CPT Lecture Notes 7: Nonlinear programming

Levent Ülkü

October 8, 2019

Nonlinear programming (NLP)

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g_1 : \mathbb{R}^n \to \mathbb{R}$, ..., $g_m : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions.

The NLP problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \text{ subject to}$$

$$g_1(x) \geq 0,$$

$$\ldots$$

$$g_m(x) \geq 0.$$

Definition: A pair $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^{n+m}$ is a Kuhn-Tucker pair for NLP if

(1) $g_i(\bar{x}) \geq 0$ for all $i = 1, ..., m$,

(2) $\bar{\lambda}_i \geq 0$ for all $i = 1, ..., m$,

(3) $\nabla f(\bar{x}) = \sum_{i=1}^{m} \nabla g_i(\bar{x})\bar{\lambda}_i$, and

(4) $\bar{\lambda}_i g_i(\bar{x}) = 0$ for all $i = 1, ..., m$.

We will say that $\bar{x}$ satisfies the Kuhn-Tucker conditions for NLP if there exists some $\bar{\lambda}$ such that $(\bar{x}, \bar{\lambda})$ is a Kuhn-Tucker pair for NLP.

We wish to resolve two questions:

(1) If $\bar{x}$ solves NLP, when does $\bar{x}$ necessarily satisfy the Kuhn-Tucker conditions for NLP? In other words, when are the KT conditions necessary for optimality in NLP?

(2) If $\bar{x}$ satisfies the Kuhn-Tucker conditions for NLP, when does $\bar{x}$ necessarily solve NLP? In other words, when are the KT conditions sufficient for optimality in NLP?

For any $x \in \mathbb{R}^n$, let $I(x) = \{i = 1, ..., m : g_i(x) = 0\}$. Hence $I(x)$ is the set of active constraints at x. The proofs of the following observations follow directly from the definition of a KT pair.

Lemma: Suppose that $(\bar{x}, \bar{\lambda})$ is a Kuhn-Tucker pair for NLP. If $I(\bar{x}) = \emptyset$, then $\nabla f(\bar{x}) = 0$. If $I(\bar{x}) \neq \emptyset$, then $\nabla f(\bar{x}) = \sum_{i \in I(\bar{x})} \nabla g_i(\bar{x}) \bar{\lambda}_i$.

Lemma: Suppose that
(1) $g_i(\bar{x}) \geq 0$ for all $i = 1, ..., m$,
(2) $\{\bar{\mu}_i\}_{i \in I(\bar{x})}$ is a collection of nonnegative numbers satisfying $\nabla f(\bar{x}) = \sum_{i \in I(\bar{x})} \nabla g_i(\bar{x}) \bar{\mu}_i$.
If $\bar{\lambda}_i = \bar{\mu}_i$ for all $i \in I(\bar{x})$ and $\bar{\lambda}_i = 0$ for all $i \notin I(\bar{x})$, then $(\bar{x}, \bar{\lambda})$ is a Kuhn-Tucker pair for NLP.

The necessity of the Kuhn-Tucker conditions

For $c, a_1, \ldots, a_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$, consider the following Linear Programming problem (LP):

$$\min_{x \in \mathbb{R}^n} c \cdot x \text{ subject to}$$

$$a_1 \cdot x \geq b_1,$$

$$\ldots$$

$$a_m \cdot x \geq b_m.$$

LP is a special case of NLP where $f(x) = c \cdot x$ and $g_i(x) = a_i \cdot x - b_i$.

The definition of a Kuhn-Tucker pair is easily adapted for LP.

Definition: A pair $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^{n+m}$ is a Kuhn-Tucker pair for LP if

(1) $a_i \cdot \bar{x} \geq b_i$ for all $i = 1, ..., m$,

(2) $\bar{\lambda}_i \geq 0$ for all $i = 1, ..., m$,

(3) $c = \sum_{i=1}^{m} a_i \bar{\lambda}_i$, and

(4) $\bar{\lambda}_i (a_i \cdot \bar{x} - b_i) = 0$ for all $i = 1, ..., m$.

Our first goal is to prove KT necessity for LP. The following result will be key.

Theorem: Suppose that $A = [a_1 \vdots \cdots \vdots a_m]$ and $a_i \in \mathbb{R}^n$ for all $i = 1, ..., m$. Then the set

$$posA := \{Au | u \in \mathbb{R}^m_+\}$$

is a nonempty, closed and convex cone.

Proof: Note that $a_i \in posA$ for every i, so $posA$ is nonempty.

If $u \in \mathbb{R}^m_+$ (and therefore $Au \in posA$) and $t \geq 0$, then $tu \in \mathbb{R}^m_+$ as well and therefore $tAu \in posA$. This establishes that $posA$ is a cone.

If $u, v \in \mathbb{R}^m_+$, then $u + v \in \mathbb{R}^m_+$. It follows that if $Au, Av \in posA$, then $Au + Av \in posA$. This establishes that $posA$ is convex. (HW: a cone C is convex if and only if $x + y \in C$ for every $x, y \in C$.)

It remains to show that $posA$ is closed. We will do so in 3 steps.

Step 1: We will show that *posA* is closed if $a_1, ..., a_m$ are linearly independent.

Assume linear independence. Choose $\{y_k\}$ in *posA* such that $y_k \to \overline{y}$. We need to show that $\overline{y} \in posA$.

If $\overline{y} = 0$, then $\overline{y} \in posA$ since *posA* is a cone.

If $\overline{y} \neq 0$, then we may assume without loss of generality that $y_k \neq 0$ for all k. (Why? Can there exist a subsequence $\{y_{k_m}\}$ with $y_{k_m} = 0$ for each m?)

Choose $\{x_k\}$ in $\mathbb{R}^m_+$ such that $y_k = Ax_k$ for every k. We need only establish that $\{x_k\}$ is a bounded sequence. (Why? If $\{x_k\}$ is a bounded sequence, then it has a convergent subsequence: there exist $\{x_{k_m}\}$ and $\bar{x}$ such that $x_{k_m} \to \bar{x}$. Since $\mathbb{R}^m_+$ is closed, $\bar{x} \in \mathbb{R}^m_+$. $y_{k_m} = Ax_{k_m}$ and linear functions are continuous, $\bar{y} = A\bar{x}$. Summarizing: $\bar{y} = A\bar{x}$ and $\bar{x} \in \mathbb{R}^m_+$ and therefore $\overline{y} \in posA$.)

Step 1 cont'd: We will show that $\{x_k\}$ is bounded.

Note that each $x_k \neq 0$ since $y_k \neq 0$. Hence

$$\frac{1}{||x_k||} y_k = A\left[\frac{1}{||x_k||} x_k\right].$$

If $\{x_k\}$ is not bounded, then there exists a subsequence $\{x_{k_m}\}$ such that $||x_{k_m}|| \geq m$ for each positive integer m. Since the sequence $\{\frac{1}{||x_{k_m}||} x_{k_m}\}$ belongs to a compact set, it contains a convergent subsequence $\{\frac{1}{||x_{k_{m_t}}||} x_{k_{m_t}}\}$ with limit z where $||z|| = 1$. Therefore

$$\frac{1}{||x_{k_{m_t}}||} y_{k_{m_t}} = A\left[\frac{1}{||x_{k_{m_t}}||} x_{k_{m_t}}\right] \rightarrow Az$$

Since the columns of A are linearly independent and $z \neq 0$, it follows that $Az \neq 0$. However, $\frac{1}{||x_{k_{m_t}}||} y_{k_{m_t}} \rightarrow 0$, an impossibility. This proves that $\{x_k\}$ is bounded.

Step 2: Suppose that $y \in posA$ and $y \neq 0$.

Let $\varnothing \neq S \subseteq \{1, .., m\}$ have the property that

(i) there exists a collection $\{x_i : i \in S\}$ of (strictly) positive numbers such that

$$y = \sum_{i \in S} a_i x_i$$

(ii) if $T \subseteq \{1, .., m\}$ and there exists a collection $\{u_i : i \in T\}$ of (strictly) positive numbers such that

$$y = \sum_{i \in T} a_i u_i$$

then $|S| \leq |T|$. Note that such S exists since $y \neq 0$ and $\{1, .., m\}$ has only finitely many subsets.

Step 2 cont'd: We will show that $\{a_i : i \in S\}$ is a linearly independent collection. To see this, suppose that $\{a_i : i \in S\}$ is not a linearly independent collection. Then there exist numbers $\{z_i : i \in S\}$ not all 0 such that $0 = \sum_{i \in S} a_i z_i$ and without loss of generality, we may assume that at least one element of $\{z_i : i \in S\}$ is negative.

Next, take strictly positive numbers x_i for every $i \in S$ such that $\sum_{i \in S} a_i x_i = y$, and note that for every real number t,

$$\sum_{i \in S} a_i x_i (1 + t \frac{z_i}{x_i}) = \sum_{i \in S} a_i x_i = y.$$

In a way that is reminiscent of the proof of CT, we will now show that for a special choice of t, $x_i(1 + t\frac{z_i}{x_i}) \geq 0$ for every $i \in S$.

Step 2 cont'd: Let

$$t = \min\{-\frac{x_i}{z_i} : i \in S \text{ and } z_i < 0\} = -\frac{x_k}{z_k}.$$

Note that $t > 0$.
If $i \in S \backslash \{k\}$ and $z_i \geq 0$, then $x_i(1 + t\frac{z_i}{x_i}) > 0$.
If $i \in S \backslash \{k\}$ and $z_i < 0$, then

$$t = -\frac{x_k}{z_k} \leq -\frac{x_i}{z_i} \Rightarrow -t\frac{z_i}{x_i} \leq 1 \Rightarrow 1 + t\frac{z_i}{x_i} \geq 0 \Rightarrow x_i(1 + t\frac{z_i}{x_i}) \geq 0$$

Finally, note that $x_k(1 + t\frac{z_k}{x_k}) = 0$.

Step 2 cont'd: Combining these observations, it follows that

$$y = \sum_{i \in S \setminus \{k\}} a_i x_i (1 + t\frac{z_i}{x_i}) \text{ where } x_i(1 + t\frac{z_i}{x_i}) \geq 0 \text{ for each } i \in S \setminus \{k\}.$$

This is an impossibility given the minimality of $|S|$. (Why? Note that the coefficients are not necessarily strictly positive here. But they must be for a subset of $S \setminus \{k\}$.)

Step 3: If A is the 0 matrix, then $posA = \{0\}$ and is obviously closed. If A is not the zero matrix, define a collection of nonempty subsets $\mathcal{C}$ of $\{1, .., m\}$ as follows: $S \in \mathcal{C}$ if and only if $\{a_i : i \in S\}$ is a linearly independent collection. For each $S \in \mathcal{C}$, let A_S be the matrix whose columns are the elements of S. Obviously,

$$\bigcup_{S \in \mathcal{C}} posA_S \subseteq posA.$$

Conversely, Step 2 shows that

$$posA \subseteq \bigcup_{S \in \mathcal{C}} posA_S$$

Therefore,

$$posA = \bigcup_{S \in \mathcal{C}} posA_S.$$

By Step 1, each $posA_S$ is closed. Hence $posA$, a finite union of closed sets, is also closed. The proof is complete.

We are now ready for the KT necessity theorem for linear programming.

Theorem: If $\bar{x}$ solves LP, then there exists some $\bar{\lambda}$ such that $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^{n+m}$ is a Kuhn-Tucker pair for LP.

Proof: Suppose $\bar{x}$ solves LP. Recalling $I(\bar{x}) = \{i : a_i \cdot \bar{x} = b_i\}$, we need only show that

(1) if $I(\bar{x}) = \emptyset$, then $c = 0$, and

(2) if $I(\bar{x}) \neq \emptyset$, then there exist nonnegative numbers $\{\bar{\mu}_i\}_{i \in I(\bar{x})}$ such that $c = \sum_{i \in I(\bar{x})} a_i \bar{\mu}_i$.

To see (1), first suppose $I(\bar{x}) = \emptyset$. Hence, for some small enough $t > 0$, $\bar{x} + te_i$ and $\bar{x} - te_i$ are feasible in LP for every i. If $c_i < 0$ for some i, then $c \cdot (\bar{x} + te_i) = c \cdot \bar{x} + tc_i < c \cdot \bar{x}$, contradicting the optimality of $\bar{x}$. If $c_i > 0$, then $c \cdot (\bar{x} - te_i) = c \cdot \bar{x} - tc_i < c \cdot \bar{x}$, leading to a similar contradiction. We conclude that $c_i = 0$ for all i.

To establish (2), suppose $I(\bar{x}) \neq \varnothing$ and let

$$K = \left\{ \sum_{i \in I(\bar{x})} \mu_i a_i : \mu_i \geq 0 \text{ for all } i \right\}.$$

K is a nonempty, closed and convex cone by the previous result. Suppose towards a contradiction that $c \notin K$. We will proceed in four steps.

Step 1: There exists $d \in \mathbb{R}^n$ such that $a_i \cdot d \geq 0$ for all $i \in I(\bar{x})$ and $c \cdot d < 0$.

To see this, apply Minkowski's theorem on K and $c \notin K$, to deduce that there exist some $d \in \mathbb{R}^n \backslash \{0\}$ and $y_0 \in K$ such that $d \cdot y \geq d \cdot y_0 > d \cdot c$ for all $y \in K$. Since K is a cone, $0 \in K$ and therefore $0 \geq d \cdot y_0$. Suppose $d \cdot y_0 < 0$. Since K is a cone, $2y_0 \in K$ and $d \cdot (2y_0) < d \cdot y_0$, a contradiction. Hence $d \cdot y_0 = 0$. It follows that $d \cdot c < 0$. Finally, noting that $a_i \in K$ for all $i \in I(\bar{x})$, we conclude that $a_i \cdot d \geq 0$ for all $i \in I(\bar{x})$.

Step 2: We will show that there exists $t^* > 0$ such that $a_i \cdot (\bar{x} + t^* d) \geq b_i$ for all $i \notin I(\bar{x})$.

To see this, note that

$$a_i \cdot (\bar{x} + td) - b_i = a_i \cdot \bar{x} - b_i + t(a_i \cdot d).$$

If $i \notin I(\bar{x})$, then $a_i \cdot \bar{x} - b_i > 0$, and $a_i \cdot \bar{x} - b_i + t(a_i \cdot d) > 0$ for all sufficiently small $t > 0$. Since there are finitely many constraints, there exists $t^* > 0$ such that $a_i \cdot (\bar{x} + td) - b_i > 0$ for all $i \notin I(\bar{x})$.

Step 3: $a_i \cdot (\bar{x} + t^* d) \geq b_i$ for all $i \in I(\bar{x})$.

This follows from the fact that $t^* > 0$ and $a_i \cdot d \geq 0$ for all $i \in I(\bar{x})$.

Step 4: By steps 2 and 3, $\bar{x} + t^* d$ is feasible. Note that $c \cdot (\bar{x} + t^* d) < c \cdot \bar{x}$ since $c \cdot d < 0$ and $t^* > 0$, an impossibility. This completes the proof.

KT Necessity for NLP

Definition: A solution $\bar{x}$ to the NLP satisfies the General Constraint Qualification (GCQ) condition if it solves the following LP:

$$\min \nabla f(\bar{x}) \cdot x$$
$$\text{subject to } g_i(\bar{x}) + \nabla g_i(\bar{x}) \cdot (x - \bar{x}) \geq 0 \text{ for all } i = 1, \ldots, m.$$

The following is the KT necessity theorem for nonlinear programming.

Theorem: If $\bar{x}$ solves the NLP and satisfies the GCQ, then it satisfies the KT conditions for NLP.

Proof: Homework.

Example: Consider the problem $\min_{(x_1,x_2)\in\mathbb{R}^2} x_1$ subject to $x_1^3 \geq x_2$ and $x_2 \geq 0$. Draw a picture and convince yourselves that $\bar{x} = (0,0)$ is the unique solution. Note that

$$\nabla f(\bar{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \nabla g_1(\bar{x}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \nabla g_2(\bar{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Furthermore $I(\bar{x}) = \{1,2\}$. However, there do not exist nonnegative numbers $\bar{\lambda}_1$ and $\bar{\lambda}_2$ such that $\nabla f(\bar{x}) = \bar{\lambda}_1 \nabla g_1(\bar{x}) + \bar{\lambda}_2 \nabla g_2(\bar{x})$. Hence, by Theorem above, $\bar{x}$ must fail GCQ. Indeed, $\bar{x}$ does not solve the local linearization $\min_{(x_1,x_2)} x_1$ subject to $-x_2 \geq 0$, $x_2 \geq 0$, which has no solution.

Definition: A solution $\bar{x}$ to the NLP satisfies the Cottle Constraint Qualification (CCQ) condition if there exists some $z \in \mathbb{R}^n$ such that $\nabla g_i(\bar{x}) \cdot z > 0$ for all $i \in I(\bar{x})$.

Theorem: Suppose $\bar{x}$ solves the NLP. If $\bar{x}$ satisfies the CCQ, then it satisfies the GCQ.

Proof: Suppose that $\overline{x}$ solves the NLP, satisfies the CCQ but not the GCQ.
Step 1: Choose $y, z \in \mathbb{R}^n$ such that $\nabla g_i(\overline{x}) \cdot z > 0$ for each $i \in I(\overline{x})$, $g_i(\overline{x}) + \nabla g_i(\overline{x}) \cdot (y - \overline{x}) \geq 0, i = 1, .., m$ and $\nabla f(\overline{x}) \cdot y < \nabla f(\overline{x}) \cdot \overline{x}$.
Next, choose $t \in (0, 1)$, define $x_t = y + tz$ and note that
$x_t - \overline{x} = y - \overline{x} + tz$.

Step 2: If $i \in I(\overline{x})$, then $g_i(\overline{x}) = 0$ and $\nabla g_i(\overline{x}) \cdot (y - \overline{x}) \geq 0$. Therefore,

$$\nabla g_i(\overline{x}) \cdot (x_t - \overline{x}) = \nabla g_i(\overline{x}) \cdot (y - \overline{x}) + t(\nabla g_i(\overline{x}) \cdot z) > 0$$

Similarly,

$$\nabla f(\overline{x}) \cdot (x_t - \overline{x}) = \nabla f(\overline{x}) \cdot (y - \overline{x}) + t(\nabla f(\overline{x}) \cdot z)$$

Since $\nabla f(\overline{x}) \cdot (y - \overline{x}) < 0$, there exists $t^* > 0$ such that

$$\begin{aligned} \nabla g_i(\overline{x}) \cdot (x_{t^*} - \overline{x}) &> 0 \text{ for all } i \in I(\overline{x}) \\ \nabla f(\overline{x}) \cdot (x_{t^*} - \overline{x}) &< 0 \end{aligned}$$

Step 3: Computing the directional derivatives, it follows that for all $i \in I(\overline{x})$

$$\begin{aligned}\lim_{\alpha \to 0+} \frac{g_i(\overline{x} + \alpha(x_{t^*} - \overline{x}))}{\alpha} &= \lim_{\alpha \to 0+} \frac{g_i(\overline{x} + \alpha(x_{t^*} - \overline{x})) - g_i(\overline{x})}{\alpha} \\ &= \nabla g_i(\overline{x}) \cdot (x_{t^*} - \overline{x}) > 0\end{aligned}$$

and

$$\lim_{\alpha \to 0+} \frac{f(\overline{x} + \alpha(x_{t^*} - \overline{x})) - f(\overline{x})}{\alpha} = \nabla f(\overline{x}) \cdot (x_{t^*} - \overline{x}) < 0.$$

Since each g_i is continuous and $g_i(\overline{x}) > 0$ if $i \notin I(\overline{x})$, we can find $\alpha^* > 0$ so that

$$\begin{aligned}
\frac{g_i(\overline{x} + \alpha^*(x_{t^*} - \overline{x}))}{\alpha^*} &> 0 \text{ for all } i \in I(\overline{x}) \\
\frac{f(\overline{x} + \alpha^*(x_{t^*} - \overline{x})) - f(\overline{x})}{\alpha^*} &< 0 \\
g_i(\overline{x} + \alpha^*(x_{t^*} - \overline{x})) &> 0 \text{ for all } i \notin I(\overline{x})
\end{aligned}$$

Summarizing, $g_i(\overline{x} + \alpha^*(x_{t^*} - \overline{x})) > 0$ for all $i = 1, .., m$ but $f(\overline{x} + \alpha^*(x_{t^*} - \overline{x})) - f(\overline{x}) < 0$. This is impossible if $\overline{x}$ solves the minimization problem.

Definition: A solution $\bar{x}$ to the NLP satisfies the Linear Independence Constraint Qualification (LICQ) condition if the collection $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ is linearly independent.

Theorem: Suppose $\bar{x}$ solves the NLP. If $\bar{x}$ satisfies the LICQ, then it satisfies the CCQ.

Proof: Without loss of generality, let $I(\bar{x}) = \{1, ..., r\}$ and let $A = [\nabla g_1(\bar{x}) \vdots \cdots \vdots \nabla g_r(\bar{x})]_{n \times r}$. If $\bar{x}$ satisfies LICQ, $rankA = r$. Hence $rankA^T = r$ as well. Since A^T is $r \times n$, $R(A^T) = \mathbb{R}^r$, i.e., there exists some $z \in \mathbb{R}^n$ such that $A^T z = 1_{r \times 1}$. Hence $\nabla g_i(\bar{x}) \cdot z > 0$ for all $i \in I(\bar{x})$.

Corollary: Suppose $\bar{x}$ solves the NLP. If $\bar{x}$ satisfies the LICQ, then it satisfies the KT conditions for NLP.

KT sufficiency theorem

When is a solution to the KT conditions a solution to NLP?

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then: (1) f is convex if and only if $f(y) - f(x) \geq \nabla f(x) \cdot (y - x)$ for every x and y. (2) f is concave if and only if $f(y) - f(x) \leq \nabla f(x) \cdot (y - x)$ for every x and y.

Proof: Homework

Theorem: Suppose $\bar{x} \in \mathbb{R}^n$ satisfies the KT conditions for NLP. If f is convex and each g_i is concave, then $\bar{x}$ solves the NLP.

Proof: Choose $x \in \mathbb{R}^n$ such that $g_i(x) \geq 0$ for all $i = 1, ..., m$. We need to show that $f(x) \geq f(\bar{x})$. Note that

$$\begin{aligned}
f(x) - f(\bar{x}) &\geq \nabla f(\bar{x}) \cdot (x - \bar{x}) \\
&= \left[\sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x})\right] \cdot (x - \bar{x}) \\
&= \sum_{i=1}^{m} \bar{\lambda}_i \left[\nabla g_i(\bar{x}) \cdot (x - \bar{x})\right] \\
&\geq \sum_{i=1}^{m} \bar{\lambda}_i \left[g_i(x) - g_i(\bar{x})\right] \\
&= \sum_{i=1}^{m} \bar{\lambda}_i g_i(x) \\
&\geq 0,
\end{aligned}$$

as desired. The proof is complete.

Definition: A function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex (quasiconcave) if the lower-contour set $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ (the upper-contour set $\{\{x \in \mathbb{R}^n : f(x) \geq \alpha\}$) is convex for every $\alpha \in \mathbb{R}$.

Homework: (1) Prove that every convex function is quasiconvex, but not vice versa. Can a strictly concave function be quasiconvex? (2) Prove that f is quasiconvex if and only if $-f$ is quasiconcave. (3) Prove that f is quasiconvex if and only if $f(tx + (1-t)y) \leq \max\{f(x), f(y)\}$ for every $x, y \in \mathbb{R}^n$ and every $t \in [0, 1]$. Provide an analogous characterization for quasiconcavity.

Theorem: Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then (1) f is quasiconcave if and only if for every x and y, $f(y) - f(x) \geq 0$ implies $\nabla f(x) \cdot (y - x) \geq 0$, and (2) f is quasiconvex if and only if for every x and y, $f(y) - f(x) \leq 0$ implies $\nabla f(x) \cdot (y - x) \leq 0$.

Proof: Homework.

Definition: A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is pseudoconvex if for every x and y, $\nabla f(x) \cdot (y - x) \geq 0$ implies $f(y) - f(x) \geq 0$. A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is pseudoconcave if for every x and y, $\nabla f(x) \cdot (y - x) \leq 0$ implies $f(y) - f(x) \leq 0$.

Homework: Show that f is pseudoconvex iff $-f$ is pseudoconcave. Show that every differentiable convex function is pseudoconvex. It can be shown that every pseudoconvex function is quasiconvex, but do not attempt to show this. Instead show that $f(x) = x^3$ is quasiconvex but not pseudoconvex.

We can now extend the KT sufficiency theorem.

Theorem: Suppose $(\bar{x}, \bar{\lambda})$ is a KT pair for NLP. If f is pseudoconvex and every g_i is quasiconcave, then x solves the NLP.

Proof: Choose $x \in \mathbb{R}^n$ such that $g_i(x) \geq 0$ for all i. If $i \in I(\bar{x})$, then $g_i(\bar{x}) = 0$ so $g_i(x) - g_i(\bar{x}) \geq 0$. This implies $\nabla g_i(\bar{x}) \cdot (x - \bar{x}) \geq 0$. Therefore $\sum_{i=1}^{m} \bar{\lambda}_i [\nabla g_i(\bar{x}) \cdot (x - \bar{x})] \geq 0$. This implies that $\nabla f(\bar{x}) \cdot (x - \bar{x}) \geq 0$. Since f is pseudoconvex, $f(x) - f(\bar{x}) \geq 0$, as desired.

Convexity is a useful assumption for the KT necessity theorem as well. In particular convexity can provide an alternate route to the CCQ.

Theorem: Suppose $\bar{x}$ solves the NLP. If each g_i is concave and if there exists x^* such that $g_i(x^*) > 0$ for all i, then $\bar{x}$ satisfies the CCQ.

Proof: Let $z = x^* - \bar{x}$ and choose $i \in I(\bar{x})$. Then

$$\begin{aligned} \nabla g_i(\bar{x}) \cdot z &= \nabla g_i(\bar{x}) \cdot (x^* - \bar{x}) \\ &\geq g_i(x^*) - g_i(\bar{x}) \\ &= g_i(x^*) \\ &> 0. \end{aligned}$$

Remark: For a NLP with concave constraints, the existence of an x^* such that $g_i(x^*) > 0$ for all i is called the Slater Constraint Qualification condition.

Comparative statics

Consider the problem $\min_{x\in\mathbb{R}^n} F(x,\alpha)$ subject to $G_i(x,\alpha)\geq 0$ for all $i=1,\ldots,m$, where $\alpha\in\mathbb{R}^p$ and all functions are differentiable. Call this problem P_α. Let $v(\alpha)$ be the value of P_α.

Definition: A vector $\bar{x}\in\mathbb{R}^n$ is a regular solution to $P_{\bar{\alpha}}$ if there exist $\varepsilon>0$, and $\hat{x}:B_\varepsilon(\bar{\alpha})\rightarrow\mathbb{R}^n$ such that (1) $\bar{x}$ satisfies the KT conditions, (2) $\hat{x}(\alpha)$ solves the problem P_α for every $\alpha\in B_\varepsilon(\bar{\alpha})$, and $\hat{x}(\bar{\alpha})=\bar{x}$, (3) $I(\hat{x}(\alpha))=I(\bar{x})$ for all $\alpha\in B_\varepsilon(\bar{\alpha})$, and (4) $\hat{x}$ is differentiable at $\bar{\alpha}$.

Theorem: Suppose that $\bar{x}$ is a regular solution to $P_{\bar{\alpha}}$ with associated multiplier vector $\bar{\lambda}$. Then

$$\nabla v(\bar{\alpha})=\nabla_\alpha F(\bar{x},\bar{\alpha})-\sum_{i=1}^m\bar{\lambda}_i\nabla_\alpha G_i(\bar{x},\bar{\alpha}).$$

Proof: Since $v(\alpha) = F(\hat{x}(\alpha), \alpha)$ for every $\alpha \in B_\varepsilon(\bar{\alpha})$ and F and $\hat{x}$ are differentiable at $\bar{\alpha}$, so is v. Writing $\hat{x}(\bar{\alpha}) = (\hat{x}_1(\bar{\alpha}), ..., \hat{x}_n(\bar{\alpha}))$,

$$\begin{aligned}
\nabla v(\bar{\alpha}) &= \nabla_\alpha F(\hat{x}(\bar{\alpha}), \bar{\alpha}) + [\nabla \hat{x}_1(\bar{\alpha}) \vdots \cdots \vdots \nabla \hat{x}_n(\bar{\alpha})] \nabla_x F(\hat{x}(\bar{\alpha}), \bar{\alpha}) \\
&= \nabla_\alpha F(\bar{x}, \bar{\alpha}) + [\nabla \hat{x}_1(\bar{\alpha}) \vdots \cdots \vdots \nabla \hat{x}_n(\bar{\alpha})] \nabla_x F(\bar{x}, \bar{\alpha}) \\
&= \nabla_\alpha F(\bar{x}, \bar{\alpha}) + [\nabla \hat{x}_1(\bar{\alpha}) \vdots \cdots \vdots \nabla \hat{x}_n(\bar{\alpha})] \sum_{i \in I(\bar{x})} \bar{\lambda}_i \nabla_x G_i(\bar{x}, \bar{\alpha}).
\end{aligned}$$

Next, observe that

$$i \in I(\bar{x}) \Rightarrow G_i(\hat{x}(\alpha), \alpha) = 0 \text{ for all } \alpha \in B_\varepsilon(\bar{\alpha}).$$

Proof cont'd.
It follows that

$$\begin{aligned} i &\in I(\bar{x}) \\ &\Rightarrow \nabla_\alpha G_i(\bar{x}, \bar{\alpha}) + [\nabla \hat{x}_1(\bar{\alpha}) \vdots \cdots \vdots \nabla \hat{x}_n(\bar{\alpha})] \nabla_x G_i(\bar{x}, \bar{\alpha}) = 0. \end{aligned}$$

Substituting, we obtain

$$\begin{aligned} \nabla v(\bar{\alpha}) &= \nabla_\alpha F(\bar{x}, \bar{\alpha}) - \sum_{i \in I(\bar{x})} \bar{\lambda}_i \nabla_\alpha G_i(\bar{x}, \bar{\alpha}) \\ &= \nabla_\alpha F(\bar{x}, \bar{\alpha}) - \sum_{i=1}^{m} \bar{\lambda}_i \nabla_\alpha G_i(\bar{x}, \bar{\alpha}) \end{aligned}$$

as desired.

Remark: Suppose that $\bar{x}$ is a regular solution to $P_{\bar{\alpha}}$ with associated multiplier vector $\bar{\lambda}$. If $m = 0$, then $\nabla v(\bar{\alpha}) = \nabla_\alpha F(\bar{x}, \bar{\alpha})$. If $p = m$, $G_i(x, \alpha) = g_i(x) - \alpha_i$, and $F(x, \alpha) = f(x)$, then $\nabla v(\bar{\alpha}) = \bar{\lambda}$.

Application: Cost minimization. Let $\bar{x}$ solve $\min_{x \in \mathbb{R}^n} \bar{w} \cdot x$ subject to $g(x) \geq \bar{y}$ and $C(\bar{w}, \bar{y}) = \bar{w} \cdot \bar{x}$. Then, by Theorem above, $\nabla_w C(\bar{w}, \bar{y}) = \bar{x}$, yielding Shephard's Lemma.

Application: Profit maximization. Let $(\bar{y}, \bar{x})$ solve $\max_{(y,x) \in \mathbb{R} \times \mathbb{R}^n} py - w \cdot x$ subject to $g(x) \geq y$ and $(y, x) \geq 0$. Let $\pi(p, w)$ be the value of the problem. Consider the minimization problem $\min -py + w \cdot x$ subject to $g(x) \geq y$ and $(y, x) \geq 0$. Let $\sigma(p, w)$ be the value of the minimization problem. Then (1) $\pi(p, w) = -\sigma(p, w)$, and (2) by Theorem above, $\nabla \sigma(p, w) = (-\bar{y}, \bar{x})$. Hotelling's Lemma follows: $\nabla \pi(p, w) = (\bar{y}, -\bar{x})$.

Example

Let us find the cost function for the technology given by the production function $f(x_1, x_2) = \min\{2x_1 + x_2, x_1 + 2x_2\}$. The cost minimization problem is

$$\min w_1 x_1 + w_2 x_2$$
$$\text{subject to } 2x_1 + x_2 \geq y,\ x_1 + 2x_2 \geq y,\ x_1 \geq 0,\ x_2 \geq 0.$$

The KT conditions are:

$$\begin{array}{ccc}
w_1 - 2\lambda_1 - \lambda_2 \geq 0 & x_1 \geq 0 & (w_1 - 2\lambda_1 - \lambda_2)x_1 = 0 \\
w_2 - \lambda_1 - 2\lambda_2 \geq 0 & x_2 \geq 0 & (w_2 - \lambda_1 - 2\lambda_2)x_2 = 0 \\
2x_1 + x_2 - y \geq 0 & \lambda_1 \geq 0 & (2x_1 + x_2 - y)\lambda_1 = 0 \\
x_1 + 2x_2 - y \geq 0 & \lambda_2 \geq 0 & (x_1 + 2x_2 - y)\lambda_2 = 0
\end{array}$$

Is there a solution with $x_1 > 0$, $x_2 > 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$? $x_1 = x_2 = y/3$, $\lambda_1 = (2w_1 - w_2)/3$, $\lambda_2 = (2w_2 - w_1)/3$. This requires $1/2 < w_1/w_2 < 2$. Is there a solution with $x_1 > 0$, $x_2 = 0$, $\lambda_1 = 0$ and $\lambda_2 > 0$? In any such solution, $w_1/w_2 \leq 1/2$, $x_1 = y$, $\lambda_2 = w_1$. Is there a solution with $x_1 = 0$, $x_2 > 0$, $\lambda_1 > 0$ and $\lambda_2 = 0$? In any such solution, $w_1/w_2 \geq 1/2$, $x_2 = y$, $\lambda_1 = w_2$. Hence the cost function is given by

$$C(w_1, w_2, y) = \begin{cases} w_1 y & \text{if } w_1/w_2 \leq 1/2 \\ (w_1 + w_2)y/3 & \text{if } 1/2 < w_1/w_2 < 2 \\ w_2 y & \text{if } 2 \leq w_1/w_2. \end{cases}$$

Maximization problems

Consider the problem $\max f(x)$ subject to $g_i(x) \leq 0$ for all $i = 1, ..., m$ where all functions are differentiable. Let $I(x) = \{i : g_i(x) = 0\}$. Prove the following results as homework.

Theorem: If $\bar{x}$ solves the maximization problem above and $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ is a linearly independent set, then there exists $\bar{\lambda} \in \mathbb{R}^m_+$ such that $\nabla f(\bar{x}) = \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(\bar{x})$ and $\bar{\lambda}_i g_i(\bar{x}) = 0$ for all i.

Theorem: Suppose that (1) $g_i(\bar{x}) \leq 0$ for all $i = 1, ..., m$, (2) there exists a set $\{\bar{\lambda}_i\}_{i \in I(\bar{x})}$ of positive numbers such that $\nabla f(\bar{x}) = \sum_{i \in I(\bar{x})} \bar{\lambda}_i \nabla g_i(\bar{x})$, (3) f is pseudoconcave, (4) each g_i is quasiconvex. Then $\bar{x}$ solves the maximization problem above.

Consumer theory

Consider the following utility maximization problem

$$\max_{x \in \mathbb{R}^n} u(x) \text{ subject to } p \cdot x \leq y \text{ and } x \geq 0.$$

The KT conditions are:

$$\begin{array}{lll} \nabla u(x) \leq \lambda p & x \geq 0 & [\nabla u(x) - \lambda p] \cdot x = 0 \\ p \cdot x \leq y & \lambda \geq 0 & (p \cdot x - y)\lambda = 0 \end{array}$$

(Why?) Note that λ is interpreted as the "marginal utility of income." Call the solutions $x(p, y)$ the demand functions and $v(p, y) := u(x(p, y))$ the indirect utility function.

Properties of v:

1. v is 0-homogeneous in (p, y).
2. v is non-decreasing in y and nonincreasing in p.
3. v is quasiconvex in (p, y). To see this, choose α and (p_1, y_1), $(p_2, y_2) \in A = \{(p, y) : v(p, y) \leq \alpha\}$. We must show that $(\bar{p}, \bar{y}) = t_1(p_1, y_1) + t_2(p_2, y_2) \in A$ if $t_1, t_2 \geq 0$ and $t_1 + t_2 = 1$. Choose, for $i = 1, 2$, $x_i \in B_i = \{x \in \mathbb{R}^n_+ : p_i \cdot x \leq y_i\}$ such that $u(x_i) = v(p_i, y_i)$. If $\bar{x}$ solves $\max\{u(x) : \bar{p} \cdot x \leq \bar{y}$ and $x \geq 0\}$, then $\bar{x} \in B_i$ for some i. (Why? Show that $\bar{x} \in B_1 \cup B_2$.) Hence $u(\bar{x}) \leq \max\{u(x_1), u(x_2)\}) \leq \alpha$ as desired.

Now consider the following expenditure minimization problem:

$$\min_{x \in \mathbb{R}^n} p \cdot x \text{ subject to } u(x) \geq \bar{u} \text{ and } x \geq 0.$$

Let $h(p, \bar{u})$ be the solution and $e(p, \bar{u}) = p \cdot h(p, \bar{u})$. The expenditure function e is mathematically identical to the cost function.

1. $e(\cdot, \bar{u})$ is 1-homogenous.
2. $e(\cdot, \bar{u})$ is nondecreasing.
3. $e(\cdot, \bar{u})$ is concave.

Furthermore:

4. By Shephard's Lemma, $\frac{\partial e}{\partial p_i} = h_i(p, \bar{u})$.

Is there an analogue of 4 for $v(p, y)$?

Theorem: (Roy's identity) $x_i(p, y) = \frac{-\frac{\partial v}{\partial p_i}}{\frac{\partial v}{\partial y}}$.

Proof: Apply Envelope Theorem on v, with parameters (p, y).

Relationships between utility maximization and expenditure minimization when preferences are strictly monotonic and the two problems have unique solutions:

1. $h(p, v(p, y)) = x(p, y)$. To see this, first note that $x(p, y)$ is feasible in the problem $\min\{p \cdot x : u(x) \geq v(p, y) \text{ and } x \geq 0\}$. Suppose there exists some $\hat{x} \geq 0$ such that $u(\hat{x}) \geq v(p, y)$ and $p \cdot \hat{x} < p \cdot x(p, y)$. Then $p \cdot \hat{x} < y$. Hence there exists $\varepsilon > 0$ such that $p \cdot [\hat{x} + (\varepsilon, ..., \varepsilon)] < y$, but $u(\hat{x} + (\varepsilon, ..., \varepsilon)) > u(\hat{x}) \geq v(p, y)$, a contradiction.
2. $x(p, e(p, \bar{u})) = h(p, \bar{u})$.
3. $e(p, v(p, y)) = y$.
4. $v(p, e(p, \bar{u})) = \bar{u}$.

Theorem: (Slutsky) $\frac{\partial x_i}{\partial p_j}(p, y) = \frac{\partial h_i}{\partial p_j}(p, v(p, y)) - x_j \frac{\partial x_i}{\partial y}(p, y)$.
Proof: Differentiate both sides of 2 above then evaluate at $\bar{u} = v(p, y)$.