# Optimization Midterm 2025 Autumn

November 23, 2025

## 1 Optimality Conditions

**1.** A useful property for symmetric definite positive matrices: Let $A \in S_d^{++}$ then we have the following inequality for all $z \in \mathbb{R}^d$:

$$\lambda_{\min}(A)\|z\|^2 \leq \langle Az, z\rangle \leq \lambda_{\max}(A)\|z\|^2$$

**Remark:** This is very useful when dealing with quadratic forms. Also note that this is very useful to prove the second order conditions for optimality. (i.e. if the Hessian $\nabla^2 f(x)$ is positive definite at a critical point $x^*$, then $x^*$ is a local minimum.)

**2. First Order Optimality Condition:** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. A necessary condition for $x^* \in \mathbb{R}^d$ to be a local minimum of $f$ is that the gradient at that point is zero:

$$\nabla f(x^*) = 0$$

**3. Second Order Optimality Condition:** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. Let $x^*$ be a critical point of $f$ (i.e. $\nabla f(x^*) = 0$). Then:

- If the Hessian $\nabla^2 f(x^*)$ is positive definite, then $x^*$ is a local minimum of $f$.
- If the Hessian $\nabla^2 f(x^*)$ has 2 eigenvalues of opposite signs, then $x^*$ is a saddle point of $f$.
- If the Hessian $\nabla^2 f(x^*) \in S_d^+$ but is not positive definite, then $x^*$ could be a local minimum or a saddle point (inconclusive).

**Remark**

**Some useful multicalculus rules:**

- **Product rule:** If $u, v : \mathbb{R}^d \to \mathbb{R}^d$ are differentiable functions, then:

$$D_x \langle u, v\rangle = \langle D_x u, v\rangle + \langle u, D_x v\rangle$$

- **Chain rule:** If $f : \mathbb{R}^m \to \mathbb{R}$ is a differentiable function and $g : \mathbb{R}^d \to \mathbb{R}^m$ is a differentiable function, then:

$$\underbrace{\frac{\partial f}{\partial x}}_{n\times 1} = \underbrace{\frac{\partial g}{\partial x}}_{n\times m} \cdot \underbrace{\frac{\partial f}{\partial g}}_{m\times 1}$$

## 2 Convexity

1. **Tangent plane property of convex functions:** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable convex ($\alpha$ strongly convex) function. Then for all $x, y \in \mathbb{R}^d$, we have:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

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

This means that the function lies above its tangent plane (plus a quadratic term for strong convexity) at any point.

2. **Monotonicity of the gradient** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable convex ($\alpha$ strongly convex) function. Then for all $x, y \in \mathbb{R}^d$, we have:

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \alpha\|y - x\|^2$$

This means that the gradient of a convex function is a monotone operator (stronger monotonicity than a quadratic function for strongly convex functions).

3. **Characterization of convexity via Hessian:** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. Then:

   - $f$ is convex if and only if its Hessian is positive semi-definite for all $x \in \mathbb{R}^d$:

   $$\nabla^2 f(x) \in S_d^+, \quad \forall x \in \mathbb{R}^d$$

   - $f$ is $\alpha$-strongly convex if and only if its Hessian's smallest eigenvalue is bounded below by $\alpha$ for all $x \in \mathbb{R}^d$:

   $$\lambda_d(\nabla^2 f(x)) \geq \alpha, \quad \forall x \in \mathbb{R}^d$$

   - If the Hessian is positive definite for all $x \in \mathbb{R}^d$, then $f$ is strictly convex, but the converse is not true. (e.g. $f(x) = x^4$ is strictly convex but its Hessian is zero at $x = 0$.)

4. **Segementation function:** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Define the segmentation function $\phi : [0, 1] \to \mathbb{R}$ as:

$$\phi(t) = f(x + t(y - x)), \quad t \in [0, 1]$$

$$\phi'(t) = (y - x)^T \cdot \nabla f(x + t(y - x)) = \langle \nabla f(x + t(y - x)), y - x \rangle$$

for fixed $x, y \in \mathbb{R}^d$. Then:

$$f(y) - f(x) = \int_0^1 \phi'(t)\, dt = \int_0^1 \langle \nabla f\left(x + t(y - x)\right), y - x \rangle\, dt$$

This is useful for proving various properties:

- If we want to prove the tangent plane property from the monotonicity of the gradient:

$$\begin{aligned}
f(y) - f(x) &= \int_0^1 \langle \nabla f\left(x + t(y-x)\right), y - x\rangle \, dt \\
&= \int_0^1 \langle \nabla f\left(x + t(y-x)\right) - \nabla f(x) + \nabla f(x), y - x\rangle \, dt \\
&= \int_0^1 \langle \nabla f\left(x + t(y-x)\right) - \nabla f(x), y - x\rangle \, dt + \langle \nabla f(x), y - x\rangle \\
&\geq 0 + \langle \nabla f(x), y - x\rangle \quad \text{(by monotonicity of the gradient)}
\end{aligned}$$

- If we want to prove that $f$ is convex from $\nabla^2 f(x) \in S_d^+$ for all $x$:

$$\begin{aligned}
f(y) - f(x) &= \int_0^1 \langle \nabla f\left(x + t(y-x)\right), y - x\rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x) + \nabla f(x), y - x\rangle \, dt
\end{aligned}$$

  Similarly as before, we can define $\varphi(t) = \langle \nabla f(x + t(y-x)), y - x\rangle$

  $\varphi'(t) = (y-x)^T \cdot \nabla^2 f(x+t(y-x)) \cdot (y-x) \geq 0 = \langle \nabla^2 f(x+t(y-x))(y-x), y-x\rangle$

$$f(y) - f(x) = \int_0^1 \int_0^t \varphi'(s) \, ds \, dt + \langle \nabla f(x), y - x\rangle$$

  Since $\nabla^2 f$ is positive semi-definite, we have $\varphi'(s) \geq 0$, which implies that $f(y) - f(x) \geq \langle \nabla f(x), y - x\rangle$.

## 2.1 General Convexity Exercises

**Reminder**: A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}^d$ and $t \in [0,1]$, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

A function $f$ is strongly convex with parameter $\alpha > 0$ if for all $x, y \in \mathbb{R}^d$ and $t \in [0,1]$, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2.$$

**Exercise 2.1.** Let $f \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ be convex. Is it true that $f$ has a unique critical point?

It is not necessarily true that a convex function has a unique critical point. As a counterexample, consider the convex function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = e^x.$$

This function is convex since its second derivative $f''(x) = e^x > 0$ for all $x \in \mathbb{R}$. However, it has no critical points because its first derivative $f'(x) = e^x$ is never zero. □

**Exercise 2.2.** Let $f \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ be convex. Show that if $f$ has a critical point at $x^* \in \mathbb{R}^d$, then $x^*$ is a global minimum of $f$.

If $\nabla f(x^*) = 0$ for some $x^* \in \mathbb{R}^d$, then for any $x \in \mathbb{R}^d$, the tangent plane property of convex functions gives us

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*) + 0 = f(x^*).$$

Thus, $f(x) \geq f(x^*)$ for all $x \in \mathbb{R}^d$, which means that $x^*$ is a global minimum of $f$. □

**Exercise 2.3.** Let $f \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ be strictly convex. Show that $f$ has at most one critical point.

Assume for the sake of contradiction that $f$ has two distinct critical points $x_1^*$ and $x_2^*$ in $\mathbb{R}^d$. By the definition of critical points, we have $\nabla f(x_1^*) = 0$ and $\nabla f(x_2^*) = 0$. Since $f$ is strictly convex, for any $\lambda \in (0,1)$, we have

$$f(\lambda x_1^* + (1-\lambda)x_2^*) < \lambda f(x_1^*) + (1-\lambda)f(x_2^*).$$

However, since both $x_1^*$ and $x_2^*$ are critical points, we have

$$f(x_1^*) = f(x_2^*).$$

Thus,

$$f(\lambda x_1^* + (1-\lambda)x_2^*) < f(x_1^*),$$

which contradicts the fact that both $x_1^*$ and $x_2^*$ are minima. Therefore, $f$ can have at most one critical point. □

## 2.2 Strong Convexcity Exercises

**Exercise 2.4.** Show that $f$ is strongly convex with parameter $\alpha > 0$, if and only if the function $g(x) = f(x) - \frac{\alpha}{2}\|x\|^2$ is convex.

The $\implies$ direction: Assume that $f$ is strongly convex with parameter $\alpha > 0$. For any $x, y \in \mathbb{R}^d$ and $t \in [0,1]$, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2.$$

Let's first deduce a magic formula to replace $\|tx + (1-t)y\|^2$

$$\begin{aligned}\|tx + (1-t)y\|^2 &= \langle tx + (1-t)y,\ tx + (1-t)y\rangle \\ &= t^2\|x\|^2 + (1-t)^2\|y\|^2 + 2t(1-t)\langle x, y\rangle.\end{aligned}$$

$$\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle \quad\implies\quad \langle x, y\rangle = \frac{\|x\|^2 + \|y\|^2 - \|x-y\|^2}{2}$$

Plugging this into the $\|tx + (1-t)y\|^2$ expression, we get

$$\boxed{\|tx + (1-t)y\|^2 = t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x-y\|^2.}$$

Now for $g(tx + (1-t)y)$ we have:

$$\begin{aligned}g(tx + (1-t)y) &= f(tx + (1-t)y) - \frac{\alpha}{2}\|tx + (1-t)y\|^2 \\ &= \underbrace{f(tx + (1-t)y)}_{\text{Strongly convex}} - \frac{\alpha}{2}\left(t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x-y\|^2\right)\end{aligned}$$

$$\begin{aligned}&\leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2 - \frac{\alpha}{2}\left(t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x-y\|^2\right) \\ &\leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}\left(t\|x\|^2 + (1-t)\|y\|^2\right) \\ &\leq tg(x) + (1-t)g(y)\end{aligned}$$

The $\impliedby$ direction follows by assuming $g$ is convex and reversing the steps above by using the magic formula for $\|tx + (1-t)y\|^2$.

**Exercise 2.5** (Some examples of strongly convex functions)**.** Determine which of the following functions are strongly convex on $\mathbb{R}^d$:

1. $f(x) = \|x\|^2$
2. $f(x) = e^{\|x\|}$
3. $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, where $A$ is a symmetric positive definite matrix.

Recall that a function $f$ is strongly convex with parameter $\alpha > 0$ if for all $x, y \in \mathbb{R}^d$ and $t \in [0,1]$, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2.$$

1. For $f(x) = \|x\|^2$, it is strongly convex with parameter $\alpha = 2$, because by the tangent plane property of strongly convex functions, we have

$$\|y\|^2 = \|x + y - x\|^2 = \|x\|^2 + 2\langle x, y - x\rangle + \|y - x\|^2$$

This shows that it is at most strongly convex with parameter 2.

2. The Hessian of $f(x) = e^{\|x\|}$ Calculate its gradient and Hessian we have

$$\nabla f(x) = e^{\|x\|} \frac{x}{\|x\|}$$

$$\nabla^2 f(x) = e^{\|x\|} \left( \frac{I}{\|x\|} + \frac{xx^\top}{\|x\|^2} - \frac{xx^\top}{\|x\|^3} \right)$$

$$\text{Eigenvalues of } \nabla^2 f(x) : \begin{cases} e^{\|x\|} & \text{(radial direction)} \\ \frac{e^{\|x\|}}{\|x\|} & \text{(any of the } n-1 \text{ orthogonal directions)} \end{cases}$$

We conclude that the function is strongly convex on $(x \neq 0)$ with parameter

$$\alpha = \begin{cases} e^{\|x\|} & \text{if } \|x\| \geq 1 \\ \frac{e^{\|x\|}}{\|x\|} & \text{if } 0 < \|x\| < 1 \end{cases}$$

3. For $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, where $A$ is a symmetric positive definite matrix, the Hessian of $f$ is given by $\nabla^2 f(x) = A$. Since $A$ is positive definite, there exists a constant $\alpha > 0$ such that for all $x \in \mathbb{R}^d$,

$$x^T Ax \geq \alpha \|x\|^2.$$

Therefore, $f$ is strongly convex with parameter $\alpha$. □

**Exercise 2.6.** Let $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ be $\alpha$-strongly convex. Does $f$ admit a unique minimizer?

By the strong convexity of $f$, we know that under the assumption of the existence of a minimizer, it must be unique. Now we show the coercivity of $f$ to ensure the existence of a minimizer.
One way to do this is to show that $f(x) \to +\infty$ as $\|x\| \to +\infty$. Since $f$ is $\alpha$-strongly convex, by the tangent plane property, we have for any $x, x_0 \in \mathbb{R}^d$:

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0\rangle + \frac{\alpha}{2}\|x - x_0\|^2.$$

For the sake of the clarity, let $x_0 = 0$, we have

$$f(x) \geq f(0) + \langle \nabla f(0), x\rangle + \frac{\alpha}{2}\|x\|^2.$$

This is a quadratic function in $\|x\|$ with a positive leading coefficient $\frac{\alpha}{2}$, which implies that $f(x) \to +\infty$ as $\|x\| \to +\infty$. Therefore, $f$ is coercive. □
A more rigorous way is to use the sublevel set definition: a function $f$ is coercive if the sublevel sets $\{x \in \mathbb{R}^d : f(x) \leq c\}$ are bounded for all $c \in \mathbb{R}$.

$$f(x) \ \geq \ f(0) + \langle \nabla f(0), x\rangle + \frac{\alpha}{2}\|x\|^2$$

Recall that Young's inequality states that for any $a, b \geq 0$ and $\epsilon > 0$, we have

$$ab \leq \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2}.$$

Now we control the linear term by the quadratic term using Cauchy-Schwarz and Young's inequality (with $\epsilon = \frac{\alpha}{2}$):

$$\langle \nabla f(0), x \rangle \;\geq\; -\|\nabla f(0)\|\|x\| \;\geq\; -\frac{1}{\alpha}\|\nabla f(0)\|^2 - \frac{\alpha}{4}\|x\|^2$$

Plugging this back we have

$$f(x) \;\geq\; f(0) - \frac{1}{\alpha}\|\nabla f(0)\|^2 + \frac{\alpha}{4}\|x\|^2$$

Thus the sublevel sets are bounded. Therefore, $f$ is coercive. $\square$

## 3 Lipschitz Continuity of the Gradient

1. **Definition:** A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ has a Lipschitz continuous gradient with constant $\mu > 0$ if for all $x, y \in \mathbb{R}^d$, we have:

$$\|\nabla f(y) - \nabla f(x)\| \leq \mu \|y - x\|$$

2. **Several Implications for Lipschitz Continuity:** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function with a Lipschitz continuous gradient with constant $\mu > 0$. Then for all $x, y \in \mathbb{R}^d$, we have:

   - Upper bound on the Hessian eigenvalues:

$$\lambda_1(\nabla^2 f(x)) \leq \mu$$

   - Quadratic upper bound using gradient at $x$:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2$$

## 4 Caratheodory theorem

1. **Convex hull:** The convex hull of a set $\Omega \subset \mathbb{R}^d$ is defined as the smallest convex set containing $\Omega$.

$$\mathrm{conv}(\Omega) := \bigcap \{C \subset \mathbb{R}^d \mid C \text{ is convex and } \Omega \subset C\}$$

   It can be expressed as:

$$\mathrm{conv}(\Omega) = \left\{ \sum_{i=1}^{N} \lambda_i x_i \mid N \in \mathbb{N}, x_i \in \Omega, \lambda_i \geq 0, \sum_{i=1}^{N} \lambda_i = 1 \right\}$$

2. **Caratheodory theorem:** Let $\Omega \subset \mathbb{R}^d$. Then any point in the convex hull of $\Omega$ can be expressed as a convex combination of at most $d + 1$ points from $\Omega$. In other words, for any $x \in \mathrm{conv}(\Omega)$, there exist points $x_1, x_2, \ldots, x_{d+1} \in \Omega$ and coefficients $\lambda_1, \lambda_2, \ldots, \lambda_{d+1} \geq 0$ such that:

$$x = \sum_{i=1}^{d+1} \lambda_i x_i \quad \text{and} \quad \sum_{i=1}^{d+1} \lambda_i = 1$$

## 5 Extreme Points and Krein-Milman Theorem

1. **Extreme Points:** Let $C$ be a convex set in a vector space. A point $x \in C$ is called an extreme point of $C$ if **it cannot be expressed as a convex combination of two distinct points in** $C$. Formally, $x$ is an extreme point if whenever $x = \lambda y + (1 - \lambda)z$ for some $y, z \in C$ and $\lambda \in (0, 1)$, it follows that $y = z = x$.

**2. Projection on closed convex sets:** Let $C \subset \mathbb{R}^d$ be a non-empty closed convex set. For any point $x \in \mathbb{R}^d$, there exists a **unique** point $P_C(x) \in C$ such that:

$$\|x - P_C(x)\| = \min_{y \in C} \|x - y\|$$

The point $P_C(x)$ is called the projection of $x$ onto the set $C$. The projection operator $P_C : \mathbb{R}^d \to C$ is **1-Lipschitz** (non-expansive), meaning that for all $x, y \in \mathbb{R}^d$:

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|$$

**3. Krein-Milman Theorem:** Let $C$ be a non-empty compact convex subset of a locally convex topological vector space. Then $C$ is the closed convex hull of its extreme points. In other words, if we denote the set of extreme points of $C$ by $\text{ext}(C)$, then:

$$C = \overline{\text{conv}}(\text{ext}(C))$$

where $\overline{\text{conv}}(\text{ext}(C))$ denotes the closure of the convex hull of the extreme points of $C$.

In other words, every compact convex set can be "reconstructed" from its extreme points by taking all possible convex combinations of those points and then taking the closure of that set.

# 6 Polyak-Lojasiewicz Inequality

**1.** Let's remind first the definition and some properties of strongly convex functions. A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is called $\alpha$-strongly convex if for all $x, y \in \mathbb{R}^d$:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2 \quad \textbf{(Function is above the Tangent Cone)}$$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \alpha \|y - x\|^2 \quad \textbf{(Curvature Condition)}$$

**2. Polyak-Lojasiewicz Inequality:** A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is said to satisfy the Polyak-Lojasiewicz (PL) inequality with constant $\mu > 0$ if for all $x \in \mathbb{R}^d$:

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq \mu(f(x) - f^*)$$

where $f^* = \inf_{x \in \mathbb{R}^d} f(x)$ is the global minimum value of $f$.

**3.** We can show that $\alpha$-strongly convex functions satisfy the PL inequality with constant $\mu = \alpha$. Let $f$ be an $\alpha$-strongly convex function and let $x^*$ be its unique minimizer. Then for any $x \in \mathbb{R}^d$:

$$\begin{aligned}
f(x) &\geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle + \frac{\alpha}{2}\|x - x^*\|^2 \\
&\geq f(x^*) + 0 + \frac{\alpha}{2}\|x - x^*\|^2 \quad (\text{since } \nabla f(x^*) = 0) \\
&\Rightarrow f(x) - f(x^*) \geq \frac{\alpha}{2}\|x - x^*\|^2
\end{aligned}$$

On the other hand, making a minor modification to the first property of strongly convex functions, we have:

$$\begin{aligned} f(x) - f(y) &\leq \langle \nabla f(x), x - y \rangle - \frac{\alpha}{2}\|x - y\|^2 \\ &\leq \|\nabla f(x)\| \|x - y\| - \frac{\alpha}{2}\|x - y\|^2 \quad \text{(by Cauchy-Schwarz inequality)} \\ &\leq \frac{\|\nabla f(x)\|^2}{2\alpha} \quad \text{(by maximizing the right-hand side over } \|x - y\|\text{)} \end{aligned}$$

Combining the two inequalities, we get:

$$f(x) - f(x^*) \leq \frac{\|\nabla f(x)\|^2}{2\alpha}$$

Rearranging this gives the PL inequality:

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq \alpha(f(x) - f(x^*))$$