Optimization Preliminaries

K.Bouanane

March 15, 2023

Introduction

Optimization Problem Setting

Consider the following problem

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a differentiable function.

Optimization Problem Setting

Consider the following problem

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a differentiable function.
Our aim is to find a minimum for the function f in the set $\mathcal{X}$.
This is called an optimization problem.

Different types of optimization problems

An optimization problem may be:

Different types of optimization problems

An optimization problem may be:

- Continuous ($\mathbf{x} \in \mathbb{R}^d$) *vs* Discrete (Combinatorial: $\mathbf{x} \in \mathbb{Z}^d$),

Different types of optimization problems

An optimization problem may be:

- Continuous ($\mathbf{x} \in \mathbb{R}^d$) *vs* Discrete (Combinatorial: $\mathbf{x} \in \mathbb{Z}^d$),
- Nonlinear ($f(\mathbf{x})$ is a nonlinear function) *vs* Linear ($f(\mathbf{x})$ is a linear function and $\mathcal{X}$ is a polytope),

Different types of optimization problems

An optimization problem may be:

- Continuous ($\mathbf{x} \in \mathbb{R}^d$) *vs* Discrete (Combinatorial: $\mathbf{x} \in \mathbb{Z}^d$),
- Nonlinear ($f(\mathbf{x})$ is a nonlinear function) *vs* Linear ($f(\mathbf{x})$ is a linear function and $\mathcal{X}$ is a polytope),
- Deterministic (all problem parameters are assumed to be exactly known) *vs* stochastic (some parameters are random variables),

Different types of optimization problems

An optimization problem may be:

- Continuous ($\mathbf{x} \in \mathbb{R}^d$) *vs* Discrete (Combinatorial: $\mathbf{x} \in \mathbb{Z}^d$),
- Nonlinear ($f(\mathbf{x})$ is a nonlinear function) *vs* Linear ($f(\mathbf{x})$ is a linear function and $\mathcal{X}$ is a polytope),
- Deterministic (all problem parameters are assumed to be exactly known) *vs* stochastic (some parameters are random variables),
- Mono-objective ($f(\mathbf{x}) \in \mathbb{R}$) *vs* multi-objective ($f(\mathbf{x}) \in \mathbb{R}^m$),

Different types of optimization problems

An optimization problem may be:

- Continuous ($\mathbf{x} \in \mathbb{R}^d$) *vs* Discrete (Combinatorial: $\mathbf{x} \in \mathbb{Z}^d$),
- Nonlinear ($f(\mathbf{x})$ is a nonlinear function) *vs* Linear ($f(\mathbf{x})$ is a linear function and $\mathcal{X}$ is a polytope),
- Deterministic (all problem parameters are assumed to be exactly known) *vs* stochastic (some parameters are random variables),
- Mono-objective ($f(\mathbf{x}) \in \mathbb{R}$) *vs* multi-objective ($f(\mathbf{x}) \in \mathbb{R}^m$),
- Convex ($f(\mathbf{x})$ is a convex function and $\mathcal{X}$ is a convex set) *vs* Non-convex ($f(\mathbf{x})$ is a non-convex function or $\mathcal{X}$ is a non-convex set).

The set $\mathcal{X} \subseteq \mathbb{R}^d$ is called the **feasible set**.

The set $\mathcal{X} \subseteq \mathbb{R}^d$ is called the **feasible set.**

If $\mathcal{X}$ is the entire domain of the function being optimized (as it often will be for our purposes), we say that the problem is **unconstrained**.

The set $\mathcal{X} \subseteq \mathbb{R}^d$ is called the **feasible set.**
If $\mathcal{X}$ is the entire domain of the function being optimized (as it often will be for our purposes), we say that the problem is **unconstrained**.
Otherwise, the problem is **constrained** and may be much harder to solve, depending on the nature of the feasible set.

The set $\mathcal{X} \subseteq \mathbb{R}^d$ is called the **feasible set.**

If $\mathcal{X}$ is the entire domain of the function being optimized (as it often will be for our purposes), we say that the problem is **unconstrained**.

Otherwise, the problem is **constrained** and may be much harder to solve, depending on the nature of the feasible set.

Notice that maximizing a function f is equivalent to minimizing $-f$.

The set $\mathcal{X} \subseteq \mathbb{R}^d$ is called the **feasible set**.

If $\mathcal{X}$ is the entire domain of the function being optimized (as it often will be for our purposes), we say that the problem is **unconstrained**.

Otherwise, the problem is **constrained** and may be much harder to solve, depending on the nature of the feasible set.

Notice that maximizing a function f is equivalent to minimizing $-f$.

Thus, the formulation given above remains valid for any optimization problem.

Local and global minima

Suppose $f : \mathbb{R}^d \to \mathbb{R}$.

Local and global minima

Suppose $f : \mathbb{R}^d \to \mathbb{R}$.
A point $\mathbf{x}$ is said to be a **local minimum**(resp. **local maximum**) of f in $\mathcal{X}$, if $f(\mathbf{x}) \leq f(\mathbf{y})$ (resp. $f(\mathbf{x}) \geq f(\mathbf{y})$) for all $\mathbf{y}$ in some neighborhood $N \subseteq \mathcal{X}$ about $\mathbf{x}$.

Local and global minima

Suppose $f : \mathbb{R}^d \to \mathbb{R}$.
A point $\mathbf{x}$ is said to be a **local minimum**(resp. **local maximum**) of f in $\mathcal{X}$, if $f(\mathbf{x}) \leq f(\mathbf{y})$ (resp. $f(\mathbf{x}) \geq f(\mathbf{y})$) for all $\mathbf{y}$ in some neighborhood $N \subseteq \mathcal{X}$ about $\mathbf{x}$.
Furthermore, if $f(\mathbf{x}) \leq f(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{X}$, then $\mathbf{x}$ is a **global minimum** of f in $\mathcal{X}$ (similarly for global maximum).

Conditions for local minima

First order necessary condition

First order necessary condition

The following proposition gives a **necessary condition** for existence of local minima of a function f.

First order necessary condition

The following proposition gives a **necessary condition** for existence of local minima of a function f.

Proposition

If $\mathbf{x}^*$ is a local minimum of f and f is continuously differentiable in a neighborhood of $\mathbf{x}^*$, then $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Proof

Proof

Let $\mathbf{x}^*$ be a local minimum of f, and suppose that $\nabla f(\mathbf{x}^*) \neq 0$.

Proof

Let $\mathbf{x}^*$ be a local minimum of f, and suppose that $\nabla f(\mathbf{x}^*) \neq 0$. Let $\mathbf{h} = -\nabla f(\mathbf{x}^*)$. Because of the continuity of f, we have

Proof

Let $\mathbf{x}^*$ be a local minimum of f, and suppose that $\nabla f(\mathbf{x}^*) \neq 0$. Let $\mathbf{h} = -\nabla f(\mathbf{x}^*)$. Because of the continuity of f, we have

$$\lim_{\alpha \to 0} -\nabla f(\mathbf{x}^* + \alpha \mathbf{h}) = -\nabla f(\mathbf{x}^*) = \mathbf{h}$$

Proof

Let $\mathbf{x}^*$ be a local minimum of f, and suppose that $\nabla f(\mathbf{x}^*) \neq 0$. Let $\mathbf{h} = -\nabla f(\mathbf{x}^*)$. Because of the continuity of f, we have

$$\lim_{\alpha \to 0} -\nabla f(\mathbf{x}^* + \alpha \mathbf{h}) = -\nabla f(\mathbf{x}^*) = \mathbf{h}$$

Hence

Proof

Let $\mathbf{x}^*$ be a local minimum of f, and suppose that $\nabla f(\mathbf{x}^*) \neq 0$. Let $\mathbf{h} = -\nabla f(\mathbf{x}^*)$. Because of the continuity of f, we have

$$\lim_{\alpha \to 0} -\nabla f(\mathbf{x}^* + \alpha \mathbf{h}) = -\nabla f(\mathbf{x}^*) = \mathbf{h}$$

Hence

$$\lim_{\alpha \to 0} \mathbf{h}^\top \nabla f(\mathbf{x}^* + \alpha \mathbf{h}) = \mathbf{h}^\top \nabla f(\mathbf{x}^*) = -\|\mathbf{h}\|^2 < 0$$

Proof

Let $\mathbf{x}^*$ be a local minimum of f, and suppose that $\nabla f(\mathbf{x}^*) \neq 0$. Let $\mathbf{h} = -\nabla f(\mathbf{x}^*)$. Because of the continuity of f, we have

$$\lim_{\alpha \to 0} -\nabla f(\mathbf{x}^* + \alpha \mathbf{h}) = -\nabla f(\mathbf{x}^*) = \mathbf{h}$$

Hence

$$\lim_{\alpha \to 0} \mathbf{h}^\top \nabla f(\mathbf{x}^* + \alpha \mathbf{h}) = \mathbf{h}^\top \nabla f(\mathbf{x}^*) = -\|\mathbf{h}\|^2 < 0$$

Thus there exists $\beta > 0$ such that $\mathbf{h}^\top \nabla f(\mathbf{x}^* + \alpha \mathbf{h}) < 0$ for all $\alpha \in [0, \beta]$.

Proof

Proof

By applying Taylor's theorem: for any $\alpha \in (0; \beta]$, there exists $t \in (0; \alpha)$ such that

Proof

By applying Taylor's theorem: for any $\alpha \in (0; \beta]$, there exists $t \in (0; \alpha)$ such that

$$f(\mathbf{x}^* + \alpha\mathbf{h}) = f(\mathbf{x}^*) + \alpha\mathbf{h}^\top \nabla f(\mathbf{x}^* + t\mathbf{h}) < f(\mathbf{x}^*)$$

Proof

By applying Taylor's theorem: for any $\alpha \in (0;\beta]$, there exists $t \in (0;\alpha)$ such that

$$f(\mathbf{x}^* + \alpha\mathbf{h}) = f(\mathbf{x}^*) + \alpha\mathbf{h}^\top \nabla f(\mathbf{x}^* + t\mathbf{h}) < f(\mathbf{x}^*)$$

whence it follows that x^* is not a local minimum, a contradiction. Hence $\nabla f(\mathbf{x}^*) = 0$.

The proof of this result shows us why the vanishing gradient is necessary for an extremum:

The proof of this result shows us why the vanishing gradient is necessary for an extremum:
If $\nabla f(\mathbf{x})$ is nonzero, there always exists a sufficiently small step $\alpha > 0$ such that $f(\mathbf{x} - \alpha \nabla f(\mathbf{x})) < f(\mathbf{x})$. For this reason, $-\nabla f(\mathbf{x})$ is called a **descent direction.**

The proof of this result shows us why the vanishing gradient is necessary for an extremum:
If $\nabla f(\mathbf{x})$ is nonzero, there always exists a sufficiently small step $\alpha > 0$ such that $f(\mathbf{x} - \alpha\nabla f(\mathbf{x})) < f(\mathbf{x})$. For this reason, $-\nabla f(\mathbf{x})$ is called a **descent direction.**
Points where the gradient vanishes are called **stationary points.**

Note that not all stationary points are extrema.

Note that not all stationary points are extrema.

For example, consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 - y^2$.

We have $f(\mathbf{0}) = \mathbf{0}$, but the point $\mathbf{0}$ is the minimum along the line $y = 0$ and the maximum along the line $x = 0$.

Note that not all stationary points are extrema.

For example, consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 - y^2$.

We have $f(\mathbf{0}) = \mathbf{0}$, but the point $\mathbf{0}$ is the minimum along the line $y = 0$ and the maximum along the line $x = 0$.

Thus it is neither a local minimum nor a local maximum of f.

Points such as these, where the gradient vanishes but there is no local extremum, are called **saddle points**.

Second order necessary condition

Second order necessary condition

A second-order necessary condition (i.e. the Hessian) for existence of local minima is given:

Proposition

If $\mathbf{x}^*$ is a local minimum of f and f is twice continuously differentiable in a neighborhood of $\mathbf{x}^*$, then $\nabla^2 f(\mathbf{x}^*)$ is positive semi-definite.

Sufficient Condition for local minima

Sufficient Condition for local minima

The sufficient condition for local minima is then given:

Sufficient Condition for local minima

The sufficient condition for local minima is then given:

Proposition

Suppose f is twice continuously differentiable with $\nabla^2 f$ positive semi-definite in a neighborhood of $\mathbf{x}^*$, and that $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Then $\mathbf{x}^*$ is a local minimum of f. Furthermore if $\nabla^2 f(\mathbf{x}^*)$ is positive definite, then $\mathbf{x}^*$ is a strict local minimum.

Proof

Let B be an open ball of radius $r > 0$ centered at $\mathbf{x}^*$ which is contained in the neighborhood. Applying Taylor's theorem, we have that for any $\mathbf{h}$ with $\|\mathbf{h}\|_2 < r$, there exists $t \in (0;1)$ such that

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \underbrace{\mathbf{h}^\top \nabla f(\mathbf{x}^*)}_{=0} + \frac{1}{2} \underbrace{\mathbf{h}^\top \nabla^2 f(\mathbf{x}^* + t\mathbf{h})\mathbf{h}}_{\geq 0} \geq f(\mathbf{x}^*)$$

Proof

Let B be an open ball of radius $r > 0$ centered at $\mathbf{x}^*$ which is contained in the neighborhood. Applying Taylor's theorem, we have that for any $\mathbf{h}$ with $\|\mathbf{h}\|_2 < r$, there exists $t \in (0;1)$ such that

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \underbrace{\mathbf{h}^\top \nabla f(\mathbf{x}^*)}_{=0} + \frac{1}{2} \underbrace{\mathbf{h}^\top \nabla^2 f(\mathbf{x}^* + t\mathbf{h})\mathbf{h}}_{\geq 0} \geq f(\mathbf{x}^*)$$

Therefore, $\mathbf{x}^*$ is a local minimum of f.

Proof

Let B be an open ball of radius $r > 0$ centered at $\mathbf{x}^*$ which is contained in the neighborhood. Applying Taylor's theorem, we have that for any $\mathbf{h}$ with $\|\mathbf{h}\|_2 < r$, there exists $t \in (0;1)$ such that

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \underbrace{\mathbf{h}^\top \nabla f(\mathbf{x}^*)}_{=0} + \frac{1}{2} \underbrace{\mathbf{h}^\top \nabla^2 f(\mathbf{x}^* + t\mathbf{h})\mathbf{h}}_{\geq 0} \geq f(\mathbf{x}^*)$$

Therefore, $\mathbf{x}^*$ is a local minimum of f.
Now, if $\nabla^2 f$ is definite positive then $\mathbf{h}^\top \nabla^2 f(\mathbf{x}^* + t\mathbf{h})\mathbf{h} > 0$ and thus $\mathbf{x}^*$ is a strict local minimum.

Note that, the conditions $f(\mathbf{x}^*) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}^*)$ positive semi-definite are not enough to guarantee a local minimum at $\mathbf{x}^*$:

Note that, the conditions $f(\mathbf{x}^*) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}^*)$ positive semi-definite are not enough to guarantee a local minimum at $\mathbf{x}^*$:

Consider the function $f(x) = x^3$. We have $f'(0) = 0$ and $f''(0) = 0$ (so the Hessian, which in this case is the 1×1 matrix $[0]$, is positive semi-definite). But f has a saddle point at $x = 0$.

Note that, the conditions $f(\mathbf{x}^*) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}^*)$ positive semi-definite are not enough to guarantee a local minimum at $\mathbf{x}^*$:

Consider the function $f(x) = x^3$. We have $f'(0) = 0$ and $f''(0) = 0$ (so the Hessian, which in this case is the 1×1 matrix $[0]$, is positive semi-definite). But f has a saddle point at $x = 0$. On the other hand, the function $f(x) = x^4$ has the same gradient and Hessian at $x = 0$, but $x = 0$ is a strict local maximum for this function.

For these reasons we require that the Hessian remains positive semi-definite as long as we are close to $\mathbf{x}^*$.
This condition is not practical to check computationally, but in some cases we can verify it analytically (usually by showing that $\nabla^2 f(\mathbf{x})$ is p.s.d. for all $\mathbf{x} \in \mathbb{R}^d$).

Convexity

Introduction

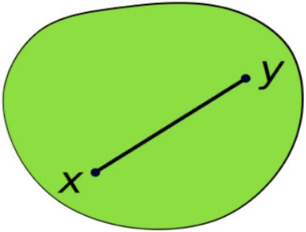
Convexity is a term that pertains to both sets and functions.

Convex sets

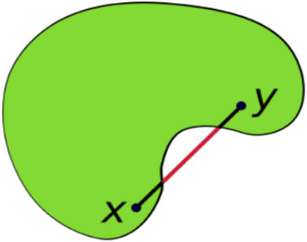
A set $\mathcal{X} \subseteq \mathbb{R}^d$ is convex if for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and all $t \in [0, 1]$

$$t\mathbf{x} + (1 - t)\mathbf{y} \in \mathcal{X}$$

Geometrically, this means that all the points on the line segment between any two points in $\mathcal{X}$ are also in $\mathcal{X}$.



(a) A convex set



(b) A non-convex set

Convex functions

For functions, there are different degrees of convexity, and how convex a function is, tells us a lot about its minima: do they exist, are they unique, how quickly can we find them using optimization algorithms, etc.

A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and all $t \in [0,1]$.
If the inequality above holds strictly for all $t \in [0,1]$ and $\mathbf{x} \neq \mathbf{y}$, then we say that f is said **strictly convex**.

A function f is **strongly convex with parameter** m (or m**-strongly convex**) if the function

$$\mathbf{x} \mapsto f(\mathbf{x}) - \frac{m}{2}\|\mathbf{x}\|_2^2$$

is convex. i.e :

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}) + \frac{m}{2}t(1-t)\|\mathbf{x} - \mathbf{y}\|_2^2$$

for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and all $t \in [0, 1]$.
These conditions are given in increasing order of strength;

A function f is **strongly convex with parameter** m (or m**-strongly convex**) if the function

$$\mathbf{x} \mapsto f(\mathbf{x}) - \frac{m}{2}\|\mathbf{x}\|_2^2$$

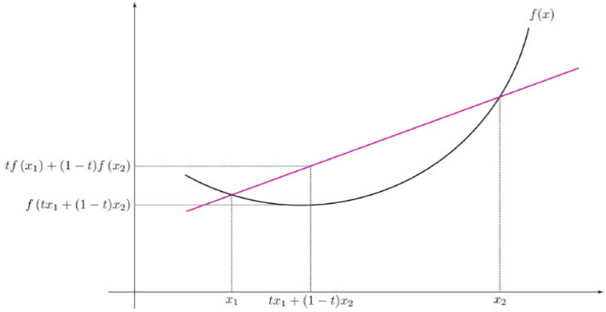
is convex. i.e :

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}) + \frac{m}{2}t(1-t)\|\mathbf{x}-\mathbf{y}\|_2^2$$

for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and all $t \in [0, 1]$.
These conditions are given in increasing order of strength; strong convexity implies strict convexity which implies convexity.

Geometrically, convexity means that the line segment between two points on the graph of f lies on or above the graph itself. Strict convexity means that the line segment lies strictly above the graph of f, except at the segment endpoints.

Geometrically, convexity means that the line segment between two points on the graph of f lies on or above the graph itself. Strict convexity means that the line segment lies strictly above the graph of f, except at the segment endpoints.



Characterization of convex, strictly convex and strongly convex functions

It is usually not easy to verify that the (strict, strong)-convexity of a given function from the definition. But the following theorems give us another way to do that.

Theorem

Suppose f is differentiable. Then

- *f is convex if and only if*

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle, \text{ for all } \mathbf{x}, \mathbf{y} \in \operatorname{dom} f$$

- *f is strictly convex if and only if*

$$f(\mathbf{x}) > f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle, \text{ for all } \mathbf{x}, \mathbf{y} \in \operatorname{dom} f, \mathbf{x} \neq \mathbf{y}$$

- *f is m-strongly convex if and only if*

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{m}{2} \|y - x\|_2^2$$

Theorem

Suppose f is twice differentiable. Then

- *f is convex if and only if $\nabla^2 f(\mathbf{x})$ is positive semi definite for all $\mathbf{x} \in \operatorname{dom} f$*
- *If $\nabla^2 f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in \operatorname{dom} f$ then f is strictly convex.*
- *f is m-strongly convex if and only if $\nabla^2 f(\mathbf{x}) - mI$ is positive semidefinite for all $\mathbf{x} \in \operatorname{dom} f$.*

Properties of convex functions

We can easily verify the following assertions

1. If f is convex and $\alpha \geq 0$, then αf is convex.

We can easily verify the following assertions

1. If f is convex and $\alpha \geq 0$, then αf is convex.
2. If f and g are convex, then $f + g$ is convex. Furthermore, if g is strictly convex, then $f + g$ is strictly convex, and if g is m-strongly convex, then $f + g$ is m-strongly convex.

We can easily verify the following assertions

1. If f is convex and $\alpha \geq 0$, then αf is convex.
2. If f and g are convex, then $f + g$ is convex. Furthermore, if g is strictly convex, then $f + g$ is strictly convex, and if g is m-strongly convex, then $f + g$ is m-strongly convex.
3. If $f_1, f_2, \ldots, f_k$ are convex and $\alpha_1, \alpha_2, \ldots, \alpha_k > 0$ then $f = \sum_{i=1}^{k} \alpha_i f_i$ is convex.

We can easily verify the following assertions

1. If f is convex and $\alpha \geq 0$, then αf is convex.
2. If f and g are convex, then $f + g$ is convex. Furthermore, if g is strictly convex, then $f + g$ is strictly convex, and if g is m-strongly convex, then $f + g$ is m-strongly convex.
3. If $f_1, f_2, \ldots, f_k$ are convex and $\alpha_1, \alpha_2, \ldots, \alpha_k > 0$ then $f = \sum_{i=1}^{k} \alpha_i f_i$ is convex.
4. If f is convex, then $g(\mathbf{x}) \equiv f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex for any appropriately sized $\mathbf{A}$ and $\mathbf{b}$.

We can easily verify the following assertions

1. If f is convex and $\alpha \geq 0$, then αf is convex.
2. If f and g are convex, then $f+g$ is convex. Furthermore, if g is strictly convex, then $f+g$ is strictly convex, and if g is m-strongly convex, then $f+g$ is m-strongly convex.
3. If $f_1, f_2, \ldots, f_k$ are convex and $\alpha_1, \alpha_2, \ldots, \alpha_k > 0$ then $f = \sum_{i=1}^{k} \alpha_i f_i$ is convex.
4. If f is convex, then $g(\mathbf{x}) \equiv f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex for any appropriately sized $\mathbf{A}$ and $\mathbf{b}$.
5. If f and g are convex, then $h(\mathbf{x}) \equiv \max\{f(\mathbf{x}); g(\mathbf{x})\}$ is convex.

Some examples

Functions that are convex but not strictly convex:

- $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + \alpha$ for any $\mathbf{w} \in \mathbb{R}^d, \alpha \in \mathbb{R}$. Such a function is called an **affine function**. Note that linear functions and constant functions are special cases of affine functions.
- $f(\mathbf{x}) = \|\mathbf{x}\|_1$.

Functions that are strictly but not strongly convex:

- $f(x) = x^4$.
- $f(x) = \exp(x)$.
- $f(x) = -\log x$.

Functions that are strongly convex:

- $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$

Convexity and existence of minima

Basically, various notions of convexity have implications about the nature of minima. It should not be surprising that the stronger conditions tell us more about the minima.

Basically, various notions of convexity have implications about the nature of minima. It should not be surprising that the stronger conditions tell us more about the minima.

Proposition

Let $\mathcal{X}$ be a convex set. If f is convex, then any local minimum of f in $\mathcal{X}$ is also a global minimum.

Proof

Suppose f is a convex function, and let $\mathbf{x}^*$ be a local minimum of f in $\mathcal{X}$. Then for some neighborhood $N(\mathbf{x}^*) \subset \mathcal{X}$, we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, for all $\mathbf{x} \in N(\mathbf{x}^*)$.

Proof

Suppose f is a convex function, and let $\mathbf{x}^*$ be a local minimum of f in $\mathcal{X}$. Then for some neighborhood $N(\mathbf{x}^*) \subset \mathcal{X}$, we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, for all $\mathbf{x} \in N(\mathbf{x}^*)$.

Suppose that there exists $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $f(\tilde{\mathbf{x}}) < f(\mathbf{x}^*)$.

Proof

Suppose f is a convex function, and let $\mathbf{x}^*$ be a local minimum of f in $\mathcal{X}$. Then for some neighborhood $N(\mathbf{x}^*) \subset \mathcal{X}$, we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, for all $\mathbf{x} \in N(\mathbf{x}^*)$.
Suppose that there exists $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $f(\tilde{\mathbf{x}}) < f(\mathbf{x}^*)$.
Consider the line segment $\mathbf{x}(t) = t\,\mathbf{x}^* + (1-t)\tilde{\mathbf{x}}$, $t \in (0,1)$.

Proof

Suppose f is a convex function, and let $\mathbf{x}^*$ be a local minimum of f in $\mathcal{X}$. Then for some neighborhood $N(\mathbf{x}^*) \subset \mathcal{X}$, we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, for all $\mathbf{x} \in N(\mathbf{x}^*)$.
Suppose that there exists $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $f(\tilde{\mathbf{x}}) < f(\mathbf{x}^*)$.
Consider the line segment $\mathbf{x}(t) = t\,\mathbf{x}^* + (1-t)\tilde{\mathbf{x}}$, $t \in (0,1)$.
Note that $\mathbf{x}(t) \in \mathcal{X}$ by the convexity of $\mathcal{X}$. Then by the convexity of f,

$$f(\mathbf{x}(t)) \leq tf(\mathbf{x}^*) + (1-t)f(\tilde{\mathbf{x}}) < tf(\mathbf{x}^*) + (1-t)f(\mathbf{x}^*) = f(\mathbf{x}^*)$$

for all $t \in (0,1)$.

Therefore, we have $f(\mathbf{x}(t)) < f(\mathbf{x}^*)$.

Therefore, we have $f(\mathbf{x}(t)) < f(\mathbf{x}^*)$.
But on the other hand, for t sufficiently close to 1, $\mathbf{x}(t) \in N(\mathbf{x}^*)$, and since $\mathbf{x}^*$ is a local minima, we have $f(\mathbf{x}(t)) \geq f(\mathbf{x}^*)$, by the definition of a neighborhood.

Therefore, we have $f(\mathbf{x}(t)) < f(\mathbf{x}^*)$.
But on the other hand, for t sufficiently close to 1, $\mathbf{x}(t) \in N(\mathbf{x}^*)$, and since $\mathbf{x}^*$ is a local minima, we have $f(\mathbf{x}(t)) \geq f(\mathbf{x}^*)$, by the definition of a neighborhood.
This is absurd, so it follows that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, thus $\mathbf{x}^*$ is a global minimum of f in $\mathcal{X}$.

Proposition

Let $\mathcal{X}$ be a convex set. If f is strictly convex, then there exists at most one local minimum of f in $\mathcal{X}$. Consequently, if it exists it is the unique global minimum of f in $\mathcal{X}$.