Convex Functions

Prof. Dan A. Simovici

UMB

Definition

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

$$f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leqslant tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y})$$

for every $\boldsymbol{x}, \boldsymbol{y} \in S$ and $t \in [0,1]$.
If $f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) < tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y})$ for every $\boldsymbol{x}, \boldsymbol{y} \in S$ and $t \in (0,1)$ then f is said to be strictly convex.
The function $g : S \longrightarrow \mathbb{R}$ is *concave* if $-g$ is convex.

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

$$\begin{aligned} f(tx_1 + (1-t)x_2) &= (tx_1 + (1-t)x_2)^2 \\ &= t^2x_1^2 + (1-t)^2x_2^2 + 2(1-t)tx_1x_2. \end{aligned}$$

Therefore,
$f((1-t)x_1 + tx_2) - tf(x_1) - (1-t)f(x_2) = -(1-t)t(x_1+x_2)^2 \leqslant 0$,
which implies that f is indeed convex.

Example

Any norm ν on $\mathbb{R}^n$ is convex. Indeed, for $t \in (0, 1)$ we have

$$\nu(t\mathbf{x} + (1-t)\mathbf{y}) \leqslant \nu(t\mathbf{x}) + \nu((1-t)\mathbf{y}) = t\nu(\mathbf{x}) + (1-t)\nu(\mathbf{y})$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Example

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. The function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by $f(\boldsymbol{x}) = \boldsymbol{x}' A \boldsymbol{x}$ is convex if and only if A is a positive semidefinite matrix. Indeed, suppose that f is convex. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have

$$(t\boldsymbol{x} + (1-t)\boldsymbol{y})' A (t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leqslant t\boldsymbol{x}' A \boldsymbol{x} + (1-t)\boldsymbol{y}' A \boldsymbol{y},$$

for $t \in (0,1)$, which amounts to

$$(t^2 - t)\boldsymbol{x}' A \boldsymbol{x} + ((1-t)^2 - (1-t))\boldsymbol{y}' A \boldsymbol{y} + (1-t)t\boldsymbol{y}' A \boldsymbol{x} + t(1-t)\boldsymbol{x}' A \boldsymbol{y} \leqslant 0.$$

Since A is symmetric, we have $(\boldsymbol{y}' A \boldsymbol{x})' = \boldsymbol{x}' A \boldsymbol{y}$ and because both terms of the last equality are scalars we have $\boldsymbol{y}' A \boldsymbol{x} = \boldsymbol{x}' A \boldsymbol{y}$. Note that $t^2 - t \leqslant 0$ because $t \in [0,1]$. Consequently,

$$\boldsymbol{x}' A \boldsymbol{x} + \boldsymbol{y}' A \boldsymbol{y} + \boldsymbol{y}' A \boldsymbol{x} + \boldsymbol{x}' A \boldsymbol{y} \geqslant 0,$$

which amounts to $(\boldsymbol{x} + \boldsymbol{y})' A (\boldsymbol{x} + \boldsymbol{y}) \geqslant 0$, so A is positive semidefinite.

Extending the definition of convex functions

Let $f : S \longrightarrow \mathbb{R}$ be a convex function, where S is a convex subset of $\mathbb{R}^n$. As a notational convenience, define the function $\hat{f} : \mathbb{R}^n \longrightarrow \hat{\mathbb{R}}$ as

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

Then, f is convex if and only if $\hat{f}$ is convex, that is, it satisfies the inequality $\hat{f}(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leqslant t\hat{f}(\boldsymbol{x}) + (1-t)\hat{f}(\boldsymbol{y})$ for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We extended the usual definition of real-number operations on $\mathbb{R}$ by $t\infty = \infty t = \infty$ for $t > 0$. If there is no risk of confusion we denote $\hat{f}$ simply by f.

The importance of convex functions for optimization problems stems from the fact that every local minimum of a strictly convex function f is the global minimum of f.
A local variant of the optimization problem is given next.

Definition

Let S be a convex subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}^n$ be a function. The *local minimization problem* $\mathcal{M}(f, \boldsymbol{g}, \boldsymbol{x}_0, \delta)$ for f at $\boldsymbol{x}_0$ is:

minimize $f(\boldsymbol{x})$

where $\boldsymbol{x} \in S \cap B(\boldsymbol{x}_0, \delta)$.

Theorem

If $\boldsymbol{x}_0$ *is a solution of the minimization problem, where* S *is a convex set and* f *is convex at* $\boldsymbol{x}_0$*, then* $\boldsymbol{x}_0$ *is also a solution of the local minimization problem.*
If S *is convex and* f *is convex at* $\boldsymbol{x}_0$*, then a solution of the local minimization problem* $\mathcal{M}(f, \boldsymbol{g}, \boldsymbol{x}_0, \delta)$ *is a solution of the minimization problem.*

Proof

Suppose that $\boldsymbol{x}_0$ is a solution of $\mathcal{M}(f, \boldsymbol{g}, \boldsymbol{x}_0, \delta)$, S is convex, and f is convex at $\boldsymbol{x}_0$. Let $y \in S - \{\boldsymbol{x}_0\}$. Since S is convex, $t\boldsymbol{x}_0 + (1-t)\boldsymbol{y} \in S$ for $t \in [0,1)$. To have $t\boldsymbol{x}_0 + (1-t)\boldsymbol{y} \in B(\boldsymbol{x}_0, \delta)$ we need to have $\parallel \boldsymbol{x}_0 - t\boldsymbol{x}_0 - (1-t)\boldsymbol{y} \parallel < \delta$, or $(1-t) \parallel \boldsymbol{x}_0 - \boldsymbol{y} \parallel < \delta$, which is the case if $t > 1 - \frac{\delta}{\|\boldsymbol{x}_0 - \boldsymbol{y}\|}$. With this condition satisfied by t we have $t\boldsymbol{x}_0 + (1-t)\boldsymbol{y} \in B(\boldsymbol{x}_0, \delta) \cap S$. Therefore,

$$\begin{aligned} f(\boldsymbol{x}_0) &\leqslant f(t\boldsymbol{x}_0 + (1-t)\boldsymbol{y}) \\ &\quad \text{(because } \boldsymbol{x}_0 \text{ is a local minimum)} \\ &\leqslant tf(\boldsymbol{x}_0) + (1-t)f(\boldsymbol{y}) \\ &\quad \text{(because } f \text{ is convex at } \boldsymbol{x}_0\text{)}, \end{aligned}$$

so $f(\boldsymbol{x}_0) \leqslant f(\boldsymbol{y})$. Thus, $\boldsymbol{x}_0$ is a solution of the minimization problem. The converse implication is immediate.

Theorem

Let S be a convex subset of $\mathbb{R}^n$, $f : S \longrightarrow \mathbb{R}^n$, and $\boldsymbol{g} : S \longrightarrow \mathbb{R}^m$, where $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}(\boldsymbol{x}) \leqslant \boldsymbol{0}_m\}$. The set of solutions of the minimization problem

minimize $f(\boldsymbol{x})$

subjected to the condition $\boldsymbol{x} \in S$

is convex.

Proof

Let $\boldsymbol{x}, \boldsymbol{y}$ be solutions of the minimization problem. Since S is convex, $t\boldsymbol{x} + (1-t)\boldsymbol{y} \in S$. Then, the convexity of f implies

$$f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leqslant tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y}) = f(\boldsymbol{x}),$$

which implies $f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) = f(\boldsymbol{x})$, so $t\boldsymbol{x} + (1-t)\boldsymbol{y}$ is also a solution.

Corollary

Under the conditions of the theorem, if f is a strictly convex function, then the solution of the optimization problem is unique.

Suppose that $\boldsymbol{x}, \boldsymbol{z}$ are two solutions of the minimization problem. Since S is convex, $t \in (0, 1)$ implies $t\boldsymbol{x} + (1-t)\boldsymbol{z} \in S$ and the strict convexity of f further implies

$$f(t\boldsymbol{x} + (1-t)\boldsymbol{z}) < tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{z}) = f(\boldsymbol{x}),$$

which contradicts the fact that $\boldsymbol{x}$ is a solution.

Theorem

Let (a, b) be an open interval of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a differentiable function on (a, b). Then, f is convex 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 $f(y) \geqslant f(x) + f'(x)(y-x)$, which is desired inequality.

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.

Theorem

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

Proof

Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be the one-argument function defined by

$$g(t) = f(t\boldsymbol{y} + (1-t)\boldsymbol{x}).$$

We have $g'(t) = (\nabla f)_{(t\boldsymbol{y}+(1-t)\boldsymbol{x}}(\boldsymbol{y} - \boldsymbol{x})$. If f is convex, then g is convex and we have $g(1) \geqslant g(0) + g'(0)$, which implies

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + (\nabla f)_{\boldsymbol{x}}(\boldsymbol{y} - \boldsymbol{x}),$$

which is the inequality we need to prove.
Conversely, suppose that for the inequality $f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + (\nabla f)'_{\boldsymbol{x}}(\boldsymbol{y} - \boldsymbol{x})$ holds for every $\boldsymbol{x}, \boldsymbol{y} \in S$. If $(1-t)\boldsymbol{x} + t\boldsymbol{y}$ and $(1-s)\boldsymbol{x} + s\boldsymbol{y}$ belong to S, then

$$f((1-t)\boldsymbol{x} + t\boldsymbol{y}) \geqslant f((1-s)\boldsymbol{x} + s\boldsymbol{y}) + (\nabla f)'_{(1-s)\boldsymbol{x}+s\boldsymbol{y}}(\boldsymbol{y} - \boldsymbol{x})(t-s),$$

so $g(t) \geqslant g(s) + g'(s)(t-s)$, so g is convex. The convexity of f follows immediately.

For functions that are twice continuously differentiable on a convex subset S of $\mathbb{R}^n$ with a non-empty interior we have the following statement:

Theorem

Let S be a convex subset of $\mathbb{R}^n$ with a non-empty interior. If $f : S \longrightarrow \mathbb{R}$ is a function in $C^2(S)$, then, f is convex on S if and only if the Hessian matrix $H_f(\boldsymbol{x})$ is positive semidefinite for every $\boldsymbol{x} \in S$.

Proof

Suppose that the Hessian matrix $H_f(\boldsymbol{x})$ is positive semidefinite for every $\boldsymbol{x} \in S$. By Taylor's theorem,

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_0) = (\nabla f)_{\boldsymbol{x}_0}(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)' H_f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))(\boldsymbol{x} - \boldsymbol{x}_0)$$

for some $t \in [0,1]$. The positive semidefiniteness of H_f means that $\frac{1}{2}(bfx - \boldsymbol{x}_0)' H_f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))(\boldsymbol{x} - \boldsymbol{x}_0) \geqslant 0$, so $f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}_0) + (\nabla f)_{\boldsymbol{x}_0}(\boldsymbol{x} - \boldsymbol{x}_0)$, which implies the convexity of f.

Proof (cont'd)

Suppose now that $H_f(\boldsymbol{x}_0)$ is not positive semidefinite at some $\boldsymbol{x}_0 \in S$. We may assume that $\boldsymbol{x}_0$ is an interior point of S since H_f is continuous. There exists $\boldsymbol{x} \in S$ such that $(\boldsymbol{x} - \boldsymbol{x}_0)' H_f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) < 0$. Applying again the continuity of the Hessian matrix, $\boldsymbol{x}$ may be selected such that $(\boldsymbol{x} - \boldsymbol{x}_0)' H_f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))(\boldsymbol{x} - \boldsymbol{x}_0) < 0$, which means that $f(\boldsymbol{x}) < f(\boldsymbol{x}_0) + f(\boldsymbol{x}_0) + (\nabla f)_{\boldsymbol{x}_0}(\boldsymbol{x} - \boldsymbol{x}_0)$, thus contradicting the convexity of f.