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 and 11.1-11.3)

General Optimization Problems

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

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

In all forms of MP, a feasible solution of a given problem is a vector that satisfies the constraints of the problem, that is, in $\mathcal{F}$.

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. Given a point $\bar{\mathbf{x}} \in R^n$, if there is 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 $\mathbf{d}$ 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 $\bar{\mathbf{x}}$.

Let's denote the set of descent directions at $\bar{\mathbf{x}}$ by $\mathcal{D}^d_{\bar{\mathbf{x}}}$, i.e.,

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

Feasible Direction

At feasible point $\bar{\mathbf{x}}$, the set of feasible directions is given by

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

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 $\bar{x}$ 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 following unconstrained problem where f is differentiable on R^n:

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

$\mathcal{D}^f_{\bar{\mathbf{x}}} = R^n$, 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 $\bar{x}$ be a (local) minimizer of (UP). If the functions f is continuously differentiable at $\bar{x}$, then*

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

Linear Equality-Constrained Problems

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

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

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

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

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

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 $\bar{x}$ 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 $\mathbf{y} \in R^n$ such that

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

The 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”. If a minimizer exists, then

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

Linear Inequality-Constrained Problems

Let us now consider the inequality-constrained problem

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

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

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

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

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 $\bar{x}$ must be empty or

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

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

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

Let $\bar{y}_i = 0$ for all remaining inactive constraints. Then we prove the theorem.

Optimization with Mixed Constraints

We now consider optimality conditions for problems having both inequality and equality constraints. That is,

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

For any feasible point $\bar{x}$ 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 Conditions again

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

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

Optimality and Complementarity Conditions

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

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

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

$$(LP) \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x}$$
$$\text{subject to} \quad A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}.$$

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

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

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:

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

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

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 $(\mathbf{x}^2, \mathbf{s}^2)$. Then we have

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

Thus, there is j such that

$$(s_j^1 - s_j^2)(x_j^1 - x_j^2) > 0.$$

If $x_j^1 > x_j^2$, then $s_j^1 < s_j^2$ since $x_j^1 s_j^1 = x_j^2 s_j^2 = \mu > 0$, which leads to $(s_j^1 - s_j^2)(x_j^1 - x_j^2) < 0$ which is a contradiction. Similarly, one cannot have $x_j^1 < x_j^2$.

[a] To see this, the second condition gives $A\mathbf{x}_1 = \mathbf{b}_1$ and $A\mathbf{x}_2 = \mathbf{b}_3$ and the third condition gives $A^T\mathbf{y}_1 + \mathbf{s}_1 = \mathbf{c}$ and $A^T\mathbf{y}_2 + \mathbf{s}_2 = \mathbf{c}$. Which imply $A(\mathbf{x}_1 - \mathbf{x}_2) = 0$ and $A^T(\mathbf{y}_1 - \mathbf{y}_2) = \mathbf{s}_2 - \mathbf{s}_1$ and thus give $0 = (\mathbf{x}_1 - \mathbf{x}_2)^T A^T(\mathbf{y}_1 - \mathbf{y}_2) = (\mathbf{x}_1 - \mathbf{x}_2)^T(\mathbf{s}_2 - \mathbf{s}_1)$.

KKT Applications: 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 s_j. The equilinitum price vector is the one that for all $j \in G$

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

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.

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

Given $(p_x, p_y) = \left(\frac{26}{3}, \frac{13}{3}\right)$, the optimal quantities are found to be

$$(x_1, y_1) = \left(\frac{1}{13}, 1\right), \quad (x_2, y_2) = \left(\frac{12}{13}, 0\right).$$

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:

$$\begin{cases} \mathbf{p}^T \mathbf{x}_i \leq w_i,\ \mathbf{x}_i \geq \mathbf{0}, & \forall i, \\ p_j y_i \geq u_{ij},\ y_i \geq 0, & \forall i, j, \\ \mathbf{u}_i^T \mathbf{x}_i = w_i y_i, & \forall i, \\ \sum_i x_{ij} = s_j, & \forall j. \end{cases}$$

Equilibrium Price Conditions (cont'd)

These conditions can be represented by

$$\left\{\begin{array}{rl} \sum_j s_j p_j \le \sum_i w_i, \; \mathbf{x}_i \ge \mathbf{0}, & \forall i, \\ \frac{\mathbf{u}_i^T \mathbf{x}_i}{w_i} \cdot p_j \ge u_{ij}, & \forall i, j, \\ \sum_i x_{ij} = s_j, & \forall j. \end{array}\right.$$

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

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

Thus, these conditions imply $\mathbf{p}^T \mathbf{x}_i = w_i, \; \forall i$.

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

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

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

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

Aggregate Social Optimization

$$\begin{array}{ll} \text{maximize} & \sum_{i\in B} w_i \log(\mathbf{u}_i^T \mathbf{x}_i) \\ \text{subject to} & \sum_{i\in B} x_{ij} \le s_j, \quad \forall j \in G \\ & x_{ij} \ge 0, \quad \forall i, j, \end{array}$$

Theorem 7 *(Eisenberg and Gale 1959) Optimal dual (Lagrange) multiplier vector of equality constraints is an equilibrium price vector.*

Optimality Conditions of the Aggregated Problem

$$\left\{\begin{array}{rcl} w_i\frac{u_{ij}}{\mathbf{u}_i^T\mathbf{x}_i} & \le & p_j, \quad \forall i,j \\ w_i\frac{u_{ij}x_{ij}}{\mathbf{u}_i^T\mathbf{x}_i} & = & p_j x_{ij}, \quad \forall i,j \\ \sum_i x_{ij} & \le & s_j, \quad \forall j \\ p_j\sum_i x_{ij} & \le & p_j s_j, \quad \forall j \\ \mathbf{x}_i,\ \mathbf{p} & \ge & \mathbf{0}. \end{array}\right.$$

Let $y_i = \mathbf{u}_i^T\mathbf{x}_i/w_i$. Then, these conditions are identical to the equilibrium price conditions, since

$$y_i = \frac{\mathbf{u}_i^T\mathbf{x}_i}{w_i} \ge \frac{u_{ij}}{p_j},\ \forall i,j.$$

Rewriting the Aggregate Social Optimization

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

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

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

Or

$$\begin{array}{ll} \text{maximize} & 5\log(u_1)+8\log(u_2) \\ \text{subject to} & 2x_1+y_1-u_1=0, \\ & 3x_2+y_2-u_2=0, \\ & x_1+x_2=1, \\ & y_1+y_2=1, \\ & x_1,x_2,y_1,y_2 \geq 0. \end{array}$$