Optimization - II

Prof. Dan A. Simovici

UMB

The greatest lower bound or infimum of a function

Definition

The greatest lower bound of f on S is the largest number m (or $-\infty$) such that $f(\boldsymbol{x}) \geqslant m$ holds for all $x \in S$.

The greatest lower bound of f on S is denoted as

$$\inf_{\boldsymbol{x} \in S} f(\boldsymbol{x}).$$

A point $\boldsymbol{x}_0$ minimizes f if and only if $\boldsymbol{x}_0 = \inf_{\boldsymbol{x} \in S} f(\boldsymbol{x})$.

Definition

The least upper bound of f on S is the smallest number M (or ∞) such that $f(\boldsymbol{x}) \leqslant M$ holds for all $x \in S$.

The least upper bound of f on S is denoted as

$$\sup_{\boldsymbol{x} \in S} f(\boldsymbol{x}).$$

A point $\boldsymbol{x}_0$ maximizes f if and only if $\boldsymbol{x}_0 = \sup_{\boldsymbol{x} \in S} f(\boldsymbol{x})$.

Example

Let $S = \mathbb{R}$.

- for $f(x) = e^x$ we have $\inf_{x \in S} e^x = 0$ and $\sup_{x \in S} e^x = \infty$;
- for $f(x) = e^x + e^{-x}$ we have $\inf_{x \in S} f(x) = 2$;
- for $f(x) = e^{-x^2}$ we have $\sup_{x \in S} f(x) = 1$.

Example

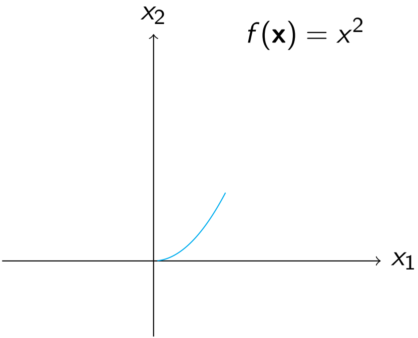
If S is the disk defined by $x^2 + y^2 \leqslant 1$ and $f(x, y) = x^2 + y^2$ then $\inf_{(x,y)\in S} f(x, y) = 0$ and $\sup_{(x,y)\in S} f(x, y) = 1$.

Example

If $S = (0,1)$ and $f : S \longrightarrow \mathbb{R}$ is $f(x) = x^2$ we have

$$\inf_{x \in S} f(x) = 0 \text{ and } \sup_{x \in S} f(x) = 1.$$

However, there is no x_0 in S such that $f(x_0) = \inf_{x \in S} f(x) = 0$ or $f(x_0) = \sup_{x \in S} f(x) = 1$.



Recall: the norm of $\boldsymbol{x} \in \mathbb{R}^n$ is

$$\| \boldsymbol{x} \| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Definition

A sequence $(\boldsymbol{x}_m)$ of $\mathbb{R}^n$ is bounded if there exists a constant K such that $\| \boldsymbol{x}_m \| \leqslant K$.

Definition

A sequence $(\boldsymbol{x}_m)$ converges to $\boldsymbol{x}_0$, written as $\lim_{m\to\infty} \boldsymbol{x}_m = \boldsymbol{x}_0$ or as $\boldsymbol{x}_m \to \boldsymbol{x}_0$ if $\lim_{m\to\infty} \| \boldsymbol{x}_m - \boldsymbol{x}_0 \| = 0$.

A set $S \subseteq \mathbb{R}^n$ is

- bounded if there exists a constant K such that $\| \boldsymbol{x} \| \leqslant K$ for all $\boldsymbol{x} \in S$;
- closed if for every convergent sequence $(\boldsymbol{x}_m)$ in S we have $\lim_{m\to\infty} \boldsymbol{x}_m \in S$.

Theorem

(Bolzano-Weierstrass Theorem) *A bounded sequence of points in $\mathbb{R}^n$ contains a convergent subsequence.*

Definition

A set S in $\mathbb{R}^n$ is compact if it is both bounded and closed.

Theorem

Let f be a continuous function on a closed set S. Suppose that there exists a number b such that $S_b = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \leqslant b\}$ is bounded and non-empty. Then, f attains its minimum at a point $\boldsymbol{x}_0$ in S, that is

$$f(\boldsymbol{x}_0) = \inf_{\boldsymbol{x} \in S} f(\boldsymbol{x}).$$

Proof

Let $m = \inf_{\mathbf{x} \in S} f(\mathbf{x})$. Then $f(\mathbf{x}) \geqslant m$ for all $\mathbf{x} \in S$. We show that there exists $\mathbf{x}_0 \in S$ such that $f(\mathbf{x}_0) = m$.

If $m = b$, then $m \leqslant f(\mathbf{x}) = b$, hence $f(\mathbf{x}) = m$ on S_b. Since $S_b \neq \emptyset$, there exists $\mathbf{x}_0 \in S_b$. This point minimizes f on S.

Suppose now that $b > m$ and let $(\mathbf{x}_q)$ be a sequence such that $\lim_{q \to \infty} f(\mathbf{x}_q) = m$. Since $b > m$ the inequality $f(\mathbf{x}_q) > b$ holds for at most a finite number of qs. Deleting these points we are left with a sequence such that $f(\mathbf{x}_q) \to m$. Since S_b is bounded, the sequence $(\mathbf{x}_q)$ is bounded and contains a convergent subsequence $(\mathbf{y}_q)$. Since S is closed the limit $\mathbf{x}_0 = \lim_{q \to \infty} \mathbf{y}_q$ belongs to S. Since f is continuous,

$$f(\mathbf{x}_0) = \lim_{q \to \infty} f(\mathbf{y}_q) = \lim_{q \to \infty} f(\mathbf{x}_q) = m.$$

Thus, $\mathbf{x}_0$ minimizes f.

Let $f : S \longrightarrow \mathbb{R}$ be a function and let S is a subset of $\mathbb{R}^n$ defined as

$$S = \{x \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) = 0 \text{ for } 1 \leqslant i \leqslant m\}.$$

S is the set of feasible points. In this context we shall refer to the following optimization problem

minimize $f(\boldsymbol{x})$

subjected to $g_i(\boldsymbol{x}) \leqslant 0$ *for* $1 \leqslant j \leqslant m.$

as the primal problem.

Definition

The associated Lagrangean of the primal problem is the function

$$L(\boldsymbol{x}, \boldsymbol{a}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} a_i g_i(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{a}'\boldsymbol{g}(\boldsymbol{x}),$$

where $a_1, \ldots, a_m$ are the Lagrange variables that range over the set $\mathbb{R}_{\geqslant 0}$.

The Dual Problem

The dual function associated to primal problem is

$$F(\boldsymbol{a}) = \inf_{\boldsymbol{x} \in S} L(\boldsymbol{x}, \boldsymbol{a}) = \inf_{\boldsymbol{x} \in S} f(\boldsymbol{x}) + \sum_{i=1}^{m} a_i g_i(\boldsymbol{x}).$$

Theorem

The function $F(\boldsymbol{a})$ is concave.

Proof

Note that if h_1, h_2 are defined on S we have:

$$\inf_{\boldsymbol{x} \in S} (h_1 + h_2)(\boldsymbol{x}) \geqslant \inf_{\boldsymbol{x} \in S} h_1(\boldsymbol{x}) + \inf_{\boldsymbol{x} \in S} h_2(\boldsymbol{x}).$$

For $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^m$ we have

$$\begin{aligned} F(t\boldsymbol{a}_1 + (1-t)\boldsymbol{a}_2) &= \inf_{\boldsymbol{x} \in S} L(\boldsymbol{x}, \boldsymbol{a}) \\ &= \inf_{\boldsymbol{x} \in S} f(\boldsymbol{x}) + (t\boldsymbol{a}_1' + (1-t)\boldsymbol{a}_2')\boldsymbol{g}(\boldsymbol{x}) \\ & \quad \inf_{\boldsymbol{x} \in S} t(f(\boldsymbol{x}) + \boldsymbol{a}_1'\boldsymbol{g}(\boldsymbol{x})) + (1-t)(f(\boldsymbol{x}) + \boldsymbol{a}_2'\boldsymbol{g}(\boldsymbol{x})) \\ &\geqslant t \inf_{\boldsymbol{x} \in S} (f(\boldsymbol{x}) + \boldsymbol{a}_1'\boldsymbol{g}(\boldsymbol{x})) + (1-t) \inf_{\boldsymbol{x} \in S} (f(\boldsymbol{x}) + \boldsymbol{a}_2'\boldsymbol{g}(\boldsymbol{x})), \end{aligned}$$

which shows that F is always concave.

The Dual Problem

The dual problem associated to the optimization problem is
maximize $F(\boldsymbol{a})$
subjected to $a_j \geqslant 0$ *for* $1 \leqslant j \leqslant m$.
The dual problem is always a concave optimization problem.

The Weak Duality Theorem

Theorem

Let $\mathbf{x}_*$ be a solution of the primal problem, $p_* = f(\mathbf{x}_*)$ and let $q_* = F(\mathbf{a}_*)$ be the optimal value of the dual problem. We have $q_* \leqslant p_*$.

Since $\mathbf{g}(\mathbf{x}_*) \leqslant \mathbf{0}_m$ it follows that

$$L(\mathbf{x}_*, \mathbf{a}) = f(\mathbf{x}_*) + \mathbf{a}'\mathbf{g}(\mathbf{x}_*) \leqslant p_*.$$

Therefore, $F(\mathbf{a}) = \inf_{\mathbf{x} \in C} L(\mathbf{x}, \mathbf{a}) \leqslant p_*$ for all $\mathbf{a}$.
Since F_* is the optimal value of F, the last inequality implies $q_* \leqslant p_*$.
The difference $p^* - q^*$ is the duality gap.
The situation when $q^* = p^*$ is designated as the strong duality.

Strong duality holds when constraints of problems satisfy constraint qualifications.

Definition

Assume that the interior of the set S is nonempty. The function $\boldsymbol{g} = \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}$ satisfies the strong qualifications if there exists a point $\bar{\boldsymbol{x}}$ in the interior of S such that $\boldsymbol{g}(\bar{\boldsymbol{x}}) < 0$.

Definition

A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is affine if there exists $\boldsymbol{w} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, $f(\boldsymbol{x}) = \boldsymbol{w}'\boldsymbol{x} + b$ for $\boldsymbol{x} \in \mathbb{R}^n$.

Definition

The weak constraint qualification are defined as

for some interior point $\boldsymbol{x} \in S$,
$g_j(\boldsymbol{x}) < 0$
or $(g_i(\boldsymbol{x}) = 0$ and g_i is affine) for $1 \leqslant i \leqslant m$.

Example

minimize $f(\boldsymbol{x}) = \boldsymbol{k}'\boldsymbol{x}$
subject to $A\boldsymbol{x} - \boldsymbol{b} \leqslant \boldsymbol{0}_m$,
where $A \in \mathbb{R}^{m \times n}$ *and* $\boldsymbol{b} \in \mathbb{R}^m$.

The Lagrangean L is

$$L(\boldsymbol{x}, \boldsymbol{a}) = \boldsymbol{k}'\boldsymbol{x} + \boldsymbol{a}'(A\boldsymbol{x} - \boldsymbol{b}) = -\boldsymbol{a}'\boldsymbol{b} + (\boldsymbol{k}' + \boldsymbol{a}'A)\boldsymbol{x},$$

which yields the dual function:

$$g(\boldsymbol{a}) = \begin{cases} -\boldsymbol{a}'\boldsymbol{b} & \text{if } \boldsymbol{k}' + \boldsymbol{a}'A = \boldsymbol{0}_m, \\ -\infty & \text{otherwise.} \end{cases}$$

and the dual problem is
maximize $-\boldsymbol{b}'\boldsymbol{a}$ *subject to* $\boldsymbol{k}' + \boldsymbol{a}'A = \boldsymbol{0}_m$
and $\boldsymbol{a} \geqslant \boldsymbol{0}$.

Example

Let $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_m \in \mathbb{R}^n$. We seek to determine a closed sphere $B[\boldsymbol{x}, r]$ of minimal radius that includes all points $\boldsymbol{k}_i$ for $1 \leqslant i \leqslant m$. This is the *minimum bounding sphere* problem, which amounts to solving the following primal optimization problem:

minimize r, where $r \geqslant 0$,

subject to $\parallel \boldsymbol{x} - \boldsymbol{k}_i \parallel \leqslant r$ for $1 \leqslant i \leqslant m$.

An equivalent formulation requires minimizing r^2 and stating the restrictions as $\parallel \boldsymbol{x} - \boldsymbol{k}_i \parallel^2 - r^2 \leqslant 0$ for $1 \leqslant i \leqslant m$. The Lagrangean of this problem is

$$\begin{aligned} L(r, \boldsymbol{x}, \boldsymbol{a}) &= r^2 + \sum_{i=1}^{m} a_i (\parallel \boldsymbol{x} - \boldsymbol{k}_i \parallel^2 - r^2) \\ &= r^2 \left(1 - \sum_{i=1}^{m} a_i\right) + \sum_{i=1}^{m} a_i \parallel \boldsymbol{x} - \boldsymbol{k}_i^2 \parallel . \end{aligned}$$

Example (cont'd)

The dual function is

$$\begin{aligned} g(\boldsymbol{a}) &= \inf_{r \in \mathbb{R}_{\geqslant 0}, \boldsymbol{x} \in \mathbb{R}^n} L(r, \boldsymbol{x}, \boldsymbol{a}) \\ &= \inf_{r \in \mathbb{R}_{\geqslant 0}, \boldsymbol{x} \in \mathbb{R}^n} r^2 \left(1 - \sum_{i=1}^{m} a_i\right) + \sum_{i=1}^{m} a_i \parallel \boldsymbol{x} - \boldsymbol{k}_i|^2 \parallel . \end{aligned}$$

Example (cont'd)

This leads to the following conditions:

$$\begin{aligned} \frac{\partial L(r, \boldsymbol{x}, \boldsymbol{a})}{\partial r} &= 2r\left(1 - \sum_{i=1}^{m} a_i\right) = 0 \\ \frac{\partial L(r, \boldsymbol{x}, \boldsymbol{a})}{\partial x_p} &= 2\sum_{i=1}^{m} a_i(\boldsymbol{x} - \boldsymbol{k}_i)_p = 0 \text{ for } 1 \leqslant p \leqslant n. \end{aligned}$$

The first equality yields $\sum_{i=1}^{m} a_i = 1$. Therefore, from the second equality we obtain $\boldsymbol{x} = \sum_{i=1}^{m} a_i \boldsymbol{k}_i$. This shows that for $\boldsymbol{x}$ is a linear combination of $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_m$. The dual function is

$$g(\boldsymbol{a}) = \sum_{i=1}^{m} a_i \left(\sum_{h=1}^{m} a_h \boldsymbol{k}_h - \boldsymbol{k}_i\right) = 0$$

because $\sum_{i=1}^{m} a_i = 1$.

Definition

A pair $(\boldsymbol{x}^*, \boldsymbol{a}^*)$ is a saddle point of the Lagrangian of the optimization problem if

$$L(\boldsymbol{x}^*, \boldsymbol{a}) \leqslant L(\boldsymbol{x}^*, \boldsymbol{a}^*) \leqslant L(\boldsymbol{x}, \boldsymbol{a}^*)$$

for $\boldsymbol{x} \in S$ and $\boldsymbol{a} \geqslant \mathbf{0}_m$.

Theorem

If $(\boldsymbol{x}^*, \boldsymbol{a}^*)$ is a saddle point of the Lagrangean, then $(\boldsymbol{x}^*, \boldsymbol{a}^*)$ is a solution of the primal problem.

Proof

Let $(\boldsymbol{x}^*, \boldsymbol{a}^*)$ be a saddle point of the Lagrangean. Since $L(\boldsymbol{x}^*, \boldsymbol{a}) \leqslant L(\boldsymbol{x}^*, \boldsymbol{a}^*)$ we have

$$f(\boldsymbol{x}^*) + \boldsymbol{a}'\boldsymbol{g}(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x}^*) + (\boldsymbol{a}^*)'\boldsymbol{g}(\boldsymbol{x}^*),$$

so $\boldsymbol{a}'\boldsymbol{g}(\boldsymbol{x}^*) \leqslant (\boldsymbol{a}^*)'\boldsymbol{g}(\boldsymbol{x}^*)$.
By taking $\boldsymbol{a} \to \infty$ we get $g(\boldsymbol{x}^*) \leqslant \mathbf{0}_m$; by taking $\boldsymbol{a} \to \mathbf{0}_m$ we get $(\boldsymbol{a}^*)'\boldsymbol{g}(\boldsymbol{x}^*) = 0$.
Thus, the inequality $L(\boldsymbol{x}^*, \boldsymbol{a}^*) \leqslant L(\boldsymbol{x}, \boldsymbol{a}^*)$ amounts to

$$f(\boldsymbol{x}^*) + (\boldsymbol{a}^*)'\boldsymbol{g}(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x}) + (\boldsymbol{a}^*)'\boldsymbol{g}(\boldsymbol{x}),$$

hence

$$f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x}) + (\boldsymbol{a}^*)'\boldsymbol{g}(\boldsymbol{x}),$$

If $\boldsymbol{x}$ satisfies the constraints $(\boldsymbol{g}(\boldsymbol{x}) \leqslant \mathbf{0}_m)$ it follows that $f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x})$.

Saddle Points for Differentiable Convex Functions

Theorem

If f and g_i are convex functions, the Slater's qualifications hold, and if $\mathbf{x}$ is a solution of the constraind optimization problem, then there exists $\mathbf{a}$ such that $(\mathbf{x}, \mathbf{a})$ is a saddle point of the Lagrangean.

If differentiablity is added, Slater's qualification may be replaced by weak Slater qualifications:

Theorem

If f and g_i are convex differentiable functions, the weak Slater's qualifications hold, and if $\mathbf{x}$ is a solution of the constraind optimization problem, then there exists $\mathbf{a}$ such that $(\mathbf{x}, \mathbf{a})$ is a saddle point of the Lagrangean.

The Karush-Kuhn-Tucker Conditions

Theorem

Let f, g_i be convex and differentiable and suppose that constraints are qualified. Then $\bar{\mathbf{x}}$ is a solution of the constrained problem if and only if there exist $\bar{\mathbf{a}} \geqslant \mathbf{0}_m$ such that

$$
\begin{aligned}
(\nabla L)_{\mathbf{x}}(\bar{\mathbf{x}}, \bar{\mathbf{a}}) &= (\nabla f)_{\mathbf{x}}(\bar{\mathbf{x}}) + \bar{\mathbf{a}}'(\nabla g)_{\mathbf{x}}(\bar{\mathbf{x}}) = 0, \\
(\nabla L)_{\mathbf{a}}(\bar{\mathbf{x}}, \bar{\mathbf{a}}) &= g(\bar{\mathbf{x}}) \leqslant \mathbf{0}_m,
\end{aligned}
$$

$$
\bar{\mathbf{a}}' g(\bar{\mathbf{x}}) = \sum_{i=1}^{m} (\bar{\mathbf{a}})'_i g_i(\bar{\mathbf{x}}) = 0.
$$

Proof

Suppose that $\bar{\boldsymbol{x}}$ is a solution. Since the constraints are qualified, there exists $\bar{\boldsymbol{a}}$ such that $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})$ is a saddle point of the Lagrangean by the previous theorem. The KKT conditions are implied by arguments in the proof of the theorem on slide 22.

Conversely, suppose that the KKT conditions are satisfied and let $\boldsymbol{x}$ such that $g(\boldsymbol{x}) \leqslant \mathbf{0}_m$. We have

$$
\begin{aligned}
f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}) &\geqslant (\nabla f)'_{\boldsymbol{x}}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \\
&\quad \text{(by the convexity of } f\text{)} \\
&\geqslant -\bar{\boldsymbol{a}}'(\nabla f)_{\boldsymbol{x}}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \\
&\quad \text{(by the first condition)} \\
&\geqslant -\bar{\boldsymbol{a}}'(\boldsymbol{x} - \bar{\boldsymbol{x}}) \\
&\quad \text{(by the convexity of } \boldsymbol{g}\text{)} \\
&\geqslant -\bar{\boldsymbol{a}}'\boldsymbol{g}(\boldsymbol{x}) \geqslant 0 \\
&\quad \text{(by the last two conditions)},
\end{aligned}
$$

which shows that $\bar{\boldsymbol{x}}$ is a minimum.