Introductory Lectures on Optimization

Foundations of Smooth Optimization (3)

Hui Qian
qianhui@zju.edu.cn

College of Computer Science, Zhejiang University

October 17, 2024

Outline

Part I
Strongly Convex Function

Definition of The Strongly Convex and The Class $\mathcal{S}_{\mu}^{1}(\mathbb{R}^n)$

We are looking for a restriction of the functional class $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$, for which we can guarantee a reasonable rate of convergence to a unique solution of the minimization problem.

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \qquad f \in \mathcal{F}^1(\mathbb{R}^n)$$

Let us try to make this non-degeneracy assumption global. Namely, let us assume that there exists some constant $\mu > 0$ such that for any $\bar{\boldsymbol{x}}$ with $\nabla f(\bar{\boldsymbol{x}}) = 0$ and any $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \frac{1}{2}\mu \left\|\boldsymbol{x} - \bar{\boldsymbol{x}}\right\|^2 .$$

Definition of The Strongly Convex and The Class $\mathcal{S}^1_\mu(\mathbb{R}^n)$

Definition 29

A continuously differentiable function $f(\boldsymbol{x})$ is called strongly convex on $\mathbb{R}^n$ (notation $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$) if there exists a constant $\mu > 0$ such that for any $x, y \in \mathbb{R}^n$ we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2}\mu \left\| \boldsymbol{y} - \boldsymbol{x} \right\|^2 . \tag{23}$$

Constant μ is called the convexity parameter of f。

Remark. $\mathcal{S}^p_\mu(\mathbb{R}^n)$ with $p = 1$. See Lemma 1.2.3 of Nesterov [2003] for geometric interpretation of strongly convex function.

Property of Strongly Convex Function

Theorem 30

If $f \in \mathcal{S}_{\mu}^{1}(\mathbb{R}^n)$ and $\nabla f(\boldsymbol{x}^*) = 0$, then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \frac{1}{2}\mu \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\|^2$$

for all $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. Since $\nabla f(\boldsymbol{x}^*) = 0$, in view of inequality (23), for any $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\begin{aligned} f(\boldsymbol{x}) &\geq f(\boldsymbol{x}^*) + \langle \nabla f(\boldsymbol{x}^*),\ \boldsymbol{x} - \boldsymbol{x}^* \rangle + \frac{1}{2}\mu \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\|^2 \\ &= f(\boldsymbol{x}^*) + \frac{1}{2}\mu \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\|^2 . \end{aligned}$$

□

Property of Strongly Convex Function

The following result justifies the addition of strongly convex functions.

Lemma 31

If $f_1 \in \mathcal{S}^1_{\mu_1}(\mathbb{R}^n)$, $f_2 \in \mathcal{S}^1_{\mu_2}(\mathbb{R}^n)$, and $\alpha, \beta \geq 0$, then

$$f = \alpha f_1 + \beta f_2 \in \mathcal{S}^1_{\alpha\mu_1+\beta\mu_2}(\mathbb{R}^n).$$

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

$$\begin{aligned} f_1(\boldsymbol{y}) &\geq f_1(\boldsymbol{x}) + \langle \nabla f_1(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2}\mu_1 \|\boldsymbol{y} - \boldsymbol{x}\|^2, \\ f_2(\boldsymbol{y}) &\geq f_2(\boldsymbol{x}) + \langle \nabla f_2(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2}\mu_2 \|\boldsymbol{y} - \boldsymbol{x}\|^2. \end{aligned}$$

It remains to add these equations multiplied respectively by α and β. $\square$

Property of Strongly Convex Function

Remark.

Note that the class $\mathcal{S}_0^1(\mathbb{R}^n)$ coincides with $\mathcal{F}^1(\mathbb{R}^n)$. Therefore addition of a convex function to a strongly convex function gives a strongly convex function with the same convexity parameter.

Other Properties

Theorem 32 (Theorem 2.1.10)

If $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$, then for any $\boldsymbol{x}$ and $\boldsymbol{y}$ from $\mathbb{R}^n$ we have

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2\mu} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\|^2, \tag{24}$$

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \leq \frac{1}{\mu} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\|^2. \tag{25}$$

Proof. Let us fix some $\boldsymbol{x} \in \mathbb{R}^n$. Consider the function

$$\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} \rangle \in \mathcal{S}^1_\mu(\mathbb{R}^n).$$

Other Properties

Remark. Consider a linear function $f(\boldsymbol{y}) = \langle \boldsymbol{a}, \ \boldsymbol{y} \rangle$. If

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

hold, then we have

$$\begin{aligned} \langle \boldsymbol{a}, \ \boldsymbol{y} \rangle &\geq \langle \boldsymbol{a}, \ \boldsymbol{x} \rangle + \langle \boldsymbol{a}, \ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2}\mu \left\| \boldsymbol{y} - \boldsymbol{x} \right\|^2 \\ &= \langle \boldsymbol{a}, \ \boldsymbol{y} \rangle + \frac{1}{2}\mu \left\| \boldsymbol{y} - \boldsymbol{x} \right\|^2 . \end{aligned}$$

Thus, $\mu = 0$, and $\langle \boldsymbol{a}, \ \boldsymbol{y} \rangle \in \mathcal{S}_0^1(\mathbb{R}^n)$.

Other Properties

$$\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} \rangle \in \mathcal{S}_{\mu}^{1}(\mathbb{R}^n).$$

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2\mu} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\|^2 \tag{24}$$

Proof. (Continued) Since $\nabla \phi(\boldsymbol{x}) = 0$, in view of (23) for any $\boldsymbol{y} \in \mathbb{R}^n$ we have that

$$\begin{aligned} \phi(\boldsymbol{x}) &= \min_{\boldsymbol{v}} \phi(\boldsymbol{v}) \geq \min_{\boldsymbol{v}} \left[\phi(\boldsymbol{y}) + \langle \nabla \phi(\boldsymbol{y}),\ \boldsymbol{v} - \boldsymbol{y} \rangle + \frac{1}{2}\mu \left\| \boldsymbol{v} - \boldsymbol{y} \right\|^2 \right] \\ &= \phi(\boldsymbol{y}) - \frac{1}{2\mu} \left\| \nabla \phi(\boldsymbol{y}) \right\|^2 \end{aligned}$$

and that is exactly (24). $\left(\nabla \phi(\boldsymbol{y}) + \mu(\boldsymbol{v} - \boldsymbol{y}) = 0, \text{ thus } \boldsymbol{v}^* = \boldsymbol{y} - \frac{\nabla \phi(\boldsymbol{y})}{\mu} \right)$

Other Properties

$$f(\boldsymbol{y}) \le f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2\mu} \| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \|^2 , \tag{24}$$

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \le \frac{1}{\mu} \| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \|^2 . \tag{25}$$

Proof. (Continued) Adding two copies of (24) with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged we get (25). □

Equivalent Definitions

Theorem 33

Let f be continuously differentiable. Both conditions below, holding for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\alpha \in [0, 1]$, are equivalent to inclusion $\mathcal{S}^1_\mu(\mathbb{R}^n)$

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \geq \mu \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2 , \tag{26}$$

$$\begin{aligned} \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \geq f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \\ + \alpha(1-\alpha)\frac{\mu}{2} \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2 . \end{aligned} \tag{27}$$

Equivalent Definitions

Finally, the second-order characterization of the class $\mathcal{S}_{\mu}^{2}(\mathbb{R}^n) \subseteq \mathcal{S}_{\mu}^{1}(\mathbb{R}^n)$ is as follows.

Theorem 34

Two times continuously differentiable function f belongs to the class $\mathcal{S}_{\mu}^{2}(\mathbb{R}^n)$ if and only if $\boldsymbol{x} \in \mathbb{R}^n$

$$\nabla^2 f(\boldsymbol{x}) \succeq \mu I_n. \tag{28}$$

Proof. Apply (26):

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \geq \mu \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2. \tag{26}$$

□

Equivalent Definitions

Remark. From (26), we have $\langle \nabla f(\boldsymbol{x} + \tau \boldsymbol{s}) - \nabla f(\boldsymbol{x}),\ \tau \boldsymbol{s} \rangle \geq \mu \|\tau \boldsymbol{s}\|^2$, and therefore

$$\frac{\langle \nabla f(\boldsymbol{x} + \tau \boldsymbol{s}) - \nabla f(\boldsymbol{x}),\ \tau \boldsymbol{s} \rangle}{\tau^2 \|\boldsymbol{s}\|^2} \geq \mu \quad \Rightarrow \quad \frac{\langle \nabla f(\boldsymbol{x} + \tau \boldsymbol{s}) - \nabla f(\boldsymbol{x}),\ \boldsymbol{s} \rangle}{\tau \|\boldsymbol{s}\|^2} \geq \mu.$$

As $\tau \to 0$, the above relation holds. Since

$$\frac{\nabla f(\boldsymbol{x} + \tau \boldsymbol{s}) - \nabla f(\boldsymbol{x})}{\tau} \to \nabla^2 f(\boldsymbol{x})\boldsymbol{s}, \quad \text{we obtain } \frac{\langle \nabla^2 f(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s} \rangle}{\|\boldsymbol{s}\|^2} \geq \mu.$$

That implies

$$\langle \nabla^2 f(\boldsymbol{x})\boldsymbol{s},\ \boldsymbol{s} \rangle \geq \mu \|\boldsymbol{s}\|^2 = \langle \mu I_n \boldsymbol{s},\ \boldsymbol{s} \rangle.$$

Examples

Example 35 (Example 2.1.1)

1. $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|^2$ belong to $\mathcal{S}_1^2(\mathbb{R}^n)$ since $\nabla^2 f(\boldsymbol{x}) = I_n$.
2. Let symmetric matrix A satisfy the condition: $\mu I_n \preceq A \preceq L I_n$. Then

$$f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a},\ \boldsymbol{x}\rangle + \frac{1}{2}\langle A\boldsymbol{x},\ \boldsymbol{x}\rangle \in \mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^n) \subset \mathcal{S}_{\mu,L}^{1,1}(\mathbb{R}^n),$$

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

Other examples can be obtained as a sum of convex and strongly convex functions.

Part II
Smooth and Strongly Convex Function

The Class $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$

For us the most interesting functional class is $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$. This class is described by the following inequalities:

$$\boxed{\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle} \geq \mu \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2, \qquad \text{(see 26)}$$

and

$$\left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\| \leq L \left\| \boldsymbol{x} - \boldsymbol{y} \right\|. \qquad \text{(see Def. of Lipschitz Continumous)}$$

or (Lipschitz and convex)

$$\boxed{\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle} \leq L \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2, \qquad \text{(see 18)}$$

The value $Q_f = \frac{L}{\mu} \geq 1$ is called the condition number of function f.

Class $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$

Remark. (1) See definition of strongly convex function:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2}\mu \|\boldsymbol{y} - \boldsymbol{x}\|^2 . \tag{23}$$

We have

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

By

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

of Lemma 1.2.3 of Nesterov [2003], we have

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

Class $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$

Remark. (2) Also, for $f \in \mathcal{S}^{1}_{\mu}(\mathbb{R}^n)$ we have

$$\boxed{\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle} \le \frac{1}{\mu} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 . \tag{25}$$

And in view of the equivalent definition of $\mathcal{F}^{1,1}_{L}(\mathbb{R}^n)$, we have

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

Property of Smooth and Strongly Convex Function

Theorem 36 (Theorem 2.1.12)

If $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, then for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have

$$\begin{aligned} \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \geq \frac{\mu L}{\mu + L} \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2 \\ + \frac{1}{\mu + L} \left\| \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \right\|^2 . \end{aligned} \tag{29}$$

Proof.

Denote $\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{1}{2}\mu \left\| \boldsymbol{x} \right\|^2$. Then $\nabla \phi(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) - \mu \boldsymbol{x}$.

Property of Smooth and Strongly Convex Function

Proof. (Continued.) We check $\phi \in \mathcal{F}_{L-\mu}^{1,1}(\mathbb{R}^n)$ as follows.

$$
\begin{aligned}
&\phi(\boldsymbol{x}) - \phi(\boldsymbol{z}) - \langle \nabla \phi(\boldsymbol{z}),\ \boldsymbol{x} - \boldsymbol{z} \rangle \\
&= f(\boldsymbol{x}) - \frac{1}{2}\mu \|\boldsymbol{x}\|^2 - \left(f(\boldsymbol{z}) - \frac{1}{2}\mu \|\boldsymbol{z}\|^2\right) - \langle \nabla f(\boldsymbol{z}) - \mu \boldsymbol{z},\ \boldsymbol{x} - \boldsymbol{z} \rangle \\
&= \underbrace{f(\boldsymbol{x}) - f(\boldsymbol{z}) - \langle \nabla f(\boldsymbol{z}),\ \boldsymbol{x} - \boldsymbol{z} \rangle}_{\leq \frac{L}{2}\|\boldsymbol{x}-\boldsymbol{z}\|^2} - \frac{\mu}{2} \underbrace{\left\{\|\boldsymbol{x}\|^2 - \|\boldsymbol{z}\|^2 + 2\langle \boldsymbol{z},\ \boldsymbol{z} - \boldsymbol{x} \rangle\right\}}_{=\|\boldsymbol{x}-\boldsymbol{z}\|^2} \\
&\leq \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{z}\|^2 - \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{z}\|^2 = \frac{L-\mu}{2} \|\boldsymbol{x} - \boldsymbol{z}\|^2 .
\end{aligned}
$$

Property of Smooth and Strongly Convex Function

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \geq \frac{\mu L}{\mu + L} \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \frac{1}{\mu + L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 . \quad (29)$$

Proof. (Continued.) If $\mu = L$ then (29) is proved (half of (26) plus half of (17))

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

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \geq \mu \|\boldsymbol{x} - \boldsymbol{y}\|^2 , \quad (26)$$

Property of Smooth and Strongly Convex Function

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\, \boldsymbol{x} - \boldsymbol{y} \rangle \geq \frac{\mu L}{\mu + L} \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \frac{1}{\mu + L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2 . \tag{29}$$

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

$$\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{1}{2}\mu \|\boldsymbol{x}\|^2$$

Proof. (Continued) If $\mu < L$, then by (17), we have

$$\langle \nabla \phi(\boldsymbol{x}) - \nabla \phi(\boldsymbol{y}),\, \boldsymbol{x} - \boldsymbol{y} \rangle \geq \frac{1}{L - \mu} \|\nabla \phi(\boldsymbol{x}) - \nabla \phi(\boldsymbol{y})\|^2 \tag{30}$$

and that is exactly (29). □

Property of Smooth and Strongly Convex Function

Remark. Consider the left side of (30). We have

$$\langle \nabla \phi(\boldsymbol{x}) - \nabla \phi(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle = \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle - \mu \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2.$$

The right side of (30) is equal to

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

Thus from (30), we have

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

Part III
Conclusion

Upper Bounds on Functional Components

- Lipschitz Continuity: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,
  - Zeroth-order Condition:
  $$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \|\boldsymbol{x} - \boldsymbol{y}\|$$
  - First-order Condition:
  $$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L \|\boldsymbol{x} - \boldsymbol{y}\|$$
  - p-order Condition:
  $$\|\nabla^p f(\boldsymbol{x}) - \nabla^p f(\boldsymbol{y})\|_* \leq L \|\boldsymbol{x} - \boldsymbol{y}\|, (p \geq 2)$$

Lower Bounds on Functional Components

- Convexity: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \alpha \in [0,1]$,
  - Zeroth-order Condition:
  $$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) - f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \geq 0.$$
  - First-order Condition:
  $$D_f(\boldsymbol{x}, \boldsymbol{y}) \triangleq f(\boldsymbol{y}) - \{f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle\} \geq 0 \text{ and}$$
  $$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \geq 0.$$
  - Second-order Condition:
  $$\nabla^2 f(x) \succeq 0.$$

Lower Bounds on Functional Components

- Strong Convexity (SC): $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,
  - Zeroth-order Condition:
  $$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) - f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \geq \alpha(1-\alpha)\frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2 .$$
  - First-order Condition:
  $$D_f(\boldsymbol{x}, \boldsymbol{y}) \geq \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2 , \text{ and } \langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x}-\boldsymbol{y}\rangle \geq \mu\|\boldsymbol{x}-\boldsymbol{y}\|^2 .$$
  - Second-order Condition:
  $$\nabla^2 f(\boldsymbol{x}) \succeq \mu I_n .$$

Other Lower Bounds on Functional Components

- Weak Strong Convexity (WSC): $D_f(\boldsymbol{x}^*, \boldsymbol{y}) \geq \frac{\mu}{2} \|\boldsymbol{x}^* - \boldsymbol{x}\|^2, \forall \boldsymbol{x} \in \mathbb{R}^n$.
- Restricted Secant Inequality (RSI): $\langle \nabla f(\boldsymbol{x}),\ \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq \mu \|\boldsymbol{x}^* - \boldsymbol{x}\|^2, \forall \boldsymbol{x} \in \mathbb{R}^n$.
- Polyak-Łojaciewicz (PL): $\frac{1}{2} \|\nabla f(\boldsymbol{x})\|^2 \geq \mu(f(\boldsymbol{x}) - f(\boldsymbol{x}^*)), \forall \boldsymbol{x} \in \mathbb{R}^n$.
- Error Bounds (EB): $\|\nabla f(\boldsymbol{x})\| \geq \mu \|\boldsymbol{x} - \boldsymbol{x}^*\|, \forall \boldsymbol{x} \in \mathbb{R}^n$.
- Quadratic Growth (QG): $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq \frac{\mu}{2} \|\boldsymbol{x}^* - \boldsymbol{x}\|^2, \forall \boldsymbol{x} \in \mathbb{R}^n$.

- (SC) $\rightarrow$ (WSC) $\rightarrow$ (RSI) $\rightarrow$ (EB) $\equiv$ (PL) $\rightarrow$ (QG)
- If f is convex, (RSI) $\equiv$ (EB) $\equiv$ (PL) $\equiv$ (QG)

Strongly Convex and Polyak-Łojaciewicz

Theorem 37

μ-Strongly convex functons are μ-PŁ.

Proof.

Suppose that f is μ-strongly convex, then for $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^n$, we have

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

Thus, we have

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

Strongly Convex and Polyak-Łojaciewicz

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

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

Proof. (Continued.)

LHS is

$$
\min_{\boldsymbol{y}} \{f(\boldsymbol{y})\} = \boxed{f(\boldsymbol{x}^*)}.
$$

And the minimum of RHS can be solved by

$$
\nabla \text{RHS} = \nabla f(\boldsymbol{x}) + \mu(\boldsymbol{y} - \boldsymbol{x}) = 0
$$

Strongly Convex and Polyak-Łojaciewicz

$$\boxed{\begin{gathered} f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{\mu}{2} \|\boldsymbol{y} - \boldsymbol{x}\|^2 . \\ \frac{1}{2} \|\nabla f(\boldsymbol{x})\|^2 \geq \mu(f(\boldsymbol{x}) - f(\boldsymbol{x}^*)). \end{gathered}}$$

Proof. (Continued.)

Thus, we have the optimal $\hat{\boldsymbol{y}} = \boldsymbol{x} - \frac{1}{\mu} \nabla f(\boldsymbol{x})$. Put back $\hat{\boldsymbol{y}}$ to RHS, we get

$$f(\boldsymbol{x}) - \frac{1}{\mu} \|\nabla f(\boldsymbol{x})\|^2 + \frac{1}{2\mu} \|\nabla f(\boldsymbol{x})\|^2 = \boxed{f(\boldsymbol{x}) - \frac{1}{2\mu} \|\nabla f(\boldsymbol{x})\|^2}.$$

Overall, we arrive the result by LHS $\geq$ RHS.

Strongly Convex and Polyak-Łojaciewicz

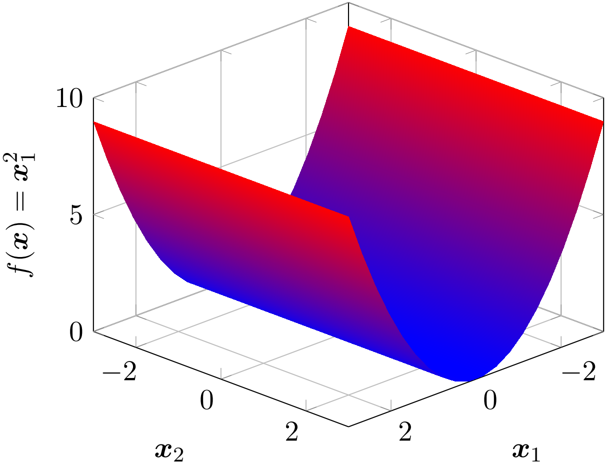
Example.
The function $f(\boldsymbol{x}) = \boldsymbol{x}_1^2$, where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^2$ is convex. However, it is not strongly convex since

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 2, 0 \\ 0, 0 \end{pmatrix}$$

is semi-positve definition, but not postive definition.
The gradient of $f(\boldsymbol{x})$ is $(2\boldsymbol{x}_1, 0)^\top$, that implies one of the optimization pointer is $x^* = (0, 0)$.
Thus set $\mu = 2$, we have that

$$\frac{1}{2}\|\nabla f(\boldsymbol{x})\|^2 = \frac{1}{2}4\boldsymbol{x}_1^2 = 2\boldsymbol{x}_1^2 \geq \boldsymbol{x}_1^2 = 2(f(\boldsymbol{x}) - f(\boldsymbol{x}^*)).$$

Strongly Convex and Polyak-Łojaciewicz



References I

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

Zhi-Quan Luo and Paul Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

Hui Zhang and Wotao Yin. Gradient methods for convex minimization: better rates under weaker conditions. *arXiv preprint arXiv:1303.4645*, 2013.

Chenxin Ma, Rachael Tappenden, and Martin Takáč. Linear convergence of randomized feasible descent methods under the weak strong convexity assumption. *The Journal of Machine Learning Research*, 17(1):8138–8161, 2016.

Ion Necoara, Yu Nesterov, and Francois Glineur. Linear convergence of first order methods for non-strongly convex optimization. *Mathematical Programming*, 175(1):69–107, 2019.

References II

Mihai Anitescu. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10(4):1116–1135, 2000.

Boris T Polyak. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics*, 3(4):864–878, 1963.

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

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

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

Thank You!

Email:qianhui@zju.edu.cn