Week 2, Lecture 1

26 January 2024 13:44

Optimization of Convex Functions

<u>Theorem 2.8</u> When f is convex, any local minimizer $\underline{x}_*$ is a global minimizer of f. If, in addition, f is differentiable, then any stationary point $(\nabla f = 0)$ is a global minimizer.

<u>Proof</u>: <u>First part</u>. Assume for contradiction that $\underline{x}_*$ is a local minimizer but that there is a second minimizer $\underline{y}$ such that:

$$\boxed{f(\underline{y}) < f(\underline{x}_*). \quad (1)}$$

By convexity,

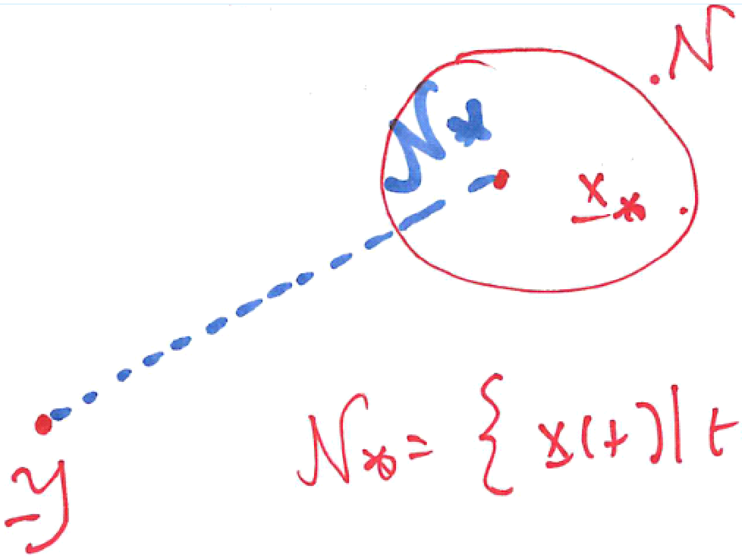
$$\underline{x}(t) = t\underline{y} + (1-t)\underline{x}_*, \quad t \in [0,1]$$

and

$$f(\underline{x}(t)) \le \underbrace{t\, f(\underline{y})}_{< t\, f(\underline{x}_*)} + \underbrace{(1-t)\, f(\underline{x}_*)}_{(1-t)\, f(\underline{x}_*)}$$

$$\Longrightarrow \boxed{f(\underline{x}(t)) < f(\underline{x}_*).}$$

Draw a picture:



$$\mathcal{N}_* = \{ \underline{x}(t) \,|\, t \in [0,1] \} \cap \mathcal{N}.$$

We have:

$$f(\underline{x}(t)) \leq f(\underline{x}_*) \quad \forall \; \underline{x}(t) \in \mathcal{N}_*.$$

So no matter the size of $\mathcal{N}$, there are points in $\mathcal{N}$ such that

$$f(\dots) \leq f(\underline{x}_*).$$

But $\underline{x}_*$ is a local minimum.
Contradiction. Hence, (1) is false. So there is no second minimizer $\underline{y}$, so $\underline{x}_*$ is the global minimum.

Second part: Assume $\underline{x}_*$ is a stationary point:

$$\nabla f(\underline{x}_*) = 0.$$

$$\Rightarrow \quad (\underline{y} - \underline{x}_*) \cdot \nabla f(\underline{x}_*) = 0.$$

But this is the $\boxed{\text{directional derivative}}$ of f at $\underline{x}_*$, in the direction $\underline{y} - \underline{x}_*$.

$$\Rightarrow \quad \frac{d}{dt} f\left(\underline{x}_* + t(\underline{y} - \underline{x}_*)\right)\Big|_{t=0} = 0$$

$$= \lim_{t \downarrow 0} \frac{f(\underline{x}_* + t(\underline{y} - \underline{x}_*)) - f(\underline{x}_*)}{t}$$

$$= \lim_{t \downarrow 0} \frac{f(t\underline{y} + (1-t)\underline{x}_*) - f(\underline{x}_*)}{t}$$

$$\leq \lim_{t \downarrow 0} \frac{tf(\underline{y}) + \cancel{(1-t)f(\underline{x}_*)} \cancel{- f(\underline{x}_*)}}{t}$$

$$\Rightarrow 0 \leq \lim_{t \downarrow 0} \frac{\cancel{t} f(\underline{y}) - \cancel{t} f(\underline{x}_*)}{\cancel{t}}.$$

$$\Rightarrow 0 \leq f(\underline{y}) - f(\underline{x}_*)$$

$$\Rightarrow f(\underline{x}_*) \leq f(\underline{y}) \quad \forall\, \underline{y} \in S.$$

Hence, $\underline{x}_*$ is a global min. $\blacksquare$

Model problem (§ 2.3)

When the cost function is ~~di~~ twice differentiable, it will "locally look like a quadratic". The quadratic cost function is the <u>model problem</u>:

$$f(\underline{x}) = c + \langle \underline{a}, \underline{x} \rangle + \frac{1}{2} \langle \underline{x}, B\underline{x} \rangle$$

where:

- c is a constant
- $\underline{a}$ is a constant vector
- B is an $n \times n$ SYMMETRIC positive-definite matrix

$$f(x_1, \dots, x_n) = c + a_i x_i + \frac{1}{2} x_i B_{ij} x_j$$

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \left(\cancel{c} + a_i x_i + \frac{1}{2} x_i B_{ij} x_j \right)$$

$$= 0 + a_i \frac{\partial x_i}{\partial x_k} + \frac{1}{2} B_{ij} \left(\frac{\partial x_i}{\partial x_k} x_j + x_i \frac{\partial x_j}{\partial x_k} \right)$$

Introduce:

$$\frac{\partial x_i}{\partial x_k} = \delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \frac{\partial f}{\partial x_k} = \underbrace{a_i \delta_{ik}} + \frac{1}{2} B_{ij} \left(\delta_{ik} x_j + x_i \delta_{jk} \right)$$

$$\Rightarrow \frac{\partial f}{\partial x_k} = a_k + \frac{1}{2} B_{ij} \delta_{ik} x_j + \frac{1}{2} B_{ij} \delta_{jk} x_i$$

$$= a_k + \frac{1}{2} B_{kj} x_j + \frac{1}{2} \underbrace{B_{ik}} x_i$$

$$\overset{\text{Symmetric}}{=} a_k + \frac{1}{2} B_{kj} x_j + \frac{1}{2} B_{ki} x_i$$

$$\overset{\text{Dummy index}}{=} a_k + \underline{\frac{1}{2} B_{kj} x_j} + \underline{\frac{1}{2} B_{kj} x_j}$$

$$= a_k + B_{kj} x_j$$

$$= a_k + [\underline{B}\,\underline{x}]_k \,.$$

$$\Rightarrow \nabla f = \underline{a} + \underline{B}\,\underline{x} \,.$$

First-order optimality : $\nabla f = 0$

$$\Rightarrow \underline{a} + \underline{B}\,\underline{x} = 0 \,.$$

$$\Rightarrow \underline{x} = -\underline{B}^{-1} \underline{a} \,, \quad \text{or}$$

$$\boxed{\underline{x}_* = -\underline{B}^{-1} \underline{a}}$$

Second-order optimality:

$$\frac{\partial f}{\partial x_k} = a_k + B_{kj} x_j$$

$$\frac{\partial^2 f}{\partial x_\ell \partial x_k} = \cancel{a_k} + B_{kj} \underbrace{\frac{\partial x_j}{\partial x_\ell}}_{\delta_{j\ell}}$$

$$= B_{kj}\delta_{j\ell}$$

$$= B_{k\ell}$$

$$\Rightarrow \frac{\partial^2 f}{\partial x_\ell \partial x_k} = B_{k\ell},$$ Hessian matrix is just $\underline{B}$.

But $\underline{B}$ is positive definite, as per the model problem, so $\underline{x}_* = -B^{-1}\underline{a}$ is a local minimizer. By convexity, this is the unique global minimizer.

Evaluation:

$$f(\underline{x}_*) = c + \langle \underline{a}, \underline{x}_* \rangle + \frac{1}{2}\langle \underline{x}_*, B\underline{x}_* \rangle$$

$$= c - \langle \underline{a}, B^{-1}\underline{a} \rangle + \frac{1}{2}\langle B^{-1}\underline{a}, BB^{-1}\underline{a} \rangle$$

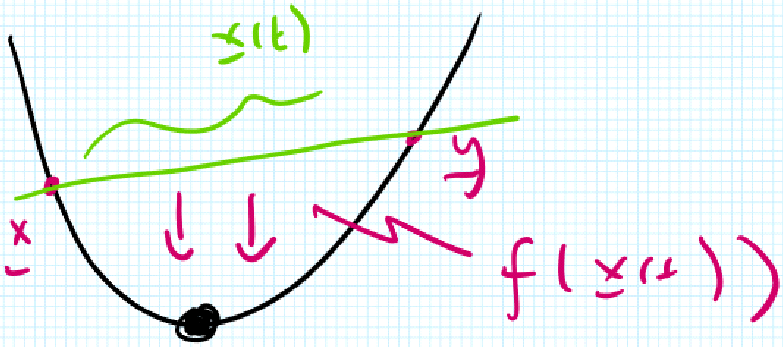
$$= c - \langle \underline{a}, B^{-1}\underline{a} \rangle + \frac{1}{2}\langle B^{-1}\underline{a}, \underline{a} \rangle$$

$$\boxed{f(\underline{x}_*) = c - \frac{1}{2}\langle \underline{a}, B^{-1}\underline{a} \rangle} = f_{min}.$$

Remark: Recall the definition of a convex function:

$$\underline{x}(t) = t\underline{x} + (1-t)\underline{y}, \quad t \in [0,1]$$

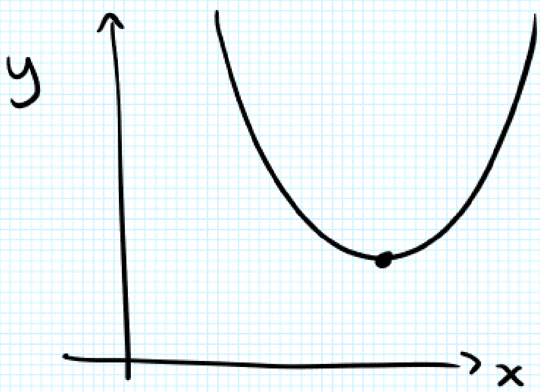
$$f(\underline{x}(t)) \leq t f(\underline{x}) + (1-t) f(\underline{y}).$$



GLOBAL MINIMIZER.

A convex function is U-shaped; the model problem involves such a function:

$$f(\underline{x}) = c + \langle \underline{a}, \underline{x} \rangle + \frac{1}{2}\langle \underline{x}, B\underline{x} \rangle$$



U-shaped surface in $\mathbb{R}^n$.

Of course, we have already proved algebraically in a previous lecture that the model problem is a convex function. The above sketches are just to supplement the analytical results with a pictorial understanding.