CONVEX FUNCTIONS

Prof. Dan A. Simovici

UMB

Convex Functions

Definition

Let $S \subseteq \mathbb{R}^n$ be a convex set. A function $f : S \longrightarrow \mathbb{R}$ is *convex* if $f(t\mathbf{x} + (1-t)\mathbf{y}) \leqslant tf(\mathbf{x}) + (1-t)f(\mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in S$ and $t \in (0, 1)$.
f is *strictly convex* if $f(t\mathbf{x} + (1-t)\mathbf{y}) < tf(\mathbf{x}) + (1-t)f(\mathbf{y})$ for each $\mathbf{x}, \mathbf{y} \in S$ and $t \in (0, 1)$.
f is *concave* (*strictly concave*) if $-f$ is convex (strictly convex).

Lemma

Lemma

Let S be a convex subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a convex function. Then, the level set

$$S_a = \{x \in S \mid f(x) \leqslant a\}$$

is a convex set.

Continuity and Convexity

Theorem

Let $S \subseteq \mathbb{R}^n$ be a convex set and let $f : S \longrightarrow \mathbb{R}$ be convex. Then, f is continuous on the interior of S.

Proof

Let $\overline{\mathbf{x}} \in \mathbf{I}(S)$. We show that for any $\epsilon > 0$ there exists $\delta > 0$ such that $\| \mathbf{x} - \overline{\mathbf{x}} \| < \delta$ implies $\| f(\mathbf{x}) - f(\overline{\mathbf{x}}) \| < \epsilon$.
Since $\overline{\mathbf{x}} \in \mathbf{I}(S)$, there exists a sphere $C(\overline{\mathbf{x}}, \delta') \subseteq S$. Define

$$\begin{aligned} t &= \max_{1 \leqslant i \leqslant n} \max\{f(\overline{\mathbf{x}} + \delta' \mathbf{e}_i) - f(\overline{\mathbf{x}}), f(\overline{\mathbf{x}} - \delta' \mathbf{e}_i) - f(\overline{\mathbf{x}})\} \\ \delta &= \min\left\{\frac{\delta'}{n}, \frac{\epsilon\delta'}{nt}\right\} \end{aligned}$$

Let $\mathbf{x}$ be such that $\mathbf{x} \in B(\overline{\mathbf{x}}, \delta)$. Define $\mathbf{z}_i \in \mathbb{R}^n$ as

$$\mathbf{z}_i = \begin{cases} \delta' \mathbf{e}_i & \text{if } x_i - \overline{x}_i \geqslant 0, \\ -\delta' \mathbf{e}_i & \text{otherwise.} \end{cases}$$

We have $\mathbf{x} - \overline{\mathbf{x}} = \sum_{i=1}^n \alpha_i \mathbf{z}_i$, where $\alpha_i \geqslant 0$ for $1 \leqslant i \leqslant n$, and

$$\| \mathbf{x} - \overline{\mathbf{x}} \| = \delta' \sqrt{\sum_{i=1}^n \alpha_i^2}.$$

This implies that $\alpha_i \leqslant \frac{1}{n}$ for $1 \leqslant i \leqslant n$.

Proof (cont'd)

Since f is convex and $0 \leqslant n\alpha_i \leqslant 1$ we obtain:

$$\begin{aligned} f(\mathbf{x}) &= f\left(\overline{\mathbf{x}} + \sum_{i=1}^{n} \alpha_i \mathbf{z}_i\right) = f\left(\frac{1}{n}\sum_{i=1}^{n}(\overline{\mathbf{x}} + n\alpha_i \mathbf{z}_i)\right) \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n} f(\overline{\mathbf{x}} + n\alpha_i \mathbf{z}_i) = \frac{1}{n}\sum_{i=1}^{n} f((1 - n\alpha_i)\overline{\mathbf{x}} + n\alpha_i(\overline{\mathbf{x}} + \mathbf{z}_i)) \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n} ((1 - n\alpha_i) f(\overline{\mathbf{x}}) + n\alpha_i f(\overline{\mathbf{x}} + \mathbf{z}_i)), \end{aligned}$$

which implies

$$f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leqslant \sum_{i=1}^{n} \alpha_i (f(\overline{\mathbf{x}} + \mathbf{z}_i) - f(\overline{\mathbf{x}})).$$

Proof (cont'd)

The definition of t,

$$t = \max_{1 \leqslant i \leqslant n} \max\{f(\overline{\mathbf{x}} + \delta' \mathbf{e}_i) - f(\overline{\mathbf{x}}), f(\overline{\mathbf{x}} - \delta' \mathbf{e}_i) - f(\overline{\mathbf{x}})\}$$

implies $f(\overline{\mathbf{x}} + \mathbf{z}_i) - f(\overline{\mathbf{x}}) \leq t$ for $1 \leqslant i \leqslant n$; since $\alpha_i \geqslant 0$, we have

$$f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leqslant t \sum_{i=1}^{n} \alpha_i.$$

Since $\alpha_i \leqslant \frac{\epsilon}{nt}$ it follows that $f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leqslant \epsilon$.

Proof (cont'd)

For $\mathbf{y} = 2\overline{\mathbf{x}} - \mathbf{x}$ we have $\| \mathbf{y} - \overline{\mathbf{x}} \| \leqslant \delta$. Therefore, $f(\mathbf{y}) - f(\overline{\mathbf{x}}) \leqslant \epsilon$.
Since $\overline{\mathbf{x}} = \frac{1}{2}(\mathbf{y} + \mathbf{x})$, we have $f(\overline{\mathbf{x}}) \leqslant \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(\mathbf{y})$. so $f(\overline{\mathbf{x}}) - f(\mathbf{x}) \leqslant \epsilon$, which concludes the proof.
Note that convex function may be discontinuous only at the borders of convex sets.

Differentiable Functions

Definition

A function $f : S \to \mathbb{R}$ (where $S \subseteq \mathbb{R}^n$) is *differentiable at* $\bar{\mathbf{x}} \in S$ if there exists a vector $L(\bar{\mathbf{x}})$ and a function $\alpha_{\bar{\mathbf{x}}} : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + (L(\bar{\mathbf{x}}))'(\mathbf{x} - \bar{\mathbf{x}}) + \parallel \mathbf{x} - \bar{\mathbf{x}} \parallel \alpha_{\bar{\mathbf{x}}}(\mathbf{x} - \bar{\mathbf{x}})$$

for each $\mathbf{x} \in S$ such that

$$\lim_{\mathbf{x} \to \bar{\mathbf{x}}} \alpha_{\bar{\mathbf{x}}}(\mathbf{x} - \bar{\mathbf{x}}) = 0.$$

f is *differentiable on an open set* T, $T \subseteq S$, if it is differentiable at each $\mathbf{x} \in T$.

For $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ the gradient of f is the function $\nabla f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

The gradient operator is the symbolic "vector"

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}.$$

Note that we can compute formally the product of the symbolic vector ∇ with itself and this yields the matrix

$$\nabla\nabla' = \nabla^2 = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} & \frac{\partial^2}{\partial x_1 x_2} & \cdots & \frac{\partial^2}{\partial x_1 x_n} \\ \frac{\partial^2}{\partial x_1 x_2} & \frac{\partial^2}{\partial x_2^2} & \cdots & \frac{\partial^2}{\partial x_1 x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2}{\partial x_n x_1} & \frac{\partial^2}{\partial x_n x_2} & \cdots & \frac{\partial^2}{\partial x_n^2} \end{pmatrix}$$

The *Hessian matrix* of f is

$$\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Theorem

Let $f : S \longrightarrow \mathbb{R}$ be a differentiable function at $\overline{\mathbf{x}} \in S$. Then, we have:

$$L(\overline{\mathbf{x}}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} (\overline{\mathbf{x}}),$$

that is, $L(\overline{\mathbf{x}}) = \nabla f(\overline{\mathbf{x}})$.

Proof

Let f be differentiable in $\overline{\mathbf{x}}$, that is,

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + (L(\overline{\mathbf{x}}))'(\mathbf{x} - \overline{\mathbf{x}}) + \parallel \mathbf{x} - \overline{\mathbf{x}} \parallel \alpha_{\overline{\mathbf{x}}}(\mathbf{x} - \overline{\mathbf{x}})$$

for each $\mathbf{x} \in S$ such that

$$\lim_{\mathbf{x} \to \overline{\mathbf{x}}} \alpha_{\overline{\mathbf{x}}}(\mathbf{x} - \overline{\mathbf{x}}) = 0.$$

Take $\mathbf{x} = \overline{\mathbf{x}} + t\mathbf{e}_i$. Then,

$$f(\overline{\mathbf{x}} + t\mathbf{e}_i) = f(\overline{\mathbf{x}}) + t(L(\overline{\mathbf{x}}))'\mathbf{e}_i + |t|\alpha_{\overline{\mathbf{x}}}(t\mathbf{e}_i),$$

where $\lim_{t \to 0} \alpha_{\overline{\mathbf{x}}}(t\mathbf{e}_i) = 0$. Therefore,

$$\lim_{t \to 0} \frac{f(\overline{\mathbf{x}} + t\mathbf{e}_i) - f(\overline{\mathbf{x}})}{t} = (L(\overline{\mathbf{x}}))'\mathbf{e}_i,$$

which shows that the i^{th} component of $L(\overline{\mathbf{x}})$ is $\frac{\partial f}{\partial x_i}(\overline{\mathbf{x}})$. Consequently, $L(\overline{\mathbf{x}}) = (\nabla f)(\overline{\mathbf{x}})$.

Directional Derivative

Definition

Let $f : S \longrightarrow \mathbb{R}$ be a function, $\overline{\mathbf{x}} \in S$, and let $\mathbf{d}$ be a vector such that $\mathbf{d} \neq \mathbf{0}$ and $\overline{\mathbf{x}} + \lambda\mathbf{d} \in S$ for $\lambda > 0$ and λ sufficiently small.
The function has the directional derivative at $\overline{\mathbf{x}}$ given by

$$\lim_{\lambda \to 0, \lambda > 0} \frac{f(\overline{\mathbf{x}} + \lambda\mathbf{d}) - f(\overline{\mathbf{x}})}{\lambda},$$

if this limit exists. The vector value of this limit is denoted by $\frac{\partial f}{\partial \mathbf{d}}(\overline{\mathbf{x}})$.

Taylor Series

By analogy with the standard Taylor series

$$f(x_0 + h) = f(x_0) + \frac{h}{1!}f'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots + \frac{h^p}{p!}f^{(p)}(x_0 + \alpha h),$$

where $\alpha \in [0, 1]$, we have the Taylor series for a function of n variables

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \frac{1}{1!}\mathbf{h}'(\nabla f)(x_0) + \frac{1}{2!}\mathbf{h}'(\nabla^2 f)(x_0)\mathbf{h} + \cdots .$$

If $\parallel \mathbf{h} \parallel$ is small, we have

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) \approx \frac{1}{1!}\mathbf{h}'(\nabla f)(x_0).$$

Thus, the largest growth of f takes place in the direction of $(\nabla f)(\mathbf{x}_0)$.

Twice Differentiable Functions

Definition

Let $f : S \longrightarrow \mathbb{R}$, where $S \subseteq \mathbb{R}^n$ and $S \neq \emptyset$.
f is twice differentiable in $\overline{\mathbf{x}} \in S$, if there exists a vector $L(\overline{\mathbf{x}})$, a symmetric matrix $H(\overline{\mathbf{x}}) \in \mathbb{R}^{n \times n}$ called the Hessian matrix, and a function $\alpha_{\overline{\mathbf{x}}}$ such that

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + L(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \overline{\mathbf{x}})' H(\overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}}) + \parallel \mathbf{x} - \overline{\mathbf{x}} \parallel^2 \alpha_{\overline{\mathbf{x}}}(\mathbf{x} - \overline{\mathbf{x}})$$

for $\mathbf{x} \in S$ and $\lim_{\mathbf{x} \to \overline{\mathbf{x}}} \alpha_{\overline{\mathbf{x}}}(\mathbf{x} - \overline{\mathbf{x}}) = 0$.
f is twice differentiable on an open set T, $T \subseteq S$, if f is twice differentiable in each point of T.

As before, $L(\overline{\mathbf{x}}) = (\nabla f)(\overline{\mathbf{x}})$ and $(H(\overline{\mathbf{x}}))_{ij} = \frac{\partial^2 f}{\partial x_i \, \partial x_j}$.

Taylor's Theorem

Theorem

If S is an open set in $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a twice differentiable function. For every $\mathbf{x}_1, \mathbf{x}_2 \in S$, there exists $\lambda \in (0, 1)$ such that for $\mathbf{x} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ we have

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + (\nabla f)(\mathbf{x}_1)'(\mathbf{x}_2 - \mathbf{x}_1) + \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)' H(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1),$$

where $H(\mathbf{x})$ is the Hessian matrix of f at $\mathbf{x}$.

Theorem

Let $S \neq \emptyset$ *be a convex subset of* $\mathbb{R}^n$, $f : S \longrightarrow \mathbb{R}$ *be a convex function, and let* $\overline{\mathbf{x}} \in S$. *If* $\mathbf{d}$ *is a vector such that* $\mathbf{d} \neq \mathbf{0}$, $\overline{\mathbf{x}} + \lambda\mathbf{d} \in S$ *for a sufficiently small* λ, *where* $\lambda > 0$, *then* $\frac{\partial f}{\partial \mathbf{d}}(\overline{\mathbf{x}})$ *exists.*

Proof

Let $\lambda_1, \lambda_2 \in \mathbb{R}_{>0}$ be two sufficiently small numbers such that $\lambda_1 < \lambda_2$. Since f is convex, we have

$$
\begin{aligned}
f(\overline{\mathbf{x}} + \lambda_1 \mathbf{d}) &= f\left(\frac{\lambda_1}{\lambda_2}(\overline{\mathbf{x}} + \lambda_2 \mathbf{d}) + \left(1 - \frac{\lambda_1}{\lambda_2}\right)\overline{\mathbf{x}}\right) \\
&\leqslant \frac{\lambda_1}{\lambda_2} f(\overline{\mathbf{x}} + \lambda_2 \mathbf{d}) + \left(1 - \frac{\lambda_1}{\lambda_2}\right) f(\overline{\mathbf{x}}),
\end{aligned}
$$

which implies

$$
\frac{f(\overline{\mathbf{x}} + \lambda_1 \mathbf{d}) - f(\overline{\mathbf{x}})}{\lambda_1} \leqslant \frac{f(\overline{\mathbf{x}} + \lambda_2 \mathbf{d}) - f(\overline{\mathbf{x}})}{\lambda_2}.
$$

Since $\frac{f(\overline{\mathbf{x}}+\lambda \mathbf{d}) - f(\overline{\mathbf{x}})}{\lambda}$ is a nondecreasing function of λ, $\lim_{\lambda \to 0, \lambda > 0} \frac{f(\overline{\mathbf{x}}+\lambda \mathbf{d}) - f(\overline{\mathbf{x}})}{\lambda}$ exists.

Differentiability and Directional Derivatives

Theorem

Let $S \neq \emptyset$ be a convex subset of $\mathbb{R}^n$, $f : S \longrightarrow \mathbb{R}$ be a convex function, and let $\overline{\mathbf{x}} \in S$.
If f is differentiable at $\overline{\mathbf{x}}$, then the directional derivative $\frac{\partial f}{\partial \mathbf{d}}(\overline{\mathbf{x}})$ exists and

$$\frac{\partial f}{\partial \mathbf{d}}(\overline{\mathbf{x}}) = (\nabla f(\overline{\mathbf{x}}))' \mathbf{d}.$$

Proof

Since f is differentiable we have

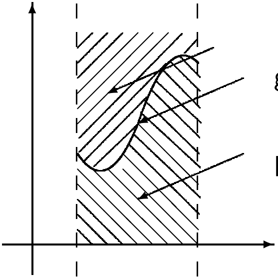
$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + (\nabla f(\overline{\mathbf{x}}))'(\mathbf{x} - \overline{\mathbf{x}}) + \parallel \mathbf{x} - \overline{\mathbf{x}} \parallel \alpha_{\overline{\mathbf{x}}}(\mathbf{x} - \overline{\mathbf{x}})$$

for each $\mathbf{x} \in S$.
Choose $\mathbf{x} = \overline{\mathbf{x}} + \theta\mathbf{d}$. Then,

$$f(\overline{\mathbf{x}} + \theta\mathbf{x}) - f(\overline{\mathbf{x}}) = \theta(\nabla f(\overline{\mathbf{x}}))'\mathbf{d} + |\theta|^2 \parallel \mathbf{d} \parallel \alpha_{\overline{\mathbf{x}}}(\mathbf{d}).$$

Taking $\theta \to 0$, the theorem follows immediately.

Epigraph, Graph and Hypograph of a Function



$$\text{epi}(f) = \left\{ \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \in \mathbb{R}^{n+1} \mid y \geqslant f(\mathbf{x}) \right\}$$

$$\text{graph}(f) = \left\{ \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \in \mathbb{R}^{n+1} \mid y = f(\mathbf{x}) \right\}$$

$$\text{hyp}(f) = \left\{ \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \in \mathbb{R}^{n+1} \mid y \leqslant f(\mathbf{x}) \right\}$$

Theorem

Let S be a non-empty convex subset of $\mathbb{R}^n$. A function $f : S \longrightarrow \mathbb{R}$ is convex if and only if epi(f) is a convex set in $\mathbb{R}^{n+1}$.

Suppose that $\mathrm{epi}(f)$ is convex and let $\mathbf{x}_1, \mathbf{x}_2 \in S$. Since

$$\begin{pmatrix} \mathbf{x}_1 \\ f(\mathbf{x}_1) \end{pmatrix}, \begin{pmatrix} \mathbf{x}_2 \\ f(\mathbf{x}_2) \end{pmatrix} \in \mathrm{epi}(f),$$

it follows that

$$\begin{aligned} & a \begin{pmatrix} \mathbf{x}_1 \\ f(\mathbf{x}_1) \end{pmatrix} + (1-a) \begin{pmatrix} \mathbf{x}_2 \\ f(\mathbf{x}_2) \end{pmatrix} \\ = & \begin{pmatrix} a\mathbf{x}_1 + (1-a)\mathbf{x}_2 \\ af(\mathbf{x}_1) + (1-a)f(\mathbf{x}_2) \end{pmatrix} \in \mathrm{epi}(f), \end{aligned}$$

so

$$f(a\mathbf{x}_1 + (1-a)\mathbf{x}_2) \leqslant af(\mathbf{x}_1) + (1-a)f(\mathbf{x}_2),$$

which means that f is convex.

Conversely, suppose that f is convex and let

$$\begin{pmatrix} \mathbf{x}_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} \mathbf{x}_2 \\ y_2 \end{pmatrix} \in \text{epi}(f),$$

that is $f(\mathbf{x}_1) \leqslant y_1$ and $f(\mathbf{x}_2) \leqslant y_2$. For $a \in [0,1]$ we have

$$a \begin{pmatrix} \mathbf{x}_1 \\ y_1 \end{pmatrix} + (1-a) \begin{pmatrix} \mathbf{x}_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a\mathbf{x}_1 + (1-a)\mathbf{x}_2 \\ ay_1 + (1-a)y_2 \end{pmatrix}.$$

Note that

$$ay_1 + (1-a)y_2 \geqslant af(\mathbf{x}_1) + (1-a)f(\mathbf{x}_2) \geqslant f(a\mathbf{x}_1 + (1-a)\mathbf{x}_2),$$

which proves that $a \begin{pmatrix} \mathbf{x}_1 \\ y_1 \end{pmatrix} + (1-a) \begin{pmatrix} \mathbf{x}_2 \\ y_2 \end{pmatrix} \in \text{epi}(f)$.

Subgradients

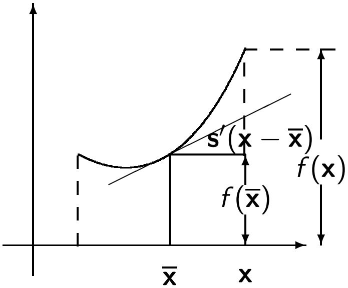
Definition

Let S be a non-empty convex subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a convex function.

A subgradient of f at $\bar{\mathbf{x}}$ is a vector $\mathbf{s}$ such that

$$f(\mathbf{x}) \geqslant f(\bar{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \bar{\mathbf{x}})$$

for all $\mathbf{x} \in S$.



Properties of Subgradients

- the collection of subgradients at $\bar{\mathbf{x}}$ is convex;
- a subgradient of a convex function at $\bar{\mathbf{x}}$ defines a supporting hyperplane of epi(f) at $\begin{pmatrix} \bar{\mathbf{x}} \\ f(\bar{\mathbf{x}}) \end{pmatrix} \in \partial\text{epi}(f)$.

Subgradients and Supporting Hyperplanes

Theorem

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and let $f : S \longrightarrow \mathbb{R}$ be a convex function. For $\overline{\mathbf{x}} \in \mathbf{I}(S)$ there exists a subgradient $\mathbf{s}$ such that the hyperplane

$$H = \left\{ \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \middle| y = f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}}) \right\} \subseteq \mathbb{R}^{n+1}$$

is a supporting hyperplane of epi(f) at $\begin{pmatrix} \overline{\mathbf{x}} \\ f(\overline{\mathbf{x}}) \end{pmatrix}$.

In particular, we have $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}})$ for $\mathbf{x} \in S$, that is, $\mathbf{s}$ is a subgradient of f at $\overline{\mathbf{x}}$.

Proof

Since f is convex, epi(f) is a convex set. Note that $\begin{pmatrix} \bar{\mathbf{x}} \\ f(\bar{\mathbf{x}}) \end{pmatrix} \in \partial\text{epi}(f)$.

From the theorem that guarantees the existence of a supporting hyperplane for every border point of a convex set (see CONV1) there is a vector $\begin{pmatrix} \mathbf{p} \\ q \end{pmatrix} \neq \mathbf{0}$ such that

$$\begin{pmatrix} \mathbf{p} \\ q \end{pmatrix}' \left(\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} - \begin{pmatrix} \bar{\mathbf{x}} \\ f(\bar{\mathbf{x}}) \end{pmatrix} \right) \leqslant 0$$

or

$$\mathbf{p}'(\mathbf{x} - \bar{\mathbf{x}}) + q(y - f(\bar{\mathbf{x}})) \leqslant 0$$

for all $\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \in \text{epi}(f)$.

Proof (cont'd)

Note that $q \leqslant 0$ (because, otherwise we could choose y sufficiently large to have $\mathbf{p}'(\mathbf{x} - \overline{\mathbf{x}}) + q(y - f(\overline{\mathbf{x}})) > 0$).
Suppose that $q = 0$. This implies $\mathbf{p}'(\mathbf{x} - \overline{\mathbf{x}}) \leqslant 0$ for $\mathbf{x} \in S$. Since $\overline{\mathbf{x}} \in S$, there exists $\lambda > 0$ such that $\overline{\mathbf{x}} + \lambda\mathbf{p} \in S$. In turn, this implies $\lambda\mathbf{p}'\mathbf{p} \leqslant 0$, and therefore, $\mathbf{p} = \mathbf{0}$ and $q = 0$ contradicting the fact that $\begin{pmatrix} \mathbf{p} \\ q \end{pmatrix} \neq \mathbf{0}$.
Thus, we conclude that $q < 0$. By dividing the inequality by $|q|$, this yields

$$\frac{1}{|q|}\mathbf{p}'(\mathbf{x} - \overline{\mathbf{x}}) - y + f(\overline{\mathbf{x}}) \leqslant 0$$

for all $\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} \in \text{epi}(f)$.

Proof (cont'd)

In particular

$$\frac{1}{|q|}\mathbf{p}'(\mathbf{x} - \overline{\mathbf{x}}) - y + f(\overline{\mathbf{x}}) = 0$$

supports epi(f) (which is in the negative half-plane of this hyperplane). By taking $y = f(\mathbf{x})$ in the last inequality we obtain

$$\frac{1}{|q|}\mathbf{p}'(\mathbf{x} - \overline{\mathbf{x}}) + f(\overline{\mathbf{x}}) \leqslant f(\mathbf{x})$$

Theorem

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and let $f : S \longrightarrow \mathbb{R}$ be a strictly convex function. For $\bar{\mathbf{x}} \in \mathbf{I}(S)$ there exists $\mathbf{s}$ such that $f(\mathbf{x}) > f(\bar{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \bar{\mathbf{x}})$ for $\mathbf{x} \in S$.

Proof

Since f is convex, there exists $\mathbf{s}$ such that

$$f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}})$$

for $\mathbf{x}, \overline{\mathbf{x}} \in S$ and $\mathbf{x} \neq \overline{\mathbf{x}}$.
Suppose that there exists $\mathbf{z}$, such that $\mathbf{z} \neq \overline{\mathbf{x}}$ and $f(\mathbf{z}) = f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{z} - \overline{\mathbf{x}})$.
By the strict convexity of f,

$$\begin{aligned} f(\lambda\overline{\mathbf{x}} + (1-\lambda)\mathbf{z}) &< \lambda f(\overline{\mathbf{x}}) + \lambda\overline{\mathbf{x}} + (1-\lambda) f(\mathbf{z}) \\ &= f(\overline{\mathbf{x}}) + (1-\lambda)\mathbf{s}'(\mathbf{z} - \overline{\mathbf{x}}). \end{aligned}$$

Taking $\mathbf{x} = \lambda\overline{\mathbf{x}} + (1-\lambda)\mathbf{z}$ in the inequality $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}})$ we obtain

$$f(\lambda\overline{\mathbf{x}} + (1-\lambda)\mathbf{z}) \geqslant f(\overline{\mathbf{x}}) + (1-\lambda)\mathbf{s}'(\mathbf{z} - \overline{\mathbf{x}}),$$

which contradicts the previous inequality.

Theorem

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and let $f : S \longrightarrow \mathbb{R}$ be a function. If for every $\overline{\mathbf{x}} \in \mathbf{I}(S)$ there exists a subgradient $\mathbf{s}$ such that

$$f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}})$$

for each $\mathbf{x} \in S$, then f is convex on $\mathbf{I}(S)$.

If $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{I}(S)$, we have $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathbf{I}(S)$ for $\lambda \in (0,1)$ due to the convexity of $\mathbf{I}(S)$.
From the existence of a subgradient $\mathbf{s}$ at $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ we have

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + (1-\lambda)\mathbf{s}'(\mathbf{x}_1 - \mathbf{x}_2) \\ f(\mathbf{x}_2) &\geq f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + \lambda\mathbf{s}'(\mathbf{x}_1 - \mathbf{x}_2) \end{aligned}$$

These inequalities imply:

$$\lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) \geqslant f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2).$$

Theorem

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and let $f : S \longrightarrow \mathbb{R}$ be a convex function. If f is differentiable at $\overline{\mathbf{x}} \in \mathbf{I}(S)$, the collection of subgradients of f equals $\{\nabla f\}$.

Proof

Let $\mathbf{s}$ be a subgradient of f at $\overline{\mathbf{x}}$. For λ sufficiently small we have

$$\begin{aligned} f(\overline{\mathbf{x}} + \lambda \mathbf{d}) &\geqslant f(\overline{\mathbf{x}}) + \lambda \mathbf{s}'\mathbf{d}, \\ f(\overline{\mathbf{x}} + \lambda \mathbf{d}) &= f(\overline{\mathbf{x}}) + \lambda (\nabla f(\overline{\mathbf{x}}))'\mathbf{d} + \lambda \parallel d \parallel \alpha_{\overline{\mathbf{x}}}(\lambda \mathbf{d}), \end{aligned}$$

which yield

$$\lambda(\mathbf{s} - \nabla f(\overline{\mathbf{x}}))'\mathbf{d} - \lambda \parallel d \parallel \alpha_{\overline{\mathbf{x}}}(\lambda \mathbf{d}) \leqslant 0.$$

Dividing by λ and taking $\lambda \to 0$ implies $\mathbf{s} - \nabla f(\overline{\mathbf{x}}))'\mathbf{d} \leqslant 0$. Choosing $\mathbf{d} = \mathbf{s} - \nabla f(\overline{\mathbf{x}})$ it follows that $\mathbf{s} = \nabla f(\overline{\mathbf{x}})$.

Convexity of Differentiable Functions

Theorem

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and let $f : S \longrightarrow \mathbb{R}$ be a differentiable function. Then, f is convex if and only if for every $\bar{\mathbf{x}} \in S$ we have

$$f(\mathbf{x}) \geqslant f(\bar{\mathbf{x}}) + (\nabla f)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}})$$

for every $\mathbf{x} \in S$.
The function f is strictly convex if for every $\bar{\mathbf{x}} \in S$ we have

$$f(\mathbf{x}) > f(\bar{\mathbf{x}}) + (\nabla f)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}})$$

for every $\mathbf{x} \in S$.

Proof

Suppose that for every $\bar{\mathbf{x}} \in S$ we have

$$f(\mathbf{x}) \geqslant f(\bar{\mathbf{x}}) + (\nabla f)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}})$$

for every $\mathbf{x} \in S$. This implies the convexity of f on $\int(S)$ by the theorem given on slide 35.

Theorem

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and let $f : S \longrightarrow \mathbb{R}$ be a differentiable function. Then, f is convex if and only if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$, we have

$$(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1) \geqslant 0.$$

Similarly, f is strictly convex if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$ such that $\mathbf{x}_1 \neq \mathbf{x}_2$ we have

$$(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1) > 0.$$

Proof

Let f be convex. Then

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}_2) + (\nabla f(\mathbf{x}_2))'(\mathbf{x}_1 - \mathbf{x}_2) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}_1) + (\nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1), \end{aligned}$$

which yield

$$(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1) \geqslant 0.$$

Proof (cont'd)

Conversely, suppose that $(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1) \geqslant 0$. By the Mean Value Theorem,

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) = (\nabla f(\mathbf{x}))'(\mathbf{x}_2 - \mathbf{x}_1),$$

where $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\lambda \in (0,1)$.
By applying the hypothesis to $\mathbf{x}$ and $\mathbf{x}_1$ we get

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}_1))'(\mathbf{x} - \mathbf{x}_1) \geqslant 0.$$

and, taking into account that $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, it follows that $(1-\lambda)(\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1) \geqslant 0$. This implies

$$\nabla f(\mathbf{x})'(\mathbf{x}_2 - \mathbf{x}_1) \geqslant \nabla f(\mathbf{x}_1))'(\mathbf{x}_2 - \mathbf{x}_1).$$

By the Mean Value Theorem,

$$f(\mathbf{x}_2) \geqslant f(\mathbf{x}_1) + (\nabla f)'(\mathbf{x}_2 - \mathbf{x}_1),$$

so f is convex.

Twice Differentiable Functions

Definition

Let $S \subseteq \mathbb{R}^n$ be a non-empty set. A function $f : S \longrightarrow \mathbb{R}$ is twice differentiable in $\bar{\mathbf{x}} \in S$ if there exists a vector $(\nabla f)(\bar{\mathbf{x}})$ and a symmetric matrix $H(\bar{\mathbf{x}})$ such that

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + (\nabla f)(\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})' H(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + \parallel \mathbf{x} - \bar{\mathbf{x}} \parallel^2 \alpha_{\bar{\mathbf{x}}}(\mathbf{x} - \bar{\mathbf{x}}),$$

for each $\mathbf{x} \in S$ and $\lim_{\mathbf{x} \to \bar{\mathbf{x}}} \alpha_{\bar{\mathbf{x}}}(\mathbf{x} - \bar{\mathbf{x}}) = 0$.

The entry $H(\mathbf{x})_{ij}$ of the matrix H is $\frac{\partial^2 f}{\partial x_i \, \partial x_j}$ for $1 \leqslant i, j \leqslant n$. $H(\mathbf{x})$ is the Hessian matrix of f in $\mathbf{x}$.

A Practical Convexity Criterion

Theorem

Let S be a nonempty open convex set in $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be twice differentiable on S. Then, f is convex if and only if its Hessian matrix is positive semidefinite at each point in S.

Note: $H(\overline{\mathbf{x}})$ is positive semidefinite if $\mathbf{x}'H(\overline{\mathbf{x}})\mathbf{x} \geqslant 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof

Suppose that f is convex and let $\overline{\mathbf{x}} \in S$. Since S is open, $\overline{\mathbf{x}} + \lambda\mathbf{x} \in S$ if λ is sufficiently small.
By convexity and double differentiability we get:

$$\begin{aligned} f(\overline{\mathbf{x}} + \lambda\mathbf{x}) &\geqslant f(\overline{\mathbf{x}}) + \lambda(\nabla f)(\overline{\mathbf{x}})'\mathbf{x} \\ f(\overline{\mathbf{x}} + \lambda\mathbf{x}) &= f(\overline{\mathbf{x}}) + \lambda(\nabla f)(\overline{\mathbf{x}})'\mathbf{x} + \frac{1}{2}\lambda^2\mathbf{x}'H(\overline{\mathbf{x}})\mathbf{x} + \lambda^2 \parallel \mathbf{x} \parallel^2 \alpha_{\overline{\mathbf{x}}}(\lambda\mathbf{x}), \end{aligned}$$

which yield

$$\frac{1}{2}\lambda^2\mathbf{x}'H(\overline{\mathbf{x}})\mathbf{x} + \lambda^2 \parallel \mathbf{x} \parallel^2 \alpha_{\overline{\mathbf{x}}}(\lambda\mathbf{x}) \geqslant 0.$$

Dividing by λ^2 and letting $\lambda \to 0$ implies $\mathbf{x}'H(\overline{\mathbf{x}})\mathbf{x} \geqslant 0$.

Proof (cont'd)

Conversely, suppose that H is positive semidefinite at each point in S and let $\mathbf{x}, \overline{\mathbf{x}} \in S$.
By Tylor Theorem (on slide 18), there exists $\lambda \in (0,1)$ and $\hat{\mathbf{x}} = \lambda \mathbf{x} + (1-\lambda)\overline{\mathbf{x}}$ such that

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + (\nabla f)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - ol\mathbf{x})' H(\hat{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}}).$$

Since $\hat{\mathbf{x}} \in S$, $H(\hat{\mathbf{x}})$ is positive semidefinite, so

$$f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + (\nabla f)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}),$$

which implies that f is convex.

Example

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be the function defined by

$$f(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + a_{10}x_1 + a_{01}x_2 + a_{00}.$$

We have

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 2a_{11}x_1 + 2a_{12}x_2 + a_{10} \\ \frac{\partial f}{\partial x_2} &= 2a_{12}x_1 + 2a_{22}x_2 + a_{01}, \end{aligned}$$

and

$$\begin{array}{ll} \frac{\partial^2 f}{\partial x_1^2} = 2a_{11} & \frac{\partial^2 f}{\partial x_1\, \partial x_2} = 2a_{12} \\ \frac{\partial^2 f}{\partial x_2\, \partial x_1} = 2a_{12} & \frac{\partial^2 f}{\partial x_1^2} = 2a_{22}. \end{array}$$

Thus

$$H(\mathbf{x}) = \begin{pmatrix} 2a_{11} & 2a_{12} \\ 2a_{12} & 2a_{11} \end{pmatrix}$$

Example (cont'd)

The eigenvalues of $H(\mathbf{x})$ are obtained from the equation $\det(\lambda I_2 - H) = 0$, or

$$\begin{vmatrix} \lambda - 2a_{11} & -2a_{12} \\ -2a_{12} & \lambda - 2a_{22} \end{vmatrix} = 0,$$

which amounts to

$$\lambda^2 - 2\lambda(a_{11} + a_{22}) + 4(a_{11}a_{22} - a_{12}^2) = 0$$

The discriminant of this equation is

$$(a_{11} + a_{12})^2 - 4(a_{11}a_{12} - a_{12}^2) = (a_{11} - a_{12})^2 + 4a_{12}^2 \geq 0,$$

so the eigenvalues are real numbers. To have a positive semidefinite matrix, we need to require the roots of this equation to be positive, that is, $a_{11} + a_{22} > 0$ and $a_{11}a_{22} - a_{12}^2 \geq 0$.

Example (cont'd)

Consider, for instance the function

$$f(x_1, x_2) = x_1^2 + 6x_1x_2 + 25x_2^2 + 8x_1 - 10x_2 + 20.$$

Since $a_{11} + a_{22} = 26$ and $a_{11}a_{22} - a_{12}^2 = 25 - 9 = 16$, H is positive semidefinite. The function f is, therefore, convex.

Optimization Problem: given a function $f : S \longrightarrow \mathbb{R}$, where $S \subseteq \mathbb{R}^n$ find $\overline{\mathbf{x}} \in S$ such that $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}})$.

- $\mathbf{x} \in S$ is a *feasible point*;
- f is the objective function;
- $\overline{\mathbf{x}}$ is an optimum point;
- if $\mathbf{x} \in S(\overline{\mathbf{x}}, \epsilon)$ and $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}})$, then $\overline{\mathbf{x}}$ is a local optimum.

Since $\max_{\mathbf{x}} f(\mathbf{x}) = -\min_{\mathbf{x}}(-f(\mathbf{x}))$, optimization problems can be cast as maximizations.

Challenges for Unconstrained Optimization

- $\bar{\mathbf{x}}$ may not exist when f is not bounded below (see $f(x) = x^3$ for $x \in \mathbb{R}$);
- $\bar{\mathbf{x}}$ may not exist even if f is bounded below (see $f(x) = e^{-x}$ for $x \in \mathbb{R}$);
- if $\bar{\mathbf{x}}$ exists it may not be unique (see $f(x) = \cos x$);
- in practice we find local minimizers and they may not be global minimizers (see $f(x) = x^2 - \cos x$);
- global minimizers are hard to find.

Why is convexity important for optimization?

Theorem

Let S be a non-empty convex subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$. Suppose that $\overline{\mathbf{x}}$ is a local optimum. The following hold:

- *if f is convex, then $\overline{\mathbf{x}}$ is a global optimum;*
- *if f is strictly convex, then $\overline{\mathbf{x}}$ is the unique global optimum.*

Proof

Let f be convex and let $\overline{\mathbf{x}}$ be a local optimum. There exists $\epsilon > 0$ such that $\mathbf{x} \in C(\overline{\mathbf{x}}, \epsilon)$ implies $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}})$.
Suppose $\overline{\mathbf{x}}$ is not a global optimum. Then, there exists $\hat{\mathbf{x}}$ such that $f(\hat{\mathbf{x}}) < f(\overline{\mathbf{x}})$. Since f is convex, we have

$$f(a\hat{\mathbf{x}} + (1-a)\overline{\mathbf{x}}) \leqslant af(\hat{\mathbf{x}}) + (1-a)f(\overline{\mathbf{x}}) < af(\overline{\mathbf{x}}) + (1-a)f(\overline{\mathbf{x}}) = f(\overline{\mathbf{x}}).$$

If a is sufficiently small we have $a\hat{\mathbf{x}} + (1-a)\overline{\mathbf{x}} \in C(\overline{\mathbf{x}}, \epsilon)$, which contradicts the local optimality of $\overline{\mathbf{x}}$.

Proof (cont'd)

Let f be strictly convex and suppose that there exists $\mathbf{z} \in S$ such that $f(\mathbf{z}) = f(\overline{\mathbf{x}})$. The strict convexity of f implies

$$f\left(\frac{\mathbf{z} + \overline{\mathbf{x}}}{2}\right) < \frac{1}{2}f(\mathbf{z}) + \frac{1}{2}f(\overline{\mathbf{x}}) = f(\overline{\mathbf{x}}),$$

which contradicts the fact that $\overline{\mathbf{x}}$ is a global optimum (because $\frac{\mathbf{z}+\overline{\mathbf{x}}}{2} \in S$ due to the convexity of S).

Let $f : S \longrightarrow \mathbb{R}$ be a convex function and S be a non-empty convex set, $S \subseteq \mathbb{R}^n$.

Optimization Problem: minimize $f(\mathbf{x})$ subjected to $\mathbf{x} \in S$.

Theorem

The point $\overline{\mathbf{x}}$ is am optimal solution if and only if f has a subgradient $\mathbf{s}$ and $\mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$ for all $\mathbf{x} \in S$.

Proof

Suppose that $\mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$ for all $\mathbf{x} \in S$, where $\mathbf{s}$ is the subgradient of f at $\overline{\mathbf{x}}$.
Since f is convex,

$$f(\mathbf{x}) \geq f(\overline{\mathbf{x}}) + \mathbf{s}'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$$

for all $\mathbf{x} \in S$, so $f(\mathbf{x}) \geq f(\overline{\mathbf{x}})$, which means that $\overline{\mathbf{x}}$ is an optimal solution.

Proof (cont'd)

Conversely, suppose that $\bar{\mathbf{x}}$ is an optimal solution and define the sets $A, B \subseteq \mathbb{R}^{n+1}$:

$$
\begin{aligned}
A &= \left\{ \begin{pmatrix} \mathbf{x} - \bar{\mathbf{x}} \\ y \end{pmatrix} \in \mathbb{R}^{n+1} \middle| y > f(\mathbf{x}) - f(\bar{\mathbf{x}}) \right\} \\
B &= \left\{ \begin{pmatrix} \mathbf{x} - \bar{\mathbf{x}} \\ y \end{pmatrix} \in \mathbb{R}^{n+1} \middle| y \leqslant 0 \right\}.
\end{aligned}
$$

Both A and B are convex. We claim $A \cap B = \emptyset$. Indeed, suppose otherwise. This would imply the existence of $\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix}$ such that

$$\mathbf{x} \in S \text{ and } 0 \geqslant y > f(\mathbf{x}) - f(\bar{\mathbf{x}}),$$

contradicting the optimality of $\bar{\mathbf{x}}$.

Proof (cont'd)

There exists a hyperplane that separates A and B, that is, there exists $\begin{pmatrix} \mathbf{a} \\ b \end{pmatrix} \in \mathbb{R}^{n+1}$ and $c \in \mathbb{R}$ such that

$$\begin{aligned} \mathbf{a}'(\mathbf{x} - \overline{\mathbf{x}}) + by &\leqslant c \text{ for } \mathbf{x} \in \mathbb{R}^n, y > f(\mathbf{x}) - f(\overline{\mathbf{x}}) \\ \mathbf{a}'(\mathbf{x} - \overline{\mathbf{x}}) + by &\geqslant c \text{ for } \mathbf{x} \in S, y \leqslant 0. \end{aligned}$$

Chosing $\mathbf{x} = \overline{\mathbf{x}}$ and $y = 0$ in the second inequality we have $c \leqslant 0$. Choosing $\mathbf{x} = \overline{\mathbf{x}}$ and $y = \epsilon > 0$ in the first inequality implies $b\epsilon \leqslant c$ for every positive ϵ, so $b \leqslant 0$ and $c \geqslant 0$. Thus, $b \leqslant 0$ and $c = 0$.

Proof (cont'd)

If $b = 0$, then $\mathbf{a}'(\mathbf{x} - \overline{\mathbf{x}}) \leq 0$ for $x \in \mathbb{R}^n$. Choosing $\mathbf{x} = \overline{\mathbf{x}} + \mathbf{a}$ yields $\mathbf{a}'\mathbf{a} = 0$, so $\mathbf{a} = \mathbf{0}$. Since $\begin{pmatrix} \mathbf{a} \\ b \end{pmatrix} \neq \mathbf{0}_{n+1}$, this implies $b < 0$, so $-b > 0$. Thus, by dividing the equations by β we get:

$$\begin{aligned} -\frac{1}{b}\mathbf{a}'(\mathbf{x} - \overline{\mathbf{x}}) - y &\leqslant 0 \text{ for } \mathbf{x} \in \mathbb{R}^n, y > f(\mathbf{x}) - f(\overline{\mathbf{x}}) \\ -\frac{1}{b}\mathbf{a}'(\mathbf{x} - \overline{\mathbf{x}}) - y &\geqslant c \text{ for } \mathbf{x} \in S, y \leqslant 0. \end{aligned}$$

Letting $y = 0$ in the second inequality we obtain

$$\mathbf{a}_1'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$$

for $\mathbf{x} \in S$, where $\mathbf{a}_1 = -\frac{1}{b}\mathbf{a}$. From the first inequality, $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + \mathbf{a}_1(\mathbf{x} - \overline{\mathbf{x}})$. Thus, $\mathbf{a}_1$ is a subgradient of f with $\mathbf{a}_1'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$.

Corollary

Corollary

If S is open, then $\mathbf{\bar{x}}$ is an optimal solution if and only if there exists a zero subgradient of f at $\mathbf{\bar{x}}$.

Proof

$\bar{\mathbf{x}}$ is an optimal solution if and only if $\mathbf{s}'(\mathbf{x} - \bar{\mathbf{x}}) \geqslant 0$ for $\mathbf{x} \in S$, where $\mathbf{s}$ is a subgradient of f at $\bar{\mathbf{x}}$. Since S is open, there exists $\lambda > 0$ such that $\mathbf{x} = \bar{\mathbf{x}} - \lambda \mathbf{s} \in S$, so $-\lambda \parallel \mathbf{s} \parallel^2 = 0$, which implies $\mathbf{s} = \mathbf{0}$.

Corollary

Corollary

If f is differentiable, then $\overline{\mathbf{x}}$ is optimal if and only if

$$(\nabla f)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$$

for $\mathbf{x} \in S$. Moreover, if S is open, then $\mathbf{x}$ is optimal if and only if $(\nabla f)(\overline{\mathbf{x}}) = \mathbf{0}$.

Example

To minimize the function f introduced on slide 50,

$$f(x_1, x_2) = x_1^2 + 6x_1x_2 + 25x_2^2 + 8x_1 - 10x_2 + 20.$$

we need to require that

$$(\nabla f)(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 + 6x_2 + 8 \\ 6x_1 + 50x_2 - 10 \end{pmatrix} = \mathbf{0},$$

which implies $x_1 = -\frac{65}{8}$ and $x_2 = \frac{11}{8}$.

Bibliography

Useful references for convex sets and convex functions are [2, 1] and [3].

D. P. Bertsekas.
Nonlinear Programming.
Athena Scientific, Belmont, MA, 1999.

S. Boyd and L. Vandenberghe.
Convex Optimization.
Cambridge, Cambridge, UK, 2004.

S. R. Lay.
Convex Sets and Their Applications.
Dover, Mineola, NY, 1982.