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

$$(\mathrm{P}) \qquad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{F}. \end{array}$$

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.

Descent Direction

Let f be a differentiable function on R^n. If point $\bar{\mathbf{x}} \in R^n$ and there exists a vector $\mathbf{d}$ such that

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

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

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

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

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

$$\mathcal{D}^d_{\bar{\mathbf{x}}} = \{\mathbf{d} \in R^n : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\}.$$

Feasible Direction

At feasible point $\bar{\mathbf{x}}$, a feasible direction is

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

Examples:

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

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

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

where the active or binding constraint set $\mathcal{A}(\bar{\mathbf{x}}) := \{i :\ A_i\bar{\mathbf{x}} = b_i\}$.

Optimality Conditions

Optimality Conditions: given a feasible solution or point $\bar{\mathbf{x}}$, what are the necessary conditions for $\bar{\mathbf{x}}$ to be a local optimizer?

A general answer would be: there exists no direction at $\bar{\mathbf{x}}$ that is both descent and feasible. Or the intersection of $\mathcal{D}^d_{\bar{\mathbf{x}}}$ and $\mathcal{D}^f_{\bar{\mathbf{x}}}$ must be empty.

Unconstrained Problems

Consider the unconstrained problem, where f is differentiable on R^n,

$$(\text{UP}) \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in R^n. \end{array}$$

$\mathcal{D}^f_{\bar{\mathbf{x}}} = R^n$, so that $\mathcal{D}^d_{\bar{\mathbf{x}}} = \{\mathbf{d} \in R^n : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\} = \emptyset$:

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

$$\nabla f(\bar{\mathbf{x}}) = \mathbf{0}.$$

Linear Equality-Constrained Problems

Consider the linear equality-constrained problem, where f is differentiable on R^n,

$$(\text{LEP}) \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}. \end{array}$$

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

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

for some $\bar{\mathbf{y}} = (\bar{y}_1; \ldots; \bar{y}_m) \in R^m$, 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

$$\mathcal{F} = \{\mathbf{x} : \; A\mathbf{x} = \mathbf{b}\} \Rightarrow \mathcal{D}^f_{\bar{\mathbf{x}}} = \{\mathbf{d} : A\mathbf{d} = 0\}.$$

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

$$A\mathbf{d} = \mathbf{0}, \; \nabla f(\bar{\mathbf{x}})\mathbf{d} \neq 0$$

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

$$\nabla f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}^T A = \sum_{i=1}^{m} \bar{y}_i A_i.$$

The Logarithmic Barrier Function Problem

Consider the problem

$$\begin{array}{ll} \text{minimize} & -\sum_{j=1}^{n} \log x_j \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

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

$$-\mathbf{e}^T \bar{X}^{-1} = \bar{\mathbf{y}}^T A = \sum_{i=1}^{m} \bar{y}_i A_i.$$

The maximizer is calle the analytic center of the feasible region.

Linear Inequality-Constrained Problems

Let us now consider the inequality-constrained problem

$$
(\text{LIP}) \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} \geq \mathbf{b}. \end{array}
$$

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

$$
\nabla f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}^T A, \ \bar{\mathbf{y}} \geq \mathbf{0}
$$

for some $\bar{\mathbf{y}} = (\bar{y}_1; \ldots; \bar{y}_m) \in R^m$, which are called Lagrange or dual multipliers, and $\bar{y}_i = 0, \quad$ if $i \notin \mathcal{A}(\bar{\mathbf{x}})$.

The geometric interpretation: the objective gradient vector is in the cone generated by the normal directions of the active-constraint hyperplanes.

Proof

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

or

$$\mathcal{D}^f_{\bar{\mathbf{x}}} = \{\mathbf{d} : \bar{A}\mathbf{d} \geq \mathbf{0}\},$$

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

$$\bar{A}\mathbf{d} \geq \mathbf{0}, \; \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0$$

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

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

when let $\bar{y}_i = 0$ for all $i \notin \mathcal{A}(\bar{\mathbf{x}})$. Then we prove the theorem.

Optimization with Mixed Constraints

We now consider optimality conditions for problems having both inequality and equality constraints. These can be denoted

$$
\text{(P)} \qquad \begin{array}{lll} \text{minimize} & f(\mathbf{x}) & \\ \text{subject to} & A\mathbf{x} & = \mathbf{b} \\ & \mathbf{x} & \geq \mathbf{0} \end{array}
$$

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

$$
\begin{aligned}
\mathcal{A}(\bar{\mathbf{x}}) &= \{j : \bar{x}_j = 0\} \\
\mathcal{D}^d_{\bar{\mathbf{x}}} &= \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\}.
\end{aligned}
$$

The KKT Theorem Again

Theorem 4 *Let $\bar{\mathbf{x}}$ be a local minimizer for (P). Then there exist multipliers $\bar{\mathbf{y}}, ^{-}$ such that*

$$
\begin{aligned}
\nabla f(\bar{\mathbf{x}}) &= \bar{\mathbf{y}}^T A + \bar{\mathbf{s}}^T \\
\bar{\mathbf{s}} &\geq \mathbf{0} \\
\bar{s}_j &= 0 \quad \text{if } j \notin \mathcal{A}(\bar{\mathbf{x}}).
\end{aligned}
$$

Optimality and Complementarity Conditions

$$
\begin{aligned}
x_j(\nabla f(\mathbf{x}) - \mathbf{y}^T A)_j &= 0, \ \forall j = 1, \ldots, n \\
A\mathbf{x} &= \mathbf{b} \\
\nabla f(\mathbf{x}) - \mathbf{y}^T A &\geq \mathbf{0} \\
\mathbf{x} &\geq \mathbf{0}.
\end{aligned}
$$

$$
\begin{aligned}
x_j s_j &= 0, \ \forall j = 1, \ldots, n \\
A\mathbf{x} &= \mathbf{b} \\
\nabla f(\mathbf{x}) - \mathbf{y}^T A - \mathbf{s}^T &= \mathbf{0} \\
\mathbf{x}, \mathbf{s} &\geq \mathbf{0}
\end{aligned}
$$

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

$$
\begin{array}{lll}
(LP) & \text{minimize} & \mathbf{c}^T\mathbf{x} \\
 & \text{subject to} & A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}.
\end{array}
$$

For any feasible $\mathbf{x}$ of (LP), it's optimal if for some $\mathbf{y}, \mathbf{s}$

$$
\begin{array}{rcl}
x_j s_j & = & 0,\ \forall j = 1, \ldots, n \\
A\mathbf{x} & = & \mathbf{b} \\
\nabla(\mathbf{c}^T\mathbf{x}) = \mathbf{c}^T & = & \mathbf{y}^T A + \mathbf{s}^T \\
\mathbf{x}, \mathbf{s} & \geq & \mathbf{0}.
\end{array}
$$

Here, $\mathbf{y}$ are Lagrange multipliers of equality constraints, and $\mathbf{s}$ (reduced cost or dual slack vector in LP) are Lagrange multipliers for $\mathbf{x} \geq \mathbf{0}$.

LCCP Examples: Barrier Optimization

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

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

$$\begin{array}{rcl} A\mathbf{x} & = & \mathbf{b} \\ c_j - \frac{\mu}{x_j} - (\mathbf{y}^T A)_j & = & 0, \forall j = 1, \ldots, n \\ \mathbf{x} & > & \mathbf{0}. \end{array}$$

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

$$\begin{array}{rcl} x_j s_j & = & \mu, \ \forall j = 1, \ldots, n, \\ A\mathbf{x} & = & \mathbf{b}, \\ A^T\mathbf{y} + \mathbf{s} & = & \mathbf{c}, \\ (\mathbf{x}, \mathbf{s}) & > & \mathbf{0}. \end{array}$$

Proof of Uniqueness

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

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

$$(\mathbf{s}^1 - \mathbf{s}^2)^T(\mathbf{x}^1 - \mathbf{x}^2) = 0.$$

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 $x^1_j < x^2_j$.

Central Path for Linear Programming

Let $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ be the KKT solutions of the barried LP problem. Then the path

$$\mathcal{C} = \{(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu)) \in \text{int}\,\mathcal{F} : \; X\mathbf{s} = \mu\mathbf{e}, \; 0 < \mu < \infty\}\,;$$

is called the (primal and dual) central path of linear programming.

Theorem 6 *Let both (LP) and (LD) have interior feasible points for the given data set* (A, b, c). *Then for any* $0 < \mu < \infty$, *the central path point pair* $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ *exists and is unique.*

KKT Application: Fisher's Equilibrium Price

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

$$\begin{array}{ll} \text{maximize} & \mathbf{u}_i^T \mathbf{x}_i := \sum_{j \in G} u_{ij} x_{ij} \\ \text{subject to} & \mathbf{p}^T \mathbf{x}_i := \sum_{j \in G} p_j x_{ij} \le w_i, \\ & x_{ij} \ge 0, \quad \forall j, \end{array}$$

Assume that the amount of each good is $\bar{s}_j$. The equilinitum price vector is the one that for all $j \in G$

$$\sum_{i \in B} x(\mathbf{p})_{ij} = \bar{s}_j$$

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.

$$\begin{array}{ll} \text{maximize} & 2x_1 + y_1 \\ \text{subject to} & p_x \cdot x_1 + p_y \cdot y_1 \leq 5, \\ & x_1, y_1 \geq 0; \\ \text{maximize} & 3x_2 + y_2 \\ \text{subject to} & p_x \cdot x_2 + p_y \cdot y_2 \leq 8, \\ & x_2, y_2 \geq 0. \end{array}$$

$$p_x = \frac{26}{3},\ p_y = \frac{13}{3},\ x_1 = \frac{1}{13},\ y_1 = 1,\ x_2 = \frac{12}{13},\ y_2 = 0$$

Equilibrium Price Conditions

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

$$\begin{array}{ll} \text{minimize} & w_i y_i \\ \text{subject to} & \mathbf{p} y_i \geq \mathbf{u}_i, \; y_i \geq 0 \end{array}$$

The necessary and sufficient conditions for an equilibrium point $\mathbf{x}_i, \mathbf{p}$ are:

$$\begin{array}{lr} \mathbf{p}^T \mathbf{x}_i = w_i, \; \mathbf{x}_i \geq \mathbf{0}, & \forall i, \\ p_j y_i \geq u_{ij}, \; y_i \geq 0, & \forall i, j, \\ \mathbf{u}_i^T \mathbf{x}_i = w_i y_i, & \forall i, \\ \sum_i x_{ij} = \bar{s}_j, & \forall j. \end{array} \quad <=> \quad \begin{array}{lr} \mathbf{p}^T \mathbf{x}_i = w_i, \; \mathbf{x}_i \geq \mathbf{0}, & \forall i, \\ p_j \geq w_i \frac{u_{ij}}{\mathbf{u}_i^T \mathbf{x}_i}, & \forall i, j, \\ \sum_i x_{ij} = \bar{s}_j, & \forall j. \end{array}$$

Equilibrium Price Conditions (continued)

These conditions can be equivalently represented by

$$\begin{array}{ll} \sum_j \bar{s}_j p_j \le \sum_i w_i, \ \mathbf{x}_i \ge \mathbf{0}, & \forall i, \\ p_j \ge w_i \frac{u_{ij}}{\mathbf{u}_i^T \mathbf{x}_i}, & \forall i, j, \\ \sum_i x_{ij} = \bar{s}_j, & \forall j. \end{array}$$

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

$$\mathbf{p}^T \mathbf{x}_i \ge w_i, \ \forall i.$$

Then, from the rest conditions

$$\sum_i w_i \ge \sum_j \bar{s}_j p_j = \sum_i \mathbf{p}^T \mathbf{x}_i \ge \sum_i w_i.$$

Thus, every inequality in the sequel has to be equal, that is, $\mathbf{p}^T \mathbf{x}_i = w_i, \ \forall i$ and $p_j x_{ij} = w_i \frac{u_{ij} x_{ij}}{\mathbf{u}_i^T \mathbf{x}_i}, \ \forall i, j$.

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

$$\log(\mathbf{u}_i^T\mathbf{x}_i) + \log(p_j) \geq \log(w_i) + \log(u_{ij}),\ \forall i, j,\ u_{ij} > 0.$$

The function on the left is (strictly) concave in $\mathbf{x}_i$ and p_j. Thus,

Theorem 7 *The equilibrium set of the Fisher Market is convex, and the equilibrium price vector is unique.*

Aggregate Social Optimization

$$\begin{array}{ll} \text{maximize} & \sum_{i \in B} w_i \log(\mathbf{u}_i^T \mathbf{x}_i) \\ \text{subject to} & \sum_{i \in B} x_{ij} = \bar{s}_j, \quad \forall j \in G, \; x_{ij} \geq 0, \quad \forall i, j. \end{array}$$

Theorem 8 *(Eisenberg and Gale 1959) Optimal dual (Lagrange) multiplier vector of equality constraints is an equilibrium price vector.*

The proof is from Optimality Conditions of the Aggregate Social Problem:

$$\begin{array}{rcl} w_i \frac{u_{ij}}{\mathbf{u}_i^T \mathbf{x}_i} & \leq & p_j, \quad \forall i, j \\ w_i \frac{u_{ij} x_{ij}}{\mathbf{u}_i^T \mathbf{x}_i} & = & p_j x_{ij}, \quad \forall i, j \quad \text{(complementarity)} \\ \sum_i x_{ij} & = & \bar{s}_j, \quad \forall j \\ \mathbf{x}_i & \geq & \mathbf{0}, \; \forall i, \end{array}$$

which is identical to the equilibrium conditions described earlier.

Rewrite Aggregate Social Optimization

$$\begin{array}{ll} \text{maximize} & \sum_{i\in B} w_i \log u_i \\ \text{subject to} & \sum_{j\in G} u_{ij}^T x_{ij} - u_i = 0, \quad \forall i \in B \\ & \sum_{i\in B} x_{ij} = \bar{s}_j, \quad \forall j \in G \\ & x_{ij} \geq 0,\ u_i \geq 0, \quad \forall i, j, \end{array}$$

This is called the weighted analytic center problem.

Question: Is the price vector $\mathbf{p}$ unique when at least one $u_{ij} > 0$ among $i \in B$ and $u_{ij} > 0$ among $j \in G$.

Aggregate Example:

$$\begin{array}{ll} \text{maximize} & 5\log(2x_1 + y_1) + 8\log(3x_2 + y_2) \\ \text{subject to} & x_1 + x_2 = 1, \\ & y_1 + y_2 = 1, \\ & x_1, x_2, y_1, y_2 \geq 0. \end{array}$$

Using the Lagrangian Function to Derive Optimality Conditions

We consider the general constrained optimization:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i(\mathbf{x}) \quad (\leq, =, \geq) \quad 0, \; i = 1, ..., m, \end{array}$$

For Lagrange Multipliers:

$$\Lambda := \{\lambda_i \quad (\leq, '\text{free}', \geq) \quad 0, \; i = 1, ..., m\},$$

the Lagrangian Function is given by

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda^T \mathbf{c}(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i c_i(\mathbf{x}), \; \lambda \in \Lambda.$$

$$\nabla_x L(\mathbf{x}, \lambda) = \mathbf{0} \quad \text{and} \quad \lambda_i c_i(\mathbf{x}) = 0, \; \forall i.$$