Optimality Conditions: More Applications of the Alternative System Theorem

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

http://www.stanford.edu/~yyye

General Optimization Problems

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

$$(\mathrm{P}) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \mathcal{F}. \end{array}$$

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.

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

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

Linear Constraint 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.

In what follows, we consider optimality conditions for Linearly Constrained Optimization Problems (LCOP).

Unconstrained Problems

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

$$(\text{UP}) \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in R^n. \end{array}$$

$\mathcal{D}^f_{\bar{\mathbf{x}}} = R^n$, 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,

$$(\text{LEP}) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}. \end{array}$$

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 $\bar{x}$ 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

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

Example: The Objective Contour Tangential to the Constraint Hyperplane

Consider the problem

$$\begin{array}{ll} \text{minimize} & (x_1-1)^2+(x_2-1)^2 \\ \text{subject to} & x_1+x_2=1. \end{array}$$

$$\bar{\mathbf{x}} = \left(\frac{1}{2};\ \frac{1}{2}\right).$$

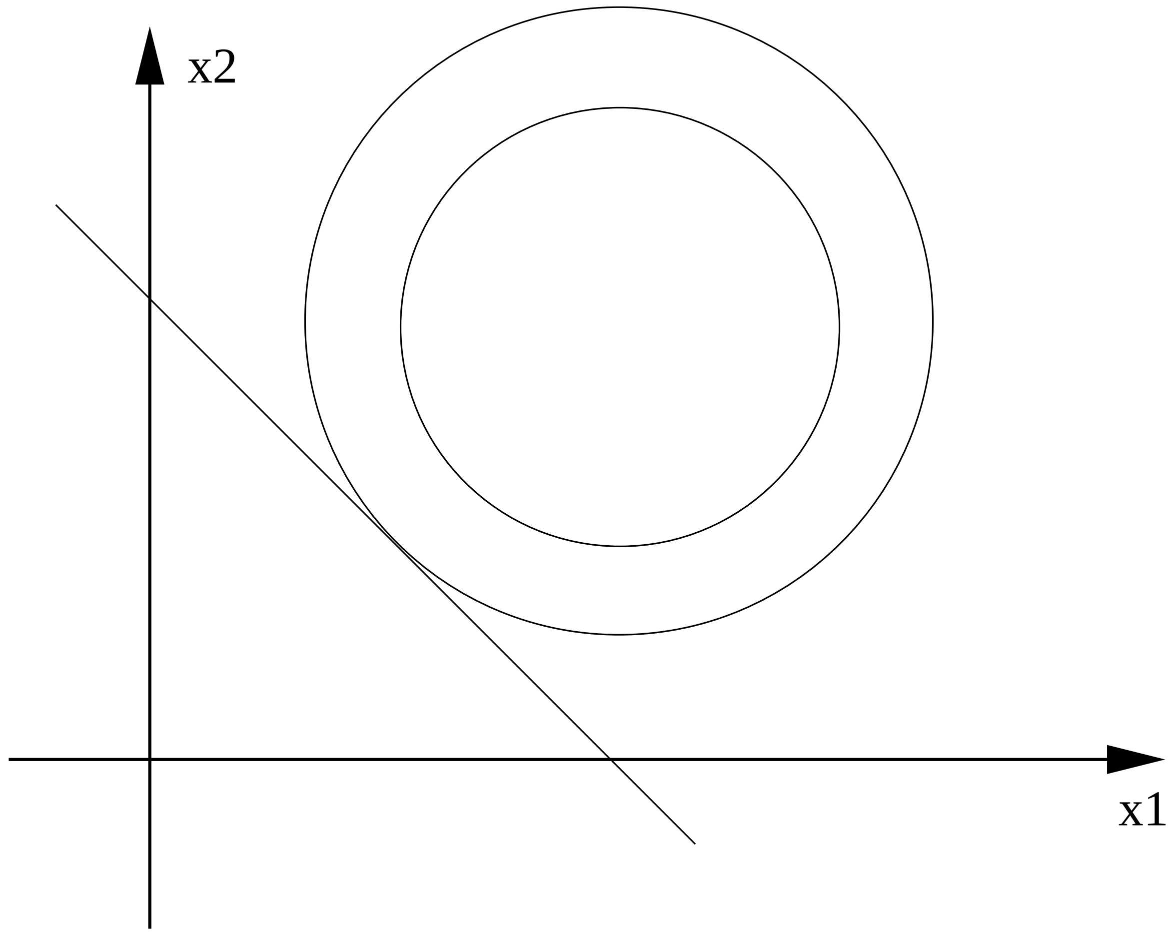

Figure 1: The Objective Contour Tangents the Constraint Hyperplane

The Barrier Optimization

Consider the problem

$$\begin{array}{ll} \min & -\sum_{j=1}^{n} \log x_j \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

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

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

Linear Inequality-Constrained Problems

Let us now consider the inequality-constrained problem

$$
\text{(LIP)} \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & A\mathbf{x} \geq \mathbf{b}. \end{array}
$$

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 $\bar{x}$ 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

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

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

Example: The Gradient is in the Normal Cone of the Half Spaces

Consider the problem

$$\begin{array}{ll} \min & (x_1-1)^2+(x_2-1)^2 \\ \text{s.t.} & -x_1-2x_2 \geq -1, \\ & -2x_1-x_2 \geq -1. \end{array}$$

$$\bar{\mathbf{x}} = \left(\frac{1}{3};\ \frac{1}{3}\right).$$

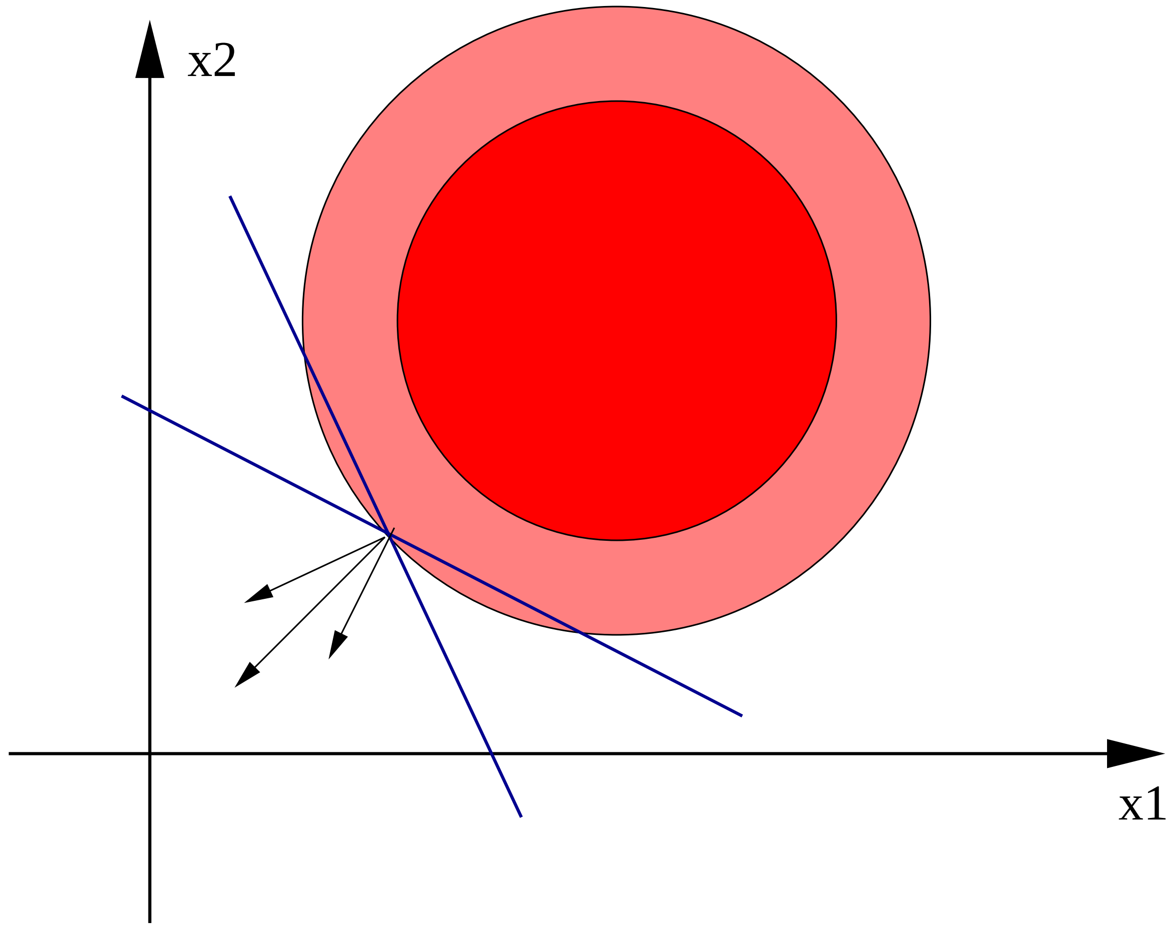

Figure 2: The objective gradient in the normal cone of the half spaces

Optimization with Mixed Linear Constraints

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

$$
(\mathrm{P}) \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{a}_i\mathbf{x} \quad (\leq, =, \geq) \quad b_i, \ i = 1, ..., m, \end{array}
$$

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

$$
\begin{array}{rcl} \mathcal{A}(\bar{\mathbf{x}}) & = & \{i : \mathbf{a}_i\mathbf{x} = b_i\} \\ \mathcal{D}^d_{\bar{\mathbf{x}}} & = & \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\}. \end{array}
$$

The KKT Theorem Again

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

$$
\begin{array}{ccc}
\nabla f(\bar{\mathbf{x}}) & = & \bar{\mathbf{y}}^T A \\
\bar{y}_i & (\leq, '\textit{free}', \geq) & 0, \; i = 1, ..., m, \\
\bar{y}_i & = & 0 \quad \textit{if}\; i \notin \mathcal{A}(\bar{\mathbf{x}}).
\end{array}
$$

When First-Order Optimality Conditions are Sufficient?

Theorem 5 *If objective f is a locally convex function in the feasible direction space at the KKT solution $\bar{\mathbf{x}}$, then the (first-order) KKT optimality conditions are sufficient for the local optimality at $\bar{\mathbf{x}}$.*

A function is locally convex in a space D means that $\phi(\alpha) := f(\bar{\mathbf{x}} + \alpha\mathbf{d})$ is a convex function of α in a sufficiently small neighborhood of 0 for all $\mathbf{d} \in D$.

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 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.

LCOP Examples: Linear Optimization

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

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, $\mathbf{y}$ (shadow prices in LP) are Lagrange or dual multipliers of equality constraints, and $\mathbf{s}$ (reduced gradient/costs in LP) are Lagrange or dual multipliers for $\mathbf{x} \geq \mathbf{0}$.

LCOP Examples : Quadratic Optimization

$$\begin{array}{lll} (QP) & \min & \mathbf{x}^T Q\mathbf{x} - 2\mathbf{c}^T\mathbf{x} \\ & \text{s.t.} & A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}. \end{array}$$

Optimality Conditions:

$$\begin{aligned} x_j s_j &= 0,\ \forall j = 1, \ldots, n \\ A\mathbf{x} &= \mathbf{b} \\ 2Q\mathbf{x} - 2\mathbf{c} - A^T\mathbf{y} - \mathbf{s} &= \mathbf{0} \\ \mathbf{x}, \mathbf{s} &\geq \mathbf{0} \end{aligned}$$

LCOP Examples: Linear Barrier Optimization

$$\min \ f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^{n} \log(x_j), \ \text{s.t.} \ A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}.$$

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

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

KKT Application: Fisher's Equilibrium Prices

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

$$\begin{array}{rl} \max & \mathbf{u}_i^T \mathbf{x}_i := \sum_{j \in G} u_{ij} x_{ij} \\ \text{s.t.} & \mathbf{p}^T \mathbf{x}_i := \sum_{j \in G} p_j x_{ij} \le w_i, \\ & x_{ij} \ge 0, \quad \forall j, \end{array}$$

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

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

where $\mathbf{x}^*(\mathbf{p})$ is a maximizer of the utility maximization problem for every buyer i.

Example of Fisher's Equilibrium Prices

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.

$$\begin{array}{rc}
\max & 2x_1 + y_1 \\
\text{s.t.} & p_x \cdot x_1 + p_y \cdot y_1 \le 5, \\
 & x_1, y_1 \ge 0; \\
\max & 3x_2 + y_2 \\
\text{s.t.} & p_x \cdot x_2 + p_y \cdot y_2 \le 8, \\
 & x_2, y_2 \ge 0.
\end{array}$$

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

Equilibrium Price Conditions

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

$$\begin{array}{ll} \min & w_i y_i \\ \text{s.t.} & \mathbf{p} y_i \geq \mathbf{u}_i, \; y_i \geq 0 \end{array}$$

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

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

Equilibrium Price Conditions continued

These conditions can be further simplified to

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

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

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

Then, from the rest conditions

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

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.*

Aggregated Social Optimization

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

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

Proof: The optimality conditions of the social problem are identical to the equilibrium conditions.

Aggregated Example

$$
\begin{aligned}
\max \quad & 5\log(2x_1 + y_1) + 8\log(3x_2 + y_2) \\
\text{s.t.} \quad & x_1 + x_2 = 1, \\
& y_1 + y_2 = 1, \\
& x_1, x_2, y_1, y_2 \geq 0.
\end{aligned}
$$

Or

$$
\begin{aligned}
\max \quad & 5\log(u_1) + 8\log(u_2) \\
\text{s.t.} \quad & 2x_1 + y_1 - u_1 = 0, \\
& 3x_2 + y_2 - u_2 = 0, \\
& x_1 + x_2 = 1, \\
& y_1 + y_2 = 1, \\
& x_1, x_2, y_1, y_2 \geq 0.
\end{aligned}
$$

Optimality Conditions of the Aggregated Problem

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

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

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

Aggregated Social Optimization Rewritten

$$\begin{array}{ll} \max & \sum_{i \in B} w_i \log u_i \\ \text{s.t.} & \sum_{j \in G} u_{ij}^T x_{ij} - u_i = 0, \quad \forall i \in B \\ & \sum_{i \in B} x_{ij} \leq s_j, \quad \forall j \in G \\ & x_{ij} \geq 0, \ s_i \geq 0, \quad \forall i, j, \end{array}$$

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 $j \in G$.