Subject 20H_Q1 Optimization Problem Date /

Attendants April 18th.

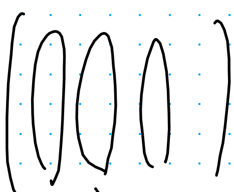
$$f(x) = \frac{1}{2} x^T A x + x^T b + a$$

as quadratic objective

$A \in \mathbb{R}^{d \times d}$: symmetric matrix

$A = U \Lambda U^T$ $\quad U \in \mathbb{R}^{d \times d}$

orthogonal matrix

$U =$ 

i-th column vector $u_i \in \mathbb{R}^d$

as eigen vector.

$\boxed{1.}$ compute grad of $f(x)$

$$\frac{\partial}{\partial x} f(x) = Ax + b \quad \in \mathbb{R}^d$$

Subject 2011_Q1 PreDoc2 Date 2

Attendants April 18th

$\boxed{2.}$ $\frac{\partial}{\partial x} f(x) = Ax + b$

stationally point x

such that $\frac{\partial}{\partial x} f(x) = 0$

$\Leftrightarrow x = -A^{-1} b$

because $\lambda_d > 0$, so A^{-1} can be calculated.

$\boxed{3.}$ Which of those stationally points are minima maxima and saddle-points? Give a mathmatically rigorous explanation why.

Hessian H is defined as follows

$$H = \frac{\partial^2}{\partial x^2} f(x) = \frac{\partial}{\partial x} (Ax + b)^T = A$$

$$= U \Lambda U^T$$

Subject 20H_Q1 PreDoc2 Date 3

Attendants April 18th

$H = A$

$= U \Lambda U^T \succeq 0$ positive definite

so $x = -A^{-1}b$ is minima of $f(x)$

$\boxed{4}$ Find the location x^* and value $f(x^*)$ of the global minimum.

$f(x)$ is quadratic function, so it is convex,

$x = -A^{-1}b$ is global minima

$f(x^*) = f(-A^{-1}b)$ $\quad (-A^{-1}b) \in \mathbb{R}^d$

$= \frac{1}{2}(-A^{-1}b)^T A(-A^{-1}b) + (-A^{-1}b)^T b + a$

$= \frac{1}{2} b^T (A^{-1})^T A A^{-1} b - b^T (A^{-1})^T b + a$

$= \frac{1}{2} b^T (A^{-1})^T b - b^T (A^{-1}) b + a$

$= -\frac{1}{2} b^T (A^{-1})^T b + a$

$= -\frac{1}{2} b^T A^{-1} b + a$

(A is symmetric
A^{-1} is also symmetric
So $(A^{-1})^T = A^{-1}$)

Subject 20H_Q1 PreDoc2 Date 4

Attendants Optimization Problem April 18th

$\boxed{5}$ Show how the k-step of gradient descent, with step-size $\tau > 0$, looks like in this case by substituting $f(x)$ with its quaderatic form above

Assume

$$f(x) = \frac{1}{2}x^T A x \leftarrow \frac{1}{2}x^T A x + x^T b + a$$

$$\nabla f(x) = Ax \,? \leftarrow Ax + b \,?$$

$$x_{k+1} = x_k - \tau \nabla f(x_k)$$

$$x_k = x_{k-1} - \tau \nabla f(x_{k-1})$$

$$= x_{k-1} - \tau A x_{k-1}$$

$$= (I - \tau A)\underbrace{x_{k-1}}_{\hookrightarrow (I-\tau A)x_{k-2}}$$

$$= (I - \tau A)^k x_0$$

Assume

$$x_k = x_{k-1} - \tau \nabla f(x_{k-1})$$

$$= x_{k-1} - \tau (A x_{k-1} + b)$$

$$= (I - \tau A)\underbrace{x_{k-1}}_{\hookrightarrow (I-\tau A)x_{k-2} - \tau b} - \tau b$$

$$\rightarrow (I - \tau A)^2 x_{k-2} - (I - \tau A)\tau b - \tau b$$

Subject 20H_Q1 Date 5

Attendants April 18th

$$x_t = (I - rA)^2 x_{t-2} - (I - rA) rb - rb$$

$$= (I - rA)^3 x_{t-3} - (I - rA)^2 rb - (I - rA) rb - rb$$

$$= (I - rA)^t x_0 - \sum_{\tau=0}^{t-1} (I - rA)^\tau rb$$

$$\left(\text{then. } x^* = -A^{-1} b \iff b = -Ax^* \right)$$

$$x_t = (I - rA)^t x_0 + rAx^* \sum_{t=0}^{t-1} (I - rA)^\tau$$

$$x_t - x^* = (I - rA)^t x_0 + rAx^* \cdot \sum_{\tau=0}^{t-1} (I - rA)^\tau - x^*$$

Subject 2011_Q1 Date 6

April 16th.

$\boxed{6}$ Give condition for choise of step $\gamma > 0$

that is sufficient to get converge

to the min x^*

From $\boxed{Q5}$

$$x_k - x^* = (I-\gamma A)^k x_0 + \gamma A x^* \cdot \sum_{j=0}^{k-1} (I-\gamma A)^j - x^*$$

$\boxed{\text{Intuition}}$

Curvature $\lambda_1 > \dots \lambda_n$

is lower bound of L (for Lipschitz continuity)

$\|\nabla f(x)\|_2^2 \leq L^2$??

$\|\nabla f(x) - \nabla f(y)\| \leq L \|x-y\| \Leftrightarrow \underline{\nabla^2 f(x) \preceq LI}$

Subject 20H_Q1 (Memo) Date 7

Attendants

$\boxed{\text{Strong convexity : } \alpha\text{-strong convex}}$

$$f(x) - \frac{\alpha}{2}\|x\|^2 \text{ is convex}$$

$$\Leftrightarrow \nabla^2 f(x) \succeq \alpha I$$

$\boxed{\beta\text{-smooth}}$

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|$$

$\boxed{L\text{-Lipschitz continuous}}$

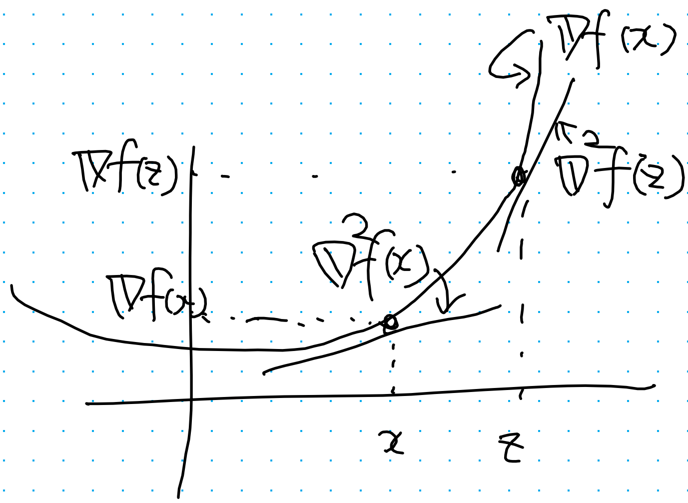
$$\|f(x) - f(y)\| \leq L \|x - y\|$$

Subject 20H_Q1 (Memo) Date 8

Attendants

$\|\nabla^2 f(x)\|_2 < L \iff \lambda_1 < L$

$\iff \frac{1}{L} < \frac{1}{\lambda_1}$



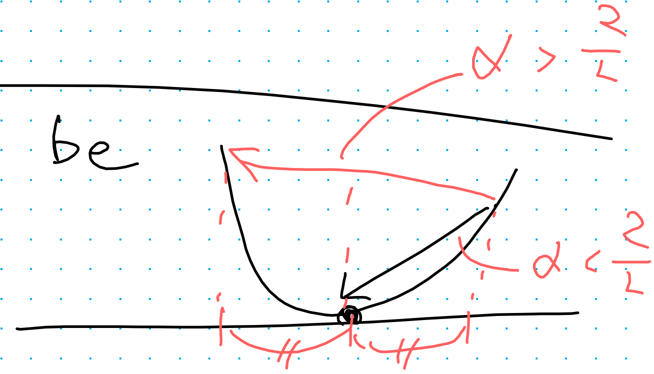
$$\|\nabla f(x) - \nabla(z)\|_2 = \left\| \int_0^1 \nabla^2 f(x + \tau(z-x))\,d\tau\,(x-z) \right\|_2$$

$$\leq \left\| \int_0^1 \nabla^2 f(x+\tau(z-x))\,d\tau \right\|_2 \|(x-z)\|_2$$

$$\leq \left(\int_0^1 L\,d\tau\right) \|x-z\|_2 = L\|x-z\|_2$$

step size α should be

$0 < \alpha < \frac{2}{L} < \frac{2}{\lambda_1}$



Subject 20H_Q1 Date 9

Attendants

$$f(x) = \frac{1}{2}x^TAx + x^Tb + a$$

$$\nabla f(x) = Ax+b$$

$$f(x_{k+1}) = x_t - \alpha\nabla f(x_k)$$

$$= f(\underbrace{x_t - \alpha(Ax_t+b)}) \rightarrow (\cdot\cdot)$$

$$= \frac{1}{2}(\cdot\cdot)^TA(\cdot\cdot) + (\cdot\cdot)^Tb + a \quad \cdots \text{①}$$

$$f(\alpha) = m\alpha^2 + n\alpha + l = m\left\{\left(\alpha^2 + 2\left(\frac{n}{2m}\right)\alpha + \left(\frac{n}{2m}\right)^2\right) + l - \frac{n^2}{4m^2}\right\}$$

$$\alpha^* = \underset{\alpha}{\mathrm{argmin}}\, f(\alpha) = m\left\{\left(\alpha + \frac{n}{2m}\right)^2 + l - \frac{n^2}{4m^2}\right\}$$

(since $f(\alpha)$ is quaderatic func)

$$= -\frac{n}{2m}$$

show next page.

where $m = \frac{1}{2}\nabla f(x)^TA\,\nabla f(x)$

$$n = -x^TA\nabla f(x) + \nabla f(x)^Tb = -\nabla f(x)\underbrace{(Ax-b)^T}_{\nabla f(x)}$$

so $$\alpha_t = \frac{\nabla f(x)\,\nabla f(x)^T}{\nabla f(x)^TA\nabla f(x)}$$

Subject 20H_Q1 Date 10

Attendants $x \leftarrow x_k$

$$f(\alpha) = \frac{1}{2}(x-\alpha Ax-\alpha b)^T A(x-\alpha Ax-\alpha b)$$

$$+(x-\alpha Ax-\alpha b)^T b + a$$

$$= \frac{1}{2}\left\{x^T x - \alpha x^T Ax - \alpha x^T b - \alpha x^T Ax + \alpha^2 x^T AAx + \alpha^2 x^T A^T b - \alpha bx + \alpha^2 Axb + \alpha^2 b^2 \right\} A$$

$$+ x^T b - \alpha x^T A^T b - \alpha b^2 + a$$

$$= \frac{1}{2}\left\{\alpha^2 (x^T AAx + x^T A^T b + b^2\right.$$

$$f(\alpha) = \frac{1}{2}(x-\alpha\nabla f(x))^T A(x-\alpha\nabla f(x))$$

$$+(x-\alpha\nabla f(x))^T b + a$$

$$= \frac{1}{2}\left\{x^T Ax - 2x^T A\alpha\nabla f(x) + \alpha^2 \nabla f(x)^T A\nabla f(x)\right\}$$

$$+ x^T b - \alpha\nabla f(x)^T b + a$$

$$= \alpha^2\left\{\frac{1}{2}\nabla f(x)^T A\nabla f(x)\right\} + \alpha(-1)\left\{x^T A\nabla f(x) + \nabla f(x)^T b\right\}$$

Subject

Date

Attendants