OPTIMIZATION (SI 416) – LECTURE 2

Harsha Hutridurga

IIT Bombay

- ♣ Take a once continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$
  - Let $x, p \in \mathbb{R}^n$
  - Then, there exists a $\ell \in (0,1)$ such that

$$f(x+p) = f(x) + \langle \nabla f(x + \ell p), p \rangle$$

- ♣ Take a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$
  - Let $x, p \in \mathbb{R}^n$
  - Then, there exists a $t \in (0,1)$ such that

$$f(x+p) = f(x) + \langle \nabla f(x), p \rangle + \frac{1}{2} \left\langle \nabla^2 f(x + tp) p, p \right\rangle$$

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

$$\boxed{\begin{array}{c} x_* \text{ is a} \\ \text{local minimizer} \\ \text{of } f \end{array} \quad \Longrightarrow \quad \nabla f(x_*) = 0}$$

♣ A $n \times n$ matrix B is called

- positive definite if

$$\langle Bv, v\rangle > 0 \qquad \text{for all } v \in \mathbb{R}^n \setminus \{0\}$$

- positive indefinite (semidefinite) if

$$\langle Bv, v\rangle \geq 0 \qquad \text{for all } v \in \mathbb{R}^n \setminus \{0\}$$

Theorem (Second order necessary condition)

If x_* is a local minimizer of f and if f is twice continuously differentiable in an open neighbourhood of x_*, then

$$\nabla^2 f(x_*) \text{ is positive semidefinite.}$$

♣ This isn't a sufficient condition: again think of $f(x) = x^3$ at $x = 0$.

- Suppose $\nabla^2 f(x_*)$ is not positive semidefinite
- Then there exists a $p \in \mathbb{R}^n$ such that

$$\left\langle \nabla^2 f(x_*)p, p \right\rangle < 0.$$

- As $\nabla^2 f$ is continuous at x_*, there exists a $T > 0$ such that

$$\left\langle \nabla^2 f(x_* + sp)p, p \right\rangle < 0 \qquad \text{for all } s \in [0, T].$$

- Taylor's theorem says for all $t_* \in (0, T]$, there exists a $s_* \in (0, t_*)$

$$f(x_* + t_* p) = f(x_*) + t_* \langle \nabla f(x_*), p \rangle + \frac{t_*^2}{2} \left\langle \nabla^2 f(x_* + s_* p)p, p \right\rangle$$

- Recall that first order necessary condition implies that $\nabla f(x_*) = 0$
- Putting it all together, we arrive at a contradiction:

$$f(x_* + t_* p) < f(x_*) \qquad \text{for all } t_* \in (0, T]$$

Theorem

Suppose f is twice continuously differentiable in an open neighbourhood of a point $x_* \in \mathbb{R}^n$.
Suppose further that

$$\nabla f(x_*) = 0 \quad \text{and} \quad \nabla^2 f(x_*) \text{ is positive definite.}$$

Then, x_* is a strict local minimizer of f.

- ♣ This isn't a necessary condition: think of $f(x) = x^4$ at $x = 0$.
- ♣ Because $\nabla^2 f$ is continuous at x_* and as it is positive definite at x_*, there exists a $r > 0$ such that

$$\nabla^2 f(z) \text{ is positive definite for all } z \in B_r(x_*).$$

- ♣ Here $B_r(x_*)$ denotes the open ball of radius r centered at x_*

SECOND ORDER SUFFICIENT CONDITION (CONTD.)

- Taking $p \in \mathbb{R}^n$ such that $\|p\| < r$ and using Taylor's theorem,

$$
\begin{aligned}
f(x_* + p) &= f(x_*) + \langle \nabla f(x_*), p \rangle + \frac{1}{2} \left\langle \nabla^2 f(x_* + sp)p, p \right\rangle \\
&= f(x_*) + \frac{1}{2} \left\langle \nabla^2 f(x_* + sp)p, p \right\rangle
\end{aligned}
$$

  for some $s \in (0, 1)$.
- Using the positive definiteness of $\nabla^2 f$ in a neighbourhood of x_* implies that

$$
f(x_* + p) > f(x_*)
$$

  for any $p \in \mathbb{R}^n$ satisfying $\|p\| < r$.
- Hence x_* is a strict local minimizer of f.

NOTION OF CONVEXITY

- A set $\Omega \subset \mathbb{R}^n$ is said to be CONVEX if

$$x, y \in \Omega \implies \alpha x + (1-\alpha) y \in \Omega \quad \text{for all } \alpha \in [0,1].$$

- Take a convex set $\Omega \subset \mathbb{R}^n$.
- A function $f : \Omega \to \mathbb{R}$ is said to be CONVEX on Ω if
  - for any two points $x, y \in \Omega$, we have

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

- f is STRICTLY CONVEX if the above inequality is strict whenever $x \neq y$ and $\alpha \in (0,1)$
- f is said to be CONCAVE if $-f$ is convex

Theorem

Suppose f is a convex function on $\mathbb{R}^n$.
Any local minimizer x_* of f is a global minimizer of f.

- Suppose x_* is a local but not a global minimizer of f
- That is, there exists a $y \in \mathbb{R}^n$ such that

$$f(y) < f(x_*)$$

- Consider a point on the line segment joining x_* and y

$$x = \lambda y + (1 - \lambda)x_* \quad \text{for some } \lambda \in (0, 1]$$

- By convexity of f, we have

$$f(x) \leq \lambda f(y) + (1 - \lambda)f(x_*) < f(x_*)$$

- The strict inequality is thanks to the earlier observation

CONVEX OPTIMIZATION (CONTD.)

- We have shown that for any point x on the line segment joining x_* and y,

$$f(x) < f(x_*)$$

- Any neighbourhood of x_* contains a piece of the aforementioned line segment
- Hence x_* is not a local minimizer
- Thus a contradiction

WORDS OF CAUTION

- We are NOT claiming that every convex function has a minimizer
  - The function $f(x) = x$ doesn't have a minimum
  - A convex function bounded below need not have a minimizer. Consider

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

SOME PROPERTIES OF CONVEX FUNCTIONS

- If $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex then so is the function $f + g$
- If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and if $\mu \geq 0$, then so is the function μf

DIFFERENTIABLE CONVEX FUNCTIONS

Theorem

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

$$f(y) \geq f(x) + \langle \nabla f(x), (y - x) \rangle \qquad \textit{for every } x, y \in \mathbb{R}^n.$$

AN IMMEDIATE COROLLARY

Theorem (Stationarity and convexity)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function on $\mathbb{R}^n$. A point x_* is a global minimizer of f if and only if $\nabla f(x_*) = 0$.

- If x_* is a global minimizer of f, then it is also a local minimizer
- Hence by the first order necessary condition, we have

$$\nabla f(x_*) = 0$$

- Let us now assume that for a point x_*, we have $\nabla f(x_*) = 0$
- For any point $y \in \mathbb{R}^n$, as f is convex, we have the differential inequality:

$$\begin{aligned} f(y) &\geq f(x_*) + \langle \nabla f(x_*), (y - x_*) \rangle \\ &= f(x_*) \end{aligned}$$

- Hence x_* is a global minimizer of f

Theorem

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

$$f(y) \geq f(x) + \langle \nabla f(x), (y - x) \rangle \qquad \textit{for every } x, y \in \mathbb{R}^n.$$

- ♣ Let us assume that the above differential inequality is satisfied
- ♣ Our goal is to show that f is convex
- ♣ Taking $w, z \in \mathbb{R}^n$, we aim to demonstrate that

$$f(\alpha w + (1-\alpha)z) \leq \alpha f(w) + (1-\alpha) f(z) \qquad \text{for } \alpha \in [0,1].$$

- ♣ Let us take

$$x = \alpha w + (1-\alpha)z \quad \text{for } \alpha \in [0,1]$$

- ♣ Observe that $w - x = (1-\alpha)(w - z)$ and $z - x = \alpha(z - w)$

♣ Applying the differential inequality for w and x yields

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

♣ Applying the differential inequality for z and x yields

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

♣ Multiplying the first inequality by α and the second inequality by $(1 - \alpha)$ and adding the resulting inequalities yields

$$\begin{aligned} \alpha f(w) + (1-\alpha) f(z) &\geq f(x) + \alpha \langle \nabla f(x), (w - x) \rangle \\ &\qquad + (1-\alpha) \langle \nabla f(x), (z - x) \rangle \\ &= f(x) = f(\alpha w + (1-\alpha) z) \end{aligned}$$

Hence f is convex.

- We have shown that the differential inequality implies convexity
- Our next objective is to show that any differentiable convex function must satisfy the following differential inequality:

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

- Pick two points $x \neq y$ in $\mathbb{R}^n$. Convexity of f implies that

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y)$$

- Take $h := y - x$. Then the above inequality rewrites as

$$f\left(x + \frac{h}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x+h)$$

- We can further rewrite the above inequality as

$$f(x+h) - f(x) \geq \frac{f\left(x + \frac{h}{2}\right) - f(x)}{2^{-1}}$$

- Observe further that

$$f\left(x+\frac{h}{4}\right) = f\left(\frac{x}{2}+\frac{1}{2}\left(x+\frac{h}{2}\right)\right) \leq \frac{1}{2}f(x)+\frac{1}{2}f\left(x+\frac{h}{2}\right)$$

- The above inequality can be rewritten as

$$f\left(x+\frac{h}{2}\right) - f(x) \geq \frac{f\left(x+\frac{h}{4}\right)-f(x)}{2^{-1}}$$

- Hence we deduce that

$$f(x+h) - f(x) \geq \frac{f\left(x+\frac{h}{4}\right)-f(x)}{2^{-2}}$$

- Arguing recursively, we arrive at

$$f(x+h) - f(x) \geq \frac{f\left(x+2^{-k}h\right)-f(x)}{2^{-k}} \qquad \text{for all } k \in \mathbb{N}.$$

LITTLE DETOUR

- ♣ For a function $g : \mathbb{R}^n \to \mathbb{R}$, its DIRECTIONAL DERIVATIVE at a point x in the direction p is given by

$$\lim_{s \to 0} \frac{g(x + sp) - g(x)}{s}$$

- ♣ If g is a differentiable function, then its directional derivative is

$$\langle \nabla g(x), p \rangle$$

PROOF OF THE THEOREM (CONTD.)

- ♣ Recall that we had

$$f(x+h) - f(x) \geq \frac{f\left(x + 2^{-k}h\right) - f(x)}{2^{-k}} \qquad \text{for all } k \in \mathbb{N}.$$

- ♣ Letting $k \to \infty$ in the above inequality yields

$$f(y) - f(x) \geq \langle \nabla f(x), h \rangle = \langle \nabla f(x), (y - x) \rangle .$$

END OF LECTURE 2

THANK YOU FOR YOUR ATTENTION