Optimality Conditions for Nonlinear Optimization

LI-PING ZHANG

Department of Mathematical Sciences

Tsinghua University, Beijing 100084

Office: New Science Building #A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Outline

- Global and local optimal solution
- Optimality conditions for unconstrained optimization
- Kuhn-Tucker condition for equality-constrained optimization
- Kuhn-Tucker condition for inequality-constrained optimization
- Lagrangian Duality Theory and Saddle-Point Theorem

An Example of Nonlinear Optimization

We now begin our whirlwind tour of nonlinear optimization. The problem that we are interested in is finding the maximum (or minimum) value taken by a real valued function defined on some subset $S \subseteq \mathcal{R}^n$. Initially, we are going to think about the unconstrained problem, whereby we are interested in finding the maximum and minimum on all of S. We are then going to move onto constrained problems, in which we are not free to choose any element in the set S, but only elements that satisfy some constraints. We will begin by thinking about equality constraints, and then inequality constraints.

先考虑无约束再一点点向约束改变

Utility Maximization

The utility maximization problem is a typical example of an optimization problem. It is a basic model in consumer theory, which concerns a single agent who consumes n commodities in nonnegative quantities. The agent's utility function is $u(\mathbf{x}) = u(x_1, \ldots, x_n)$. The agent's objective is to maximize $u(\mathbf{x})$ over the budget set

$$B(\mathbf{p}, I) = \{\mathbf{x} \in \mathcal{R}^n_+ | \mathbf{px} \leq I\}.$$

That is, to solve

$$\begin{array}{ll} \textbf{maximize} & u(\mathbf{x}) \\ \textbf{subject to} & \mathbf{x} \in B(\mathbf{p}, I). \end{array}$$

Optimization Problems

The **question**: How does one recognize an optimal solution to a nonlinear objective and/or nonlinearly constrained optimization problem?

Let the problem have the form

$$(\mathbf{P}) \qquad \begin{array}{ll} \textbf{minimize} & f(\mathbf{x}) \quad \text{目标函数} \\ \textbf{subject to} & \mathbf{x} \in S. \end{array}$$

In all forms of mathematical programming, a feasible solution of a given problem is a vector that satisfies the constraints of the problem.

Global and Local Optimizers

A global minimizer for (P) is a vector $\bar{\mathbf{x}}$ such that

$$\bar{\mathbf{x}} \in S \quad \textbf{and} \quad f(\bar{\mathbf{x}}) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in S.$$

Unlike linear programming, sometimes one has to settle for a local minimizer, that is, a vector $\bar{\mathbf{x}}$ such that

$$\bar{\mathbf{x}} \in S \quad \textbf{and} \quad f(\bar{\mathbf{x}}) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in S \cap N(\bar{\mathbf{x}})$$

凸集上的凸规划才方
便找全局最优解

where $N(\bar{\mathbf{x}})$is a neighborhood of $\bar{\mathbf{x}}$. Typically, $N(\bar{\mathbf{x}})$ is a ball centered at $\bar{x}$ having suitably small radius.

The value of the objective function f at a global minimizer or a local minimizer is also of interest. The global minimum value or a local minimum value, according to whether $\bar{\mathbf{x}}$ is a global minimizer or a local minimizer, respectively.

Optimality Conditions in $\mathcal{R}$

Considering the case of $\mathcal{R}$, we have the following theorem.

Theorem 1 *Let $n = 1$ and f be differentiable.*

- *If $x^* \in int(S)$ is a local minimizer of (P) then $f'(x^*) = 0$. Moreover, if $f''(x^*)$ exists, then $f''(x^*) \geq 0$.* 梯度为0hessian半正定
- *If $f'(x^*) = 0$ and $f''(x^*) > 0$, then x^* is a strict local minimizer of (P).*

We are now going to extend this result in two ways. First, we are going to allow for functions on some arbitrary subset of $\mathcal{R}^n$. Second, we are going to generate a result that allows us to say something about any point in the feasible set. In order to do so we first need to define the concepts of a feasible direction and a descent direction.

方向d使∃α>0

而 $x_0+\alpha d$ 亦可行

$f(x_0+\alpha d) \leq f(x_0)$

对α充分小

Descent directions

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

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

这里是内积小于

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 descent at $\bar{\mathbf{x}}$.

Denote by $\mathcal{D}_{\bar{\mathbf{x}}}$ the **cone of descent directions** at $\bar{\mathbf{x}}$ or

下降方向锥

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

这里也有d不等于0

Feasible directions

At feasible point $\bar{\mathbf{x}}$, the **cone of feasible directions** at $\bar{\mathbf{x}}$ is the set 可行方向锥

$$\mathcal{F}_{\bar{\mathbf{x}}} := \{\mathbf{d} \in R^n : \mathbf{d} \neq \mathbf{0},\ \bar{\mathbf{x}} + \lambda \mathbf{d} \in S \text{ for all } \lambda \in (0, \bar{\gamma}) \text{ for some } \bar{\gamma} > 0\}.$$

Examples:

$$S = R^n \Rightarrow \mathcal{F}_{\bar{\mathbf{x}}} = R^n.$$

$$S = \{\mathbf{x} :\ A\mathbf{x} = \mathbf{b}\} \Rightarrow \mathcal{F}_{\bar{\mathbf{x}}} = \{\mathbf{d} : A\mathbf{d} = 0\}.$$

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

原式取等的部分(积极约束)的部分要保证加上去的非负，其它部分(非积极约束)无限制

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

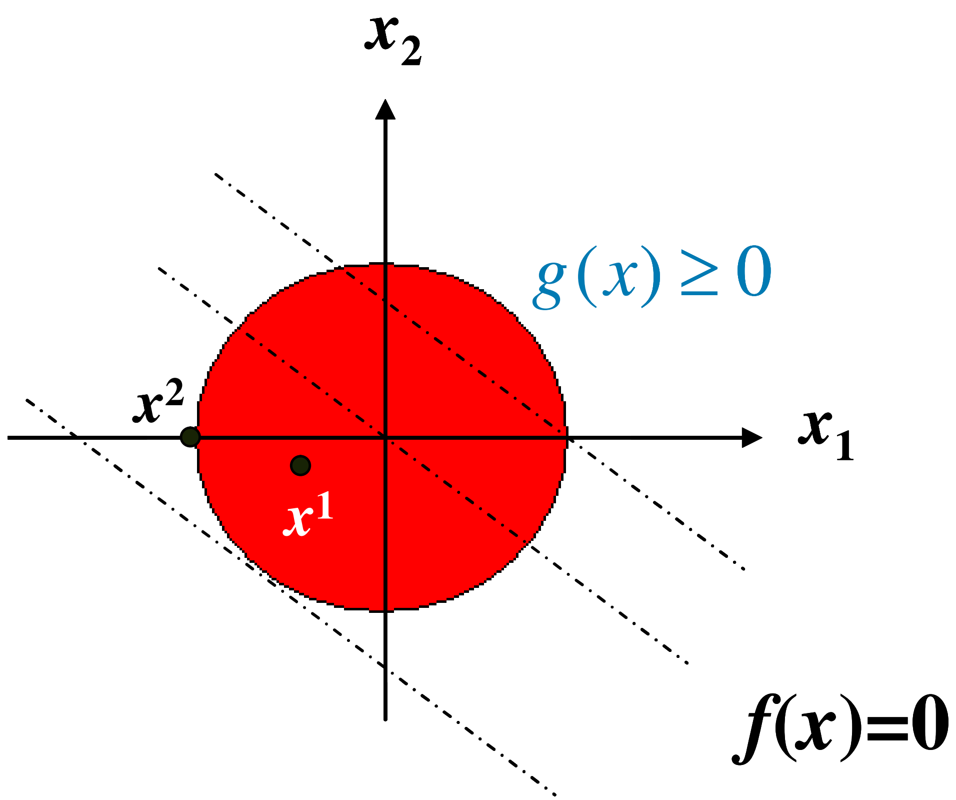
Example for Feasible Direction

Consider

$$\begin{array}{ll} \min & x_1 + x_2 \\ s.t. & g(\mathbf{x}) = 1 - x_1^2 - x_2^2 \geq 0 \end{array}$$

For the interior point $\mathbf{x}^1$, $\mathcal{F}_{\mathbf{x}^1} = R^2$.

For the boundary point $\mathbf{x}^2 = (-1; 0)$, $\mathcal{F}_{\mathbf{x}^2} = \{\mathbf{d} \in R^2 | \, d_1 > 0\}$.


x_2
$g(x) \geq 0$
x^2
x_1
x^1
$f(x)=0$

First-Order Necessary Condition

Theorem 2 *Assume f is $\mathcal{C}^1$ on S. If $x^* \in S$ is a local min of f, then*

$$\nabla f(x^*)^T d \geq 0$$

可行方向与下降方向交集为空

for all feasible directions d at x^*.

必要非充分条件

Proof. Let $d \in \mathcal{R}^n$ be a feasible direction at x^* and define $g(t) = f(x^* + td)$ on $[0, \bar{t}]$. We have $g(t) \geq g(0)$ for t sufficiently small, and

$$g'(t) = \nabla f(x^* + td)^T d.$$

Using Taylor Theorem,

$$g(t) = g(0) + t\nabla f(x^*)^T d + o(t) \quad \Rightarrow \quad \nabla f(x^*)^T d \geq 0.$$

□

Second-Order Necessary Condition

Theorem 3 *Assume f is $\mathcal{C}^2$ on S. If $x^* \in S$ is a local min of f, then for any feasible direction $d \in \mathcal{R}^n$ at x^**

1) $\nabla f(x^*)^T d \geq 0$. 必要非充分条件

2) $\nabla f(x^*)^T d = 0 \Rightarrow d^T H(x^*) d \geq 0$, *where $H(x^*)$ is the Hessian matrix.*

Proof. For 2), using Taylor's second-order approximation, we have

$$g(t) = g(0) + g'(0)t + \frac{1}{2} g''(0)t^2 + o(t^2).$$

Since $g'(0) = \nabla f(x^*)^T d = 0$ and $g''(0) = d^T H(x^*) d$,

$$\frac{1}{2} d^T H(x^*) d \geq \frac{o(t^2)}{t^2}.$$

So we are done. $\square$

Unconstrained Problems

Consider the unconstrained problem

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

We have the following necessary conditions.

Theorem 4 *Let $\bar{\mathbf{x}}$ be a (local) minimizer of (UP). If the functions f is continuously differentiable at $\bar{\mathbf{x}}$, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$.* 一阶必要条件

Theorem 5 *Let $\bar{\mathbf{x}}$ be a (local) minimizer of (UP). If f is twice differentiable at $\bar{\mathbf{x}}$, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H(\bar{\mathbf{x}})$ is positive semidefinite.*

二阶必要条件

Sufficient Conditions

Theorem 6 *Suppose that f is twice differentiable at $\bar{\mathbf{x}}$. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H(\bar{\mathbf{x}})$ is positive definite, then $\bar{\mathbf{x}}$ is a strictly local minimizer of (UP).*

二阶 充分条件

Proof. Since f is twice differentiable at $\bar{\mathbf{x}}$ and $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, we must have, for each $\mathbf{x} \in \mathcal{R}^n$,

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^T H(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + o(\|\mathbf{x} - \bar{\mathbf{x}}\|^2), \tag{1}$$

as $\mathbf{x} \to \bar{\mathbf{x}}$. Suppose, by contradiction, that there exists a sequence $\{\mathbf{x}^k\}$ converging to $\bar{\mathbf{x}}$ such that $f(\mathbf{x}^k) \leq f(\bar{\mathbf{x}})$, $\mathbf{x}^k \neq \bar{\mathbf{x}}$, for each k. Denoting $(\mathbf{x}^k - \bar{\mathbf{x}})/\|\mathbf{x}^k - \bar{\mathbf{x}}\|$ by $\mathbf{d}^k$, (1) implies that

$$\frac{1}{2}(\mathbf{d}^k)^T H(\bar{\mathbf{x}})\mathbf{d}^k + o(\|\mathbf{d}^k\|^2) \leq 0, \quad \textbf{for each } k. \tag{2}$$

Since $\|\mathbf{d}^k\| = 1$, there exists an index set $\mathcal{K}$ such that $\mathbf{d}^k \to \mathbf{d}$ with $\|\mathbf{d}\| = 1$ as $k \in \mathcal{K}$ and $k \to \infty$. Hence, it follows from (2) that $\mathbf{d}^T H(\bar{\mathbf{x}})\mathbf{d} \le 0$. This contradicts the assumption that $H(\bar{\mathbf{x}})$ is positive definite since $\|\mathbf{d}\| = 1$.

Therefor, $\bar{\mathbf{x}}$ is a strictly local minimum.

f凸时，局部最优等价于全局最优

Theorem 7 *If the functions f is convex and continuously differentiable, then $\bar{\mathbf{x}}$ is a global minimizer of (UP) if and only if*

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

由此结合凸函数梯度不等式易见为局部最优解

Theorem 8 *Suppose that f is twice differentiable at $\bar{\mathbf{x}}$. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H(\mathbf{x})$ is positive semidefinite for any $\mathbf{x} \in N(\bar{\mathbf{x}})$, then $\bar{\mathbf{x}}$ is a local minimizer of (UP).*

在一个小临域内都是半正定的

（二阶必要条件的一个反向）

（或者说二阶充分条件的一个弱化形式）

Examples

Consider

$$\min \quad f(\mathbf{x}) = \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - x_2^2 - x_1$$

Gradient

$$\nabla f(\mathbf{x}) = \begin{cases} x_1^2 - 1 \\ x_2^2 - 2x_2 \end{cases}$$

Hessian

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2x_1 & 0 \\ 0 & 2x_2 - 2 \end{pmatrix}$$

We find $\mathbf{x}^* = (1; 2)$ is a strictly local minimizer.

Examples continued

Consider

$$\min \quad f(\mathbf{x}) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

Gradient

$$\nabla f(\mathbf{x}) = \begin{cases} 4(x_1 - 2)^3 + 2(x_1 - 2x_2) \\ -4(x_1 - 2x_2) \end{cases}$$

Let $\nabla f(\mathbf{x}) = \mathbf{0}$ and we obtain $\mathbf{x}^* = (2; 1)$. Hessian

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 12(x_1 - 2)^2 + 2 & -4 \\ -4 & 8 \end{pmatrix} \Rightarrow \nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}$$

Thus, $\mathbf{x}^*$ is a local minimizer.

Equality-Constrained Problems

Consider the classical equality-constrained problem

$$
(\textbf{EP}) \qquad \begin{array}{ll} \textbf{minimize} & f(\mathbf{x}) \\ \\ \textbf{subject to} & \mathbf{h}(\mathbf{x}) = \mathbf{0}. \end{array}
$$

The vector function $\mathbf{h}(\mathbf{x})$ from R^n to R^m is

$$
\mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_m(\mathbf{x}) \end{bmatrix}.
$$

Suppose the functions $f, h_1, \ldots, h_m$ are differentiable on R^n. Accordingly, each of these functions has a gradient at every point $\mathbf{x} \in R^n$. The Jacobian matrix of the mapping $\mathbf{h}$ will be

$$\left[\frac{\partial h_i(\mathbf{x})}{\partial x_j}\right].$$

The Jacobian of $\mathbf{h}$ is the matrix whose m rows are the gradients of functions $h_1, \ldots, h_m$ and is denoted by $\nabla \mathbf{h}(\mathbf{x})$.

If $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, then $\nabla \mathbf{h}(\mathbf{x}) = \mathbf{A}$. In this special case, the Jacobian is a constant matrix, whereas in general it is a variable matrix in $R^{m \times n}$.

We assume $m \leq n$ because we don't want the problem to be over-constrained.

The Lagrange Theorem

Theorem 9 *(Lagrange) Let $\bar{\mathbf{x}}$ be a local minimizer of (EP). If the functions f and $h_1, \ldots, h_m$ are continuously differentiable at $\bar{\mathbf{x}}$ and the Jacobian matrix $\nabla \mathbf{h}(\bar{\mathbf{x}})$ has rank m when $\mathbf{h}(\mathbf{x})$ are nonlinear, then there exist scalars $\bar{y}_1, \ldots, \bar{y}_m$ such that* 对线性约束时没有要求

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \bar{y}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

利用Lagrange乘子得必要条件

The numbers $\bar{y}_1, \ldots, \bar{y}_m$ are called Lagrange multipliers.

The function

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \sum_{i=1}^{m} y_i h_i(\mathbf{x})$$

is called the Lagrangian function, or simply Lagrangian, for (EP). The Lagrange theorem says that the gradient vector, with respect to $\mathbf{x}$, is $\mathbf{0}$ at a local optimizer.

Proof when Constraints are Linear

Consider linearly constrained case:

$$S = \{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\} \Rightarrow \mathcal{F}_{\bar{\mathbf{x}}} = \{\mathbf{d} : A\mathbf{d} = 0\}.$$

If $\bar{\mathbf{x}}$ is a local optimizer, then by First-Order Necessary Condition (Theorem 2), the system

可行方向　下降方向

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

has no feasible solution. By Farkas' lemma it must be true

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

for some $\bar{y}_1, \ldots, \bar{y}_m$. □

Example 1: The Need for a Regularity Condition

The property "$\nabla \mathbf{h}(\bar{\mathbf{x}})$ has full rank" is called a regularity condition or constraint qualification. Lagrange's theorem is not valid without a regularity condition when constraints are nonlinear. 约束规范

正则性条件这里的x称为正则点

Consider the problem

$$\begin{array}{ll} \textbf{minimize} & x_1 \\ \textbf{subject to} & x_1^2 + (x_2 - 1)^2 - 1 = 0 \\ & x_1^2 + (x_2 + 1)^2 - 1 = 0 \end{array}$$

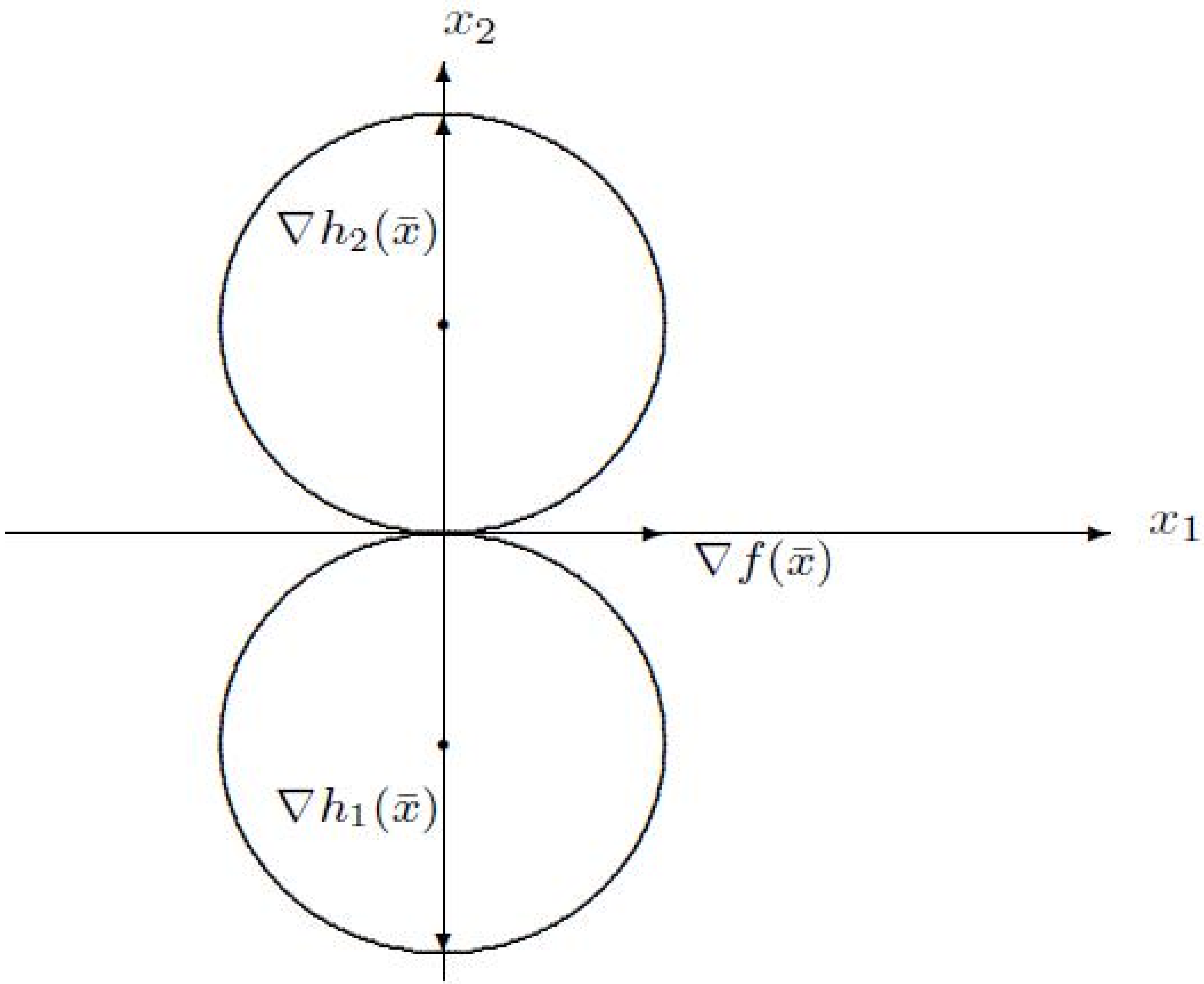


Figure 1: Example 1

Example 1 continued

This problem has just one feasible point: $\bar{\mathbf{x}} = (0, 0)$.

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) &= (1, 0) \\ \nabla h_1(\bar{\mathbf{x}}) &= (0, -2) \\ \nabla h_2(\bar{\mathbf{x}}) &= (0, 2). \end{aligned}$$

One can see that the Lagrange Theorem does not hold.

Proof when Constraints are Nonlinear

In order to take on the proof, we are going to require finding out what a “feasible direction” is for any $\mathbf{x} \in S = \{\mathbf{x} \in \mathcal{R}^n : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$. We need to define two new objects.

Definition 1 *A curve on S is a family of points $x(t) \in S$ continuously parameterized by t for $a \le t \le b$. The curve is differentiable if*

为了描述可行方向而引入曲线的概念

$$\dot{x}(t) = \frac{d}{dt} x(t)$$

exist for all $t \in [a, b]$.

The curve $x(t)$ is said to pass through x^* if there exists a $t^* \in [a, b]$ such that $x(t^*) = x^*$.

The Tangent Plane

Definition 2 *The tangent plane to S at x^* is* 切平面

$$T(x^*) = \{\dot{x}(t^*) |\ x : [a, b] \to S, x(t^*) = x^*\}.$$

The tangent plane at x^* is defined as the collection of the derivatives at x^* of all these differentiable curves on S passing through x^*. The tangent plane is a subspace of $\mathcal{R}^n$.

类似隐函数定理

Theorem 10 *At a regular point x^* of the surface S defined by $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ the tangent plane is equal to*

$$H = \{d \,|\, \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{d} = \mathbf{0}\}.$$

Proof. It is clear that $T(x^*) \subset H$ whether x^* is regular or not, for any curve $x(t)$ passing through x^* at $t = t^*$ having derivative $\dot{x}(t^*)$ such that $\nabla \mathbf{h}(\mathbf{x}^*)\dot{\mathbf{x}}(\mathbf{t}^*) \neq \mathbf{0}$ would not lie on S.

To prove that $H \subset T(x^*)$ we must show that if $d \in H$ then there is a curve on S passing through x^* with derivative d. To construct such a curve we consider the equations

构造路径使其为切向量

$$\mathbf{h}(\mathbf{x}^* + \mathbf{t}\mathbf{d} + \nabla \mathbf{h}(\mathbf{x}^*)^{\mathbf{T}}\mathbf{u}(\mathbf{t})) = \mathbf{0},$$

为了保证括号里的在S中

where for fixed t we consider $u(t) \in \mathcal{R}^m$ to be the unknown. This is a nonlinear system of m equations and m unknowns, parameterized continuously, by t. At $t = 0$ there is a solution $u(0) = 0$. The Jacobian matrix of the system with respect to u at $t = 0$ is the $m \times m$ matrix

$$\nabla \mathbf{h}(\mathbf{x}^*)\nabla \mathbf{h}(\mathbf{x}^*)^{\mathbf{T}},$$

which is nonsingular, since $\nabla \mathbf{h}(\mathbf{x}^*)$ is of full rank if x^* is a regular point. Thus, by the Implicit Function Theorem, there is a continuously differentiable solution $u(t)$ in some region $-\delta \leq t \leq \delta$.

隐函数定理

Let $x(t) = x^* + td + \nabla \mathbf{h}(\mathbf{x}^*)^{\mathbf{T}} \mathbf{u}(\mathbf{t})\ \ (\mathbf{t} \in [-\delta, \delta])$. Then $x(t)$ is a curve on S passing through x^*. By differentiating the system $\mathbf{h}(\mathbf{x}(\mathbf{t})) = \mathbf{0}$ with respect to t at $t = 0$ we obtain

$$0 = \frac{d}{dt}\mathbf{h}(\mathbf{x}(\mathbf{t}))|_{\mathbf{t}=\mathbf{0}} = \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{d} + \nabla \mathbf{h}(\mathbf{x}^*)\nabla \mathbf{h}(\mathbf{x}^*)^{\mathbf{T}} \dot{\mathbf{u}}(\mathbf{0}),$$

which implies that $\dot{u}(0) = 0$. Hence,

$$\dot{x}(0) = d + \nabla \mathbf{h}(\mathbf{x}^*)^{\mathbf{T}} \dot{\mathbf{u}}(\mathbf{0}) \quad \Rightarrow \quad \dot{\mathbf{x}}(\mathbf{0}) = \mathbf{d}.$$

Thus, $d \in T(x^*)$. $\square$

Remarks

It is important to recognize that the condition of being a regular point is not a condition on the constraint surface itself but on its representation in terms of an $\mathbf{h}$. The tangent plane is defined independently of the representation, while H is not.

Example 1 *In $\mathcal{R}^2$ let $h(x_1, x_2) = x_1$. Then $h(x) = 0$ yields the x_2 axis, and every point on that axis is regular. If instead we put $h(x_1, x_2) = x_1^2$, again S is the x_2 axis but now no point on the axis is regular. Indeed in this case $H = \mathcal{R}^2$, while the tangent plane is the x_2 axis.*

由此说明正则点的必要性

Proof when Constraints are Nonlinear (cont.)

We will now use the concept of the tangent plane, along with Farkas' Lemma to derive The Lagrange Theorem.

h满秩，在x处正则

Lemma 1 *Suppose that hypotheses of The Lagrange Theorem hold, the system*

$$\nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{d} = \mathbf{0}, \quad \nabla \mathbf{f}(\bar{\mathbf{x}})^{\mathbf{T}}\mathbf{d} < \mathbf{0} \tag{3}$$

has no solution.

By the above lemma, we know that the system (3) has no solution. Then, by Farkas' Lemma, this implies that the alternative system must have a solution, and this derives The Lagrange Theorem.

于是只需要证明这个引理就有了
Lagrange定理成立

Proof. Assume by contradiction that $\mathbf{d}$ solves the above system. Then there exists a curve $\mathbf{x}(t)$ on S such that $\bar{\mathbf{x}} = \mathbf{x}(\bar{t})$ and $\dot{\mathbf{x}}(\bar{t}) = \mathbf{d}$. Then

$$
\begin{aligned}
f(\mathbf{x}(t)) &= f(\mathbf{x}(\bar{t})) + \tfrac{d}{dt} f(\mathbf{x}(t))|_{t=\bar{t}}(t-\bar{t}) + o(t-\bar{t}) \\
&= f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^T \dot{\mathbf{x}}(\bar{t})(t-\bar{t}) + o(t-\bar{t}) \\
&= f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^T \mathbf{d}(t-\bar{t}) + o(t-\bar{t}) \\
\Rightarrow & \quad \tfrac{f(\mathbf{x}(t)) - f(\bar{\mathbf{x}})}{(t-\bar{t})} = \nabla f(\bar{\mathbf{x}})^T \mathbf{d} + \tfrac{o(t-\bar{t})}{(t-\bar{t})},
\end{aligned}
$$

which is less than zero for t close to $\bar{t}$, contradicting the idea that $\bar{\mathbf{x}}$ is a local minimizer. □

Second-Order Necessary Conditions

By an argument analogous to that used for the unconstrained case, we can also derive the corresponding second-order conditions for constrained problems. It is assumed that $f, \mathbf{h} \in \mathcal{C}^2$.

Theorem 11 *Suppose that $\bar{\mathbf{x}}$ is a local minimum of (EP) and that it is a regular point of these constraints. Then there is a $\bar{y} \in E^m$*

二阶必要条件

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \bar{y}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

Moreover, the matrix $\nabla_x^2 L(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is positive semidefinite on the tangent plane H, that is, $\mathbf{d}^T \nabla_x^2 L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \mathbf{d} \geq 0$ for all $\mathbf{d} \in H$.

先从曲线上的
极值着手考虑

proof. From elementary calculus it is clear that for every twice differentiable curve $\mathbf{x}(t)$ on the constraint surface S through $\bar{\mathbf{x}}$ (with $\mathbf{x}(0) = \bar{\mathbf{x}}$) we have

$$\boxed{0 \leq \frac{d^2}{dt^2} f(\mathbf{x}(t))|_{t=0}} = \dot{\mathbf{x}}(0)^T \nabla^2 f(\bar{\mathbf{x}}) \dot{\mathbf{x}}(0) + \nabla f(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(0). \tag{4}$$

Furthermore, differentiating the relation $\bar{\mathbf{y}}^T \mathbf{h}(\mathbf{x}(\mathbf{t})) = \mathbf{0}$ twice, we obtain

$$\dot{\mathbf{x}}(0)^T \bar{\mathbf{y}}^T H(\bar{\mathbf{x}}) \dot{\mathbf{x}}(0) + \bar{\mathbf{y}}^T \nabla \mathbf{h}(\bar{\mathbf{x}}) \ddot{\mathbf{x}}(\mathbf{0}) = \mathbf{0}. \tag{5}$$

Combining (4) and (5), we have

$$\dot{\mathbf{x}}(0)^T \nabla_x^2 L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \dot{\mathbf{x}}(0) \geq 0.$$

Since $\dot{\mathbf{x}}(0)$ is arbitrary in H, we immediately obtain the stated conclusion. □

Second-Order Sufficient Conditions

Theorem 12 *Let $\bar{\mathbf{x}}$ be a feasible point of (EP) such that the Jacobian matrix $\nabla \mathbf{h}(\bar{\mathbf{x}})$ has rank m. Suppose there exist scalars $\bar{y}_1, \ldots, \bar{y}_m$ such that*

二阶充分条件

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \bar{y}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

Define

$$\mathcal{D}(\bar{\mathbf{x}}) = \{\mathbf{d} \in \mathcal{R}^n | \mathbf{d} \neq \mathbf{0}, \nabla h(\bar{\mathbf{x}})\mathbf{d} = 0\}.$$

If $\mathbf{d}^T \nabla_x^2 L(\bar{\mathbf{x}}, \bar{\mathbf{y}})\mathbf{d} > 0$ holds for all $\mathbf{d} \in \mathcal{D}(\bar{\mathbf{x}})$, then $\bar{\mathbf{x}}$ is a strict local minimum of (EP).

某种意义上的严格正定

对比下页的一般二阶条件

Second-Order Necessary Condition

Theorem 3 *Assume f is $\mathcal{C}^2$ on S. If $x^* \in S$ is a local min of f, then for any feasible direction $d \in \mathcal{R}^n$ at x^**

1) $\nabla f(x^*)^T d \geq 0$. 必要非充分条件

2) $\nabla f(x^*)^T d = 0 \Rightarrow d^T H(x^*) d \geq 0$, *where $H(x^*)$ is the Hessian matrix.*

Proof. For 2), using Taylor's second-order approximation, we have

$$g(t) = g(0) + g'(0)t + \frac{1}{2} g''(0)t^2 + o(t^2).$$

Since $g'(0) = \nabla f(x^*)^T d = 0$ and $g''(0) = d^T H(x^*) d$,

$$\frac{1}{2} d^T H(x^*) d \geq \frac{o(t^2)}{t^2}.$$

So we are done. $\square$

Unconstrained Problems

Consider the unconstrained problem

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

We have the following necessary conditions.

Theorem 4 *Let $\bar{\mathbf{x}}$ be a (local) minimizer of (UP). If the functions f is continuously differentiable at $\bar{\mathbf{x}}$, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$.* 一阶必要条件

Theorem 5 *Let $\bar{\mathbf{x}}$ be a (local) minimizer of (UP). If f is twice differentiable at $\bar{\mathbf{x}}$, then $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H(\bar{\mathbf{x}})$ is positive semidefinite.*

二阶必要条件

Sufficient Conditions

Theorem 6 *Suppose that f is twice differentiable at $\bar{\mathbf{x}}$. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H(\bar{\mathbf{x}})$ is positive definite, then $\bar{\mathbf{x}}$ is a strictly local minimizer of (UP).* 二阶 充分条件

Proof. Since f is twice differentiable at $\bar{\mathbf{x}}$ and $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, we must have, for each $\mathbf{x} \in \mathcal{R}^n$,

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^T H(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + o(\|\mathbf{x} - \bar{\mathbf{x}}\|^2), \tag{1}$$

as $\mathbf{x} \to \bar{\mathbf{x}}$. Suppose, by contradiction, that there exists a sequence $\{\mathbf{x}^k\}$ converging to $\bar{\mathbf{x}}$ such that $f(\mathbf{x}^k) \leq f(\bar{\mathbf{x}})$, $\mathbf{x}^k \neq \bar{\mathbf{x}}$, for each k. Denoting $(\mathbf{x}^k - \bar{\mathbf{x}})/\|\mathbf{x}^k - \bar{\mathbf{x}}\|$ by $\mathbf{d}^k$, (1) implies that

$$\frac{1}{2}(\mathbf{d}^k)^T H(\bar{\mathbf{x}})\mathbf{d}^k + o(\|\mathbf{d}^k\|^2) \leq 0, \quad \textbf{for each } k. \tag{2}$$

Proof. If $\bar{\mathbf{x}}$ is not a strict relative minimum point, there exists a sequence of feasible points $\{\mathbf{y}_k\}$ converging to $\bar{\mathbf{x}}$ such that for each k, $f(\mathbf{y}_k) \leq f(\bar{\mathbf{x}})$. Write each $\mathbf{y}_k$ in the form $\mathbf{y}_k = \bar{\mathbf{x}} + \delta_k \mathbf{d}_k$ where $\mathbf{d}_k \in E^n$, $\|d_k\| = 1$, and $\delta_k > 0$ for each k. Clearly, $\delta_k \to 0$ and the sequence $\{\mathbf{d}_k\}$, being bounded, must have a convergent subsequence converging to some $\bar{\mathbf{d}}$. For convenience of notation, we assume that the sequence $\{\mathbf{d}_k\}$ is itself convergent to $\bar{\mathbf{d}}$. We also have $\mathbf{h}(\mathbf{y_k}) - \mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and dividing by δ_k and letting $k \to \infty$ we see that $\nabla \mathbf{h}(\bar{\mathbf{x}})\bar{\mathbf{d}} = \mathbf{0}$. So, $\bar{\mathbf{d}} \in \mathcal{D}(\bar{\mathbf{x}})$.

Now by Taylor's theorem, we have for each j

$$0 = h_j(\mathbf{y}_k) = h_j(\bar{\mathbf{x}}) + \delta_k \nabla h_j(\bar{\mathbf{x}})\mathbf{d}_k + \frac{\delta_k^2}{2}\mathbf{d}_k^T \nabla^2 h_j(\mathbf{t}_j)\mathbf{d}_k \qquad (6)$$

and

$$0 \geq f(\mathbf{y}_k) - f(\bar{\mathbf{x}}) = \delta_k \nabla f(\bar{\mathbf{x}})\mathbf{d}_k + \frac{\delta_k^2}{2}\mathbf{d}_k^T \nabla^2 f(\mathbf{t}_0)\mathbf{d}_k, \qquad (7)$$

where each t_j is a point on the line segment joining $\bar{\mathbf{x}}$ and $\mathbf{y}_k$. Multiplying (6) by $-\bar{y}_j$ and adding these to (7) we obtain

$$0 \geq \frac{\delta_k^2}{2}\mathbf{d}_k^T \nabla_x^2 L(\bar{\mathbf{y}}_k, \bar{\mathbf{y}})\mathbf{d}_k,$$

which yields a contradiction as $k \to \infty$. $\square$

Sensitivity: The Meaning of the Lagrangian Multipliers

The i-th multiplier $\bar{y}_i$ measures, in a sense, the sensitivity of the value of the objective function at the local optimum $\bar{\mathbf{x}}$ to a small relaxation of the i-th constraint h_i. We assume that that the constrained set is given by

$$\{\mathbf{x} \in \mathcal{R}^n | h(\mathbf{x}) = c\},$$

where $c = (c_1, \ldots, c_m)$. Let $\bar{\mathbf{x}}(c)$ be a local optimum of (EP) for each c and the constrained qualification holds. By Lagrange Theorem, there exists $\bar{\mathbf{y}}(c) \in \mathcal{R}^m$ such that

$$\nabla f(\bar{\mathbf{x}}(c)) - \sum_{i=1}^{m} \bar{y}_i(c) \nabla h_i(\bar{\mathbf{x}}(c)) = \mathbf{0}.$$

Suppose that $\bar{\mathbf{x}}(c)$ is a differentiable function, then we have

$$\begin{aligned}
\partial f(\bar{\mathbf{x}}(c))/\partial c_i &= \left(\frac{\partial f(\bar{\mathbf{x}}(c))}{\partial x_1}, \ldots, \frac{\partial f(\bar{\mathbf{x}}(c))}{\partial x_n}\right) \begin{pmatrix} \frac{\partial \bar{\mathbf{x}}_1(c)}{\partial c_i} \\ \vdots \\ \frac{\partial \bar{\mathbf{x}}_n(c)}{\partial c_i} \end{pmatrix} \\
&= (\bar{y}_1(c), \ldots, \bar{y}_m(c)) \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix} \begin{pmatrix} \frac{\partial \bar{\mathbf{x}}_1(c)}{\partial c_i} \\ \vdots \\ \frac{\partial \bar{\mathbf{x}}_n(c)}{\partial c_i} \end{pmatrix} \\
&= (\bar{y}_1(c), \ldots, \bar{y}_m(c)) \begin{pmatrix} \frac{\partial h_1}{\partial c_i} \\ \vdots \\ \frac{\partial h_m}{\partial c_i} \end{pmatrix} = \bar{y}_i(c).
\end{aligned}$$

$\frac{\partial h_j}{\partial c_i} = \delta_{ij}$

This states precisely that a small relaxation in constraint i will raise the minimized value of the objective function by $\bar{y}_i(c)$. Therefore, $\bar{y}_i(c)$ is the marginal value of the shadow price of constraint i at c.

Inequality-Constrained Problems

Let us now consider the inequality-constrained problem

$$
(\mathbf{IP}) \qquad \begin{array}{ll} \textbf{minimize} & f(\mathbf{x}) \\ \textbf{subject to} & \mathbf{c}(\mathbf{x}) \geq \mathbf{0}. \end{array}
$$

If, at some vector $\bar{\mathbf{x}}$, we have $c_i(\bar{\mathbf{x}}) = 0$, then the i-th constraint is said to be active or binding at $\bar{\mathbf{x}}$. Relative to this problem, we again define the (possibly empty) set

$$
\mathcal{A}(\bar{\mathbf{x}}) := \{i : \ c_i(\bar{\mathbf{x}}) = 0\}.
$$

The KKT Theorem

Theorem 13 *(Karush [1939]; Kuhn & Tucker [1951]) If $\bar{\mathbf{x}}$ is a local minimizer for (IP), and the (KKT) constraint qualification,* 约束规范 或者是线性约束

$$(\nabla c_i(\bar{\mathbf{x}}),\ i \in \mathcal{A}(\bar{\mathbf{x}})) \text{ is linearly independent,}$$

约束规范+局部极小

i.e., **LICQ**, *is satisfied at $\bar{\mathbf{x}}$, then there exist numbers $\bar{y}_1, \ldots, \bar{y}_m$ such that*

唯一

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \bar{y}_i \nabla c_i(\bar{\mathbf{x}}) = \mathbf{0}$$ FOC

KKT条件

$$\left.\begin{aligned} \bar{y}_i &\geq 0 \\ \bar{y}_i = 0 \quad &\textbf{if } i \notin \mathcal{A}(\bar{\mathbf{x}}) \end{aligned}\right\} \quad \textbf{for } i = 1, \ldots, m$$ 互补松弛条件

也即 $\bar{y}_i \geq 0$
$c_i(\bar{x}) \geq 0$
$y_i c_i(\bar{x}) = 0$
$\forall\ 1 \leq i \leq m$

Proof when Constraints are Linear

Consider linearly constrained case:

$$S = \{\mathbf{x} : \; A\mathbf{x} \geq \mathbf{b}\} \Rightarrow \mathcal{F}_{\bar{\mathbf{x}}} = \{\mathbf{d} : A_{i.}\mathbf{d} \geq 0, \; \forall i \in \mathcal{A}(\bar{\mathbf{x}})\},$$

or

$$\mathcal{F}_{\bar{\mathbf{x}}} = \{\mathbf{d} : \bar{A}\mathbf{d} \geq \mathbf{0}\}.$$

If $\bar{\mathbf{x}}$ is a local optimizer, then by Theorem 2, 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 Farkas' lemma it must be true

$$\nabla f(\bar{\mathbf{x}}) = \sum_{i \in \mathcal{A}(\bar{\mathbf{x}})} \bar{y}_i A_{i.}, \quad \bar{y}_i \geq 0.$$

Let $\bar{y}_i = 0$ for all remaining $i \notin \mathcal{A}(\bar{\mathbf{x}})$. Then we complete the proof. $\square$

The KKT Conditions

The necessary conditions of local optimality (in the theorem above) are called the Karush-Kuhn-Tucker conditions.

The vector $\bar{\mathbf{x}}$ is called a KKT stationary point, and $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is called a KKT pair.

To say that $\bar{\mathbf{x}}$ is a KKT stationary point means that there exists a vector $\bar{\mathbf{y}}$ such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a KKT pair, i.e., satisfies the KKT first-order necessary conditions of local optimality.

一阶必要条件

有最优解的时候：

必要性+KKT点唯一+线性约束时 => 充分性

Example 2: The KKT Conditions is not Sufficient

Consider the optimization problem

$$\begin{aligned} \textbf{minimize} \quad & x_2 \\ \textbf{subject to} \quad & x_1^2 + x_2 \geq 0 \end{aligned}$$

Its KKT points satisfy

$$\begin{cases} \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \lambda \begin{pmatrix} 2x_1 \\ 1 \end{pmatrix} = 0 \\ \lambda(x_1^2 + x_2) = 0, \quad x_1^2 + x_2 \geq 0, \quad \lambda \geq 0 \end{cases}$$

This shows that the problem has the unique KKT point $\mathbf{x}^* = (0; 0)$ with $\lambda^* = 1$, and the LICQ is satisfied at $\mathbf{x}^*$. But $\mathbf{x}^*$ is not optimal.

Example 3: The need for a CQ

Consider the optimization problem

$$\begin{aligned} \textbf{minimize} \quad & (x_1 - 1)^2 + (x_2 - 1)^2 \\ \textbf{subject to} \quad & (1 - x_1 - x_2)^3 \geq 0 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

换成线性约束?

非线性约束时约束规范的必要性

C_2? C_3?

This problem has a unique optimal solution: $\bar{x} = (\frac{1}{2}, \frac{1}{2})$.

Clearly, the (KKT) constraint qualification, $(\nabla c_i(\mathbf{x}),\ i \in \mathcal{A}(\mathbf{x}))$ is linearly independent, is not satisfied at $\bar{x}$, and $\bar{x}$ is not a KKT point.

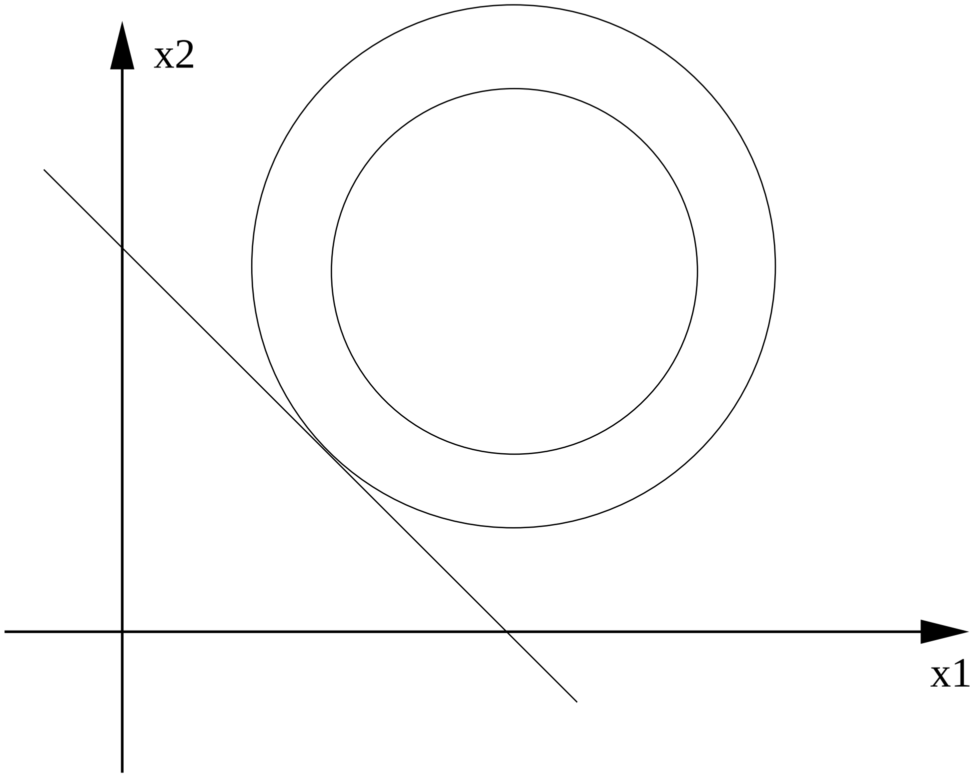


Figure 2: Example 3

Example 3 continued

For (IP), if $\bar{\mathbf{x}}$ is an optimal solution then by Theorem 2, we have

$$(GOC_0) \qquad \mathcal{D}^0_{\bar{\mathbf{x}}} \cap \mathcal{F}^0_{\bar{\mathbf{x}}} = \emptyset,$$

where 下降方向 可行方向

$$\mathcal{D}^0_{\bar{\mathbf{x}}} = \{\mathbf{d} \in R^n : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\}, \quad \mathcal{F}^0_{\bar{\mathbf{x}}} = \{\mathbf{d} \in R^n : \nabla c_i(\bar{\mathbf{x}})\mathbf{d} > 0, \forall i \in \mathcal{A}(\bar{x})\}.$$

Example 3 shows that (GOC_0) is not sufficient.

只是必要条件

The feasible region of Example 3 is the same as that with constraints

$$x_1 + x_2 \leq 1, \quad x_1 \geq 0, \quad x_2 \geq 0.$$

Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)$ be any point satisfying $\tilde{x}_1 + \tilde{x}_2 = 1$ with $\tilde{x}_1 \neq 0$ and $\tilde{x}_2) \neq 0$. With $c_1(x) = (1 - x_1 - x_2)^3$, we have

$$\nabla c_1(\tilde{x}) = 3(1 - \tilde{x}_1 - \tilde{x}_2)^2(-1, \ -1) = (0, \ 0).$$

Clearly, $\mathcal{D}^0_{\tilde{x}} \cap \mathcal{F}^0_{\tilde{x}} = \emptyset$. This illustrates that the condition (GOC_0) can be satisfied by infinitely many nonoptimal points in the feasible region as well as by the optimal solution, and hence it is not sufficient.

Fritz John's Theorem

Theorem 14 *(F.John [1948]) If $\bar{\mathbf{x}}$ is a local minimizer for (IP) in which the functions f and $c_i, i = 1, \ldots, m$ are differentiable, then there exists a set of nonnegative scalars $\bar{y}_0, \bar{y}_1, \ldots, \bar{y}_m$ not all of which are zero such that*

$$\bar{y}_0 \nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \bar{y}_i \nabla c_i(\bar{\mathbf{x}}) = \mathbf{0}$$

目前为止还不需要约束规范

$$\bar{y}_i c_i(\bar{\mathbf{x}}) = 0, \quad i = 1, \ldots, m$$

Proof

We may assume that $\mathcal{A}(\bar{\mathbf{x}}) \neq \emptyset$. Since $\bar{\mathbf{x}}$ is a local minimizer of (IP), the system

$$\begin{aligned} & v^T \nabla f(\bar{\mathbf{x}}) < 0 \\ & v^T \nabla c_i(\bar{\mathbf{x}}) > 0 \quad \textbf{for all } \ i \in \mathcal{A}(\bar{\mathbf{x}}) \end{aligned}$$

has no solution. Accordingly, by Gordan's Theorem, there exist nonnegative scalars $\bar{y}_0, \bar{y}_i (i \in \mathcal{A}(\bar{\mathbf{x}}))$ not all of which are zero such that

$$\bar{y}_0 \nabla f(\bar{\mathbf{x}}) - \sum_{i \in \mathcal{A}(\bar{\mathbf{x}})} \bar{y}_i \nabla c_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

If $i \in \{1, \ldots, m\} \backslash \mathcal{A}(\bar{\mathbf{x}})$, define $\bar{y}_i = 0$. This completes the proof. $\square$

Remark

Notice that Fritz John's Theorem requires no constraint qualification, yet we know that a constraint qualification is required. So what's the catch? The catch is that the scalar $\bar{y}_0$ might equal zero. Notice, though, that if $\bar{y}_0 > 0$, that the conclusion of the Karush-Kuhn-Tucker theorem holds with the multipliers $\bar{y}_1/\bar{y}_0, \ldots, \bar{y}_m/\bar{y}_0$. Notice also that if $\bar{y}_0 = 0$, then the vectors $\nabla c_i(\bar{\mathbf{x}})$ must be (positively) linearly dependent. Hence any condition that rules out such linear dependence will imply $\bar{y}_0 > 0$. In addition, the proof of Fritz John's Theorem gives the proof of the KKT theorem when constraints are nonlinear.

于是利用约束规范得不等于0

Optimization with Mixed Constraints

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

$$
(\mathbf{P}) \qquad \begin{array}{ll} \textbf{minimize} & f(\mathbf{x}) \\ \textbf{subject to} & c_i(\mathbf{x}) \geq 0 \quad i \in \mathcal{I} \\ & h_i(\mathbf{x}) = 0 \quad i \in \mathcal{E} \end{array}
$$

Typically, we take

$$
\mathcal{I} = \{1, \ldots, m\}, \quad \mathcal{E} = \{1, \ldots, \ell\},
$$

and

$$
\mathbf{c}(\mathbf{x}) = (c_1(\mathbf{x}); ...; c_m(\mathbf{x})), \quad \mathbf{h}(\mathbf{x}) = (\mathbf{h_1}(\mathbf{x}); ...; \mathbf{h}_\ell(\mathbf{x})).
$$

The KKT Theorem for (P)

Our aim is to establish an Karush-Kuhn-Tucker necessary conditions of optimality in (P).

Theorem 15 *Let $\bar{\mathbf{x}}$ be a local minimizer for (P). Assume the functions c_i are differentiable at $\bar{\mathbf{x}}$ for all $i \in \mathcal{I}$, and the functions h_i are continuously differentiable at $\bar{\mathbf{x}}$ for all $i \in \mathcal{E}$. If all the vectors $\nabla c_i(\bar{\mathbf{x}})$* **for** *$i \in \mathcal{A}(\bar{\mathbf{x}})$ and $\nabla h_i(\bar{\mathbf{x}})$* **for** *$i \in \mathcal{E}$ are linearly independent, then there exist (unique) multipliers $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_\ell$ such that*

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \lambda_i \nabla c_i(\bar{\mathbf{x}}) - \sum_{i=1}^{\ell} \mu_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$\lambda_i \geq 0 \qquad \forall i = 1, \ldots, m$$

$$\lambda_i = 0 \qquad \text{if } i \notin \mathcal{A}(\bar{\mathbf{x}}).$$

对μi没有限制

Proof of The KKT Theorem for (P)

Lemma 2 *Let $\bar{\mathbf{x}}$ be a local minimizer for (P) and a regular point of the surface $\{\mathbf{x} \in \mathcal{R}^n | \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$. If the functions c_i are continuous at $\bar{\mathbf{x}}$ for all $i \notin \mathcal{A}(\bar{\mathbf{x}})$, the functions c_i are differentiable at $\bar{\mathbf{x}}$ for all $i \in \mathcal{A}(\bar{\mathbf{x}})$, and the functions h_i are continuously differentiable at $\bar{\mathbf{x}}$ for all $i \in \mathcal{E}$, then*

$$\mathcal{D} \cap \mathcal{F} = \emptyset,$$

where

$$\begin{aligned}
\mathcal{A}(\bar{\mathbf{x}}) &= \{i \in \mathcal{I} : c_i(\bar{\mathbf{x}}) = 0\} \\
\mathcal{D} &= \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\} \\
\mathcal{F} &= \{\mathbf{d} : \ \nabla c_i(\bar{\mathbf{x}})\mathbf{d} > 0 \ \ \textbf{for all } i \in \mathcal{A}(\bar{x}), \ \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{d} = \mathbf{0}\}.
\end{aligned}$$

可行方向

Proof. Assume by contradiction that there exists $d \in \mathcal{D} \cap \mathcal{F}$. By Theorem 10, there is a curve $x(t)$ on the surface passing through $\bar{\mathbf{x}}$ such that $x(0) = \bar{\mathbf{x}}$ and $\dot{x}(0) = d$.

? ?

For $i \in \mathcal{A}(\bar{\mathbf{x}})$, we have

$$\frac{d}{dt} c_i(x(t))|_{t=0} = \nabla c_i(\bar{\mathbf{x}})^T \dot{x}(0) = \nabla c_i(\bar{\mathbf{x}})^T d > 0.$$

Hence, there exists $\delta_1 > 0$ such that $c_i(x(t)) \geq 0$ for all $t \in [0, \delta_1)$.

For $i \notin \mathcal{A}(\bar{\mathbf{x}})$, since $c_i(\bar{x}) > 0$ and c_i is continuous at $\bar{\mathbf{x}}$, there exists $\delta_2 > 0$ such that $c_i(x(t)) \geq 0$ for all $t \in [0, \delta_2)$.

Since

$$\frac{d}{dt} f(x(t))|_{t=0} = \nabla f(\bar{\mathbf{x}})^T d < 0,$$

there exists $\delta_3 > 0$ such that

$$f(x(t)) < f(\bar{\mathbf{x}}) \quad \forall t \in [0, \delta_3).$$

Let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Then for any $t \in [0, \delta)$, $x(t)$ is feasible for (P) and $f(x(t)) < f(\bar{x})$. This is a contradiction. □

Lemma 3 *Let $A \in \mathcal{R}^{m\times n}$, $B \in \mathcal{R}^{p\times n}$. If the system $A\mathbf{x} < \mathbf{0},\ B\mathbf{x} = \mathbf{0}$ is infeasible, then there exist $\mathbf{y} \geq \mathbf{0}$, $\mathbf{u}$ not all zero such that $A^T\mathbf{y} + B^T\mathbf{u} = 0$.*

By the above two lemmas, there exist multipliers $\lambda_0, \lambda_1, \ldots, \lambda_m, u_1, \ldots, u_\ell$ not all zero such that

$$\lambda_0 \nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \lambda_i \nabla c_i(\bar{\mathbf{x}}) - \sum_{i=1}^{\ell} u_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$\lambda_0,\ \lambda_i \geq 0 \quad \forall i = 1, \ldots, m.$$

By the LICQ, we prove the theorem. □

Second-Order Conditions for (P)

The second-order conditions, both necessary and sufficient, for problems with inequality constraints, are derived essentially by consideration only of the equality constrained problem that is implied by the active constraints. The appropriate tangent plane for these problems is the plane tangent to the active constraints.

Theorem 16 (Second-Order Necessary Condition) *Suppose the functions $f, \mathbf{c}, \mathbf{h} \in \mathcal{C}^{\mathbf{2}}$ and the hypotheses in Theorem 15 hold, then there exist (unique) multipliers $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_\ell$ such that $\bar{\mathbf{x}}$ is a KKT point and the Lagrange Hessian matrix $\nabla_x^2 L(\bar{\mathbf{x}}, \lambda, \mu)$ is positive semidefinite on the tangent subspace $\mathcal{T}(\bar{\mathbf{x}})$ of the active constraints at $\bar{\mathbf{x}}$, i.e.,*

$$\mathcal{T}(\bar{\mathbf{x}}) = \{\mathbf{d} : \ \nabla c_i(\bar{\mathbf{x}})\mathbf{d} = 0 \quad \textbf{for all } i \in \mathcal{A}(\bar{x}), \ \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{d} = \mathbf{0}\}.$$

Theorem 17 (Second-Order Sufficient Condition) *Let the functions $f,\ c_i, i \in \mathcal{I},\ h_i, i \in \mathcal{E}$ be twice continuously differentiable and $\mathbf{x}^*$ be a feasible solution to (P). If there exist vectors w^* and v^* such that $(\mathbf{x}^*, w^*, v^*)$ is a KKT-pair for this problem and the corresponding Lagrange Hessian matrix $\nabla_x^2 L(\mathbf{x}^*, w^*, v^*)$ is positive definite on the set*

$$G = \left\{ \mathbf{d} \neq \mathbf{0} \;\middle|\; \begin{array}{ll} \nabla c_i(\mathbf{x}^*)^T \mathbf{d} = 0, & i \in \mathcal{A}(\mathbf{x}^*) \text{ with } w_i^* > 0 \\ \nabla c_i(\mathbf{x}^*)^T \mathbf{d} \geq 0, & i \in \mathcal{A}(\mathbf{x}^*) \text{ with } w_i^* = 0 \\ \nabla h_i(\mathbf{x}^*)^T \mathbf{d} = 0, & i \in \mathcal{E} \end{array} \right\},$$

then $\mathbf{x}^*$ is a strict local minimum of (P).

Proof

If $\mathbf{x}^*$ is not a strict local minimum point, then there exists a sequence $\{\mathbf{x}^k\} \subset S$ such that $\mathbf{x}^k \neq \mathbf{x}^*$, $\mathbf{x}^k \to \mathbf{x}^*$ and $f(\mathbf{x}^k) \leq f(\mathbf{x}^*)$. Let $\mathbf{d}^k = (\mathbf{x}^k - \mathbf{x}^*)/\|\mathbf{x}^k - \mathbf{x}^*\|$, we assume, without loss of generality, that $\mathbf{d}^k \to \mathbf{d}$ with $\|\mathbf{d}\| = 1$. Since

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^T(\mathbf{x}^k - \mathbf{x}^*) + o(\|\mathbf{x}^k - \mathbf{x}^*\|) \leq 0,$$

we have $\nabla f(\mathbf{x}^*)^T\mathbf{d} \leq 0$. For $i \in \mathcal{A}(\mathbf{x}^*)$, $c_i(\mathbf{x}^*) = 0$. Since $\mathbf{x}^k \in S$,

$$c_i(\mathbf{x}^k) = c_i(\mathbf{x}^*) + \nabla c_i(\mathbf{x}^*)^T(\mathbf{x}^k - \mathbf{x}^*) + o(\|\mathbf{x}^k - \mathbf{x}^*\|) \geq 0,$$

i.e.,

$$\nabla c_i(\mathbf{x}^*)^T\mathbf{d} \geq 0, \quad i \in \mathcal{A}(\mathbf{x}^*).$$

Similarly, we have $\nabla h_i(\mathbf{x}^*)^T\mathbf{d} = 0, i \in \mathcal{E}$. By the KKT condition,

$$\nabla f(\mathbf{x}^*)^T\mathbf{d} = \sum_{i \in \mathcal{I}(\mathbf{x}^*)} w_i^* \nabla c_i(\mathbf{x}^*)^T\mathbf{d} + \sum_{i \in \mathcal{E}} v_i^* \nabla h_i(\mathbf{x}^*)^T\mathbf{d}.$$

This yields $\nabla f(\mathbf{x}^*)^T\mathbf{d} = 0$ and $\nabla c_i(\mathbf{x}^*)^T\mathbf{d} = 0$ for $i \in \mathcal{A}(\mathbf{x}^*)$ with $w_i^* > 0$. That is, $\mathbf{d} \in G$ and hence $\mathbf{d}^T\nabla_x^2 L(\mathbf{x}^*, w^*, v^*)\mathbf{d} > 0$.

On the other hand, since $(\mathbf{x}^*, w^*, v^*)$ is a KKT-pair, $L(\mathbf{x}^*, w^*, v^*) = f(\mathbf{x}^*)$,

$$L(\mathbf{x}^k, w^*, v^*) = f(\mathbf{x}^k) - \sum_{i\in\mathcal{I}} w_i^* c_i(\mathbf{x}^k) - \sum_{i\in\mathcal{E}} v_i^* h_i(\mathbf{x}^k) \le f(\mathbf{x}^k),$$

and $\nabla_x L(\mathbf{x}^*, w^*, v^*) = 0$. Since $f(\mathbf{x}^*) \ge f(\mathbf{x}^k)$ and

$$L(\mathbf{x}^k, w^*, v^*) = L(\mathbf{x}^*, w^*, v^*) + \frac{1}{2}(\mathbf{x}^k - \mathbf{x}^*)^T\nabla_x^2 L(\mathbf{x}^*, w^*, v^*)(\mathbf{x}^k - \mathbf{x}^*) + o(\|\mathbf{x}^k - \mathbf{x}^*\|^2),$$

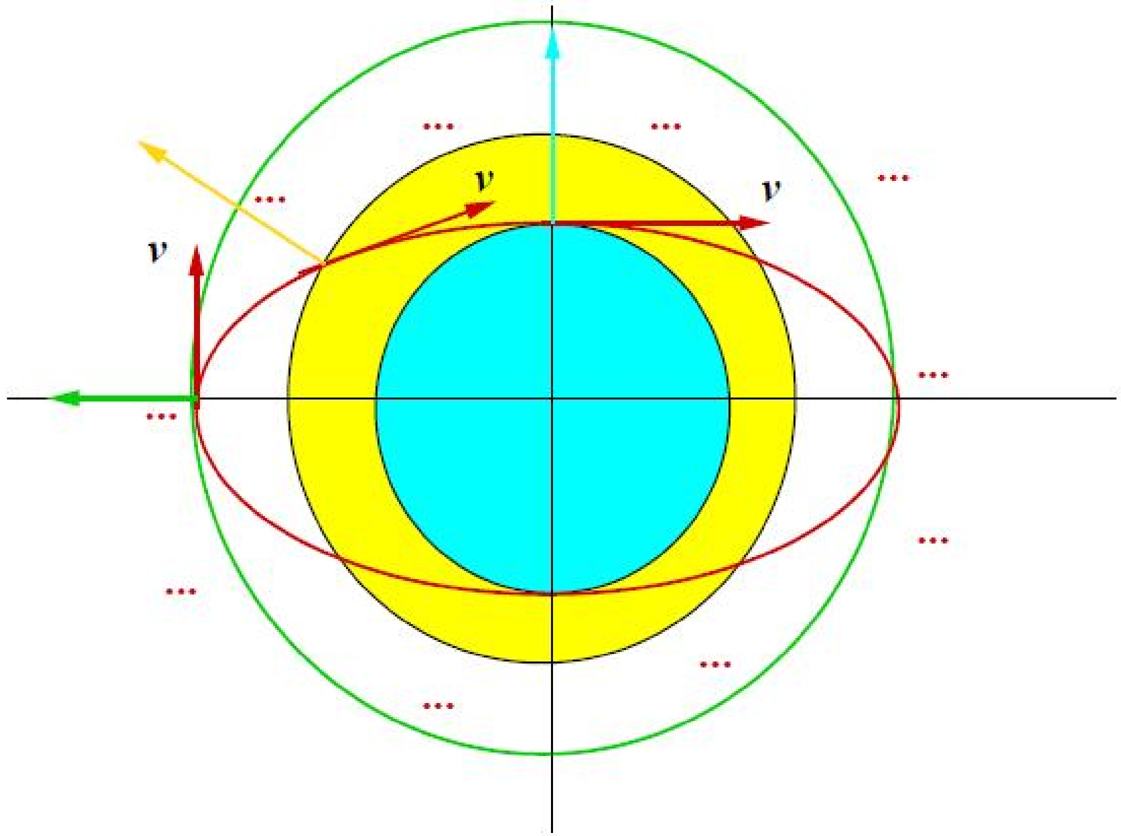
we have

$$\mathbf{d}^T\nabla_x^2 L(\mathbf{x}^*, w^*, v^*)\mathbf{d} \le 0.$$

This is a contradiction. Hence, $\mathbf{x}^*$ is a strict local minimum point. $\square$

Example 1

$$\min \quad (x_1)^2 + (x_2)^2 \quad \text{s.t.} \quad -(x_1)^2/4 - (x_2)^2 + 1 \le 0$$



$$L(x,v)=x_1^2+x_2^2-v(\frac{1}{4}x_1^2+x_2^2-1),$$

$$\nabla_x L(x,v)=\begin{pmatrix} 2x_1(1-\frac{v}{4}) \\ 2x_2(1-v) \end{pmatrix}, \quad \nabla_x^2 L(x,v)=\begin{pmatrix} 2(1-\frac{v}{4}) & 0 \\ 0 & 2(1-v) \end{pmatrix}.$$

$$T(x)=\{(z_1,z_2):\ \frac{1}{4}x_1z_1+x_2z_2=0\}.$$

We see that there are two possible values for v: either 4 or 1, which lead to total four KKT points:

$$\begin{pmatrix} x_1 \\ x_2 \\ v \end{pmatrix}=\begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ 0 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Consider the first KKT point:

$$\nabla_x^2 L(2,0,4) = \begin{pmatrix} 0 & 0 \\ 0 & -6 \end{pmatrix}, \quad G = \{(z_1, z_2) : z_1 = 0\}.$$

Then the Hessian is not possible semidefinite on G since

$$z^T \nabla_x^2 L(2,0,4) z = -6z_2^2 \leq 0.$$

Consider the third KKT point:

$$\nabla_x^2 L(0,1,1) = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & 0 \end{pmatrix}, \quad G = \{(z_1, z_2) : z_2 = 0\}.$$

Then the Hessian is positive definite on G since

$$z^T \nabla_x^2 L(0,1,1) z = \frac{3}{2} z_1^2 > 0, \quad \forall 0 \neq z \in G.$$

So, $(0,1,1)^T$ and $(0,-1,1)^T$ are two optimal solutions.

Example 2

Consider the problem

$$\begin{array}{ll} \min & (x_1-1)^2+x_2^2 \\ s.t. & -x_1+\beta x_2^2=0. \end{array}$$

Please answer the question: $x^*=(0;0)$ is a local optimal solution for what values of β?

Solution: Its Lagrange function

$$L(x,v)=(x_1-1)^2+x_2^2-v(-x_1+\beta x_2^2),$$

and

$$\nabla_x L(x,v)=\begin{pmatrix} 2(x_1-1)+v \\ 2x_2-2v\beta x_2 \end{pmatrix}, \quad \nabla_x^2 L(x,v)=\begin{pmatrix} 2 & 0 \\ 0 & 2-2v\beta \end{pmatrix}.$$

It follows from $\nabla_x L(x^*, v^*) = 0$ that $v^* = 2$. Hence, (x^*, v^*) is a KKT-pair. Let $h(x) = -x_1 + \beta x_2^2$, then by $\nabla h(x^*)^T d = 0$ we obtain $d = (0; d_2)$ with $d_2 \neq 0$, and hence the set

$$G = \{d = (0; d_2) |\ d_2 \neq 0\}.$$

For any $d \in G$, $d^T \nabla_x^2 L(x^*, v^*) d = (2 - 4\beta) d_2^2$. Hence, when $\beta < 1/2$, by the second-order sufficient condition, x^* is a strict local optimal solution.

For $\beta = 1/2$, the original problem is rewritten as

$$\min\ f(x_1) := (x_1 - 1)^2 + 2x_1$$

Obviously, $x_1 = 0$ is a stationary point and $f''(x_1) = 2$. Hence, x^* is a local minimum point of the original problem. Thus, $x^* = (0; 0)$ is is a local optimal solution of this problem with $\beta \leq 1/2$.

Convex Optimization

Let us now consider the inequality-constrained problem

$$(\text{CP}) \qquad \begin{array}{ll} \textbf{minimize} & f(\mathbf{x}) \\ \textbf{subject to} & \mathbf{c}_i(\mathbf{x}) \geq \mathbf{0} \quad i \in \mathcal{I}, \end{array}$$

where f and $-c_i, i \in \mathcal{I}$ are differentiable convex functions. Let $\mathcal{F}$ be the feasible region. If a point $\mathbf{x}^* \in \mathcal{F}$ satisfying

$$\nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{F},$$

the point $\mathbf{x}^*$ is called a stationary point of (CP).

用于验证的充要条件

Theorem 18 *Consider the convex program* $\min\{f(\mathbf{x}),\ \mathbf{x} \in \Omega\}$*, where* Ω *is a closed convex set. Then* $\mathbf{x}^* \in \Omega$ *is an optimal solution if and only if*

$$\nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0, \quad \forall\, \mathbf{x} \in \Omega.$$

Proof. For any $\mathbf{x} \in \Omega$, if $\nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$, then we have by the convexity of f,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}^*).$$

This implies that $\mathbf{x}^* \in \Omega$ is an optimal solution.

Conversely, since $\mathbf{x}^* \in \Omega$ is an optimal solution, there does not exists a descent and feasible direction at $\mathbf{x}^*$. For any $\mathbf{x} \in \Omega$, $\mathbf{x} - \mathbf{x}^*$ is a feasible direction since Ω is a closed convex set. Hence, we have $\nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$. The proof is complete. $\square$

It is clear that the stationary point and the optimal solution is equivalent for (CP).

Sufficient Optimality Conditions

Theorem 19 *If f and $-c_1, \ldots, -c_m$ are differentiable convex functions, then the (first-order) KKT optimality conditions are sufficient for the global optimality of a feasible solution in (CP).*

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

对于可微凸规划用于求解的充要条件

Proof

Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ be a KKT pair for (CP) in which $\bar{\mathbf{x}}$ is a feasible solution. Consider the Lagrangian function $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x})$ associated with (IP) where $\mathbf{y}$ be nonnegative. Then, by our hypotheses, L is a convex and differentiable function of $\mathbf{x}$. Hence by the gradient inequality applied to L

$$L(\mathbf{x}, \bar{\mathbf{y}}) \geq L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \nabla_x L(\bar{\mathbf{x}}, \bar{\mathbf{y}})\,(\mathbf{x} - \bar{\mathbf{x}}) \text{ for all feasible } \mathbf{x}.$$

More explicitly,

$$f(\mathbf{x}) - \bar{\mathbf{y}}^T \mathbf{c}(\mathbf{x}) \geq f(\bar{\mathbf{x}}) - \bar{\mathbf{y}}^T \mathbf{c}(\bar{\mathbf{x}}) + [\nabla f(\bar{\mathbf{x}}) - \bar{\mathbf{y}}^T \nabla c(\bar{\mathbf{x}})](\mathbf{x} - \bar{\mathbf{x}}).$$

Hence,

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \bar{\mathbf{y}}^T \mathbf{c}(\mathbf{x}) \geq f(\bar{\mathbf{x}}).$$

This proves that $\bar{\mathbf{x}}$ is a global minimizer for (CP). $\square$

Remark

Notice that the preceding theorem does not mention the constraint qualification (CQ). It is simply given that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a KKT pair, and this being the case, no CQ assumption is required. But the converse is not true.

Example:

$$\begin{array}{ll} \min & x_1 + x_2 \\ s.t. & (x_1 - 1)^2 + x_2^2 \leq 1, \quad (x_1 + 2)^2 + x_2^2 \leq 4. \end{array}$$

This problem has a unique optimal solution $x^* = (0, 0)$, but it is not a KKT point.

It is clear that KKT point is optimal for (CP), and the converse is not true. The convexity plays a big role in Optimization.

Slater CQ for (CP)

If the set

$$\{\mathbf{x} \in R^n |\, c_i(\mathbf{x}) > 0,\ i \in \mathcal{I}\} \neq \emptyset,$$ 有一个内点就好

it is said that Slater CQ holds for (CP).

Theorem 20 *If Slater CQ is satisfied for (CP), then for any $\mathbf{x} \in \mathcal{F}$ there exists a feasible direction.*

Proof. By hypothesis, there exists $\hat{\mathbf{x}} \in \mathcal{F}$ such that $c_i(\hat{\mathbf{x}}) > 0$ for all $i \in \mathcal{I}$. For any $\mathbf{x} \in \mathcal{F}$, let $\mathbf{d} = \hat{\mathbf{x}} - \mathbf{x}$. The convexity of $-c_i, i \in \mathcal{I}$ implies

$$0 < c_i(\hat{\mathbf{x}}) \leq c_i(\mathbf{x}) + \nabla c_i(\mathbf{x})^T \mathbf{d} = \nabla c_i(\mathbf{x})^T \mathbf{d} \quad \forall i \in \mathcal{A}(\mathbf{x}).$$

□

Theorem 21 *If Slater CQ holds for (CP), then the optimal solution is the KKT point.*

Proof. Let $\mathbf{x}^*$ be the optimal solution of (CP). By Theorem 13 and the geometric optimality condition,

$$\{\mathbf{d} \in R^n | \nabla c_i(\mathbf{x}^*)^T \mathbf{d} > 0, \quad i \in \mathcal{A}(\mathbf{x}^*)\} \neq \emptyset,$$

and

$$S_1 = \{\mathbf{d} \in R^n | \nabla f(\mathbf{x}^*)^T \mathbf{d} < 0, \quad \nabla c_i(\mathbf{x}^*)^T \mathbf{d} > 0, \quad i \in \mathcal{A}(\mathbf{x}^*)\} = \emptyset.$$

It is to show that the set

$$S_2 = \{\mathbf{d} \in R^n | \nabla f(\mathbf{x}^*)^T \mathbf{d} < 0, \quad \nabla c_i(\mathbf{x}^*)^T \mathbf{d} \geq 0, \quad i \in \mathcal{A}(\mathbf{x}^*)\} = \emptyset.$$

Since there exists $\bar{d}$ such that $\nabla c_i(\mathbf{x}^*)^T \bar{d} > 0$, for all $i \in \mathcal{A}(\mathbf{x}^*)$. Take an arbitrary $\mathbf{d} \in R^n$ with $\nabla c_i(\mathbf{x}^*)^T \mathbf{d} \geq 0$ for all $i \in \mathcal{A}(\mathbf{x}^*)$. Then, for any sufficiently $\varepsilon > 0$,

$$(\mathbf{d} + \varepsilon \bar{d})^T \nabla c_i(\mathbf{x}^*) > 0, \quad i \in \mathcal{A}(\mathbf{x}^*).$$

Since $S_1 = \emptyset$,

$$(\mathbf{d} + \varepsilon \bar{d})^T \nabla f(\mathbf{x}^*) \geq 0.$$

Let $\varepsilon \to 0$, we have $\nabla f(\mathbf{x}^*)^T \mathbf{d} \geq 0$. Hence $S_2 = \emptyset$. By Farkas Lemma,

$$A^T \mathbf{x} \geq \mathbf{0}, \ \mathbf{b}^T \mathbf{x} < 0 \quad \textbf{and} \quad A\mathbf{y} = \mathbf{b}, \ \mathbf{y} \geq \mathbf{0}$$

are the infeasible certificate systems. This shows that $\mathbf{x}^*$ is a KKT point of (CP).

$\square$

Exercise

Consider the following problem

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ s.t. \quad & A\mathbf{x} = \mathbf{0}, \\ & \mathbf{x}^T\mathbf{x} \leq 1, \end{aligned}$$

where $A \in \mathcal{R}^{m\times n}$ with **rank**$(A) = m$, and $\|A^T(AA^T)^{-1}A\mathbf{c} - \mathbf{c}\| \neq 0$.
Please write the analytic expression of its optimal solution.

Lagrangian Duality

Consider the nonlinear programming problem (P), which we call the primal problem here.

Primal Problem P

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ s.t. \quad & \mathbf{c}(\mathbf{x}) \geq \mathbf{0}, \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ & \mathbf{x} \in \mathcal{X}. \text{ 约束集} \end{aligned}$$

Lagrangian Dual Problem D

$$\begin{aligned} \max \quad & \theta(\mathbf{u}, \mathbf{v}) \\ s.t. \quad & \mathbf{u} \geq \mathbf{0}, \end{aligned}$$

where

$$\theta(\mathbf{u}, \mathbf{v}) = \inf_{\mathbf{x} \in \mathcal{X}} \{L(\mathbf{x}, \mathbf{u}, \mathbf{v})\}$$

and the Lagrangian function

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) - \mathbf{u}^T \mathbf{c}(\mathbf{x}) - \mathbf{v}^T \mathbf{h}(\mathbf{x}).$$

Example for Lagrange Dual

Consider the following primal problem

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ s.t. \quad & x_1 + x_2 - 4 \geq 0, \\ & x_1, x_2 \geq 0. \end{aligned}$$

Note that the optimal solution occurs at $(x_1, x_2) = (2, 2)$, whose objective value is equal to 8.

Let $\mathcal{X} = \{\mathbf{x} \in R^2 |\, x_1, x_2 \geq 0\}$. The dual objective function is given by

$$\begin{aligned} \theta(u) &= \inf\{x_1^2 + x_2^2 - u(x_1 + x_2 - 4) : \ (x_1, x_2) \in \mathcal{X}\} \\ &= \inf\{x_1^2 - ux_1 : \ x_1 \geq 0\} + \inf\{x_2^2 - ux_2 : \ x_2 \geq 0\} + 4u \\ &= \begin{cases} -\frac{1}{2}u^2 + 4u & \text{for } u \geq 0 \\ 4u & \text{for } u < 0. \end{cases} \end{aligned}$$

Then its Lagrangian dual problem is

$$\begin{array}{ll} \max & -\frac{1}{2}u^2 + 4u \\ s.t. & u \geq 0, \end{array}$$

whose optimal solution is $u = 4$ and the objective value is also 8.

Let $\mathcal{X} = R^2$. The dual objective function is given by

$$\begin{aligned}\theta(u,v,w) &= \inf\{x_1^2 + x_2^2 - u(x_1 + x_2 - 4) - vx_1 - wx_2 : (x_1, x_2) \in \mathcal{X}\} \\ &= \inf\{x_1^2 - (u+v)x_1\} + \inf\{x_2^2 - (u+w)x_2\} + 4u,\end{aligned}$$

which implies that for $u, v, w \geq 0$ we have

$$\theta(u,v,w) = -\frac{1}{4}(u+v)^2 - \frac{1}{4}(u+w)^2 + 4u.$$

Then its Lagrangian dual problem is

$$\begin{array}{ll}\max & -\frac{1}{4}(u+v)^2 - \frac{1}{4}(u+w)^2 + 4u \\ s.t. & u \geq 0,\ v \geq 0,\ w \geq 0.\end{array}$$

其KKT条件?

Its optimal solution is $u = 4,\ v = 0,\ w = 0$, and the objective value is also 8.

Weak Duality Theorem

Theorem 22 *Let* $\mathbf{x}$ *be a feasible solution of* **Problem P** *and let* $(\mathbf{u}, \mathbf{v})$ *be a feasible solution of* **Problem D**. *Then* $f(\mathbf{x}) \geq \theta(\mathbf{u}, \mathbf{v})$.

Proof. By the definition of θ and since $\mathbf{x} \in \mathcal{F}_p$, we have

$$\theta(\mathbf{u}, \mathbf{v}) \leq f(\mathbf{x}) - \mathbf{u}^T \mathbf{c}(\mathbf{x}) - \mathbf{v}^T \mathbf{h}(\mathbf{x}) \leq \mathbf{f}(\mathbf{x})$$

since $\mathbf{u} \geq \mathbf{0}$, $\mathbf{c}(\mathbf{x}) \geq \mathbf{0}$, and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. This completes the proof. $\square$

Corollary

(i) $\inf\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{F}_p\} \geq \sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\}$.

(ii) If $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in \mathcal{F}_p$ and $\bar{\mathbf{u}} \geq \mathbf{0}$, then $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solve the primal and dual problems, respectively.

(iii) If $\inf\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{F}_p\} = -\infty$, then $\theta(\mathbf{u}, \mathbf{v}) = -\infty$ for each $\mathbf{u} \geq \mathbf{0}$.

(iv) If $\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\} = \infty$, then the primal problem has no feasible solution.

Duality Gap

From Corollary to Weak Duality Theorem, the optimal objective value of the primal problem is greater than or equal to the optimal objective value of the dual problem. If strict inequality holds true, a duality gap is said to exist.

Example: Consider the following problem

一般的强对偶定理未必成立

$$\begin{aligned} \min \quad & -2x_1 + x_2 \\ s.t. \quad & x_1 + x_2 - 3 = 0, \\ & (x_1, x_2) \in \mathcal{X}, \end{aligned}$$

where $\mathcal{X} = \{(0,0), (0,4), (4,4), (4,0), (1,2), (2,1)\}$.

It is easy to verify that $(2,1)$ is the primal optimal solution with objective value equal to -3.

The dual objective function is given by

$$\begin{aligned}\theta(v) &= \inf\{-2x_1 + x_2 - v(x_1 + x_2 - 3) : (x_1, x_2) \in \mathcal{X}\} \\ &= \begin{cases} -4 - 5v & \text{for } v \geq 1 \\ -8 - v & \text{for } -2 \leq v \leq 1 \\ 3v & \text{for } v \leq -2. \end{cases}\end{aligned}$$

Then, the dual optimal solution is $\bar{v} = -2$ with objective value -6.

Note that there exists a duality gap in this example.

Strong Duality Theorem

Consider **Problem P** in the inequality-constrained form.

Theorem 23 *Let $\mathcal{X} \subset \mathcal{R}^n$ be a nonempty convex set, and let f and $-c_1, \ldots, -c_m$ be convex functions. Suppose that the following CQ holds true:*

$$\{\mathbf{x} \in \mathcal{X} : \ \mathbf{c}(\mathbf{x}) > \mathbf{0}\} \neq \emptyset,$$ 有内点

which is called the Slater Constraint Qualification. Then

$$\inf\{f(\mathbf{x}) : \ \mathbf{c}(\mathbf{x}) \geq \mathbf{0}, \mathbf{x} \in \mathcal{X}\} = \sup\{\theta(\mathbf{u}) : \ \mathbf{u} \geq \mathbf{0}\}. \tag{8}$$

Furthermore, if the $\inf$ is finite, then $\sup\{\theta(\mathbf{u}) : \ \mathbf{u} \geq \mathbf{0}\}$ is achieved at $\bar{\mathbf{u}}$ with $\bar{\mathbf{u}} \geq \mathbf{0}$. If the $\inf$ is achieved at $\bar{\mathbf{x}}$, then $\bar{\mathbf{u}}^T \mathbf{c}(\bar{\mathbf{x}}) = 0$.

Lemma 4 *Let $\mathcal{X} \subset \mathcal{R}^n$ be a nonempty convex set, and let f and $-c_1, \ldots, -c_m$ be convex functions. Consider the following two systems:*

System I: $f(\mathbf{x}) < 0,\ \mathbf{c}(\mathbf{x}) \geq \mathbf{0}$ **for some** $\mathbf{x} \in \mathcal{X}$

System II: $u_0 f(\mathbf{x}) - \mathbf{u}^T \mathbf{c}(\mathbf{x}) \geq 0$ **for all** $\mathbf{x} \in \mathcal{X}$

If System I has no solution $\mathbf{x}$, then System II has a solution $(u_0, \mathbf{u})$ with $(u_0, \mathbf{u}) \geq \mathbf{0}$ and $(u_0, \mathbf{u}) \neq \mathbf{0}$. The converse holds true if $u_0 > 0$.

Proof: Suppose that System I has no solution, then

$$(0, \mathbf{0}) \notin \Omega := \{(\alpha, \mathbf{y}) :\ \alpha > f(\mathbf{x}), \mathbf{y} \leq \mathbf{c}(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathcal{X}\}.$$

Noting that $\mathcal{X}, f, -\mathbf{c}$ are convex, it can be easily shown that Ω is convex. Since $(0, \mathbf{0}) \notin int(\Omega)$, by the Separating Hyperplane Theorem, there exists a vector $(u_0, \mathbf{u}) \neq \mathbf{0}$ such that

$$u_0 \alpha - \mathbf{u}^T \mathbf{y} \geq 0 \quad \forall (\alpha, \mathbf{y}) \in cl(\Omega). \tag{9}$$

Now fix an $\mathbf{x} \in \mathcal{X}$. Since α and $-\mathbf{y}$ can be made arbitrarily large, (9) holds true only if $(u_0, \mathbf{u}) \geq \mathbf{0}$. Furthermore, it follows from $(f(\mathbf{x}), \mathbf{c}(\mathbf{x})) \in cl(\Omega)$ and (9) that $u_0 f(\mathbf{x}) - \mathbf{u}^T \mathbf{c}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}$.

To prove the converse, assume that System II has a solution $(u_0, \mathbf{u})$ with $u_0 > 0$ and $\mathbf{u} \geq \mathbf{0}$. Let $\mathbf{x} \in \mathcal{X}$ such that $\mathbf{c}(\mathbf{x}) \geq \mathbf{0}$. Then $u_0 f(\mathbf{x}) \geq 0$ since $\mathbf{u} \geq \mathbf{0}$. Thus, $f(\mathbf{x}) \geq 0$ since $u_0 > 0$. Hence, System I has no solution. This completes the proof. $\square$

Proof of Strong Duality Theorem

Let $\beta = \inf\{f(\mathbf{x}) : \ \mathbf{c}(\mathbf{x}) \geq \mathbf{0}, \mathbf{x} \in \mathcal{X}\}$. By assumption, $\beta < \infty$. If $\beta = -\infty$, then by Corollary to Weak Duality Theorem, $\sup\{\theta(\mathbf{u}) : \ \mathbf{u} \geq \mathbf{0}\} = -\infty$, and hence (8) holds true. Suppose that β is finite, and consider the following system:

$$f(\mathbf{x}) - \beta < 0, \quad \mathbf{c}(\mathbf{x}) \geq \mathbf{0}, \quad \mathbf{x} \in \mathcal{X}.$$

By the definition of β, this system has no solution. Hence, by the above lemma, there exists a nonzero vector $(u_0, \mathbf{u})$ with $(u_0, \mathbf{u}) \geq \mathbf{0}$ such that

$$u_0(f(\mathbf{x}) - \beta) - \mathbf{u}^T\mathbf{c}(\mathbf{x}) \geq 0 \quad \textbf{for all } \mathbf{x} \in \mathcal{X}. \tag{10}$$

We now show that $u_0 > 0$. By contradiction, suppose that $u_0 = 0$. Since the Slater CQ holds, there exists $\bar{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{c}(\bar{\mathbf{x}}) > \mathbf{0}$. It follows from (10) that $\mathbf{u} = \mathbf{0}$. This is impossible since $(u_0, \mathbf{u}) \neq \mathbf{0}$. This implies that $u_0 > 0$.

Hence, dividing (10) by u_0 and denoting $\mathbf{u}/u_0$ by $\bar{\mathbf{u}}$, we get

$$f(\mathbf{x}) - \bar{\mathbf{u}}^T \mathbf{c}(\mathbf{x}) \geq \beta \quad \textbf{for all } \mathbf{x} \in \mathcal{X}. \tag{11}$$

This shows that $\theta(\bar{\mathbf{u}}) = \inf\{f(\mathbf{x}) - \bar{\mathbf{u}}^T \mathbf{c}(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} \geq \beta$. By Weak Duality Theorem and the definition of β, $\theta(\bar{\mathbf{u}}) = \beta$ and $\bar{\mathbf{u}}$ solves the dual problem.

Let $\bar{\mathbf{x}}$ be an optimal solution to the primal problem, i.e., $\bar{\mathbf{x}} \in \mathcal{F}_p$ and $f(\bar{\mathbf{x}}) = \beta$. Then, (11) implies that $\bar{\mathbf{u}}^T \mathbf{c}(\bar{\mathbf{x}}) = 0$ since $\bar{\mathbf{u}} \geq \mathbf{0}$ and $\mathbf{c}(\bar{\mathbf{x}}) \geq \mathbf{0}$. This completes the proof. $\square$

Example

Consider the following problem

$$\begin{array}{ll} \min & f(x) = \frac{1}{x} \\ s.t. & c(x) = x - 1 \geq 0, \quad x \in \mathcal{X} = (0, +\infty). \end{array}$$

Clearly, this problem has no optimal solution. However, the Slater CQ is satisfied for this problem and $f_{\inf} = 0$.

Its Lagrangian dual problem is $\max\{\theta(\lambda) |\, \lambda \geq 0\}$, where

$$\theta(\lambda) = \inf\{\frac{1}{x} - \lambda(x-1) |\, x \in \mathcal{X}\} = \begin{cases} -\infty, & \lambda > 0, \\ 0, & \lambda = 0, \\ 2\sqrt{-\lambda} + \lambda, & \lambda < 0. \end{cases}$$

Obviously, its optimal solution is $\bar{\lambda} = 0$ with optimal value is $\theta_{\max} = 0 = f_{\inf}$.

Remark: Wolfe Dual for Convex Program

For the convex program, when $\mathcal{X} = \mathcal{R}^n$, its Lagrangian dual can be rewritten as the following simple form: 因为此时 $f(x) - u^T c(x)$ 为凸函数(无限制)

故最小值等价于梯度为0

$$
\begin{aligned}
\max \quad & L(\mathbf{x}, \mathbf{u}) \\
s.t. \quad & \nabla_x L(\mathbf{x}, \mathbf{u}) = \mathbf{0}, \\
& \mathbf{u} \geq \mathbf{0},
\end{aligned}
$$

since $\bar{\mathbf{x}}$ is a minimizer of $L(\mathbf{x}, \mathbf{u})$ for any given $\mathbf{u}$ iff $\nabla_x L(\mathbf{x}, \mathbf{u})|_{\mathbf{x}=\bar{\mathbf{x}}} = \mathbf{0}$.

Example for Wolfe Dual for CP

Consider the following convex primal problem

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ s.t. & x_1 + x_2 - 4 \geq 0, \\ & x_1, x_2 \geq 0. \end{array}$$

Let

$$L(\mathbf{x}, \mathbf{u}) = x_1^2 + x_2^2 - \mu_1(x_1 + x_2 - 4) - \mu_2 x_1 - \mu_3 x_2.$$

It follows from

$$\nabla_x L(\mathbf{x}, \mathbf{u}) = \begin{pmatrix} 2x_1 - \mu_1 - \mu_2 \\ 2x_2 - \mu_1 - \mu_3 \end{pmatrix} = 0$$

that

$$x_1 = \frac{1}{2}(\mu_1 + \mu_2), \quad x_2 = \frac{1}{2}(\mu_1 + \mu_3).$$

Then, its Wlofe dual is

$$\begin{aligned} \max \quad & L(\mathbf{x}, \mathbf{u}) \\ s.t. \quad & \nabla_x L(\mathbf{x}, \mathbf{u}) = \mathbf{0}, \\ & \mathbf{u} \geq \mathbf{0}, \end{aligned}$$

i.e.,

$$\begin{aligned} \max \quad & -\tfrac{1}{4}(\mu_1 + \mu_2)^2 - \tfrac{1}{4}(\mu_1 + \mu_3)^2 + 4\mu_1 \\ s.t. \quad & \mathbf{u} \geq 0. \end{aligned}$$

Its optimal solution is $\mathbf{u} = (4; 0; 0)$ and the objective value 8.

Lagrangian Dual Problem of LP

Consider the following primal linear program:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ s.t. \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

The Lagrangian dual of this problem is to maximize $\theta(\mathbf{v})$, where

$$\theta(\mathbf{v}) = \inf\{\mathbf{c}^T\mathbf{x} - \mathbf{v}^T(A\mathbf{x} - \mathbf{b}) : \ \mathbf{x} \geq \mathbf{0}\}.$$

Clearly,

$$\theta(\mathbf{v}) = \begin{cases} \mathbf{b}^T\mathbf{v} & \text{if } A^T\mathbf{v} - \mathbf{c} \leq \mathbf{0} \\ -\infty & \text{otherwise.} \end{cases}$$

Hence, the Lagrangian dual problem can be written as follows:

$$\begin{array}{ll} \max & \mathbf{b}^T\mathbf{v} \\ s.t. & A^T\mathbf{v} \leq \mathbf{c}. \end{array}$$

This is precisely the LP dual problem. Thus, in the case of linear programs, the Lagrangian dual problem does not involve the primal variables. Furthermore, and itself is a linear program.

Saddlepoint Problems

Let $F : A \times B \to \mathcal{R}$ be a given function. If $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in A \times B$ and

$$F(\bar{\mathbf{x}}, \mathbf{y}) \leq F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \leq F(\mathbf{x}, \bar{\mathbf{y}}) \ \textbf{for all} \ (\mathbf{x}, \mathbf{y}) \in A \times B,$$

then $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is called a **saddlepoint** of F on $A \times B$.

Now consider the inequality-constrained problem (IP) and define its associated Lagrangian function

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T c(\mathbf{x})$$

on the set $A \times B = \mathcal{R}^n \times \mathcal{R}^m_+$.

Theorem 24 *If $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a saddle point of L, then $\bar{\mathbf{x}}$ solves (IP) and $\bar{\mathbf{y}}$ solves its Lagrangian dual.*

Proof. The vector $\bar{\mathbf{x}}$ is feasible for (IP). Indeed, if $c(\bar{\mathbf{x}})$ has a negative component, then the inequality $L(\bar{\mathbf{x}}, \mathbf{y}) \leq L(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ for all $\mathbf{y} \geq \mathbf{0}$ cannot hold. Moreover, $\bar{\mathbf{y}}^T c(\bar{\mathbf{x}}) = 0$ since

$$0 \leq \bar{\mathbf{y}}^T c(\bar{\mathbf{x}}) \leq \mathbf{y}^T c(\bar{\mathbf{x}}) \quad \textbf{for all } \mathbf{y} \geq \mathbf{0}.$$特别取y为0

Hence the vector $\bar{\mathbf{x}}$ is a global minimizer for (IP) since for any feasible solution $\mathbf{x}$ of (IP) we have

$$f(\bar{\mathbf{x}}) = L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \leq L(\mathbf{x}, \bar{\mathbf{y}}) = f(\mathbf{x}) - \bar{\mathbf{y}}^T c(\mathbf{x}) \leq f(\mathbf{x}).$$

$\square$

Remark

The condition of being a saddle point of the Lagrangian function L for the problem (IP) is obviously very strong, for it yields *sufficient* conditions for a vector to be a global minimizer using

- no differentiability assumption,
- no regularity assumption, and
- no convexity assumption.

Example: Nonexistence of a Saddlepoint

This example exhibits a nonlinear program having a global optimal solution but no saddlepoint for the Lagrangian function.

Consider the problem

$$\begin{aligned} \textbf{minimize} \quad & -x_1 \\ \textbf{subject to} \quad & x_2 - x_1^2 \geq 0 \\ & -x_2 \geq 0. \end{aligned}$$

Then $\bar{\mathbf{x}} = (0; 0)$ is the unique global minimizer. Indeed, since $0 \geq x_2 \geq x_1^2 \geq 0$, it is the only feasible solution. The associate Lagrangian function is

$$L(x_1, x_2, y_1, y_2) = -x_1 - y_1(x_2 - x_1^2) + y_2 x_2.$$

Note that $L(\bar{\mathbf{x}}, \mathbf{y}) = 0$ for all $\mathbf{y}$. If there exists a saddlepoint $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, then

$$0 \leq L(\mathbf{x}, \bar{\mathbf{y}}) = -x_1 - \bar{y}_1(x_2 - x_1^2) + \bar{y}_2 x_2 \quad \textbf{for all } \mathbf{x}.$$

However, for such vector $\mathbf{x}$ with $x_2 = 0$ and $x_1 > 0$ sufficiently small, we get $L(\mathbf{x}, \bar{\mathbf{y}}) < 0$, so we have a contradiction.

Theorem 25 *For (CP), the KKT point is equivalent to the saddle point.*

Proof. Let

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T c(\mathbf{x})$$

be the associated Lagrangian function with (CP), and let $(\mathbf{x}^*, \mathbf{y}^*)$ be a saddle point of L. It is to show that $(\mathbf{x}^*, \mathbf{y}^*)$ is a KKT-pair of (CP).

$$\mathbf{x}^* = arg \min_{\mathbf{x} \in R^n} L(\mathbf{x}, \mathbf{y}^*) \Rightarrow \nabla_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{y}^*) = 0,$$

i.e.,

$$\nabla f(\mathbf{x}^*) - \sum_{i \in \mathcal{I}} y_i^* \nabla c_i(\mathbf{x}^*) = 0.$$

Since

$$\mathbf{y}^* = arg \max_{y_i \geq 0,\ i \in \mathcal{I}} L(\mathbf{x}^*, \mathbf{y}) = arg \min_{\mathbf{e}_i^T \mathbf{y} \geq 0,\ i \in \mathcal{I}} -L(\mathbf{x}^*, \mathbf{y}),$$

which is a linear constrained optimization problem, by KKT theorem, there exist $\lambda_i \geq 0, i \in \mathcal{I}$ such that

此时 $g(y) = y^T c(x^*) + f(x^*)$

$e_i^T y \geq 0$

且线性约束 故KKT条件自然成立

$$c_i(\mathbf{x}^*) = \lambda_i, \quad \lambda_i y_i^* = 0, \quad i \in \mathcal{I}.$$

That is,

$$y_i^* c_i(\mathbf{x}^*) = 0 \quad \forall i \in \mathcal{I}.$$

Hence, $(\mathbf{x}^*, \mathbf{y}^*)$ is a KKT-pair of (CP).

Conversely, let $(\mathbf{x}^*, \mathbf{y}^*)$ be a KKT-pair of (CP). Since $L(\mathbf{x}, \mathbf{y}^*)$ is convex,

$$L(\mathbf{x}, \mathbf{y}^*) \geq L(\mathbf{x}^*, \mathbf{y}^*).$$

利用梯度不等式再结合kkt条件得梯度为0

For any $\mathbf{y}$ with $y_i \geq 0, i \in \mathcal{I}$,

$$L(\mathbf{x}^*, \mathbf{y}) - L(\mathbf{x}^*, \mathbf{y}^*) = -\sum_{i \in \mathcal{I}} y_i c_i(\mathbf{x}^*) + \sum_{i \in \mathcal{I}} y_i^* c_i(\mathbf{x}^*) = -\sum_{i \in \mathcal{I}} y_i c_i(\mathbf{x}^*) \leq 0,$$

i.e.,

$$L(\mathbf{x}^*, \mathbf{y}) \leq L(\mathbf{x}^*, \mathbf{y}^*).$$

This shows that $(\mathbf{x}^*, \mathbf{y}^*)$ is a saddle point. $\square$

Applications of Nonlinear Optimization and KKT Conditions

Recall Fisher's Exchange Market

Buyers have money (w_i) to buy goods and maximize their individual utility functions; Producers sell their goods for money. The equilibrium prices is an assignment of prices to goods so as when every buyer buys an maximal bundle of goods then the market clears, meaning that all the money is spent and all goods are sold.

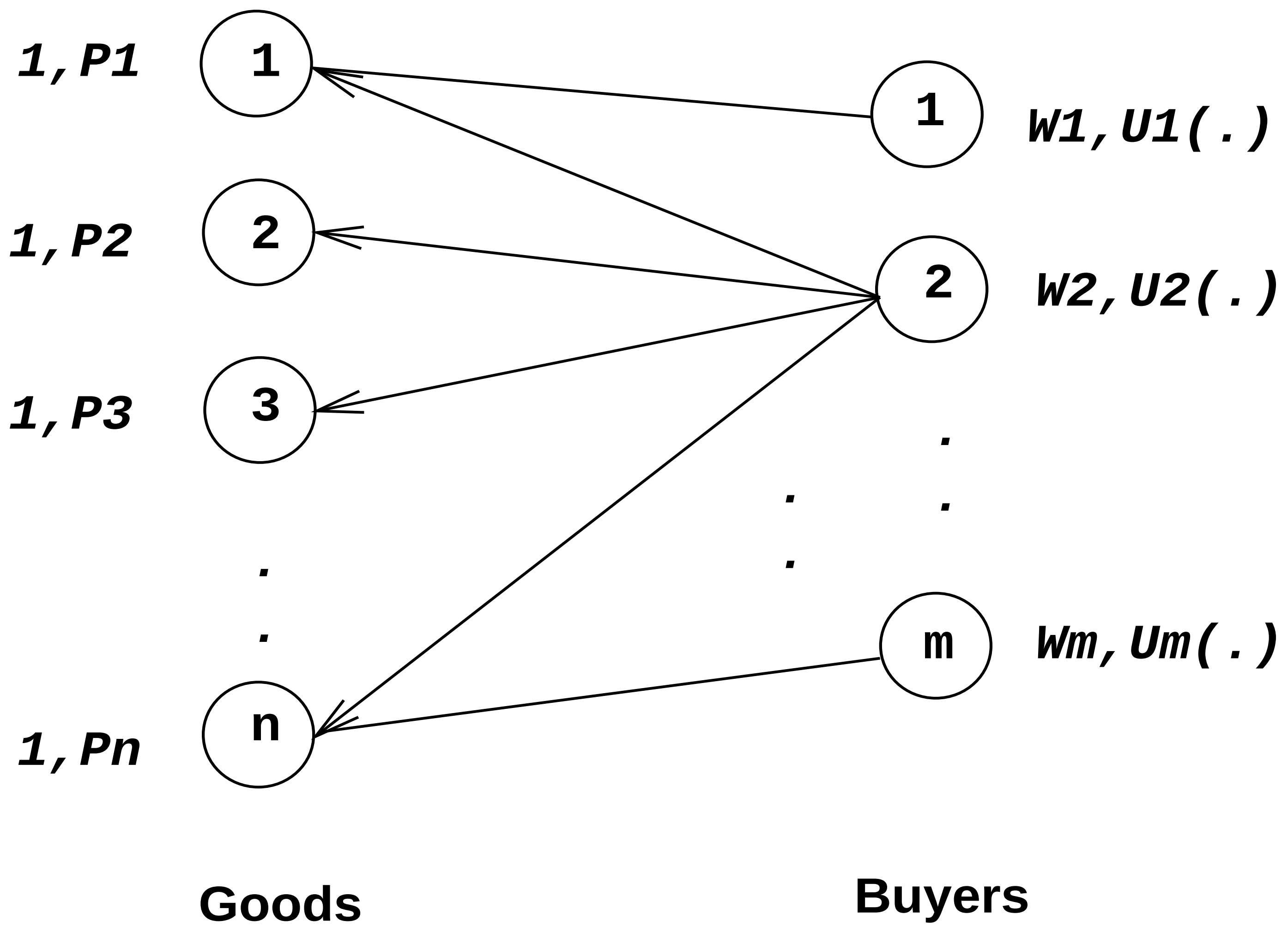


Figure 3: Fisher's Exchange Market Model

Each buyer's strategy and the equilibrium price

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

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

Without loss of generality, assume that the amount of each good is 1. The equilibrium price vector is the one that for all $j \in G$

$$\sum_{i\in B} x(\mathbf{p})_{ij} = 1$$

where $\mathbf{x}(\mathbf{p})_i$ is an optimal solution to the player i's problem.

Aggregate social optimization

$$
\begin{array}{ll}
\textbf{maximize} & \sum_{i\in B} w_i \log(\mathbf{u}_i^T \mathbf{x}_i) \\
\textbf{subject to} & \sum_{i\in B} x_{ij} = 1, \quad \forall j \in G \\
& x_{ij} \geq 0, \quad \forall i, j.
\end{array}
$$

Theorem 26 *(Eisenberg and Gale 1959) Optimal dual vector of equality constraints is an equilibrium price vector.*

Optimality Conditions of the aggregated problem

$$\begin{aligned} w_i \frac{u_{ij}}{\mathbf{u}_i^T \mathbf{x}_i} &\leq p_j, \quad \forall i, j \\ \sum_{i \in B} x_{ij} &= 1, \quad \forall j \\ w_i \frac{u_{ij} x_{ij}}{\mathbf{u}_i^T \mathbf{x}_i} &= p_j x_{ij}, \quad \forall i, j \end{aligned}$$

Individual optimality

For any i, summing the third condition over j

$$\sum_{j \in G} p_j x_{ij} = w_i;$$

also,

$$u_{ij} \le \frac{\mathbf{u}_i^T \mathbf{x}_i}{w_i} p_j, \quad \forall j.$$

These imply that $\mathbf{x}_i$ is individually optimal for buyer i.

The optimality condition for the player i's problem is

$$\begin{aligned}
\mathbf{p}^T \mathbf{x}_i \le w_i,\ \mathbf{x}_i \ge \mathbf{0}, &\quad \forall i,\\
\mathbf{p} y_i \ge \mathbf{u}_i,\ y_i \ge 0, &\quad \forall i,\\
(\mathbf{p}^T \mathbf{x}_i - w_i) y_i = 0, &\quad \forall i,\\
\textstyle\sum_i x_{ij} = 1, &\quad \forall j.
\end{aligned}$$

Example of Fisher's Model

Buyer $1, 2$'s optimization problems for given prices p_x, p_y.

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

The aggregate social optimization is

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

$$p_x = 26/3, \;\; p_y = 13/3, \;\; x_1 = 1/13, \;\; y_1 = 1, \;\; x_2 = 12/13, \;\; y_2 = 0.$$