# Introductory Lectures on Optimization

## Foundations of Smooth Optimization (2)

Hui Qian
qianhui@zju.edu.cn

College of Computer Science, Zhejiang University

September 25, 2024

# Outline

# Part I
# Convex Function

# Difficulty in General Unconstrained Minimization Problem

Consider the unconstrained minimization problem

$$\boxed{\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \qquad (9)}$$

where the objective function $f(\boldsymbol{x})$ is smooth enough.

Under very weak assumptions on the function, we cannot do too much

- It is impossible to guarantee convergence even to a local minimum, and
- it is impossible to get acceptable bounds on the global performance of minimization schemes.

Let us try to introduce some reasonable assumptions on the function $f$ in order to make our problem more tractable.

# Assumptions That Make Sense

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \tag{9}$$

From the results of the previous, we could come to the conclusion that the main reason for our troubles is the weakness of the first-order optimality condition.

The first additional property we definitely need is as follows.

**Assumption 12**

For any $f \in \mathcal{F}$, the first-order optimality condition is sufficient for a point to be a global solution to (9).

# Assumptions That Make Sense

Further, the main feature of any tractable functional class $\mathcal{F}$ is :

The possibility to verify the inclusion $f \in \mathcal{F}$ in a simple way.

Usually, this is ensured by a set of basic elements of the class, endowed with a list of possible operations with elements of $\mathcal{F}$ ( such operations are called invariant).

## Assumption 13

If $f_1, f_2 \in \mathcal{F}$ and $\alpha, \beta \geq 0$, then $\alpha f_1 + \beta f_2 \in \mathcal{F}$.

## Assumption 14

Any linear function $f(\boldsymbol{x}) = \alpha + \langle a, \ \boldsymbol{x} \rangle$ belongs to $\mathcal{F}$.

# Definition of The Convex Function

Consider $f \in \mathcal{F}$. Let us fix some $\boldsymbol{x}_0 \in \mathbb{R}^n$ and observe the following function:

$$\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle \nabla f(\boldsymbol{x}_0),\ \boldsymbol{y} \rangle.$$

In view of Assumptions 13 and 14, $\phi \in \mathcal{F}$. Note that $\nabla \phi(\boldsymbol{y})|_{\boldsymbol{y}=\boldsymbol{x}_0} = \nabla f(\boldsymbol{x}_0) - \nabla f(\boldsymbol{x}_0) = 0$. Therefore, in view of Assumption 12, $\boldsymbol{x}_0$ is the global minimum of function $\phi$, and for any $\boldsymbol{y} \in \mathbb{R}^n$ we have

$$\boxed{\phi(\boldsymbol{y})} \geq \phi(\boldsymbol{x}_0) = \boxed{f(\boldsymbol{x}_0) - \langle \nabla f(\boldsymbol{x}_0),\ \boldsymbol{x}_0 \rangle}.$$

Hence,

$$\boxed{f(\boldsymbol{y}) \geq f(\boldsymbol{x}_0) + \langle \nabla f(\boldsymbol{x}_0),\ \boldsymbol{y} - \boldsymbol{x}_0 \rangle.}$$

# Definition of The Convex Function

**Definition 15 (Convex Set)**

A set $Q \subseteq \mathbb{R}^n$ is called convex if for any $\boldsymbol{x}, \boldsymbol{y} \in Q$ and $\alpha$ from $[0, 1]$ we have

$$\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in Q. \tag{10}$$

We denote by $\mathcal{F}^k(Q)$ the class we discussed above, and call it class of the convex function:

- Any $f \in \mathcal{F}^k(Q)$ is a convex function (see Definition 16), and
- any $f \in \mathcal{F}^k(Q)$ is $k$ times continuously differentiable on $Q$.
- We assume $Q = \mathbb{R}^n$ in this chapter.

# Definition of The Convex Function

**Definition 16 (Convex Function)**

A continuously differentiable function $f(\cdot)$ is called convex on a convex set $Q$ (notation $f \in \mathcal{F}^1(Q^n)$) if for any $\boldsymbol{x}, \boldsymbol{y} \in Q$ we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle. \tag{11}$$

If $-f(\boldsymbol{x})$ is convex, we call $f(\boldsymbol{x})$ concave.

# Properties of The Convex Function

## Global Property:

**Theorem 17**
If $f \in \mathcal{F}^1(\mathbb{R}^n)$ and $\nabla f(\boldsymbol{x}^*) = 0$, then $\boldsymbol{x}^*$ is the global minimum of $f(\boldsymbol{x})$ on $\mathbb{R}^n$.

**Proof.**
In view of inequality (11), for any $\boldsymbol{x} \in \mathbb{R}^n$ we have

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

□

# Properties of The Convex Function

## A. Conic Combination:

**Lemma 18**

If $f_1$ and $f_2$ belong to $\mathcal{F}^1(\mathbb{R}^n)$, and $\alpha, \beta \geq 0$, then the function $f = \alpha f_1 + \beta f_2$ also belong to $\mathcal{F}^1(\mathbb{R}^n)$.

**Proof.**

For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\begin{aligned} f_1(\boldsymbol{y}) &\geq f_1(\boldsymbol{x}) + \langle \nabla f_1(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle, \\ f_2(\boldsymbol{y}) &\geq f_2(\boldsymbol{x}) + \langle \nabla f_2(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle. \end{aligned}$$

It remains to multiply the first equation by $\alpha$, the second one by $\beta$, and add the results. $\square$

# Properties of The Convex Function

## B. Affine Composition:

**Lemma 19**
If $f \in \mathcal{F}^1(\mathbb{R}^n)$, $b \in \mathbb{R}^m$, and $A : \mathbb{R}^n \to \mathbb{R}^m$, then

$$\phi(\boldsymbol{x}) = f(A\boldsymbol{x} + \boldsymbol{b}) \in \mathcal{F}^1(\mathbb{R}^m).$$

Proof.
Indeed, let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Define $\bar{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{b}$, $\bar{\boldsymbol{y}} = A\boldsymbol{y} + \boldsymbol{b}$. since $\nabla\phi(\boldsymbol{x}) = A^\top \nabla f(A\boldsymbol{x} + \boldsymbol{b})$, we have

$$\begin{aligned} \phi(\boldsymbol{y}) = f(\bar{\boldsymbol{y}}) &\geq f(\bar{\boldsymbol{x}}) + \langle \nabla f(\bar{\boldsymbol{x}}),\ \bar{\boldsymbol{y}} - \bar{\boldsymbol{x}} \rangle \text{ We have } \bar{\boldsymbol{y}} - \bar{\boldsymbol{x}} = A(\boldsymbol{y} - \boldsymbol{x}) \\ &= \phi(\boldsymbol{x}) + \langle \nabla f(\bar{\boldsymbol{x}}),\ A(\boldsymbol{y} - \boldsymbol{x}) \rangle = \phi(\boldsymbol{x}) + \langle A^\top \nabla f(\bar{\boldsymbol{x}}),\ \boldsymbol{y} - \boldsymbol{x} \rangle \\ &= \phi(\boldsymbol{x}) + \langle \nabla \phi(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle. \end{aligned}$$

# Properties of The Convex Function

## C. Pointwise maximum and supremum:

**Lemma 20**

If $f_i(\boldsymbol{x})$, $i \in I$, are convex, then

$$g(\boldsymbol{x}) = \max_{i \in I} f_i(\boldsymbol{x})$$

is also convex.

Remark. The property extends to the pointwise supremum over a infinite set. If $f(\boldsymbol{x}, \omega)$ is convex in $\boldsymbol{x}$, for $\omega \in \Omega$, then

$$g(\boldsymbol{x}) = \sup_{\omega \in \Omega} f(\boldsymbol{x}, \omega)$$

is convex.

# Properties of The Convex Function

## D. Convex monotone composition:

**Lemma 21**

1. If $f$ is a convex function on $\mathbb{R}^n$ and $F(\cdot)$ is a convex and non-decreasing function on $\mathbb{R}$, then $g(\boldsymbol{x}) = F(f(\boldsymbol{x}))$ is convex.
2. If $f_i, i = 1, \ldots, m$ are convex functions on $\mathbb{R}^n$ and $F(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$ is convex and non-decreasing (component-wise) in each argument, then

$$g(\boldsymbol{x}) = F(f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))$$

is convex.

# Properties of The Convex Function

## E. Partial minimization:

**Lemma 22**

If $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n$ and $Y$ is a convex set, then

$$g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in Y} f(\boldsymbol{x}, \boldsymbol{y})$$

is convex.

# Equivalent Definitions

**Theorem 23**

A continuously differentiable function $f$ belongs to the class $\mathcal{F}^1(\mathbb{R}^n)$ if and only if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\alpha \in [0, 1]$ we have

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}). \tag{12}$$

Proof. Define $\boldsymbol{x}_\alpha = \alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}$. let $f \in \mathcal{F}^1(\mathbb{R}^n)$. Then

$$f(\boldsymbol{x}_\alpha) \leq f(\boldsymbol{y}) + \langle \nabla f(\boldsymbol{x}_\alpha), \ \boldsymbol{y} - \boldsymbol{x}_\alpha \rangle = f(\boldsymbol{y}) + \alpha \langle \nabla f(\boldsymbol{x}_\alpha), \ \boldsymbol{y} - \boldsymbol{x} \rangle$$
$$f(\boldsymbol{x}_\alpha) \leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}_\alpha), \ \boldsymbol{x} - \boldsymbol{x}_\alpha \rangle = f(\boldsymbol{x}) - (1-\alpha) \langle \nabla f(\boldsymbol{x}_\alpha), \ \boldsymbol{y} - \boldsymbol{x} \rangle.$$

Multiplying the first inequality by $(1-\alpha)$, the second one by $\alpha$, and adding the results, we get (12).

# Equivalent Definitions

**Theorem 23**

A continuously differentiable function $f$ belongs to the class $\mathcal{F}^1(\mathbb{R}^n)$ if and only if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\alpha \in [0,1]$ we have

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}). \tag{12}$$

Proof. (Continued) Let (12) be true for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and $\alpha \in [0,1]$. Let us choose some $\alpha \in [0,1)$. Then

$$\begin{aligned} f(\boldsymbol{y}) \geq \frac{1}{1-\alpha}[f(\boldsymbol{x}_\alpha) - \alpha f(\boldsymbol{x})] &= f(\boldsymbol{x}) + \frac{1}{1-\alpha}[f(\boldsymbol{x}_\alpha) - f(\boldsymbol{x})] \\ &= f(\boldsymbol{x}) + \frac{1}{1-\alpha}[f(\boldsymbol{x} + (1-\alpha)(\boldsymbol{y}-\boldsymbol{x})) - f(\boldsymbol{x})]. \end{aligned}$$

Letting $\alpha$ tend to 1, we get (11).第二等式右边第二项是方向导数表达

# Equivalent Definitions

**Theorem 24**

A continuously differentiable function $f$ belongs to the class $\mathcal{F}^1(\mathbb{R}^n)$ if and only if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \geq 0. \tag{13}$$

*Proof.* Let $f$ be a convex and continuously differentiable function. Then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle,$$
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle.$$

Adding these inequalities, we get (13).

# Equivalent Definitions

Proof. (Continued)

Let (13) hold for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Define $\boldsymbol{x}_\tau = \boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})$. Then

$$
\begin{aligned}
f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \int_0^1 \langle \nabla f(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})),\ \boldsymbol{y} - \boldsymbol{x} \rangle d\tau \\
&= f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \int_0^1 \langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle d\tau \\
&= f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \int_0^1 \frac{1}{\tau} \boxed{\langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{x}_\tau - \boldsymbol{x} \rangle} d\tau \\
&\geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle.
\end{aligned}
$$

□

# Equivalent Definitions

**Theorem 25**

A twice continuously differentiable function $f$ belongs to the class $\mathcal{F}^2(\mathbb{R}^n)$ if and only if for any $x \in \mathbb{R}^n$ we have

$$\nabla^2 f(\boldsymbol{x}) \succeq 0. \tag{14}$$

*Proof.* Let a function $f \in C^2(\mathbb{R}^n)$ be convex . Let $\boldsymbol{x}_\tau = \boldsymbol{x} + \tau \boldsymbol{s}$, for $\tau > 0$ . Then in view of (13), we have

$$0 \leq \frac{1}{\tau^2}\langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{x}_\tau - \boldsymbol{x}\rangle = \frac{1}{\tau}\langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{s}\rangle$$
$$= \frac{1}{\tau}\int_0^\tau \langle \nabla^2 f(\boldsymbol{x} + \lambda \boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle d\lambda \quad \text{(explained in the next pages.)},$$

and we get (14) by letting $\tau$ tend to zero.

# Equivalent Definitions

(1) Let $\boldsymbol{y} = \boldsymbol{x}_\tau$. We have

$$\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}) = \int_0^1 \nabla^2(\boldsymbol{x} + p(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x})dp.$$

Let $p = \frac{\lambda}{\tau}$, we have

$$\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}) = \int_0^1 \nabla^2(\boldsymbol{x} + \frac{\lambda}{\tau}(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x})d\frac{\lambda}{\tau} = \int_0^\tau \nabla^2(\boldsymbol{x} + \lambda\boldsymbol{s})\boldsymbol{s}d\lambda.$$

Thus, we arrive at

$$\langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{s}\rangle = \int_0^\tau \langle \nabla^2 f(\boldsymbol{x} + \lambda\boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle d\lambda$$

# Equivalent Definitions

(2) For $G(\lambda) = \langle \nabla^2 f(\boldsymbol{x} + \lambda \boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle$ and $\Phi(y) = \int_0^y G(\lambda)d\lambda$,

$$
\begin{aligned}
\lim_{\tau\to 0} \boxed{\frac{1}{\tau}\int_0^\tau \langle \nabla^2 f(\boldsymbol{x} + \lambda \boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle d\lambda} &= \lim_{\tau\to 0} \frac{1}{\tau}\left(\int_0^{y+\tau} G(\lambda)d\lambda - \int_0^y G(\lambda)d\lambda\right)\Big|_{y=0} \\
&= \lim_{\tau\to 0} \frac{\Phi(y+\tau) - \Phi(y)}{\tau}\Big|_{y=0} \\
&= \Phi'(y)|_{y=0} = G(y)|_{y=0} = \boxed{\langle \nabla^2 f(\boldsymbol{x})s,\ s\rangle}.
\end{aligned}
$$

$$\Rightarrow \langle \nabla^2 f(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s}\rangle \geq 0.$$

# Equivalent Definitions

**Theorem 25**
A twice continuously differentiable function $f$ belongs to the class $\mathcal{F}^2(\mathbb{R}^n)$ if and only if for any $x \in \mathbb{R}^n$ we have

$$\nabla^2 f(\boldsymbol{x}) \succeq 0. \tag{14}$$

Proof. (Continued) Let (14) hold for all $\boldsymbol{x} \in \mathbb{R}^n$. Then for $\boldsymbol{y} \in \mathbb{R}^n$ we have

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \\ &\quad + \underbrace{\int_0^1 \int_0^\tau \langle \nabla^2 f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle d\lambda d\tau}_{\geq 0 \quad \text{(explained in the next page.)}} \\ &\geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle. \end{aligned}$$

# Equivalent Definitions

(3) For $\Phi(t) = f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$, we have

$$\Phi(t) = \Phi(0) + \int_0^t \Phi'(\tau) d\tau, \quad \text{and} \quad \Phi'(\tau) = \Phi'(0) + \int_0^\tau \Phi''(\lambda) d\lambda.$$

It is clear that

$$\Phi(t) = \Phi(0) + \int_0^t \Phi'(0) d\tau + \int_0^t \int_0^\tau \Phi''(\lambda) d\lambda d\tau.$$

Thus, we obtain that

$$\begin{aligned} f(\boldsymbol{y}) = \Phi(1) &= \Phi(0) + \int_0^1 \Phi'(0) d\tau + \int_0^1 \int_0^\tau \Phi''(\lambda) d\lambda d\tau \\ &= f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \int_0^1 \int_0^\tau \langle \nabla^2 f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle d\lambda d\tau \end{aligned}$$

# Examples

Let us look at some examples of differentiable convex functions on $\mathbb{R}^n$.

## Example 26

1. Every linear function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \ \boldsymbol{x} \rangle$ is convex.
2. Let matrix $A$ be symmetric and positive semidefinite. Then the quadratic function

$$f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \ \boldsymbol{x} \rangle + \frac{1}{2} \langle A\boldsymbol{x}, \ \boldsymbol{x} \rangle$$

is convex ( since $\nabla^2 f(\boldsymbol{x}) = A \succeq 0$ ).

# Examples

## Example 26

3 The following functions of one variable belong to $\mathcal{F}^1(\mathbb{R})$:

$$
\begin{aligned}
f(x) &= e^x, \\
f(x) &= |x|^p, \ p > 1, \\
f(x) &= \frac{x^2}{1+|x|}, \\
f(x) &= |x| - \ln(1+|x|).
\end{aligned}
$$

We can check this using Theorem 25.

# Examples

## Example 26

4. Functions arising in Geometric Optimization, like

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} e^{\alpha_i + \langle \boldsymbol{a}_i,\, \boldsymbol{x} \rangle},$$

are convex (see Lemma19 of Nesterov [2003]).

5. Similarly, functions arising in $\ell_p$-norm approximation problems, like

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} |\langle \boldsymbol{a}_i,\, \boldsymbol{x} \rangle - \boldsymbol{b}_i|^p,$$

are convex too.

# Part II
# Smooth and Convex Function

# The Class $\mathcal{F}_L^{k,l}(\mathbb{R}^n)$

As with general nonlinear functions, the differentiability itself cannot ensure any special topological properties of convex functions. Therefore we need to consider the problem classes with Lipschitz continuous derivatives of a certain order.

We introduce a new function type $\mathcal{F}_L^{k,l}(\mathbb{R}^n)$, and remark as follows.

1. Any function $f \in \mathcal{F}_L^{k,l}(\mathbb{R}^n)$ is convex, and
2. the meaning of the index is the same as $C_L^{k,l}(\mathbb{R}^n)$.
3. The most important class of that type is $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$ , the dass of convex functions with Lipschitz continuous gradient.

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

Let us provide it with several necessary and sufficient conditions.

**Theorem 27**

All conditions below, holding for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\alpha$ from $[0, 1]$, are equivalent to the inclusion $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$:

$$0 \leq f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle \leq \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{15}$$

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq f(\boldsymbol{y}), \tag{16}$$

$$\frac{1}{L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle, \tag{17}$$

$$0 \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \leq L \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{18}$$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

**Theorem 27**

(Continued) All conditions below, holding for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\alpha$ from $[0, 1]$, are equivalent to inclusion $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$:

$$\begin{aligned} \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \geq f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \\ + \frac{\alpha(1-\alpha)}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 , \end{aligned} \tag{19}$$

$$\begin{aligned} 0 \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) - f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \\ \leq \alpha(1-\alpha)\frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2 . \end{aligned} \tag{20}$$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$\boxed{0 \leq f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \leq \frac{L}{2} \|x - \boldsymbol{y}\|^2 , \qquad (15)}$$

Proof. Condition (15)

Indeed, (15) follows from the definition of convex functions and Lemma (1.2.3) of Nesterov [2003]. That is the first (left) inequality can be derived from

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

and the second (right) is from

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle| \leq \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2 .$$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$0 \leq f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \leq \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2, \quad (15)$$

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq f(\boldsymbol{y}), \quad (16)$$

Proof. Condition (15) $\Rightarrow$ (16)

Let us fix $\boldsymbol{x} \in \mathbb{R}^n$. Consider the function $\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} \rangle$. Note that $\phi \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$, and its optimal point is $\boldsymbol{y}^* = \boldsymbol{x}$. Therefore, in the view of (15), we have

$$\boxed{\phi(\boldsymbol{y}^*)} \underbrace{\leq}_{\boldsymbol{y}^*} \phi\left(\boldsymbol{y} - \frac{1}{L}\nabla\phi(\boldsymbol{y})\right) \underbrace{\leq}_{(15)} \boxed{\phi(\boldsymbol{y}) - \frac{1}{2L}\|\nabla\phi(\boldsymbol{y})\|^2}.$$

$$\phi\left(\boldsymbol{y} - \frac{1}{L}\nabla\phi(\boldsymbol{y})\right) \leq \phi(\boldsymbol{y}) + \left\langle \nabla\phi(\boldsymbol{y}), \boldsymbol{y} - \frac{1}{L}\nabla\phi(\boldsymbol{y}) - \boldsymbol{y} \right\rangle + \frac{1}{2L}\|\nabla\phi(\boldsymbol{y})\|^2.$$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$0 \le f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \le \frac{L}{2} \|x - \boldsymbol{y}\|^2, \tag{15}$$

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \le f(\boldsymbol{y}), \tag{16}$$

Proof. (Continued.) Thus we have

$$\boxed{\phi(\boldsymbol{y}^*)} \le \boxed{\phi(\boldsymbol{y}) - \frac{1}{2L} \|\nabla \phi(\boldsymbol{y})\|^2}$$

$$\Rightarrow f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{x} \rangle \le f(\boldsymbol{y}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} \rangle - \frac{1}{2L} \|\nabla \phi(\boldsymbol{y})\|^2.$$

And we get (16), since $\nabla \phi(\boldsymbol{y}) = \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x})$. □

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq f(\boldsymbol{y}), \tag{16}$$

$$\frac{1}{L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle, \tag{17}$$

Proof.

Condition (16) $\Rightarrow$ (17)

We obtain (17) from inequality (16) by adding two copies of it with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged.

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$\boxed{\frac{1}{L}\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y}\rangle, \qquad (17)}$$

Proof.

$((17) \Rightarrow f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n))$

Applying now to (17) Cauchy-Schwartz inequality we get Lipschitz condition. (By the way, the fact that the right side of (17) is greater or equal to zero implies the convexity) :

$$\boxed{\frac{1}{L}\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2} \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y}\rangle \leq \boxed{\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \, \|\boldsymbol{x} - \boldsymbol{y}\|}.$$

□

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$0 \leq f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(x), \boldsymbol{y} - \boldsymbol{x} \rangle \leq \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{15}$$

$$0 \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \leq L \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{18}$$

Proof.

Condition (15) $\Rightarrow$ (18)

We obtain (18) from inequality (15) by adding two copies of it with $x$ and $\boldsymbol{y}$ interchanged. $\square$

# Necessary and Sufficient Conditions

$$0 \le f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \le \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{15}$$

$$0 \le \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \le L \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{18}$$

Proof. (Continued)

Condition (18) $\Rightarrow$ (15) In order to get (15) from (18) we apply integration:

$$\begin{aligned} f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle &= \int_0^1 \langle \nabla f(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle d\tau \\ &\le \frac{1}{2} L \|\boldsymbol{y} - \boldsymbol{x}\|^2 . \end{aligned}$$

$\square$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\|^2 \leq f(\boldsymbol{y}), \tag{16}$$

$$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \geq f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) + \frac{\alpha(1-\alpha)}{2L} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\|^2, \tag{19}$$

Proof. Condition (16) $\Rightarrow$ (19) Denote $\boldsymbol{x}_\alpha = \alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}$. Then, using (16) we get

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_\alpha) + \langle \nabla f(\boldsymbol{x}_\alpha),\ (1-\alpha)(\boldsymbol{x} - \boldsymbol{y}) \rangle + \frac{1}{2L} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}_\alpha) \right\|^2,$$

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}_\alpha) + \langle \nabla f(\boldsymbol{x}_\alpha),\ \alpha(\boldsymbol{y} - \boldsymbol{x}) \rangle + \frac{1}{2L} \left\| \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}_\alpha) \right\|^2.$$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 \leq f(\boldsymbol{y}), \tag{16}$$

$$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \geq f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) + \frac{\alpha(1-\alpha)}{2L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2, \tag{19}$$

Proof. (Continued.) Adding these inequalities multiplied by $\alpha$ and $1-\alpha$ respectively, and using inequality

$$\alpha \|\boldsymbol{g}_1 - \boldsymbol{u}\|^2 + (1-\alpha) \|\boldsymbol{g}_2 - \boldsymbol{u}\|^2 \geq \alpha(1-\alpha) \|\boldsymbol{g}_1 - \boldsymbol{g}_2\|^2,$$

we get (19). □

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

**Remark.** Consider

$$\underbrace{(\alpha\|\boldsymbol{x}\|)^2}_{a^2} + \underbrace{((1-\alpha)\|\boldsymbol{y}\|)^2}_{b^2} \geq \underbrace{2\alpha(1-\alpha)\|\boldsymbol{x}\|\|\boldsymbol{y}\|}_{2ab}. \tag{21}$$

We add $\alpha(1-\alpha)(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2)$ to both sides of (21), and we arrive at

$$\begin{aligned}\alpha\|\boldsymbol{x}\|^2 + (1-\alpha)\|\boldsymbol{y}\|^2 &\geq \alpha(1-\alpha)(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 \\ &\geq \alpha(1-\alpha)\|\boldsymbol{x}-\boldsymbol{y}\|^2.\end{aligned}$$

The second one follows from the triangle inequality for norms.

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$

$$0 \le f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \le \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{15}$$

$$0 \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) - f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha(1-\alpha)\frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2. \tag{20}$$

Proof. Condition (15) $\Rightarrow$ (20) Similarly, from (15) we get

$$0 \le f(\boldsymbol{x}) - f(\boldsymbol{x}_\alpha) - \langle \nabla f(\boldsymbol{x}_\alpha), (1-\alpha)(\boldsymbol{x} - \boldsymbol{y}) \rangle \le \frac{L}{2} \|(1-\alpha)(\boldsymbol{x} - \boldsymbol{y})\|^2,$$

$$0 \le f(\boldsymbol{y}) - f(\boldsymbol{x}_\alpha) - \langle \nabla f(\boldsymbol{x}_\alpha), \alpha(\boldsymbol{y} - \boldsymbol{x}) \rangle \le \frac{L}{2} \|\alpha(\boldsymbol{y} - \boldsymbol{x})\|^2.$$

Adding these inequalities multiplied by $\alpha$ and $1-\alpha$ respectively, we obtain (20). $\square$

# Necessary and Sufficient Conditions for The Class $\mathcal{F}_L^{2,1}(\mathbb{R}^n)$

**Theorem 28**

Twice continuously differentiable function $f$ belong to $\mathcal{F}_L^{2,1}(\mathbb{R}^n)$ if and only if for any $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$0 \preceq \nabla^2 f(\boldsymbol{x}) \preceq L I_n. \tag{22}$$

*Proof.* The statement follows from Theorem 2.1.4 of Nesterov [2003] and condition (18). Specifically, Theorem 2.1.4 solve the first inequality, and based on (18) we have

$$L \geq \frac{1}{\tau}\langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{s}\rangle, \quad \text{where } \boldsymbol{x}_\tau = \boldsymbol{x} + \tau \boldsymbol{s}, \quad \|\boldsymbol{s}\| = 1.$$

Remark,

$$0 \leq \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y}\rangle \leq L \left\|\boldsymbol{x} - \boldsymbol{y}\right\|^2. \tag{18}$$

# Necessary and Sufficient Conditions

(1) We have

$$\langle \nabla f(\boldsymbol{x}_\tau) - \nabla f(\boldsymbol{x}),\ \boldsymbol{s}\rangle = \langle \nabla f(\boldsymbol{x}+\lambda \boldsymbol{s}),\ \boldsymbol{s}\rangle\Big|_{\lambda=0}^{\tau} = \int_0^\tau \langle \nabla^2 f(\boldsymbol{x}+\lambda\boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle d\lambda$$

(2) For $G(\lambda) = \langle \nabla^2 f(\boldsymbol{x}+\lambda\boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle$ and $\Phi(y) = \int_0^y G(\lambda)d\lambda$,

$$\begin{aligned}\lim_{\tau\to 0}\boxed{\frac{1}{\tau}\int_0^\tau \langle \nabla^2 f(\boldsymbol{x}+\lambda\boldsymbol{s})\boldsymbol{s},\ \boldsymbol{s}\rangle d\lambda} &= \lim_{\tau\to 0}\frac{1}{\tau}\left(\int_0^{y+\tau} G(\lambda)d\lambda - \int_0^y G(\lambda)d\lambda\right)\Big|_{y=0}\\ &= \lim_{\tau\to 0}\frac{\Phi(y+\tau)-\Phi(y)}{\tau}\Big|_{y=0}\\ &= \Phi'(y)|_{y=0} = G(y)|_{y=0} = \boxed{\langle \nabla^2 f(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s}\rangle}.\end{aligned}$$

$\Rightarrow \langle \nabla^2 f(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s}\rangle \leq L$, since (18).

# References I

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87. Springer Science & Business Media, 2003.

Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

Stephen J Wright and Benjamin Recht. *Optimization for data analysis*. Cambridge University Press, 2022.

# Thank You!

Email:qianhui@zju.edu.cn