Optimality Conditions for Linearly Constrained Optimization

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye
(LY: Chapters 7.1-7.2, 11.1-11.3)
General Optimization Problems

Let the problem have the general mathematical programming (MP) form

\[
\begin{align*}
\text{(P)} & \quad \text{minimize} \quad f(x) \\
\text{subject to} & \quad x \in \mathcal{F}.
\end{align*}
\]

In all forms of mathematical programming, a feasible solution of a given problem is a vector that satisfies the constraints of the problem, that is, in \( \mathcal{F} \).

First question: How does one recognize or certify an optimal solution to a generally constrained and objectived optimization problem?

Answer: Optimality Condition Theory again.
Let $f$ be a differentiable function on $\mathbb{R}^n$. If point $\bar{x} \in \mathbb{R}^n$ and there exists a vector $d$ such that

$$\nabla f(\bar{x})d < 0,$$

then there exists a scalar $\bar{\tau} > 0$ such that

$$f(\bar{x} + \tau d) < f(\bar{x}) \text{ for all } \tau \in (0, \bar{\tau}).$$

The vector $d$ (above) is called a descent direction at $\bar{x}$. If $\nabla f(\bar{x}) \neq 0$, then $\nabla f(\bar{x})$ is the direction of steepest ascent and $-\nabla f(\bar{x})$ is the direction of steepest descent at $\bar{x}$.

Denote by $\mathcal{D}^d_{\bar{x}}$ the set of descent directions at $\bar{x}$, that is,

$$\mathcal{D}^d_{\bar{x}} = \{d \in \mathbb{R}^n : \nabla f(\bar{x})d < 0\}.$$
At feasible point $\bar{x}$, a feasible direction is

$$D^f_{\bar{x}} := \{ d \in \mathbb{R}^n : d \neq 0, \bar{x} + \lambda d \in \mathcal{F} \text{ for all small } \lambda > 0 \}.$$ 

Examples:

$$\mathcal{F} = \mathbb{R}^n \Rightarrow D^f = \mathbb{R}^n.$$ 

$$\mathcal{F} = \{ x : Ax = b \} \Rightarrow D^f = \{ d : Ad = 0 \}.$$ 

$$\mathcal{F} = \{ x : Ax \geq b \} \Rightarrow D^f = \{ d : A_i d \geq 0, \forall i \in A(\bar{x}) \},$$

where the active or binding constraint set $A(\bar{x}) := \{ i : A_i \bar{x} = b_i \}$. 
Optimality Conditions: given a feasible solution or point $\bar{x}$, what are the necessary conditions for $\bar{x}$ to be a local optimizer?

A general answer would be: there exists no direction at $\bar{x}$ that is both descent and feasible. Or the intersection of $D^d_{\bar{x}}$ and $D^f_{\bar{x}}$ must be empty.
Consider the unconstrained problem, where $f$ is differentiable on $\mathbb{R}^n$,

\[
\begin{align*}
\text{(UP)} & \\
\text{minimize} & & f(x) \\
\text{subject to} & & x \in \mathbb{R}^n.
\end{align*}
\]

$D_f^x = \mathbb{R}^n$, so that $D^d_f = \{d \in \mathbb{R}^n : \nabla f(\bar{x})d < 0\} = \emptyset$:

**Theorem 1** Let $\bar{x}$ be a (local) minimizer of (UP). If the functions $f$ is continuously differentiable at $\bar{x}$, then

$$\nabla f(\bar{x}) = 0.$$
Linear Equality-Constrained Problems

Consider the linear equality-constrained problem, where $f$ is differentiable on $\mathbb{R}^n$,

\[
\text{(LEP)} \quad \begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad Ax = b.
\end{align*}
\]

**Theorem 2** (the Lagrange Theorem) Let $\bar{x}$ be a (local) minimizer of (LEP). If the functions $f$ is continuously differentiable at $\bar{x}$, then

\[
\nabla f(\bar{x}) = \bar{y}^T A
\]

for some $\bar{y} = (\bar{y}_1; \ldots; \bar{y}_m) \in \mathbb{R}^m$, which are called Lagrange or dual multipliers.

The geometric interpretation: the objective gradient vector is perpendicular to or the objective level set tangents the constraint hyperplanes.
Proof

Consider feasible direction space

$$\mathcal{F} = \{ x : Ax = b \} \Rightarrow D_{x}^{f} = \{ d : Ad = 0 \}.$$ 

If $\bar{x}$ is a local optimizer, then the intersection of the descent and feasible direction sets at $\bar{x}$ must be empty or

$$Ad = 0, \nabla f(\bar{x})d \neq 0$$

has no feasible solution for $d$. By the Alternative System Theorem it must be true that its alternative system has a solution, that is, there is $\tilde{y} \in R^{n}$ such that

$$\nabla f(\bar{x}) = \tilde{y}^{T}A = \sum_{i=1}^{m} \tilde{y}_{i}A_{i}.$$
The Logarithmic Barrier Function Problem

Consider the problem

\[
\text{minimize} \quad - \sum_{j=1}^{n} \log x_j \\
\text{subject to} \quad Ax = b, \quad x \geq 0
\]

The non-negativity constraint can be removed if the feasible region has an "interior", that is, there is a feasible solution such that \( x > 0 \). Thus, if a minimizer \( \bar{x} \) exists, then \( \bar{x} > 0 \) and

\[
-e^T \bar{X}^{-1} = \bar{y}^T A = \sum_{i=1}^{m} \bar{y}_i A_i.
\]
Let us now consider the inequality-constrained problem

\[
\text{(LIP)} \quad \begin{array}{l}
\text{minimize} & f(x) \\
\text{subject to} & Ax \geq b.
\end{array}
\]

**Theorem 3** (the KKT Theorem) Let \( \bar{x} \) be a (local) minimizer of \( \text{(LIP)} \). If the functions \( f \) is continuously differentiable at \( \bar{x} \), then

\[
\nabla f(\bar{x}) = \bar{y}^T A, \quad \bar{y} \geq 0
\]

for some \( \bar{y} = (\bar{y}_1; \ldots; \bar{y}_m) \in R^m \), which are called Lagrange or dual multipliers, and \( \bar{y}_i = 0 \), if \( i \not\in A(\bar{x}) \).

The geometric interpretation: the objective gradient vector is in the cone generated by the normal directions of the active-constraint hyperplanes.
\[ \mathcal{F} = \{ x : Ax \geq b \} \Rightarrow \mathcal{D}_{\bar{x}}^f = \{ d : A_i d \geq 0, \forall i \in \mathcal{A}(\bar{x}) \}, \]

or

\[ \mathcal{D}_{\bar{x}}^f = \{ d : \bar{A} d \geq 0 \}, \]

where \( \bar{A} \) corresponds to those active constraints. If \( \bar{x} \) is a local optimizer, then the intersection of the descent and feasible direction sets at \( \bar{x} \) must be empty or

\[ \bar{A} d \geq 0, \ \nabla f(\bar{x}) d < 0 \]

has no feasible solution. By the Alternative System Theorem it must be true that its alternative system has a solution, that is, there is \( \bar{y} \geq 0 \) such that

\[ \nabla f(\bar{x}) = \bar{y}^T \bar{A} = \sum_{i \in \mathcal{A}(\bar{x})} \bar{y}_i A_i = \sum_i \bar{y}_i A_i, \]

when let \( \bar{y}_i = 0 \) for all \( i \notin \mathcal{A}(\bar{x}) \). Then we prove the theorem.
We now consider optimality conditions for problems having both inequality and equality constraints. These can be denoted

\[
\begin{align*}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & Ax = b \\
& x \geq 0
\end{align*}
\]

For any feasible point \( \bar{x} \) of (P) we have the sets

\[
\mathcal{A}(\bar{x}) = \{ j : \bar{x}_j = 0 \}
\]

\[
\mathcal{D}^d_x = \{ d : \nabla f(\bar{x})d < 0 \}.
\]
Theorem 4  Let $\bar{x}$ be a local minimizer for (P). Then there exist multipliers $\bar{y}, \bar{s}$ such that

$$\nabla f(\bar{x}) = \bar{y}^T A + \bar{s}^T$$

$$\bar{s} \geq 0$$

$$\bar{s}_j = 0 \text{ if } j \notin A(\bar{x}).$$
Optimality and Complementarity Conditions

\[ x_j (\nabla f(x) - y^T A)_j = 0, \quad \forall j = 1, \ldots, n \]
\[ Ax = b \]
\[ \nabla f(x) - y^T A \geq 0 \]
\[ x \geq 0. \]

\[ x_j s_j = 0, \quad \forall j = 1, \ldots, n \]
\[ Ax = b \]
\[ \nabla f(x) - y^T A - s^T = 0 \]
\[ x, s \geq 0 \]
Sufficient Optimality Conditions

**Theorem 5** If $f$ is a differentiable convex function in the feasible region and the feasible region is a convex set, then the (first-order) KKT optimality conditions are sufficient for the global optimality of a feasible solution.

**Corollary 1** If $f$ is differentiable convex function in the feasible region, then the (first-order) KKT optimality conditions are sufficient for the global optimality of a feasible solution for linearly constrained optimization.

How to check convexity, say $f(x) = x^3$?

- Hessian matrix is PSD in the feasible region.
- Epigraph is a convex set.
LCCP Examples: Linear Optimization

\[ \text{(LP)} \quad \text{minimize} \quad c^T x \]
\[ \text{subject to} \quad Ax = b, \ x \geq 0. \]

For any feasible \( x \) of (LP), it’s optimal if for some \( y, s \)

\[ x_j s_j = 0, \ \forall j = 1, \ldots, n \]
\[ Ax = b \]
\[ \nabla (c^T x) = c^T = y^T A + s^T \]
\[ x, s \geq 0. \]

Here, \( y \) are Lagrange multipliers of equality constraints, and \( s \) (reduced cost or dual slack vector in LP) are Lagrange multipliers for \( x \geq 0 \).
LCCP Examples: Barrier Optimization

\[ f(x) = c^T x - \mu \sum_{j=1}^{n} \log(x_j), \]

for some fixed \( \mu > 0 \). Assume that interior of the feasible region is not empty:

\[
\begin{align*}
Ax &= b \\
c_j - \frac{\mu}{x_j} - (y^T A)_j &= 0, \forall j = 1, \ldots, n \\
x &> 0.
\end{align*}
\]

Let \( s_j = \frac{\mu}{x_j} \) for all \( j \) (note that this \( s \) is not the \( s \) in the KKT condition of \( f(x) \)). Then

\[
\begin{align*}
x_j s_j &= \mu, \forall j = 1, \ldots, n, \\
Ax &= b, \\
A^T y + s &= c, \\
(x, s) &> 0.
\end{align*}
\]
Proof of Uniqueness

Solution pair of \((x, s)\) of the barrier optimization problem is unique.

Suppose there two different pair \((x^1, s^1)\) and \((x^2, s^2)\). Note that

\[(s^1 - s^2)^T(x^1 - x^2) = 0.\]

Thus, there is \(j\) such that

\[(s^1_j - s^2_j)(x^1_j - x^2_j) > 0.\]

If \(x^1_j > x^2_j\), then \(s^1_j < s^2_j\) since \(x^1_j s^1_j = x^2_j s^2_j = \mu > 0\), which leads to \((s^1_j - s^2_j)(x^1_j - x^2_j) < 0\) – a contradiction. Similarly, one cannot have \(x^1_j < x^2_j\).
KKT Application: Fisher’s Equilibrium Price

Player $i \in B$’s optimization problem for given prices $p_j, j \in G$.

\[
\text{maximize} \quad u_i^T x_i := \sum_{j \in G} u_{ij} x_{ij}
\]
\[
\text{subject to} \quad p_i^T x_i := \sum_{j \in G} p_j x_{ij} \leq w_i,
\]
\[
x_{ij} \geq 0, \quad \forall j,
\]

Assume that the amount of each good is $s_j$. The equilibrium price vector is the one that for all $j \in G$

\[
\sum_{i \in B} x(p)_{ij} = s_j
\]
Example of Fisher’s Equilibrium Price

There two goods, x and y, each with 1 unit on the market. Buyer 1, 2's optimization problems for given prices $p_x, p_y$.

\[
\begin{align*}
\text{maximize} & \quad 2x_1 + y_1 \\
\text{subject to} & \quad p_x \cdot x_1 + p_y \cdot y_1 \leq 5, \\
& \quad x_1, y_1 \geq 0;
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad 3x_2 + y_2 \\
\text{subject to} & \quad p_x \cdot x_2 + p_y \cdot y_2 \leq 8, \\
& \quad x_2, y_2 \geq 0.
\end{align*}
\]

\[
p_x = \frac{26}{3}, \quad p_y = \frac{13}{3}, \quad x_1 = \frac{1}{13}, \quad y_1 = 1, \quad x_2 = \frac{12}{13}, \quad y_2 = 0
\]
Equilibrium Price Conditions

Player $i \in B$’s dual problem for given prices $p_j, j \in G$.

minimize $w_i y_i$

subject to $p y_i \geq u_i, y_i \geq 0$

The necessary and sufficient conditions for an equilibrium point $x_i, p$ are:

$p^T x_i \leq w_i, x_i \geq 0, \forall i,$

$p_j y_i \geq u_{ij}, y_i \geq 0, \forall i, j,$

$u_i^T x_i = w_i y_i, \forall i,$

$\sum_i x_{ij} = s_j, \forall j.$
Equilibrium Price Conditions (continued)

These conditions can be represented by

$$
\sum_j s_j p_j \leq \sum_i w_i, \quad x_i \geq 0, \quad \forall i,
$$

$$
\frac{u_i^T x_i}{w_i} \cdot p_j \geq u_{ij}, \quad \forall i, j,
$$

$$
\sum_i x_{ij} = s_j, \quad \forall j.
$$

since from the second inequality (after multiplying $x_{ij}$ to both sides and take sum over $j$) we have

$$
p^T x_i \geq w_i, \quad \forall i.
$$

Then, from the rest conditions

$$
\sum_i w_i \geq \sum_j s_j p_j = \sum_i p^T x_i \geq \sum_i w_i.
$$

Thus, these conditions imply $p^T x_i = w_i, \quad \forall i$. 


Equilibrium Price Property

If $u_{ij}$ has at least one positive coefficient for every $j$, then we must have $p_j > 0$ for every $j$ at every equilibrium. Moreover, the second inequality can be rewritten as

$$\log(u_i^T x_i) + \log(p_j) \geq \log(w_i) + \log(u_{ij}), \ \forall i, j, u_{ij} > 0.$$ 

The function on the left is (strictly) concave in $x_i$ and $p_j$. Thus,

**Theorem 6** The equilibrium set of the Fisher Market is convex, and the equilibrium price vector is unique.
Aggregate Social Optimization

maximize \[ \sum_{i \in B} w_i \log(u_i^T x_i) \]
subject to \[ \sum_{i \in B} x_{ij} = s_j, \quad \forall j \in G \]
\[ x_{ij} \geq 0, \quad \forall i, j, \]

**Theorem 7** (Eisenberg and Gale 1959) Optimal dual (Lagrange) multiplier vector of equality constraints is an equilibrium price vector.
Let $y_i = \frac{u_i^T x_i}{w_i}$. Then, these conditions are identical to the equilibrium price conditions, since

$$y_i = \frac{u_i^T x_i}{w_i} \geq \frac{u_{ij}}{p_j}, \forall i, j.$$
Rewrite Aggregate Social Optimization

maximize \[ \sum_{i \in B} w_i \log u_i \]
subject to \[ \sum_{j \in G} u_{ij}^T x_{ij} - u_i = 0, \quad \forall i \in B \]
\[ \sum_{i \in B} x_{ij} \leq s_j, \quad \forall j \in G \]
\[ x_{ij} \geq 0, \quad s_i \geq 0, \quad \forall i, j, \]

This is called the weighted analytic center problem.

**Question:** Is the price vector \( p \) unique when at least one \( u_{ij} > 0 \) among \( i \in B \) and \( u_{ij} > 0 \) among \( j \in G \).
Aggregate Example

maximize \[ 5 \log(2x_1 + y_1) + 8 \log(3x_2 + y_2) \]
subject to \[ x_1 + x_2 = 1, \]
\[ y_1 + y_2 = 1, \]
\[ x_1, x_2, y_1, y_2 \geq 0. \]

Or

maximize \[ 5 \log(u_1) + 8 \log(u_2) \]
subject to \[ 2x_1 + y_1 - u_1 = 0, \]
\[ 3x_2 + y_2 - u_2 = 0, \]
\[ x_1 + x_2 = 1, \]
\[ y_1 + y_2 = 1, \]
\[ x_1, x_2, y_1, y_2 \geq 0. \]