Karush-Kuhn-Tucker Theorem

We seek first order conditions for local optimality for the following formulation of non-linear programming:

\[
\begin{align*}
\text{Minimize } & \ f(x) \\
\text{Subject to } & \ h(x) = 0 \\
& \ g(x) \leq 0
\end{align*}
\]

where \( x \) is a vector in \( \mathbb{R}^n \), \( h \) maps into \( \mathbb{R}^k \), and \( g \) maps into \( \mathbb{R}^m \). We will assume all the functions are \( C^2 \) everywhere of interest. If \( g \) is not present, the necessary conditions become the Lagrange Multiplier Theorem that we saw last time.

The strategy will be the same. Given a candidate point, \( x^* \), for local optimality, we linearize the active constraints about \( x^* \). We then establish conditions that cannot occur at a local optimum for the linearized problem. We then have to argue that the conditions cannot occur at a local optimum for the original problem. This is usually the difficult point. That is why (1) is much easier to deal with when \( h \) and \( g \) are linear. This gives rise to various kinds of "constraint qualifications." We will duck most of these issues. In any case, given the "impossible" conditions we use a theorem of alternatives to obtain the conditions that we seek.

A couple of observations before we start. First, we let \( I = \{i|g_i(x) = 0\} \). \( I \) is called the active set. The other inequalities are said to be inactive. It is important to notice that if \( g_i(x) \) is inactive then by continuity there exists an \( \varepsilon > 0 \), so that for \( ||x - x^*|| < \varepsilon \), we have \( g(x) < 0 \). This essentially means that inactive constraints do not have to be checked when testing \( x^* \) to be a local minimum.

For the moment we will blithely replace the constraint functions by their first order Taylor expansion. That is:

\[
\begin{align*}
& f(x) \approx f(x^*) + \nabla f(x^*) y \\
& h_i(x) \approx h_i(x^*) + \nabla h_i(x^*) y \\
& g_i(x) \approx g_i(x^*) + \nabla g_i(x^*) y
\end{align*}
\]

Where \( y = x - x^* \).

Assuming the linear representation we have:

**Lemma 1:** Let \( x^* \) be a local minimum for (1). Then for all \( y \in \mathbb{R}^n \) such that:

\[
\begin{align*}
& \nabla h_i(x^*) y = 0 \text{ for } i = 1, \ldots, m, \text{ and} \\
& \nabla g_i(x^*) y \leq 0 \text{ for } i \in I
\end{align*}
\]

must also satisfy

\[ \nabla f(x^*) y \geq 0 \]
“Proof:” The two conditions simply guarantee that the approximations to the equality constraints and the active equality constraints, respectively, satisfy (1). But if \( \nabla f(x^*)y < 0 \), for every \( y > 0 \), no matter how small, the approximation to \( f \) satisfies \( f(x^*) + \nabla f(x^*)y < f(x^*) \), so that \( x^* \) cannot be a local minimum.

**Lemma 2 (Theorem of Alternatives):**

Either

There exists \( x \) such that

\[
A_i x = 0, \ i \in E
\]
\[
A_i x \leq 0, \ i \in L
\]
\[
cx < 0
\]

Or

There exists \( \lambda \) and \( \mu \geq 0 \) such that

\[
c = \sum_{i \in E} \lambda_i A_i - \sum_{i \in L} \mu_i A_i
\]

**Proof:** Use duality theory in the usual way. Consider the linear program:

Minimize \( cx \)  
Subject to \( A_i x = 0 \)  
and \( -A_i x \geq 0 \)

Observe that \( x = 0 \) satisfies the constraints. If there exists an \( x \) satisfying the constraints with \( cx < 0 \), then the solution is unbounded below. In this case, the dual does not have a solution. If the minimum value of \( cx \) is 0, then the dual has a solution, which gives the other alternative.

Applying Lemma 2 to Lemma 1 yields:

**Karush-Kuhn-Tucker Theorem:** Let \( x^* \) be a relative minimum point for (1). Then there exists vectors \( \lambda \in \mathbb{R}^k \), and \( \mu \in \mathbb{R}^m \), \( \mu \geq 0 \) such that

\[
\nabla f(x^*) = \lambda^T \nabla h(x^*) - \mu^T \nabla g(x^*), \quad \mu^T g(x^*) = 0.
\]

We still have the problem with the hand waving in the proof of Lemma 1. This is the issue of constraint qualifications alluded to earlier. The subject is very technical, so we will not pursue it further. To see it in all its gory detail see Bertsekas, Dimitri P., *Nonlinear Programming*, Athena Scientific, 1995. Remember, however, for linear constraints our argument is exact. In particular, the Markowitz model for portfolio optimization has linear constraints. One popular constraint qualification is to assume \( x^* \) is regular.

**Definition:** A feasible solution, \( x^* \), to (1) is regular if the gradients of the equality constraints and the active inequality constraints are linearly independent.

**(Correct) Karush-Kuhn-Tucker Theorem:** Let \( x^* \) be a relative minimum point for (1) and regular. Then there exists vectors \( \lambda \in \mathbb{R}^k \), and \( \mu \in \mathbb{R}^m \), \( \mu \geq 0 \) such that

\[
\nabla f(x^*) = \lambda^T \nabla h(x^*) - \mu^T \nabla g(x^*), \quad \mu^T g(x^*) = 0.
\]
**Sufficient Conditions:** If the equality constraints, \( h(x) \), are linear, and the inequality constraints, \( g(x) \), and the objective are convex, then the KKT conditions are sufficient.

**Proof:** Consider \( L(x; \lambda, \mu) = f(x) - \lambda h(x) + \mu g(x) \) for \( \mu \geq 0 \). \( L \) is convex in \( x \). Moreover, if \( \nabla L(x^*; \lambda, \mu) = 0 \) then \( L \) is minimized as an unconstrained function of \( x \) at \( x^* \) (this follows immediately from A.11 in Notes, p. 13). Next, \( L(x; \lambda, \mu) = f(x) - \lambda h(x) + \mu g(x) \leq f(x) \) for all feasible \( x \). Finally, \( L(x^*; \lambda, \mu) = f(x^*) \) because \( \mu^* g(x^*) = 0 \). Then \( L \) and \( f \) satisfy the conditions of the relaxation lemma (Notes p. 2) and \( x^* \) minimizes \( f(x) \) in (1).