

Statistical Optimization: Lecture 4

Optimality Condition and Strict/Strong Convexity

Zijian Guo

Zhejiang University
Center for data science

March 23, 2026

Outline

Convex Optimization

Optimality Condition

Strict Convexity

Strong Convexity

Convex functions generate convex sets

A convex function naturally produces convex sets through its sublevel sets.

Proposition. If $f : \text{dom}(f) \rightarrow \mathbb{R}$ is convex, then for every $\alpha \in \mathbb{R}$,

$$C_\alpha := \{x \in \text{dom}(f) : f(x) \leq \alpha\}$$

is a convex set.

Proof. Let $x, y \in C_\alpha$ and $\lambda \in [0, 1]$. Since f is convex,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \leq \lambda\alpha + (1 - \lambda)\alpha = \alpha.$$

Hence

$$\lambda x + (1 - \lambda)y \in C_\alpha.$$

So C_α is convex. □

Operations on convex sets

Intersections. If $C_1, \dots, C_m \subseteq \mathbb{R}^d$ are convex, then

$$\bigcap_{i=1}^m C_i$$

is convex.

Proof. If $x, y \in \bigcap_{i=1}^m C_i$, then $x, y \in C_i$ for every i . Since each C_i is convex,

$$\lambda x + (1 - \lambda)y \in C_i \quad \forall i, \forall \lambda \in [0, 1].$$

Hence

$$\lambda x + (1 - \lambda)y \in \bigcap_{i=1}^m C_i.$$



From optimization to convex optimization

Consider the problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t. } \mathbf{x} \in \Omega,$$

where f is the objective function and Ω is the feasible set.

When $\Omega = \mathbb{R}^d$, the problem is **unconstrained**. When $\Omega \subsetneq \mathbb{R}^d$, the problem is **constrained**.

A standard convex program has the form

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \\ & A\mathbf{x} = \mathbf{b}, \end{aligned}$$

where f_0, \dots, f_m are convex and the equality constraints are affine.

Examples of convex programs

- Linear programming:

$$\min_{\mathbf{x}} \mathbf{c}^\top \mathbf{x} \quad \text{s.t. } A\mathbf{x} \leq \mathbf{b}.$$

- Quadratic programming with $Q \succeq 0$:

$$\min_{\mathbf{x}} \mathbf{x}^\top Q\mathbf{x} + \mathbf{c}^\top \mathbf{x} \quad \text{s.t. } A\mathbf{x} \leq \mathbf{b}.$$

- Regularized least squares:

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_2^2, \quad \lambda \geq 0.$$

- Lasso:

$$\min_{\mathbf{x}} \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad \lambda \geq 0.$$

Why convexity matters for optimization

Consider

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}),$$

where f is convex and Ω is convex.

Key fact. Every local minimum is also a global minimum.

A point \mathbf{x}^* is called a **stationary point** if

$$\nabla f(\mathbf{x}^*) = 0.$$

For a general nonconvex problem, a stationary point need not be a global minimum. But in convex optimization, first-order conditions do not merely identify stationary points – they characterize global optima.

Outline

Convex Optimization

Optimality Condition

Strict Convexity

Strong Convexity

Sufficient Condition (Unconstrained Case)

Theorem. Suppose that $f : \text{dom}(f) \rightarrow \mathbb{R}$ is convex and differentiable over an open domain $\text{dom}(f) \subseteq \mathbb{R}^d$. Let $\mathbf{x} \in \text{dom}(f)$. If $\nabla f(\mathbf{x}) = 0$, then \mathbf{x} is a global minimum.

Proof. Suppose that $\nabla f(\mathbf{x}) = 0$. For any $\mathbf{y} \in \text{dom}(f)$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}).$$

Since $\nabla f(\mathbf{x}) = 0$, this becomes

$$f(\mathbf{y}) \geq f(\mathbf{x}).$$

Hence \mathbf{x} is a global minimum. □

The converse is also true and does not even require convexity.

Necessary Condition (Unconstrained Case)

Theorem. Suppose that $f : \text{dom}(f) \rightarrow \mathbb{R}$ is differentiable over an open domain $\text{dom}(f) \subseteq \mathbb{R}^d$. Let $\mathbf{x} \in \text{dom}(f)$. If \mathbf{x} is a global minimum, then

$$\nabla f(\mathbf{x}) = 0.$$

Proof. Suppose that $\nabla f(\mathbf{x})_i \neq 0$ for some i . For $t \in \mathbb{R}$, we define

$$\mathbf{x}(t) = \mathbf{x} + t\mathbf{e}_i,$$

where \mathbf{e}_i is the i -th unit vector. For $|t|$ sufficiently small, we have $\mathbf{x}(t) \in \text{dom}(f)$ since $\text{dom}(f)$ is open. Let $z(t) = f(\mathbf{x}(t))$. By the chain rule,

$$z'(0) = \nabla f(\mathbf{x})^\top \mathbf{e}_i = \nabla f(\mathbf{x})_i \neq 0.$$

Hence, z decreases in one direction as we move away from 0, and this yields $f(\mathbf{x}(t)) < f(\mathbf{x})$ for some t , so \mathbf{x} is not a global minimum. \square

More general optimality condition

So far, we considered the unconstrained problem

$$\min_{\mathbf{x} \in \text{dom}(f)} f(\mathbf{x}),$$

where $\text{dom}(f) \subseteq \mathbb{R}^d$ is open.

More generally, we are often interested in minimizing f only over a subset $X \subseteq \text{dom}(f)$. In that case, the optimization problem takes the form

$$\arg \min \{f(\mathbf{x}) : \mathbf{x} \in X\}. \quad (2.10)$$

or equivalently

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in X. \end{aligned} \quad (2.11)$$

This leads to a more general first-order optimality condition.

Constrained minimization

Frequently, we are interested in minimizing a convex function only over a subset X of its domain.

Definition. Let $f : \text{dom}(f) \rightarrow \mathbb{R}$ be convex and let $X \subseteq \text{dom}(f)$ be a convex set. A point $\mathbf{x} \in X$ is a minimizer of f over X if

$$f(\mathbf{x}) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in X.$$

If f is differentiable, minimizers of f over X have a very useful characterization.

Theorem. Suppose that $f : \text{dom}(f) \rightarrow \mathbb{R}$ is convex and differentiable over an open domain $\text{dom}(f) \subseteq \mathbb{R}^d$, and let $X \subseteq \text{dom}(f)$ be a convex set. Then $\mathbf{x}^* \in X$ is a minimizer of f over X if and only if

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in X.$$

Proof

(\Rightarrow) Suppose that \mathbf{x}^* is a minimizer of f over X . Fix any $\mathbf{x} \in X$, and for $t \in [0, 1]$, define

$$\mathbf{x}(t) := \mathbf{x}^* + t(\mathbf{x} - \mathbf{x}^*).$$

Since X is convex, $\mathbf{x}(t) \in X$ for all $t \in [0, 1]$. Hence

$$f(\mathbf{x}(t)) \geq f(\mathbf{x}^*).$$

So the right derivative at $t = 0$ must be nonnegative:

$$\left. \frac{d}{dt} f(\mathbf{x}(t)) \right|_{t=0} = \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0.$$

Proof

(\Leftarrow) Now suppose that

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}.$$

Since f is convex and differentiable, its first-order characterization gives

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \quad \forall \mathbf{x} \in \mathcal{X}.$$

By the assumed inequality,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \forall \mathbf{x} \in \mathcal{X}.$$

Therefore, \mathbf{x}^* is a minimizer of f over \mathcal{X} . □

Constrained minimization

The theorem extends the first-order optimality condition from the unconstrained case to the constrained case.

If $X = \text{dom}(f)$, then the condition

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in X$$

reduces to

$$\nabla f(\mathbf{x}^*) = 0,$$

which is exactly the unconstrained optimality condition.

If X does not contain the global minimum, then theorem has a geometric interpretation: X is contained in the halfspace

$$\left\{ \mathbf{x} \in \mathbb{R}^d : \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \right\},$$

whose normal vector is $\nabla f(\mathbf{x}^*)$.

Constrained minimization

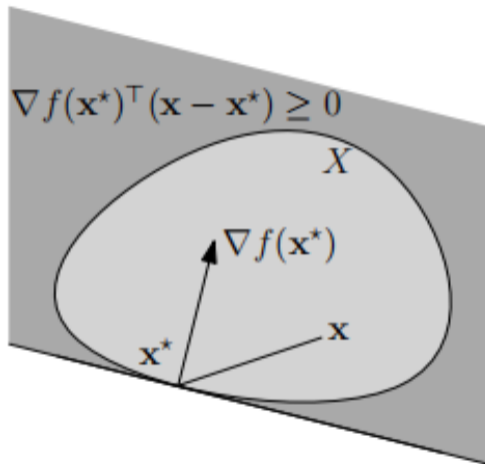


Figure: 2.8: Optimality condition for constrained optimization

Outline

Convex Optimization

Optimality Condition

Strict Convexity

Strong Convexity

Strict Convexity

Definition. A function $f : \text{dom}(f) \rightarrow \mathbb{R}$ is *strictly convex* if (i) $\text{dom}(f)$ is convex and (ii) for all $\mathbf{x} \neq \mathbf{y} \in \text{dom}(f)$ and all $\lambda \in (0, 1)$, we have

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}). \quad (2.8)$$

Theorem. Let $f : \text{dom}(f) \rightarrow \mathbb{R}$ be strictly convex. Then f has at most one global minimum.

Proof. Suppose $x^* \neq y^*$ are two global minima with $f_{\min} = f(x^*) = f(y^*)$, and let

$$z = \frac{1}{2}x^* + \frac{1}{2}y^*.$$

By (2.8),

$$f(z) < \frac{1}{2}f_{\min} + \frac{1}{2}f_{\min} = f_{\min},$$

a contradiction to x^* and y^* being global minima. □

Characterizations of convexity

Suppose that f is differentiable on a convex set $\Omega \subseteq \mathbb{R}^d$.

First-order characterization. The function f is convex on Ω if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Thus, the tangent hyperplane at any point lies below the graph.

Second-order characterization. If f is twice continuously differentiable, then f is convex on Ω if and only if

$$\nabla^2 f(\mathbf{x}) \succeq 0, \quad \forall \mathbf{x} \in \Omega.$$

Characterizations of strict convexity

Suppose that f is differentiable on a convex set $\Omega \subseteq \mathbb{R}^d$.

First-order characterization. The function f is strictly convex on Ω if and only if

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \Omega, \mathbf{x} \neq \mathbf{y}.$$

Thus, the tangent hyperplane at any point lies strictly below the graph at every other point.

Second-order sufficient condition. If f is twice continuously differentiable and

$$\nabla^2 f(\mathbf{x}) \succ 0, \quad \forall \mathbf{x} \in \Omega,$$

then f is strictly convex on Ω .

The converse is not always true.

Proof of “strict convexity \Rightarrow first-order inequality”

Suppose that f is strictly convex. Then f is convex, so by the first-order characterization of convexity,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

We now show that the inequality is in fact strict when $\mathbf{x} \neq \mathbf{y}$.

Assume for contradiction that for some $\mathbf{x} \neq \mathbf{y}$,

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}).$$

Let

$$\mathbf{z} = (1 - \lambda)\mathbf{x} + \lambda\mathbf{y}, \quad \lambda \in (0, 1).$$

Since f is strictly convex,

$$f(\mathbf{z}) < (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}) = f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}).$$

Proof of “strict convexity \Rightarrow first-order inequality”

On the other hand, since f is convex,

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{z} - \mathbf{x}) = f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}),$$

which is a contradiction. Therefore,

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x} \neq \mathbf{y}.$$

Proof of “first-order inequality \Rightarrow strict convexity”

Now assume that

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x} \neq \mathbf{y}.$$

Let $\mathbf{x} \neq \mathbf{y}$ and $\lambda \in (0, 1)$, and define

$$\mathbf{z} = \lambda \mathbf{x} + (1 - \lambda) \mathbf{y}.$$

Then $\mathbf{z} \neq \mathbf{x}$ and $\mathbf{z} \neq \mathbf{y}$. Applying the assumed inequality to (\mathbf{z}, \mathbf{x}) and (\mathbf{z}, \mathbf{y}) , we get

$$f(\mathbf{x}) > f(\mathbf{z}) + \nabla f(\mathbf{z})^\top (\mathbf{x} - \mathbf{z}),$$

$$f(\mathbf{y}) > f(\mathbf{z}) + \nabla f(\mathbf{z})^\top (\mathbf{y} - \mathbf{z}).$$

Proof of “first-order inequality \Rightarrow strict convexity”

Multiply the first inequality by λ , the second by $1 - \lambda$, and add:

$$\lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) > f(\mathbf{z}) + \nabla f(\mathbf{z})^\top (\lambda(\mathbf{x} - \mathbf{z}) + (1 - \lambda)(\mathbf{y} - \mathbf{z})).$$

Since

$$\mathbf{z} = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y},$$

we have

$$\lambda(\mathbf{x} - \mathbf{z}) + (1 - \lambda)(\mathbf{y} - \mathbf{z}) = \mathbf{0}.$$

Therefore,

$$f(\mathbf{z}) < \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}),$$

that is,

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

So f is strictly convex on Ω .



Proof of “second-order inequality \Rightarrow strict convexity”

Assume that f is twice continuously differentiable and

$$\nabla^2 f(\mathbf{x}) \succ 0 \quad \forall \mathbf{x} \in \Omega.$$

Fix $\mathbf{x} \neq \mathbf{y}$, and define

$$h(t) := f((1-t)\mathbf{x} + t\mathbf{y}), \quad t \in [0, 1].$$

Since Ω is convex, $(1-t)\mathbf{x} + t\mathbf{y} \in \Omega$ for all $t \in [0, 1]$.

By the chain rule,

$$h''(t) = (\mathbf{y} - \mathbf{x})^\top \nabla^2 f((1-t)\mathbf{x} + t\mathbf{y}) (\mathbf{y} - \mathbf{x}).$$

Proof of “second-order inequality \Rightarrow strict convexity”

Because $\mathbf{y} - \mathbf{x} \neq 0$ and $\nabla^2 f \succ 0$,

$$h''(t) > 0 \quad \forall t \in [0, 1].$$

Hence h is strictly convex on $[0, 1]$, so for every $\lambda \in (0, 1)$,

$$h(\lambda) < (1 - \lambda)h(0) + \lambda h(1).$$

Substituting back,

$$f((1 - \lambda)\mathbf{x} + \lambda\mathbf{y}) < (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}).$$

Therefore, f is strictly convex on Ω . □

Outline

Convex Optimization

Optimality Condition

Strict Convexity

Strong Convexity

Strong Convexity

Definition. Let $f: \mathbb{R}^d \rightarrow \mathbb{R}$ be a function. We say that f is μ -**strongly convex** (with $\mu > 0$) if for all $x, y \in \mathbb{R}^d$ and all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{\mu}{2}\lambda(1 - \lambda)\|x - y\|^2.$$

Compared with ordinary convexity, strong convexity adds the extra quadratic term

$$\frac{\mu}{2}\lambda(1 - \lambda)\|x - y\|^2,$$

which means that f lies below the chord by a uniform positive margin.

Another Definition of Strong Convexity

Define

$$g(\mathbf{x}) := f(\mathbf{x}) - \frac{\mu}{2} \|\mathbf{x}\|^2.$$

Then

$$f \text{ is } \mu\text{-strongly convex on } \Omega \iff g \text{ is convex on } \Omega.$$

Indeed, by definition, g is convex if and only if for all $\mathbf{x}, \mathbf{y} \in \Omega$ and $\lambda \in [0, 1]$,

$$g(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda g(\mathbf{x}) + (1 - \lambda)g(\mathbf{y}).$$

Substituting $g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2} \|\mathbf{x}\|^2$ gives

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) - \frac{\mu}{2} \|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\|^2 \leq \lambda \left(f(\mathbf{x}) - \frac{\mu}{2} \|\mathbf{x}\|^2 \right) + (1 - \lambda) \left(f(\mathbf{y}) - \frac{\mu}{2} \|\mathbf{y}\|^2 \right),$$

which is exactly the definition of μ -strong convexity.

Another Definition of Strong Convexity

To make the equivalence explicit, rearrange the previous inequality as

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\mu}{2} \left(\lambda \|\mathbf{x}\|^2 + (1 - \lambda)\|\mathbf{y}\|^2 - \|\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}\|^2 \right).$$

Now use the identity

$$\|\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}\|^2 = \lambda^2 \|\mathbf{x}\|^2 + (1 - \lambda)^2 \|\mathbf{y}\|^2 + 2\lambda(1 - \lambda)\mathbf{x}^\top \mathbf{y},$$

so

$$\begin{aligned} \lambda \|\mathbf{x}\|^2 + (1 - \lambda)\|\mathbf{y}\|^2 - \|\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}\|^2 &= \lambda(1 - \lambda)\|\mathbf{x}\|^2 + \lambda(1 - \lambda)\|\mathbf{y}\|^2 - 2\lambda(1 - \lambda)\mathbf{x}^\top \mathbf{y} \\ &= \lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Hence

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\mu}{2} \lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

which is exactly the definition of μ -strong convexity.

Characterizations of strong convexity

Suppose that f is differentiable on a convex set $\Omega \subseteq \mathbb{R}^d$.

First-order characterization. The function f is μ -strongly convex on Ω if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Thus, strong convexity strengthens convexity by adding a uniform quadratic lower bound.

Second-order characterization. If f is twice continuously differentiable, then f is μ -strongly convex on Ω if and only if

$$\nabla^2 f(\mathbf{x}) \succeq \mu I, \quad \forall \mathbf{x} \in \Omega.$$

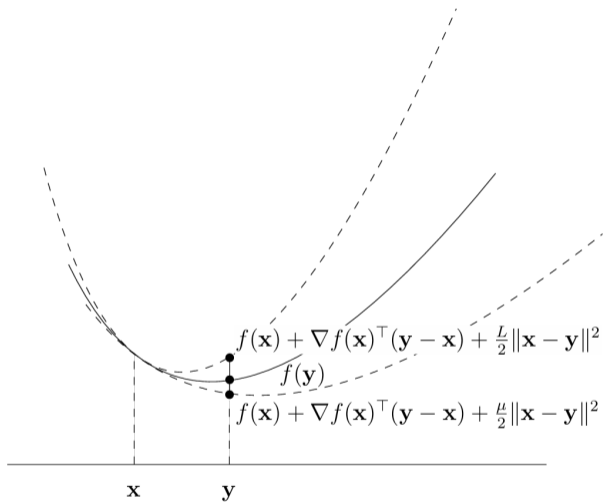


Figure: Strong convexity

Proof of first-order characterization of strong convexity

By the first-order characterization of convexity, $g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$ is convex if and only if

$$g(\mathbf{y}) \geq g(\mathbf{x}) + \nabla g(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Now

$$\nabla g(\mathbf{x}) = \nabla f(\mathbf{x}) - \mu\mathbf{x},$$

so

$$f(\mathbf{y}) - \frac{\mu}{2}\|\mathbf{y}\|^2 \geq f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2 + (\nabla f(\mathbf{x}) - \mu\mathbf{x})^\top (\mathbf{y} - \mathbf{x}).$$

Rearranging gives

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{\mu}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

Hence the first-order characterization holds. □

Another Definition of first-order characterization

More intuitively, strong convexity can also be written in the following equivalent gradient form (Gradient monotonicity): for all $x, y \in \mathbb{R}^d$,

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \mu \|x - y\|^2.$$

This means that the gradient mapping ∇f is μ -**strongly monotone**.

Proof: Strong convexity \iff Gradient monotonicity

(1) Strong convexity \Rightarrow gradient monotonicity.

Assume f is μ -strongly convex, i.e. for all x, y ,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2} \|y - x\|^2.$$

Apply this once with (x, y) and once with (y, x) , then add:

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2} \|y - x\|^2,$$

$$f(x) \geq f(y) + \nabla f(y)^\top (x - y) + \frac{\mu}{2} \|x - y\|^2.$$

Subtract $f(x) + f(y)$ and use $x - y = -(y - x)$:

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \mu \|x - y\|^2,$$

so ∇f is μ -strongly monotone.

Proof: Strong convexity \iff Gradient monotonicity

(2) Gradient monotonicity \Rightarrow strong convexity.

$$\begin{aligned} f(x) - f(y) - \nabla f(y)^\top (x - y) &= \int_0^1 \frac{d}{dt} f(t(x - y) + y) dt - \nabla f(y)^\top (x - y) \\ &= \int_0^1 \nabla f(t(x - y) + y)^\top (x - y) dt - \nabla f(y)^\top (x - y) \\ &= \int_0^1 (\nabla f(t(x - y) + y) - \nabla f(y))^\top (x - y) dt \\ &\geq \int_0^1 \mu \|t(x - y)\|^2 \frac{1}{t} dt \\ &= \frac{\mu}{2} \|x - y\|^2. \end{aligned}$$

so f is μ -strongly convex. □

Proof of second-order characterization of strong convexity

Assume now that f is twice continuously differentiable, and with

$$g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2} \|\mathbf{x}\|^2.$$

Then

$$\nabla^2 g(\mathbf{x}) = \nabla^2 f(\mathbf{x}) - \mu I.$$

By the second-order characterization of convexity, g is convex on Ω if and only if

$$\nabla^2 g(\mathbf{x}) \succeq 0 \quad \forall \mathbf{x} \in \Omega.$$

Substituting the expression for $\nabla^2 g(\mathbf{x})$, this is equivalent to

$$\nabla^2 f(\mathbf{x}) - \mu I \succeq 0 \quad \forall \mathbf{x} \in \Omega,$$

that is,

$$\nabla^2 f(\mathbf{x}) \succeq \mu I \quad \forall \mathbf{x} \in \Omega.$$

Relation between strong, strict, and convexity

We have the implication chain

$$\text{strong convexity} \implies \text{strict convexity} \implies \text{convexity}.$$

Neither converse is true.

Examples:

- $f(x) = x$ is convex, but not strictly convex;
- $f(x) = x^4$ is strictly convex on \mathbb{R} , but not strongly convex on \mathbb{R} .

Thus, strict convexity rules out flat line segments, while strong convexity imposes a uniform positive curvature condition.

Example: $f(x) = x^4$ is strictly convex

Consider

$$f(x) = x^4, \quad x \in \mathbb{R}.$$

We prove strict convexity using the first-order characterization. Since

$$f'(x) = 4x^3,$$

for any $x \neq y$,

$$f(y) - f(x) - f'(x)(y - x) = y^4 - x^4 - 4x^3(y - x).$$

Expanding and factorizing,

$$\begin{aligned} y^4 - x^4 - 4x^3(y - x) &= (y - x)(y^3 + y^2x + yx^2 + x^3) - 4x^3(y - x) \\ &= (y - x)(y^3 + y^2x + yx^2 - 3x^3) \\ &= (y - x)^2(y^2 + 2xy + 3x^2). \end{aligned}$$

Example: $f(x) = x^4$ is strictly convex

Now

$$y^2 + 2xy + 3x^2 = (y + x)^2 + 2x^2 > 0$$

whenever $x \neq y$. Hence

$$f(y) - f(x) - f'(x)(y - x) > 0 \quad \forall x \neq y.$$

Therefore,

$$f(y) > f(x) + f'(x)(y - x) \quad \forall x \neq y,$$

so $f(x) = x^4$ is strictly convex on \mathbb{R} . □

Example: $f(x) = x^4$ is not strongly convex

Since f is twice continuously differentiable,

$$f''(x) = 12x^2.$$

If f were μ -strongly convex for some $\mu > 0$, then by the second-order characterization of strong convexity, we would have

$$f''(x) \geq \mu \quad \forall x \in \mathbb{R}.$$

But at $x = 0$,

$$f''(0) = 0.$$

Therefore, there does not exist any $\mu > 0$ such that f is μ -strongly convex on \mathbb{R} . □

Quadratic functions

Let

$$f(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x} + \mathbf{b}^\top \mathbf{x} + \mathbf{c},$$

where $A = A^\top$.

Since

$$\nabla f(\mathbf{x}) = 2A\mathbf{x} + \mathbf{b}, \quad \nabla^2 f(\mathbf{x}) = 2A,$$

the curvature of f is completely determined by the matrix A .

In particular:

- f is convex if and only if $A \succeq 0$;
- f is strictly convex if and only if $A \succ 0$;
- f is μ -strongly convex if and only if

$$2A \succeq \mu I \quad \text{equivalently} \quad A \succeq \frac{\mu}{2} I.$$

Strict vs. strong convexity for quadratic functions

For a quadratic function

$$f(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c, \quad A = A^\top,$$

defined on \mathbb{R}^d , we have

$$f \text{ is strictly convex} \iff A \succ 0 \iff f \text{ is strongly convex.}$$

Indeed, since

$$\nabla^2 f(\mathbf{x}) = 2A,$$

strict convexity and strong convexity are both determined by the same matrix A .