Regression - III

Prof. Dan A. Simovici

UMB

If $S \subseteq \mathbb{R}^n$ and $\mathcal{C}_S$ is the collection of convex sets that contain S, then

1. $\mathcal{C}_S \neq \emptyset$ because $\mathbb{R}^n$ is a convex set that contains S.
2. Any intersection of subsets of $\mathcal{C}_S$ is a convex set that contains S.

Thus $\bigcap \mathcal{C}_S$ is the least convex set that contains S.

Definition

The convex closure of the subset S of $\mathbb{R}^n$ is the set $\mathbf{K}_{\mathrm{conv}}(S) = \bigcap \mathcal{C}_S$.

The convex closure of S is denoted by $\mathbf{K}_{\mathrm{conv}}(S)$.

Note that

- $S \subseteq \mathbf{K}_{\mathrm{conv}}(S)$;
- $S_1 \subseteq S_2$ implies $\mathbf{K}_{\mathrm{conv}}(S_1) \subseteq \mathbf{K}_{\mathrm{conv}}(S_2)$;
- $\mathbf{K}_{\mathrm{conv}}(\mathbf{K}_{\mathrm{conv}}(S)) = \mathbf{K}_{\mathrm{conv}}(S)$.

Definition

Let $f : \mathbb{R}^n \longrightarrow \hat{\mathbb{R}}$ be a function. Its *epigraph* is the set

$$\text{epi}(f) = \{(\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leqslant y\}.$$

The *hypograph* of f is the set

$$\text{hyp}(f) = \{(\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R} \mid y \leqslant f(\mathbf{x})\}.$$

The epigraph of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is the dotted area in $\mathbb{R}^2$ located above the graph of the function f and it is shown in Figure **??**(a); the hypograph of f is the dotted area below the graph shown in Figure **??**(b).

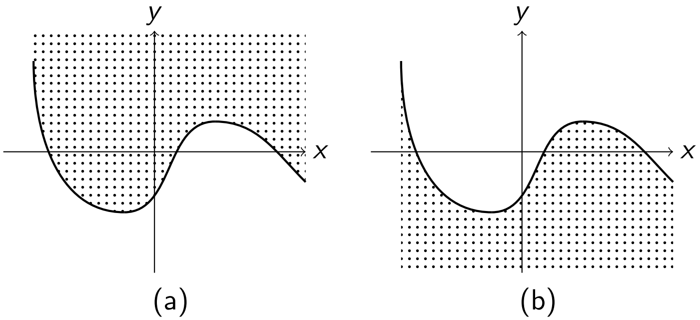

Figure : Epigraph (a) and hypograph (b) of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$

Note that the intersection

$$\mathrm{epi}(f) \cap \mathrm{hyp}(f) = \{(x, y) \in S \times \mathbb{R} \mid y = f(x)\}$$

is the graph of the function f.
If $f(x) = \infty$, then $(x, \infty) \notin \mathrm{epi}(f)$. Thus, for the function f_∞ defined by $f_\infty(x) = \infty$ for $x \in S$ we have $\mathrm{epi}(f_\infty) = \emptyset$.

Definition

Let $f : \mathbb{R}^n \longrightarrow \hat{\mathbb{R}}$ be a function and let $a \in \hat{\mathbb{R}}$. The *level set for f at a* is the set

$$L_{f,a} = \{x \in S \mid f(x) \leqslant a\}.$$

Definition

Let C, D be two subsets of $\mathbb{R}^n$. A hyperplane $\mathbf{w}'\mathbf{x} - a = 0$ *separates* C, D if $\mathbf{w}'\mathbf{x} \leqslant a$ for every $x \in C$ and $\mathbf{w}'\mathbf{x} \geqslant a$ for every $x \in D$.
If a separating hyperplane H exists for two subsets C, D of L we say that C, D are *separable*.

C and D are separated by a hyperplane H if C and D are located in distinct closed half-spaces associated to H. The sets C and D are *linearly separable* if there exists a hyperplane that separates them.

Definition

The subsets C and D of $\mathbb{R}^n$ are *strictly separated* by a hyperplane $\mathbf{w}'\mathbf{x} = a$ if we have either $\mathbf{w}'\mathbf{x} > a > \mathbf{w}'\mathbf{y}$ for $\mathbf{x} \in C$ and $\mathbf{y} \in D$, or $\mathbf{w}'\mathbf{y} > a > \mathbf{w}'\mathbf{x}$ for $\mathbf{x} \in C$ and $\mathbf{y} \in D$.
The sets C and D are *strictly linearly separable* if there exists a hyperplane that strictly separates them.

Theorem

Let C be a convex subset in $\mathbb{R}^n$ such that $\mathbf{I}(C) \neq \emptyset$ and let V be an affine subspace such that $V \cap \mathbf{I}(C) = \emptyset$.
There exists a closed hyperplane H, $\mathbf{w}'\mathbf{x} = a$, in $\mathbb{R}^n$ such that

1. *$V \subseteq H$ and $H \cap \mathbf{I}(C) = \emptyset$, and*
2. *there exists $c \in \mathbb{R}$ such that $\mathbf{w}'\mathbf{x} = c$ for all $\mathbf{x} \in V$ and $\mathbf{w}'\mathbf{x} < c$ for all $x \in \mathbf{I}(C)$.*

Definition

Let C be a convex set in $\mathbb{R}^n$. A hyperplane H is a *supporting hyperplane of C* if the following conditions are satisfied:

1. H is closed;
2. C is included in one of the closed half-spaces determined by H;
3. $H \cap \mathbf{K}(C) \neq \emptyset$.

Theorem

Let C be a convex set in a linear space L. If $\mathbf{I}(C) \neq \emptyset$ and $x_0 \in \partial C$, then there exists a supporting hyperplane H of C such that $x_0 \in H$.

Theorem

(Separation Theorem) Let C_1, C_2 be two non-empty convex sets in $\mathbb{R}^n$ such that $\mathbf{I}(C_1) \neq \emptyset$ and $C_2 \cap \mathbf{I}(C_1) = \emptyset$.
There exists a closed hyperplane H, $\mathbf{w}'\mathbf{x} = a$, separating C_1 and C_2. In other words, there exists a linear functional $f \in L^*$ such that

$$\sup\{\mathbf{w}'\mathbf{x} \mid \mathbf{x} \in C_1\} \leqslant \inf\{\mathbf{w}'\mathbf{x} \mid x \in C_2\},$$

which means that C_1 and C_2 are located in distinct half-spaces determined by H.

Corollary

Let C_1, C_2 be two disjoint subsets of $\mathbb{R}^n$. If C_1 is open, then C_1 and C_2 are separable, that is, that

$$\sup\{\mathbf{w}'\mathbf{x} \mid \mathbf{x} \in C_1\} \leqslant \inf\{\mathbf{w}'\mathbf{x} \mid \mathbf{x} \in C_2\},$$

Definition

Let X be an open set in $\mathbb{R}^n$ and let $f : X \longrightarrow \mathbb{R}$ be a function.
The function f is *Gâteaux differentiable* in $\mathbf{x}_0$, where $\mathbf{x}_0 \in X$ if there exists a linear operator $(D_x f)(\mathbf{x}_0) : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that

$$(D_x f)(x_0)(u) = \lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{u}) - f(\mathbf{x}_0)}{t}$$

for every $\mathbf{u}$ such that $x_0 + t\mathbf{u} \in X$. The linear operator $(D_x f)(x_0)$ is the *Gâteaux derivative* of f in x_0.
The *Gâteaux differential* of f at x_0 is the the linear operator $\delta f(x_0; h)$ given by

$$\delta f(x_0; u) = \lim_{t \to 0} \frac{f(x_0 + tu) - f(x_0)}{t}.$$

Example

Let $\mathbf{a}$ be a vector in $\mathbb{R}^n$. Define $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ as $f(\mathbf{x}) = \mathbf{x}'\mathbf{a}$. We have:

$$
\begin{aligned}
(D_x f)(\mathbf{x}_0)(\mathbf{u}) &= \lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{u}) - f(\mathbf{x}_0)}{t} \\
&= \lim_{t \to 0} \frac{(\mathbf{x}_0 + t\mathbf{u})'\mathbf{a} - \mathbf{x}_0'\mathbf{a}}{t} \\
&= \lim_{t \to 0} \frac{t\mathbf{u}'\mathbf{a}}{t} = \mathbf{u}'\mathbf{a}.
\end{aligned}
$$

Example

Let $A \in \mathbb{R}^{n\times n}$ be a matrix and let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be the functional $f(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$. We have $(Df)(\mathbf{x}_0) = \mathbf{x}_0'(A + A')$.
By applying the definition of Gâteaux differential we have

$$
\begin{aligned}
(Df)(\mathbf{x}_0)(\mathbf{u}) &= \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{u}) - f(\mathbf{x}_0)}{t} \\
&= \lim_{t\to 0} \frac{(\mathbf{x}_0' + t\mathbf{u}')A(\mathbf{x}_0 + t\mathbf{u}) - \mathbf{x}_0'A\mathbf{x}_0}{t} \\
&= \lim_{t\to 0} \frac{t\mathbf{u}'A\mathbf{x}_0 + t\mathbf{x}_0'A\mathbf{u} + t^2\mathbf{u}'A\mathbf{u}}{t} \\
&= \mathbf{u}'A\mathbf{x}_0 + \mathbf{x}_0'A\mathbf{u} = \mathbf{x}_0'A'\mathbf{u} + \mathbf{x}_0'A\mathbf{u} \\
&= \mathbf{x}_0'(A + A')\mathbf{u},
\end{aligned}
$$

which yields $(Df)(\mathbf{x}_0) = \mathbf{x}_0'(A + A')$.
If $A \in \mathbb{R}^{n\times n}$ is symmetric and $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is the functional $f(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$, then $(Df)(\mathbf{x}_0) = 2\mathbf{x}_0'A$.

Example

The norm $\| \cdot \| : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geqslant 0}$ is not Gâteaux differentiable in $\mathbf{0}_n$.
Indeed, suppose that $\| \cdot \|$ were differentiable in $\mathbf{0}_n$, which would mean that the limit:

$$\lim_{t \to 0} \frac{\| tu \|}{t} = \lim_{t \to 0} \frac{|t|}{t} \| u \|$$

exists for every $u \in \mathbb{R}^n$, which is contradictory.
However, the square of the norm, $\| \cdot \|^2$ is differentiable in $\mathbf{0}_n$ because

$$\lim_{t \to 0} \frac{\| t\mathbf{u} \|^2}{t} = \lim_{t \to 0} t \| \mathbf{u} \| = 0.$$

Example

Consider the norm $\| \cdot \|_1$ on $\mathbb{R}^n$ given by

$$\| \mathbf{x} \|_1 = |x_1| + \ldots + |x_n|$$

for $\mathbf{x} \in \mathbb{R}^n$. This norm is not Gâteaux differentiable in any point $\mathbf{x}_0$ located on an axis. Indeed, let $\mathbf{x}_0 = a\mathbf{e}_i$ be a point on the i^{th} axis. The limit

$$\begin{aligned} &\lim_{t \to 0} \frac{\| \mathbf{x}_0 + t\mathbf{u} \|_1 - \| \mathbf{x}_0 \|_1}{t} \\ = \quad &\lim_{t \to 0} \frac{\| a\mathbf{e}_i + t\mathbf{u} \|_1 - \| a\mathbf{e}_i \|_1}{t} \\ = \quad &\lim_{t \to 0} \frac{|t||u_1| + \cdots + |t||u_{i-1}| + (|t||u_i| - |a|) + |t||u_{i+1}| + \cdots + |t||u_n|}{t} \end{aligned}$$

does not exists, so the norm $\| \cdot \|_1$ is not differentiable in any of these points.

Definition

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function and let $\mathbf{h} \in \mathbb{R}^n - \{\mathbf{0}_n\}$.
The *directional derivative at $\mathbf{x}_0$ in the direction* $\mathbf{h}$ is the function $\frac{\partial f}{\partial h}(x_0)$ given by

$$\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}_0) = \lim_{t \downarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{h}) - f(\mathbf{x}_0)}{t}.$$

f is Gâteaux differentiable at $\mathbf{x}_0$ if its directional derivative exists in every direction.
Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function differentiable at $\mathbf{x}_0 \in \mathbb{R}^n$. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$, then $(Df)(\mathbf{x}_0)(\mathbf{e}_i)$ is known as the *partial derivative* of f with respect to x_i and is denoted by $\frac{\partial f}{\partial x_i}(\mathbf{x}_0)$.

Theorem

Let X be an open set in $\mathbb{R}^n$ and let $f : X \longrightarrow \mathbb{R}$ be a function. If f is Gâteaux differentiable on X, then

$$\| f(\mathbf{u}) - f(\mathbf{v}) \| \leqslant \| \mathbf{u} - \mathbf{v} \| \sup\{f'(a\mathbf{u} + (1-a)\mathbf{v}) \mid a \in [0,1]\}.$$

Let $w \in X$ such that $\| \mathbf{w} \| = 1$ and $\| f(\mathbf{u}) - f(\mathbf{v}) \| = (\mathbf{w}, f(\mathbf{u}) - f(\mathbf{v}))$. Define the the real-valued function g as $g(t) = (\mathbf{w}, f(\mathbf{u} + t(\mathbf{v} - \mathbf{u})))$ for $t \in [0, 1]$. We have the inequality

$$\| f(\mathbf{u}) - f(\mathbf{v}) \| = (\mathbf{w}, f(\mathbf{v}) - f(\mathbf{u})) = |g(1) - g(0)| \leqslant \sup\{|g'(t)| \mid t \in [0, 1]\}.$$

Since

$$\begin{aligned} g'(t) &= (\mathbf{w}, \mathrm{DER}f(\mathbf{u} + t(\mathbf{v} - \mathbf{u}))t) \\ &= \left(\mathbf{w}, \lim_{r \to 0} \frac{f(\mathbf{u} + (t + r)(\mathbf{v} - \mathbf{u})) - f(\mathbf{u} + t(\mathbf{v} - \mathbf{u}))}{r}\right) \\ &= \left(\mathbf{w}, f'_{\mathbf{u}+t(\mathbf{v}-\mathbf{u})}(\mathbf{v} - \mathbf{u})\right), \end{aligned}$$

we have $|g'(t)| \leqslant \| f'_{\mathbf{u}+t(\mathbf{v}-\mathbf{u})}(\mathbf{v} - \mathbf{u}) \|$, hence

$$\begin{aligned} |g'(t)| &\leqslant \| f'_{\mathbf{u}+t(\mathbf{v}-\mathbf{u})}(\mathbf{v} - \mathbf{u}) \| \\ &\leqslant \| f'_{\mathbf{u}+t(\mathbf{v}-\mathbf{u})} \| \| \mathbf{v} - \mathbf{u} \| . \end{aligned}$$

Recall that for $u, v \in \mathbb{R} \cup \{\infty\}$, the sum $u + v$ is always defined.
It is useful to extend the notion of convex function by allowing ∞ as a value. Thus, if a function f is defined on a subset S of a linear space L, $f : S \longrightarrow \mathbb{R}$, the *extended-value function* of f is the function $\hat{f} : L \longrightarrow \mathbb{R} \cup \{\infty\}$ defined by

$$\hat{f}(x) = \begin{cases} f(x) & \text{if } x \in S, \\ \infty & \text{otherwise,} \end{cases}$$

If a function $f : S \longrightarrow \mathbb{R}$ is convex, where $S \subseteq L$ is a convex set, then its extended-value function $\hat{f}$ satisfies the inequality that defines convexity $\hat{f}((1-t)x + ty) \leqslant (1-t)\hat{f}(x) + t\hat{f}(y)$ for every $x, y \in L$ and $t \in [0,1]$, if we adopt the convention that $0 \cdot \infty = 0$.

Definition

The trivial convex function is the function $f_\infty : S \longrightarrow \mathbb{R} \cup \{\infty\}$ defined by $f(x) = \infty$ for every $x \in S$.

A extended-value convex function $\hat{f} : S \longrightarrow \mathbb{R} \cup \{\infty\}$ is *properly convex* or a *proper function* if $\hat{f} \neq f_\infty$.
The *domain* of a function $f : S \longrightarrow \mathbb{R} \cup \{\infty\}$ is the set $\text{Dom}(f) = \{x \in S \mid f(\mathbf{x}) < \infty\}$.

Example

Let $f : (0, \infty) \longrightarrow \mathbb{R}$ be defined by $f(x) = x^2$. The definition domain of f is clearly convex and we have:

$$\begin{aligned} f((1-t)x_1 + tx_2) &= ((1-t)x_1 + tx_2)^2 \\ &= (1-t)^2 x_1^2 + t^2 x_2^2 + 2(1-t)tx_1x_2. \end{aligned}$$

Therefore,

$$\begin{aligned} & f((1-t)x_1 + tx_2) - (1-t)f(x_1) - tf(x_2) \\ &= (1-t)^2 x_1^2 + t^2 x_2^2 + 2(1-t)tx_1x_2 - (1-t)x_1^2 - tx_2^2 \\ &= -t(1-t)(x_1 - x_2)^2 \leqslant 0, \end{aligned}$$

which implies that f is indeed convex.

Example

The function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = |a - xb|$ is convex because

$$\begin{aligned}
f((1-t)x_1 + tx_2) &= |a - ((1-t)x_1 + tx_2)b| \\
&= |a(1-t) + at - ((1-t)x_1 + tx_2)b| \\
&= |(1-t)(a - x_1 b) + t(a - x_2 b)| \\
&\leqslant |(1-t)(a - x_1 b)| + |t(a - x_2 b)| = (1-t)f(x_1) + tf
\end{aligned}$$

for $t \in [0, 1]$.

Example

Any norm ν on a real linear space L is convex. Indeed, for $t \in [0, 1]$ we have

$$\nu(tx + (1 - t)y) \leqslant \nu(tx) + \nu((1 - t)y) = t\nu(x) + (1 - t)\nu(y)$$

for $x, y \in L$.

It is easy to verify that any linear combination of convex functions with non-negative coefficients defined on a real linear space L (of functions convex at $x_0 \in L$) is a convex function (a function convex at x_0).

Example

Let $A \in \mathbb{R}^{n\times n}$ be a matrix. If A is a positive matrix then the function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by $f(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$ is convex on $\mathbb{R}^n$.
Let $t \in [0,1]$ and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By hypothesis we have

$$(t - t^2)(\mathbf{x} - \mathbf{y})'A(\mathbf{x} - \mathbf{y}) \geqslant 0$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ because $t - t^2 \geqslant 0$. Therefore,

$$\begin{aligned}
&(1-t)\mathbf{x}'A\mathbf{x} + t\mathbf{y}'A\mathbf{y} \\
&\quad = \mathbf{x}'A\mathbf{x} + t\mathbf{x}'A(\mathbf{y}-\mathbf{x}) + t(\mathbf{y}-\mathbf{x})'A\mathbf{x} + t(\mathbf{y}-\mathbf{x})'A(\mathbf{y}-\mathbf{x}) \\
&\quad \geqslant \mathbf{x}'A\mathbf{x} + t\mathbf{x}'A(\mathbf{y}-\mathbf{x}) + t(\mathbf{y}-\mathbf{x})'A\mathbf{x} + t^2(\mathbf{y}-\mathbf{x})'A(\mathbf{y}-\mathbf{x}) \\
&\quad = (\mathbf{x} + t(\mathbf{y}-\mathbf{x}))'A(\mathbf{x} + t(\mathbf{y}-\mathbf{x})
\end{aligned}$$

for $t \in [0,1]$, which proves the convexity of f.

Theorem

Let (a, b) *be an open interval of* $\mathbb{R}$ *and let* $f : (a, b) \longrightarrow \mathbb{R}$ *be a differentiable function on* (a, b)*. Then,* f *is convex on* (a, b) *if and only if* $f(y) \geqslant f(x) + f'(x)(y - x)$ *for every* $x, y \in (a, b)$*.*

Proof

Suppose that f is convex on (a, b). Then, for $x, y \in (a, b)$ we have

$$f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y)$$

for $t \in [0, 1]$. Therefore, for $t < 1$ we have

$$f(y) \geqslant f(x) + \frac{f(x + t(y-x)) - f(x)}{t(y-x)}(y-x).$$

When $t \to 0$ we obtain $f(y) \geqslant f(x) + f'(x)(y-x)$.
Conversely, suppose that $f(y) \geqslant f(x) + f'(x)(y-x)$ for every $x, y \in (a, b)$ and let $z = (1-t)x + ty$. We have

$$f(x) \geqslant f(z) + f'(z)(x-z),$$
$$f(y) \geqslant f(z) + f'(z)(y-z).$$

By multiplying the first inequality by $1-t$ and the second by t we obtain

$$(1-t)f(x) + tf(y) \geqslant f(z),$$

which shows that f is convex.

Extension of Previous Theorem

Theorem

Let S be a convex subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a Gâteaux differentiable function on S. Then, f is convex on S if and only if $f(\mathbf{y}) \geqslant f(\mathbf{x}) + (\nabla f)(\mathbf{x})'(\mathbf{y} - \mathbf{x})$ for every $\mathbf{x}, \mathbf{y} \in S$.

Corollary

Let S be an convex subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a Gâteaux differentiable function on S. If $(\nabla f)(\mathbf{x}_0)'(\mathbf{x} - \mathbf{x}_0) \geqslant 0$ for every $\mathbf{x} \in S$, then $f(\mathbf{x}_0)$ is a minimum for f in S.

Example

Let $S = \mathbf{K}_{\mathrm{conv}}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq \mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be the linear function defined by $f(\mathbf{x}) = \mathbf{c}'\mathbf{x}$. We have $(\nabla f)(\mathbf{x}) = \mathbf{c}$.
If $\mathbf{c}'(\mathbf{x} - \mathbf{x}_0) \geqslant 0$ for every $\mathbf{x} \in S$, then $\mathbf{x}_0$ is a minimizer for f. Note that $\mathbf{x} \in S$ if and only if $\mathbf{x} = \sum_{i=1}^m b_i \mathbf{a}_i$, where $b_i \geqslant 0$ for $1 \leqslant i \leqslant m$ and $\sum_{i=1}^m b_i = 1$. Thr previous inequality can be written as

$$\mathbf{c}' \left(\sum_{i=1}^m b_i \mathbf{a}_i - \mathbf{x}_0 \right) = \mathbf{c}' \sum_{i=1}^m b_i (\mathbf{a}_i - \mathbf{x}_0) \geqslant 0$$

for $b_i \geqslant 0$, $1 \leqslant i \leqslant m$, and $\sum_{i=1}^m b_i = 1$. When $\mathbf{x}_0 = \mathbf{a}_i$ and

$$b_j = \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{otherwise} \end{cases}$$

this condition is satisfied. Thus, there exists a point $\mathbf{a}_i$ that is a minimizer for f on S.

Theorem

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a convex, differentiable function. Any critical point $\mathbf{x}_0$ of f is a global minimum for f.

Proof

Let $\mathbf{x}_0$ be a critical point for f. Suppose that $\mathbf{x}_0$ is not a global minimum for f. Then, there exists $\mathbf{z}$ such that $f(\mathbf{z}) < f(\mathbf{x}_0)$. Since f is *differentiable* in $\mathbf{x}_0$, we have

$$
\begin{aligned}
(\nabla f)'_{\mathbf{x}_0}(\mathbf{z}-\mathbf{x}_0) &= \frac{d}{dt} f(\mathbf{x}_0 + t(\mathbf{z}-\mathbf{x}_0))_{t=0} \\
&= \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t(\mathbf{z}-\mathbf{x}_0)) - f(\mathbf{x}_0)}{t} \\
&= \lim_{t\to 0} \frac{f(t\mathbf{z} + (1-t)\mathbf{x}_0))) - f(\mathbf{x}_0)}{t} \\
&\leqslant \frac{tf(\mathbf{z}) + (1-t)f(\mathbf{x}_0) - f(\mathbf{x}_0)}{t} \\
&= \frac{t(f(\mathbf{z}) - tf(\mathbf{x}_0))}{t} < 0,
\end{aligned}
$$

which implies $(\nabla f)_{\mathbf{x}_0} \neq \mathbf{0}_n$, thus contradicting the fact that $\mathbf{x}_0$ is a critical point.