OPTIMIZATION (SI 416) – LECTURE 3

Harsha Hutridurga

IIT Bombay

STORY SO FAR

♣ Take a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$

$$\boxed{\begin{array}{ccc} x_* \text{ is a} & & \nabla f(x_*) = 0 \\ \text{local minimizer} & \Longrightarrow & \text{and} \\ \text{of } f & & \big\langle \nabla^2 f(x_*)p, p \big\rangle \geq 0 \text{ for all } p \in \mathbb{R}^n \end{array}}$$

$$\boxed{\begin{array}{ccc} \nabla f(x_*) = 0 & & x_* \text{ is a} \\ \text{and} & \Longrightarrow & \text{strict local minimizer} \\ \big\langle \nabla^2 f(x_*)p, p \big\rangle > 0 \quad \forall p \in \mathbb{R}^n \setminus \{0\} & & \text{of } f \end{array}}$$

STORY SO FAR (CONTD.)

♣ Take a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ which is also convex on $\mathbb{R}^n$

$$\boxed{\begin{array}{ccc} x_* \text{ is a} & & x_* \text{ is a} \\ \text{local minimizer} & \Longleftrightarrow & \text{global minimizer} \\ \text{of } f & & \text{of } f \end{array}}$$

$$\boxed{\begin{array}{ccc} x_* \text{ satisfies} & & x_* \text{ satisfies} \\ f(x_*) \leq f(x) \text{ for all } x \in \mathbb{R}^n & \Longleftrightarrow & \nabla f(x_*) = 0 \end{array}}$$

STORY SO FAR (CHARACTERIZATION OF CONVEX FUNCTIONS)

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function.

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x, y \in \mathbb{R}^n, \\ \text{convex} & \iff & f(y) \geq f(x) + \langle \nabla f(x), (y - x) \rangle \end{array}}$$

Theorem

A twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the Hessian matrix $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$.

- Take a convex function f and let $x \in \mathbb{R}^n$.
- Define a function $g : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$g(y) := f(y) - \langle \nabla f(x), (y - x) \rangle$$

- Note that $y \mapsto -\langle \nabla f(x), (y - x) \rangle$ is linear. Hence is convex
- Thus g is a convex function
- Observe that for all $y \in \mathbb{R}^n$, we have

$$\nabla g(y) = \nabla f(y) - \nabla f(x) \quad \text{and} \quad \nabla^2 g(y) = \nabla^2 f(y)$$

- Note in particular that $\nabla g(x) = 0$

- Recall that g is a convex function and we have shown $\nabla g(x) = 0$
- Hence x is a global minimizer of g
- Second order necessary condition then implies that the Hessian $\nabla^2 g(x)$ is positive semidefinite
- Recall that we had $\nabla^2 g(x) = \nabla^2 f(x)$ and that x was arbitrary
- We have thus demonstrated that $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$
- Next, we assume that $\nabla^2 f(x)$ is positive semidefinite and then show that f is convex
- Taking $x, y \in \mathbb{R}^n$ and employing the Taylor's theorem we obtain

$$
\begin{aligned}
f(y) &= f(x + y - x) \\
&= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x + s(y - x))(y - x), (y - x) \rangle \\
&\geq f(x) + \langle \nabla f(x), (y - x) \rangle
\end{aligned}
$$

This proves convexity of f.

STRICT CONVEXITY

♣ A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be STRICTLY CONVEX on $\mathbb{R}^n$ if for all $x, y \in \mathbb{R}^n$, $x \neq y$,

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y) \quad \text{for } \alpha \in (0,1).$$

Theorem

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if and only if

$$f(y) > f(x) + \langle \nabla f(x), (y-x) \rangle \qquad \textit{for all } x, y \in \mathbb{R}^n,\ x \neq y$$

♣ The proof of the above result is exactly similar to the convex case wherein we replace inequalities by strict inequalities

Theorem

A strictly convex function $f : \mathbb{R}^n \to \mathbb{R}$ has at most one global minimizer

- Above statement doesn't guarantee a global minimizer
- It says: if there is a global minimizer of f, then it must be unique
- Suppose there are two distinct global minimizers of f, say x and y
- That is,

$$f(x) = f(y) \leq f(z) \qquad \text{for all } z \in \mathbb{R}^n.$$

- Let us take in particular $z = \frac{x+y}{2}$. Then by strict convexity,

$$f(z) = f\left(\frac{x+y}{2}\right) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = f(x)$$

- Thus we arrive at a contradiction

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function such that its Hessian matrix $\nabla^2 f(x)$ is positive definite for all $x \in \mathbb{R}^n$.
Then, the function f is strictly convex.

- Take $x, y \in \mathbb{R}^n$ such that $x \neq y$.
- Employing the Taylor's theorem we obtain

$$\begin{aligned} f(y) &= f(x + y - x) \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x + s(y-x))(y-x), y - x \rangle \\ &> f(x) + \langle \nabla f(x), (y - x) \rangle \end{aligned}$$

  This proves strict convexity of f.
- Not every strictly convex function has a positive definite Hessian
- Consider $f(x) = x^4$ whose $f''(0) = 0$

STRICT CONVEXITY OF x^4

- Note: $g(x) = x^2$ is strictly convex as $g''(x) = 2 > 0$ for all $x \in \mathbb{R}$
- Goal is show that for any $x, y \in \mathbb{R}$ with $x \neq y$, there holds

$$(\alpha x + (1-\alpha)y)^4 < \alpha x^4 + (1-\alpha)y^4 \qquad \text{for all } \alpha \in (0,1).$$

- Note that strict convexity of x^2 implies

$$(\alpha x + (1-\alpha)y)^2 < \alpha x^2 + (1-\alpha)y^2 \qquad \text{for all } \alpha \in (0,1).$$

- Squaring on both sides and using the fact that x^2 is an increasing function on $[0,\infty)$, we get

$$(\alpha x + (1-\alpha)y)^4 < \left(\alpha x^2 + (1-\alpha)y^2\right)^2$$

- Again using the strict convexity of x^2, we arrive at

$$(\alpha x + (1-\alpha)y)^4 < \alpha x^4 + (1-\alpha)y^4.$$

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be STRONGLY CONVEX if there exists a $\lambda > 0$ such that

$$f(x) - \lambda \|x\|^2 \quad \text{is convex.}$$

- Here

$$\|x\|^2 := \sum_{i=1}^{n} x_i^2$$

- Observe that $g(x) = \|x\|^2$ is strictly convex as

$$\nabla^2 g(x) = 2I, \qquad \text{where } I \text{ is the identity matrix}$$

- Given a convex function $f : \mathbb{R}^n \to \mathbb{R}$, we can build a strongly convex function

$$f(x) + \mu \|x\|^2 \qquad \text{for any } \mu > 0$$

♣ Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable strongly convex function

i.e. $g(x) := f(x) - \lambda\|x\|^2$ is convex for some $\lambda > 0$.

i.e. $\nabla^2 g(x) = \nabla^2 f(x) - 2\lambda I$ is positive semidefinite.

i.e. $\left\langle \nabla^2 g(x)p, p \right\rangle = \left\langle \nabla^2 f(x)p, p \right\rangle - \langle 2\lambda I p, p\rangle \geq 0$ for all $p \in \mathbb{R}^n$.

$\implies \left\langle \nabla^2 f(x)p, p \right\rangle \geq 2\lambda\|p\|^2$ for all $p \in \mathbb{R}^n$.

Lemma

If $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable strongly convex function, then there exists a $\lambda > 0$ such that

$$\left\langle \nabla^2 f(x)p, p \right\rangle \geq 2\lambda\|p\|^2 \quad \textit{for all } p \in \mathbb{R}^n.$$

- ♣ Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable strongly convex function
- ♣ Taking $x, y \in \mathbb{R}^n$ and employing the Taylor's theorem we obtain

$$\begin{aligned} f(y) &= f(x + y - x) \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x + s(y - x))(y - x), y - x \rangle \end{aligned}$$

  for some $s \in (0, 1)$.
- ♣ Employing the lemma from the previous slide, we deduce that there exists a $\lambda > 0$ such that

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

Lemma

If $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable strongly convex function, then there exists a $\lambda > 0$ such that for all $x, y \in \mathbb{R}^n$,

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

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable strongly convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \lambda \|y - x\|^2 \qquad \text{for all } x, y \in \mathbb{R}^n$$

for some $\lambda > 0$.

- Cauchy-Schwarz inequality says: for any $u, v \in \mathbb{R}^n$,

$$|\langle u, v \rangle| \leq \|u\| \, \|v\|$$

i.e.

$$-\|u\| \, \|v\| \leq \langle u, v \rangle \leq \|u\| \, \|v\|$$

- Using this in the earlier property of strong convexity, we deduce

$$f(y) \geq f(x) - \|\nabla f(x)\| \, \|y - x\| + \lambda \|y - x\|^2 \qquad \text{for all } x, y \in \mathbb{R}^n$$

- Suppose x_* is a global minimizer of f
- Thanks to the inequality from the previous slide, we have

$$f(x_*) \geq f(x) - \|\nabla f(x)\| \, \|x_* - x\| + \lambda \, \|x_* - x\|^2 \qquad \text{for all } x \in \mathbb{R}^n$$

- Thanks to x_* being a global minimizer, the above inequality yields

$$\|x_* - x\| \leq \frac{1}{\lambda} \|\nabla f(x)\| \qquad \text{for all } x \in \mathbb{R}^n$$

- From the last inequality, we can conclude that
  - smaller the gradient of f at a point, closer it is to a global minimizer
  - there can at most be one global minimizer of a strongly convex function

- Consider a function $f : \mathbb{R}^n \to \mathbb{R}$.

f is strongly convex	$\implies$	f is strictly convex	$\implies$	f is convex

- Strict convexity implying convexity is obvious
- Suppose f is strongly convex. Then, there exists a $\lambda > 0$ such that

$$
\begin{aligned}
f(\alpha x + (1-\alpha)y) - \lambda \left\|\alpha x + (1-\alpha)y\right\|^2 & \\
&\leq \alpha \left(f(x) - \lambda \|x\|^2 \right) + (1-\alpha)\left(f(y) - \lambda \|y\|^2 \right) \\
&= \alpha f(x) + (1-\alpha) f(y) - \lambda \left(\alpha \|x\|^2 + (1-\alpha)\|y\|^2 \right)
\end{aligned}
$$

- Rearranging the above inequality yields

$$
\begin{aligned}
f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) & \\
&+ \lambda \Big(\|\alpha x + (1-\alpha)y\|^2 - \alpha\|x\|^2 - (1-\alpha)\|y\|^2 \Big)
\end{aligned}
$$

STRONGLY STRICTLY AND JUST (CONTD.)

$$
\begin{aligned}
f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \\
+ \lambda\Big(\|\alpha x + (1-\alpha)y\|^2 - \alpha\|x\|^2 - (1-\alpha)\|y\|^2 \Big)
\end{aligned}
$$

- As $\|x\|^2$ is strictly convex, the term in red is strictly negative
- Hence f is strictly convex

REVERSE IMPLICATIONS ARE NOT TRUE

- $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x$ is convex but not strictly convex
- $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^4$ is strictly convex but not strongly convex
  - There does not exist a $\lambda > 0$ such that $x^4 - \lambda x^2$ is convex on $\mathbb{R}$

RECAP (CHARACTERIZATION OF CONVEX FUNCTIONS)

♣ Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function.

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x, y \in \mathbb{R}^n, \\ \text{convex} & \iff & f(y) \geq f(x) + \nabla f(x) \cdot (y - x) \end{array}}$$

♣ Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function.

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x \in \mathbb{R}^n, \\ \text{convex} & \iff & \left\langle \nabla^2 f(x) p, p \right\rangle \geq 0 \text{ for all } p \in \mathbb{R}^n \end{array}}$$

♣ A function f is said to be strictly convex if for all $x, y \in \mathbb{R}^n$, $x \neq y$,

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y) \quad \text{for} \quad \alpha \in (0,1)$$

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x, y \in \mathbb{R}^n,\ x \neq y, \\ \text{strictly convex} & \iff & f(y) > f(x) + \nabla f(x) \cdot (y - x) \end{array}}$$

♣ Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function.

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x \in \mathbb{R}^n, \\ \text{strictly convex} & \Longleftarrow & \left\langle \nabla^2 f(x) p, p \right\rangle > 0 \text{ for all } p \in \mathbb{R}^n \setminus \{0\} \end{array}}$$

♣ A strictly convex function has at most one global minimizer

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be STRONGLY CONVEX if there exists a $\lambda > 0$ such that

$$f(x) - \lambda\|x\|^2 \quad \text{is convex.}$$

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function.

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x \in \mathbb{R}^n, \\ \text{strongly convex} & \Longrightarrow & \left\langle \nabla^2 f(x)p, p \right\rangle \geq 2\lambda\|p\|^2 \quad \text{for all } p \in \mathbb{R}^n \end{array}}$$

$$\boxed{\begin{array}{ccc} f \text{ is} & & \text{for every } x, y \in \mathbb{R}^n, \\ \text{strongly convex} & \Longrightarrow & f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \lambda \|y - x\|^2 \end{array}}$$

- Consider a function $f : \mathbb{R}^n \to \mathbb{R}$.

$$\boxed{\begin{array}{ccccc} f \text{ is} & & f \text{ is} & & f \text{ is} \\ \text{strongly convex} & \Longrightarrow & \text{strictly convex} & \Longrightarrow & \text{convex} \end{array}}$$

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex, i.e. there exists $\lambda > 0$ s.t.

$$
\begin{aligned}
f(\alpha x + (1-\alpha)y) &- \lambda \left\| \alpha x + (1-\alpha)y \right\|^2 \\
&\leq \alpha \left(f(x) - \lambda \|x\|^2 \right) + (1-\alpha) \left(f(y) - \lambda \|y\|^2 \right) \\
&= \alpha f(x) + (1-\alpha) f(y) - \lambda \left(\alpha \|x\|^2 + (1-\alpha)\|y\|^2 \right)
\end{aligned}
$$

- Rearranging the above inequality yields

$$
\begin{aligned}
f(\alpha x + (1-\alpha)y) \leq \alpha f(x) &+ (1-\alpha) f(y) \\
&+ \lambda \Big(\left\| \alpha x + (1-\alpha)y \right\|^2 - \alpha \|x\|^2 - (1-\alpha)\|y\|^2 \Big)
\end{aligned}
$$

- Note that

$$
\left\| \alpha x + (1-\alpha)y \right\|^2 = \alpha^2 \left\| x \right\|^2 + (1-\alpha)^2 \left\| y \right\|^2 + 2\alpha(1-\alpha) \langle x, y \rangle
$$

- Hence we get

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

- Putting it all together, we obtain

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) - \lambda\alpha(1-\alpha)\|x-y\|^2$$

Lemma

If $f : \mathbb{R}^n \to \mathbb{R}$ is a strongly convex function, then there exists a $\lambda > 0$ such that for all $x, y \in \mathbb{R}^n$ and $\alpha \in [0, 1]$,

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) - \lambda\alpha(1-\alpha)\|x-y\|^2$$

END OF LECTURE 3

THANK YOU FOR YOUR ATTENTION