unique Lagrange mult. vectors \( \lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) \), \( \mu^* = (\mu_1^*, \ldots, \mu_r^*) \), such that

\[
\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \\
\mu_j^* \geq 0, \quad j = 1, \ldots, r, \\
\mu_j^* = 0, \quad \forall \ j \notin A(x^*).
\]

If \( f, h, \) and \( g \) are twice cont. differentiable, then

\[
y' \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) y \geq 0, \quad \text{for all } y \text{ such that,} \\
\nabla h(x^*)' y = 0, \quad \nabla g_j(x^*)' y = 0, \ j \in A(x^*).
\]

- Similar sufficiency conditions and sensitivity results. They require strict complementarity, i.e., \( \mu_j^* > 0 \) for all \( j \in A(x^*) \), as well as regularity of \( x^* \).
Proof of Karush-Kuhn-Tucker Conditions

Use equality-constraints result to obtain all the conditions except for
\[ \mu_j^* \geq 0 \text{ for } j \in A(x^*) \]. Introduce the penalty functions
\[ g_j^+(x) = \max\{0, g_j(x)\}, \quad j = 1, \ldots, r, \]
and for \( k = 1, 2, \ldots \), let \( x^k \) minimize
\[ f(x) + \frac{k}{2} \|h(x)\|^2 + \frac{k}{2} \sum_{j=1}^{r} (g_j^+(x))^2 + \frac{1}{2} \|x - x^*\|^2 \]
over a closed sphere of \( x \) such that \( f(x^*) \leq f(x) \). Using the same argument as for equality constraints,
\[ \lambda_i^* = \lim_{k \to \infty} kh_i(x^k), \quad i = 1, \ldots, m, \]
\[ \mu_j^* = \lim_{k \to \infty} kg_j^+(x^k), \quad j = 1, \ldots, r. \]
Since \( g_j^+(x^k) \geq 0 \), we obtain \( \mu_j^* \geq 0 \) for all \( j \).
General Sufficiency Condition

Consider the problem \( \min_{x \in X} g(x) \leq 0 f(x) \). Let \( x^* \) be feasible and
\[
\mu^*_j \geq 0, \quad j = 1, \ldots, r, \quad \mu^*_j = 0, \quad \forall j \notin A(x^*),
\]
\[
x^* = \arg \min_{x \in X} L(x, \mu^*).
\]

Then \( x^* \) is a global minimum of the problem.

**Proof:** We have
\[
f(x^*) = f(x^*) + \mu^* g(x^*) = \min_{x \in X} \{ f(x) + \mu^* g(x) \}
\]
\[
\leq \min_{x \in X, g(x) \leq 0} \{ f(x) + \mu^* g(x) \} \leq \min_{x \in X, g(x) \leq 0} f(x),
\]
where the first eq. follows from the hypothesis, which implies that
\( \mu^* g(x^*) = 0 \), and the last ineq. follows from the nonnegativity of \( \mu^* \).

- **Special Case:** Let \( X = \mathbb{R}^n \), \( f \) and \( g_j \) be convex and differentiable.
  Then first order KKT conditions sufficient for global optimality.
Fritz John Necessary Conditions: If $x^*$ is a local minimum, there exists a scalar $\mu_0^*$ and Lagrange multipliers $\lambda_1^*, \ldots, \lambda_m^*$ and $\mu_1^*, \ldots, \mu_r^*$ satisfying the following conditions:

$$
\mu_0^* \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0.
$$

$$
\mu_j^* \geq 0, \quad j = 0, 1, \ldots, r.
$$

$\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*$ are not all equal to 0.

In every neighborhood $N$ of $x^*$ there is an $x \in N$ that simultaneously violates all the constraints that correspond to nonzero multipliers.

Note 1: The last condition implies that $\mu_j^* = 0$ for all $j$ with $g_j(x^*) < 0$ (complementary slackness condition).

Note 2: If we can show that $\mu_0^* > 0$, we can divide by $\mu_0^*$ and obtain a Lagrange multiplier theorem.
Proof of Fritz John Conditions

Let $x^k$ minimize

$$f(x) + \frac{k}{2} \sum_{i=1}^{m} (h_i(x))^2 + \frac{k}{2} \sum_{j=1}^{r} (g^+_j(x))^2 + \frac{1}{2} \|x - x^*\|^2$$

over a closed sphere around $x^*$, $S$, such that $f(x^*) \leq f(x)$ for all feasible $x$ with $x \in S$ and $g^+_j(x) = \max\{0, g_j(x)\}$. Then $x^k \to x^*$ and

$$\nabla f(x^k) + \sum_{i=1}^{m} \xi^k_i \nabla h_i(x^k) + \sum_{j=1}^{r} \zeta^k_j \nabla g_j(x^k) + (x^k - x^*) = 0$$

where $\xi^k_i = kh_i(x^k)$ and $\zeta^k_j = kg^+_j(x^k)$. Denote,

$$\delta^k = \sqrt{1 + \sum_{i=1}^{m} (\xi^k_i)^2 + \sum_{j=1}^{r} (\zeta^k_j)^2},$$

$$\mu^k_0 = \frac{1}{\delta^k}, \quad \lambda^k_i = \frac{\xi^k_i}{\delta^k}, \quad \mu^k_j = \frac{\zeta^k_j}{\delta^k},$$

and take limit as $k \to \infty$. 
• Consider the problem $\min_{g(x)\leq 0} f(x)$ where all the $g_j$ are concave.

**Proposition:** If $x^*$ is a local minimum, we can take $\mu_0^* = 1$ in the Fritz John conditions.

**Proof:** Assume to reach a contradiction that $\mu_0^* = 0$. By concavity,

$$g_j(x) \leq g_j(x^*) + \nabla g_j(x^*)'(x - x^*), \quad j = 1, \ldots, r, \forall x$$

Multiply with $\mu_j^*$ and add over $j$:

$$\sum_{j=1}^{r} \mu_j^* g_j(x) \leq \sum_{j=1}^{r} \mu_j^* g_j(x^*) + \left( \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) \right)'(x - x^*) = 0$$

since $\mu_j^* g_j(x^*) = 0$ for all $j$ [by FJ cond (iv)], and $\mu_j^* \nabla g_j(x^*) = 0$ [by FJ cond (i)]. On the other hand, there are $j$ for which $\mu_j^* > 0$ and by FJ cond. (iv), there is an $x$ satisfying $g_j(x) > 0$ for all these $j$. For this $x$, we have $\sum_{j=1}^{r} \mu_j^* g_j(x) > 0$ – contradiction.
Mangasarian-Fromovitz Condition

Let $x^*$ be a local minimum of the problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h_i(x) = 0, \quad i = 1, \ldots, m \\
& \quad g_j(x) \leq 0, \quad j = 1, \ldots, r.
\end{align*}
\]

Assume that the gradients $\nabla h_i(x^*), i = 1, \ldots, m$ are linearly independent, and that there exists a vector $d$ such that

\[
\nabla h_i(x^*)'d = 0, \quad \forall \ i = 1, \ldots, m,
\]

\[
\nabla g_j(x^*)'d < 0, \quad \forall \ j \in A(x^*).
\]

Then we can take $\mu_0^* = 1$ in the FJ conditions.
Corollaries:

- **Slater Condition** ($h_i$: linear, $g_j$: convex, and there exists a feasible $\bar{x}$ such that $g_j(\bar{x}) < 0$ for all $j$).

- Conditions for minimax problems.
Alternative Development for Linear Constraints

• Consider the problem \( \min_{a_j'x \leq b_j, j=1,...,r} f(x) \).

• Remarkable property: No need for regularity.

• Proposition: If \( x^* \) is a local minimum, there exist \( \mu_1^*, \ldots, \mu_r^* \) with \( \mu_j^* \geq 0, j = 1, \ldots, r \), such that

\[
\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* a_j = 0, \quad \mu_j^* = 0, \quad \forall \ j \notin A(x^*).
\]
• The proof uses Farkas Lemma: Consider the cone $C$ “generated” by $a_j$, $j \in A(x^*)$, and the “polar” cone $C^\perp$ shown below.

$$C^\perp = \{y | a_j'y \leq 0, j=1,\ldots,r\}$$

$$C = \{x | x = \sum_{j=1}^{r} m_j a_j, m_j \geq 0 \}$$

Then, $(C^\perp)^\perp = C$, i.e.,

$$x \in C \quad \text{iff} \quad x'y \leq 0, \quad \forall \ y \in C^\perp.$$
Proof of Farkas Lemma

Proof: First show that $C$ is closed (nontrivial). Then, let $x$ be such that $x'y \leq 0, \forall y \in C^\perp$, and consider its projection $\hat{x}$ on $C$. We have

$$x'(x - \hat{x}) = \|x - \hat{x}\|^2,$$

\(\ast\)

$$(x - \hat{x})'a_j \leq 0, \quad \forall j.$$  

Hence, $(x - \hat{x}) \in C^\perp$, and using the hypothesis,

$$x'(x - \hat{x}) \leq 0.$$  

\(\ast\ast\)

From ($\ast$) and ($\ast\ast$), we obtain $x = \hat{x}$, so $x \in C$. 

12
Proof of Lagrange Multiplier Result

The local min \( x^* \) of the original problem is also a local min for the problem \( \min_{a'_j x \leq b_j, j \in A(x^*)} f(x) \). Hence

\[
\nabla f(x^*)'(x - x^*) \geq 0, \quad \forall \ x \text{ with } a'_j x \leq b_j, \ j \in A(x^*).
\]

Since a constraint \( a'_j x \leq b_j, j \in A(x^*) \) can also be expressed as \( a'_j (x - x^*) \leq 0 \), we have

\[
\nabla f(x^*)'y \geq 0, \quad \forall \ y \text{ with } a'_j y \leq 0, \ j \in A(x^*).
\]
From Farkas’ lemma, \(-\nabla f(x^*)\) has the form

\[
\sum_{j \in A(x^*)} \mu_j^* a_j, \quad \text{for some } \mu_j^* \geq 0, \; j \in A(x^*).
\]

Let \(\mu_j^* = 0\) for \(j \notin A(x^*)\).