پاسخ تمرین سری اول

الگوریتم‌های بهینه‌سازی

حسام مومیوند فرد

810803063

بهار ۱۴۰۴

Nocedal

2.1 $f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2$

a) $\nabla f(x)$, $\nabla^2 f(x)$

b) $x^* = (1,1)^T$ is the only local minimizer & $\nabla^2 f(x^*)$ is positive definite.

a) $\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \end{bmatrix} = \begin{bmatrix} 200(x_2 - x_1^2) \times 2x_1 + 2(1-x_1)(-1) \\ 200(x_2-x_1^2) \times (1) + 0 \end{bmatrix} = \begin{bmatrix} -400x_1(x_2-x_1^2) - 2(1-x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$

$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f(x) \\ \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) \end{bmatrix} = \begin{bmatrix} -400(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$

b) $\nabla f(x) = 0 \Rightarrow \begin{bmatrix} -400x_1(x_2-x_1^2) - 2(1-x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \xrightarrow{x_2 - x_1^2 = 0} -2(1-x_1) = 0$

$\Rightarrow -2 + 2x_1 = 0 \Rightarrow x_1 = 1$

$\Rightarrow x_2 = 1$

$\Rightarrow x^* = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = [1 \quad 1]^T$ ☒

$\nabla^2 f(x^*) = \begin{bmatrix} -400(1-1) + 800 \times 1 + 2 & -400 \times 1 \\ -400 \times 1 & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$

$\det(\nabla^2 f(x^*) - \lambda I) = 0 \Rightarrow \det\begin{bmatrix} 802 - \lambda & -400 \\ -400 & 200 - \lambda \end{bmatrix} = 0$

$\Rightarrow (802 - \lambda)(200 - \lambda) - (-400)(-400) = 0$

$\Rightarrow 802 \times 200 - 802\lambda - 200\lambda + \lambda^2 - 160000 = 0$

$\Rightarrow \dots \Rightarrow \lambda \simeq 1001, 1$ ☒

Nocedal

2.2

$$f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$$

a) Only one stationary point
b) that point is saddle
c) contour lines

a)

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \end{bmatrix} = \begin{bmatrix} 8+2x_1 \\ 12-4x_2 \end{bmatrix} = [0 \;\; 0]^T \Rightarrow \begin{cases} 8+2x_1 = 0 \to x_1 = -4 \\ 12-4x_2 = 0 \to x_2 = 3 \end{cases}$$

$$\Rightarrow x^* = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$$

$$\frac{\partial}{\partial x_1} f(x) = 8 + 2x_1$$

$$\frac{\partial}{\partial x_2} f(x) = 12 - 4x_2$$

b)

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f(x) \\ \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}$$

$$\frac{\partial^2}{\partial x_1^2} f(x) = \frac{\partial}{\partial x_1}(8+2x_1) = 2$$

$$\frac{\partial^2}{\partial x_1 \partial x_2} f(x) = \frac{\partial}{\partial x_2}(8+2x_1) = 0$$

$$\frac{\partial^2}{\partial x_2 \partial x_1} f(x) = \frac{\partial}{\partial x_1}(12-4x_2) = 0$$

$$\frac{\partial^2}{\partial x_2^2} f(x) = \frac{\partial}{\partial x_2}(12-4x_2) = -4$$

$$\det(\nabla^2 f(x) - \lambda I) = 0 \Rightarrow \det \begin{vmatrix} 2-\lambda & 0 \\ 0 & -4-\lambda \end{vmatrix} = 0 \Rightarrow (2-\lambda)(-4-\lambda) = 0 \Rightarrow \begin{cases} \lambda = 2 \\ \lambda = -4 \end{cases}$$

λ ها مثبت و منفی هستند که نقطه مورد نظر (x^*) از نوع saddle است.

c)

$$f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2 = x_1^2 + 8x_1 + 16 - 16 - 2(x_2^2) - 2(-6x_2) - 2(9-9)$$

$$= (x_1+4)^2 - 2(x_2-3)^2 - 16 + 18 = C \Rightarrow (x_1+4)^2 - 2(x_2-3)^2 = C - 2$$

x_1, x_2 ریشه های درجه دو و مختلف العلامت هستند که کانتورهای ما به شکل هذلولی خواهند بود.

Nocedal

2.4 a) second-order Taylor $f_{(x)} = \cos(1/x)$ around nonzero

b) Third-order Taylor $g_{(x)} = \cos(x)$ around any point

c) second-order for $x=1$

a) $f_{(x)} \approx f_{(x_0)} + f'_{(x)}(x-x_0) + \frac{f''_{(x)}}{2}(x-x_0)^2$

$f_{(x)} = \cos(1/x)$

$f'_{(x)} = -\sin(1/x) \times -\frac{1}{x^2} = \frac{\sin(1/x)}{x^2}$

$f''_{(x)} = \frac{\partial}{\partial x}\frac{\sin(1/x)}{x^2} = \frac{\cos(1/x)\times -\frac{1}{\cancel{x^2}}\times \cancel{x^2} - 2x\sin(1/x)}{x^4} = -\frac{\cos(1/x)+2x\sin(1/x)}{x^4}$

$\Rightarrow f_{(x)} \approx \cos(1/x_0) + \frac{\sin(1/x_0)}{x_0^2}(x-x_0) - \frac{\cos(1/x_0)+2x_0\sin(1/x_0)}{2x_0^4}(x-x_0)^2$

b) $f'''_{(x)} = \frac{\partial}{\partial x} - \frac{\cos(1/x)+2x\sin(1/x)}{x^4} = -\left[\frac{\partial}{\partial x}\frac{\cos(1/x)}{x^4} + \frac{\partial}{\partial x}\frac{2\sin(1/x)}{x^3}\right]$

$= -\left[\frac{-\sin(1/x)\times -\frac{1}{\cancel{x^2}}\times x^{\cancel{4}\,2} - 4x^3\cos(1/x)}{x^8} + \frac{2\cos(1/x)\times -\frac{1}{\cancel{x^2}}\times x^{\cancel{3}} - 3x^2\times 2\sin(1/x)}{x^6}\right]$

$= -\left[\frac{\sin(1/x) - 4x\cos(1/x) - 2x\cos(1/x) - 6x^2\sin(1/x)}{x^6}\right]$

$= \frac{6x^2\sin(1/x) + 6x\cos(1/x) - \sin(1/x)}{x^6}$

$\Rightarrow f_{(x)} = \cos(1/x_0) + \frac{\sin(1/x_0)}{x_0^2}(x-x_0) - \frac{\cos(1/x_0)+2x_0\sin(1/x_0)}{2x_0^4}(x-x_0)^2 + \frac{6x_0^2\sin(1/x_0)+6x_0\cos(1/x_0)-\sin(1/x_0)}{6x_0^6}(x-x_0)^3$

c) $f_{(x)} = \cos(1) + \sin(1)(x-1) + \frac{1}{2}\left[\cos(1) + 2\sin(1)\right](x-1)^2$

Nocedal 2.5

$$f(x) = \|x\|^2 \qquad \mathbb{R}^2 \to \mathbb{R}$$

$$x_k = \left(1+\frac{1}{2^k}\right)\begin{bmatrix}\cos k\\ \sin k\end{bmatrix}$$

a) $f(x_{k+1}) < f(x_k) \qquad k=0,1,2,\dots$

b) $\{x \mid \|x\|^2 = 1\}$ is the Limit point for $\{x_k\}$

Hint $\forall \theta \in [0, 2\pi] \qquad \xi_k = k \pmod{2\pi} = k - 2\pi \left\lfloor \frac{k}{2\pi} \right\rfloor$

a) $f(x) = x_1^2 + x_2^2$

$$x_k = \left(1+\frac{1}{2^k}\right)\begin{bmatrix}\cos k\\ \sin k\end{bmatrix} = \begin{bmatrix}\left(1+\frac{1}{2^k}\right)\cos k\\ \left(1+\frac{1}{2^k}\right)\sin k\end{bmatrix}$$

$$f(x_k) = \left(\left(1+\frac{1}{2^k}\right)\cos k\right)^2 + \left(\left(1+\frac{1}{2^k}\right)\sin k\right)^2$$

$$f(x_{k+1}) = \left(\left(1+\frac{1}{2^{k+1}}\right)\cos(k+1)\right)^2 + \left(\left(1+\frac{1}{2^{k+1}}\right)\sin(k+1)\right)^2$$

$$f(x_k) = \left(1+\frac{1}{2^k}\right)^2\left(\cos^2 k + \sin^2 k\right) = \left(1+\frac{1}{2^k}\right)^2$$

$$f(x_{k+1}) = \left(1+\frac{1}{2^{k+1}}\right)\left(\cos^2(k+1) + \sin^2(k+1)\right) = \left(1+\frac{1}{2^{k+1}}\right)^2$$

واضحه که توان k ما در مخرج به این معناست که هر چقدر k ما بزرگتر بشود عبارت ما کوچکتر شده و به سمت 1 میل می‌کنه پس نامعادله

$$f(x_k) > f(x_{k+1})$$

همواره برقرار است. $\blacksquare$

b)

طبق صحبتی که با خودتون داشتم در مورد این تمرین به علت اینکه $\|x\|^2$ به فرم یک دایره با مرکز $(0,0)$ محدب است و زیر دنباله‌های x^k با افزایش k ها به سمت $\begin{bmatrix}\cos k\\ \sin k\end{bmatrix}$ میل می‌کنند در نتیجه می‌توان گفت به سمت یک نقطه روی $x_1^2+x_2^2=1$ در حال میل کردن هستند.

Nocedal

2.6 Prove that all local minimizers are strict.

$$p \implies q$$

Hint: x^* is an isolated local minimizer. $\mathcal{N}$ (neighborhood). for any $x \in \mathcal{N}$, $x \neq x^*$; $f(x) > f(x^*)$

خب فرض کنیم که $\begin{cases} f(x) \geq f(x^*) \\ f(x) \leq f(x^*) \end{cases}$ $\forall x \in \mathcal{N}, x \neq x^* \to$ از این نتیجه میگیریم که برای نقیض $f(x) > f(x^*)$ ، $f(x) = f(x^*)$ است.

$$\bar{q} \implies \bar{p}$$

خب حالا درصورتی که $f(x) = f(x^*)$ باشد، $x \neq x^* \Leftarrow$ نتیجه x^* نمی‌تونه لزوما isolate باشه چون که $\underbrace{\text{همسایگیش یه نقطه برابر خودش وجود داره.}}_{\bar{p}}$

اثبات از روش نقیض کامل شد ☑

Nocedal 2.10

$\tilde{f}(z) = f(x)$ s.t $x = Sz + s$ for some $S \in \mathbb{R}^{n \times n}$ and $s \in \mathbb{R}^n$. show that

$$\begin{cases} \nabla \tilde{f}(z) = S^T \nabla f(x) & \text{(a} \\ \nabla^2 \tilde{f}(z) = S^T \nabla^2 f(x) S & \text{(b} \end{cases}$$

a) $x \in \mathbb{R}^n$ $1 \times n$ vector

$S \in \mathbb{R}^{n \times n}$ $n \times n$ Matrix

$s \in \mathbb{R}^n$ $1 \times n$ vector

$z \in \mathbb{R}^n$ $1 \times n$ vector

$\tilde{f}(z) = f(x) = f(Sz + s)$

$\nabla \tilde{f}(z) = \nabla f(Sz + s) \overset{?}{=} S^T \nabla f(x)$

$$\underset{x}{\begin{bmatrix} \vdots \\ x_i \\ \vdots \end{bmatrix}} = i\underset{S}{\begin{bmatrix} \quad \end{bmatrix}} \underset{z}{\begin{bmatrix} \quad \end{bmatrix}} + \underset{s}{\begin{bmatrix} \quad \end{bmatrix}} \Rightarrow x_i = \sum_{k=1}^{n} S_{ik} z_k + s_i \Rightarrow \frac{\partial x_i}{\partial z_j} = S_{ij}$$

. s_i که ثابت در سطر i ام ستون j ام

$$\nabla \tilde{f}(z) \Rightarrow \frac{\partial \tilde{f}(z)}{\partial z_j} = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} \times \frac{\partial x_i}{\partial z_j} = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} \times S_{ij}$$

$$\Rightarrow \nabla \tilde{f}(z) = \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} S_{i1} \\ \vdots \\ \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} S_{in} \end{bmatrix} = S^T \nabla f(x) \checkmark$$

b) $$\nabla^2 \tilde{f}(z) = \frac{\partial^2 \tilde{f}(z)}{\partial z^2} = \frac{\partial^2 \tilde{f}(z)}{\partial z_i \, \partial z_j} = \frac{\partial}{\partial z_i}\left(\frac{\partial \tilde{f}(z)}{\partial z_j}\right) = \underbrace{\frac{\partial}{\partial z_i}}_{\frac{\partial}{\partial x_m}\frac{\partial x_m}{\partial z_i}}\left(\sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k} S_{kj}\right)$$

$$\Rightarrow \frac{\partial^2 \tilde{f}(z)}{\partial z_i \partial z_j} = \sum_{k=1}^{n}\sum_{m=1}^{n} \frac{\partial}{\partial x_m}\left(\frac{\partial f(x)}{\partial x_k}\right) \underbrace{\frac{\partial x_m}{\partial z_i}}_{S_{mi}} S_{kj} = \sum_{k=1}^{n}\sum_{m=1}^{n} \frac{\partial}{\partial x_m}\left(\frac{\partial f(x)}{\partial x_k}\right) S_{mi} S_{kj}$$

$$\Rightarrow \frac{\partial^2 \tilde{f}(z)}{\partial z_i \partial z_j} = \sum_{k=1}^{n}\sum_{m=1}^{n} \frac{\partial^2 f(x)}{\partial x_m} S_{mi} S_{kj} = S^T \nabla^2 f(x) S \checkmark$$

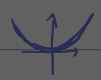
Nocedal

3.4 Show that one dimentional minimizer of strongly convex quadratic function always satisfies the Goldstine conditions. (3.11)

$$f(x_k) + (1-C)\,\alpha_k \nabla f_k^T p_k \le f(x_k + \alpha_k p_k) \le f(x_k) + C\alpha_k \nabla f_k^T p_k \qquad (3.11)$$

$f_{(x)}$ is strongly convex $\Longrightarrow$ $\left|\begin{array}{l} f_{(x)} = \frac{1}{2} x^T A x - b^T x + d \\ v^T A v > 0 \end{array}\right.$

به این فرم است



یعنی A مثبت معین است — دیگه مینیمم داره ⇐ از نظر ظاهری شبیه x^2:

$\Rightarrow \nabla f_{(x)} = Qx + b \Rightarrow \nabla f_k = \nabla f(x_k) = Qx_k + b$

گرادیان توی نقطه x_k در جهت کاهشی p_k داریم:

$x = x_k + \alpha p_k$

$\Rightarrow f_{(x_k + \alpha p_k)} = \frac{1}{2}(x_k + \alpha p_k)^T Q (x_k + \alpha p_k) + b^T (x_k + \alpha p_k) + c$

$= \frac{1}{2}(x_k^T + \alpha p_k^T) Q (x_k + \alpha p_k) + b^T (x_k + \alpha p_k) + c$

$= \frac{1}{2} x_k^T Q x_k + \alpha x_k^T Q p_k + \alpha p_k^T Q x_k + \alpha^2 p_k^T Q p_k + b^T (x_k + \alpha p_k) + c$

$= \frac{1}{2}(x_k^T Q x_k + 2\alpha x_k^T Q p_k + \alpha^2 p_k^T Q p_k) + b^T (x_k + \alpha p_k) + c$

$= \phi(\alpha)$ ← هدف ما مینیمم کردن اینه...

$\left|\begin{array}{l} \phi_{(\alpha)} = \frac{1}{2}(p_k^T Q p_k)\alpha^2 + \nabla f_k^T p_k \alpha + f_{(x_k)} \\ \\ b + x_k Q = \nabla f_k \end{array}\right.$

$$\frac{\partial \phi_{(\alpha)}}{\partial \alpha} = 0 \Rightarrow (p_k^T Q p_k)\alpha + \nabla f_k^T p_k = 0 \Rightarrow \alpha_k = -\frac{\overset{\text{جهت نزولی}}{\nabla f_k^T p_k}}{\underset{\text{مثبت معین}}{p_k^T Q p_k}} \Rightarrow \alpha > 0$$

بعد از محاسبه‌ی α_k حالا می‌تونیم بریم سراغ بررسی نامساوی گلدشتاین

a) $f(x) + (1-c)\,\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha_k p_k)$

b) $f(x_k + \alpha_k p_k) \leq f(x_k) + c\,\alpha_k \nabla f_k^T p_k$

$$a:\ \phi(\alpha_k) = f(x_k + \alpha_k p_k) = \frac{1}{2}(p_k^T Q p_k)\left(-\frac{\nabla f_k^T p_k}{p_k^T Q p_k}\right)^2 + \nabla f_k^T p_k\left(-\frac{\nabla f_k^T p_k}{p_k^T Q p_k}\right) + f(x_k)$$

$$= \frac{1}{2}\frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k} - \frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k} + f(x_k) = -\frac{1}{2}\frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k} + f(x_k) \Rightarrow c \leq \frac{1}{2},\ c \in (0, 0.5)$$

$$b:\ f(x_k) + c\,\alpha_k \nabla f_k^T p_k = f(x_k) + c\left(-\frac{\nabla f_k^T p_k}{p_k^T Q p_k}\right)\nabla f_k^T p_k = f(x_k) - c\,\frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k}$$

$$\Rightarrow \quad c \leq \frac{1}{2}, \quad c \in (0, 0.5)$$

Nocedal

3.6 Consider the steepest descent method with exact line searches applied to the convex quadratic function (3.24). Using the properties given in this chapter, show that if the initial point is such that $x_0 - x^*$ is parallel to an eigenvector of Q, then the steepest descent method will find the solution in one step.

$$f(x) = \frac{1}{2} x^T Q x - b^T x \qquad (3.24)$$

$f(x) = \frac{1}{2} x^T Q x + b^T x + c$; $x_0 - x^* \parallel$ Eigenvector of Q

steepest descent $\rightarrow P_k = -\nabla f(x_k)$, α_k

$\hookrightarrow x^*$ s.t. $\nabla f(x^*) = 0$

$Q : \underset{\text{Eigenvalue}}{\lambda}, \overset{\text{Eigenvector}}{v}$; $\lambda v = Qv$

$\Rightarrow x_0 - x^* = \beta v \Rightarrow x_0 = x^* + \beta v$

$\nabla f(x) = Qx + b \longrightarrow Qx^* + b = 0$

$\hookrightarrow Q x_0 + b = Q(x^* + \beta v) + b = Qx^* + \beta Q v + b \Rightarrow \nabla f(x_0) = \beta Q v = Q \lambda v$

$\left.\begin{array}{l} \phi(\alpha) = f(x_0 + \alpha P_0) \\ P_0 = -\nabla f(x_0) = -\beta \lambda v \end{array}\right| \Rightarrow x_0 + \alpha P_0 = x_0 + \alpha(-\beta \lambda v) = x^* + \beta v - \alpha \beta \lambda v = x^* + \beta(1 - \alpha\lambda) v$

$\Rightarrow \phi(\alpha) = f(x_0 + \alpha P_0) = \frac{1}{2}(x_0 + \alpha P_0)^T Q (x_0 + \alpha P_0) + b^T(x_0 + \alpha P_0) + c$

$\Rightarrow \phi(\alpha) = f(x_0 + \alpha P_0) = \frac{1}{2}(x^* + \beta(1-\alpha\lambda) v)^T Q (x^* + \beta(1-\alpha\lambda) v) + b^T(x^* + \beta(1-\alpha\lambda) v) + c$

$= \frac{1}{2}(x^{*T} + \beta(1-\alpha\lambda) v^T) Q (x^* + \beta(1-\alpha\lambda) v) + b^T(x^* + \beta(1-\alpha\lambda) v) + c$

$= x^{*T} Q x^* + 2\beta(1-\alpha\lambda) x^{*T} Q v + \beta^2 (1-\alpha\lambda)^2 v^T Q v$

$$\begin{cases} v^T Q v = v^T(\lambda v) = \lambda \|v\|^2 \\ Qx^* + b = 0 \Rightarrow Qx^* = -b \\ (x^*)^T Q v = (x^*)^T(\lambda v) = \lambda (x^*)^T v \end{cases}$$

$$\Rightarrow f_{(x_0+\alpha p_0)} = \frac{1}{2}(x^*)^T Q x^* + \beta(1-\alpha\lambda)\lambda x^{*T} v + \frac{1}{2}\beta^2(1-\alpha\lambda)^2\lambda\|v\|^2 - x^{*T}Qx^* - \beta(1-\alpha\lambda)\lambda x^{*T} v + C$$

$$= -\frac{1}{2}x^{*T}Qx^* + \frac{1}{2}\beta^2(1-\alpha\lambda)^2\lambda\|v\|^2 + C = \frac{1}{2}\beta^2\lambda\|v\|^2(1-\alpha\lambda)^2 + C = \phi_{(\alpha)}$$

$$\frac{\partial\phi_{(\alpha)}}{\partial\alpha} = \beta^2\lambda\|v\|^2(1-\alpha\lambda)(-\lambda) = -\beta^2\lambda^2\|v\|^2(1-\alpha\lambda) = 0 \Rightarrow \alpha_0 = \frac{1}{\lambda}$$

$$x_0 = x^* + \beta v \rightarrow x_1 = x^* + \beta v - \beta v = x^* \longrightarrow x_1 = x^* \quad \checkmark$$

Nocedal

4.5

show that τ_k defined by (4.12) does identify the minimizer of m_k along the direction $-g_k$.

$$\tau_k = \begin{cases} 1 & \text{if } g_k^T B_k g_k \leq 0; \\ \min\left(\|g_k\|^3 / (\Delta_k g_k^T B_k g_k), 1 \right) & \text{o.w.} \end{cases} \qquad (4.12)$$

$$m_k(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$

$$\nabla f(x_k) = g_k$$

$$p = \tau(-g_k)$$

$$\Rightarrow m_k(x_k + \tau(-g_k)) = f(x_k) + \underbrace{g_k^T(-\tau g_k)}_{= -\tau(g_k^T g_k) = -\tau \|g_k\|^2} + \frac{1}{2} \overbrace{(-\tau g_k)^T B_k (-\tau g_k)}^{= \tau^2 g_k^T B_k g_k}$$

$$\Rightarrow m_k(x_k + \tau(-g_k)) = f(x_k) - \tau \|g_k\|^2 + \frac{1}{2} \tau^2 (g_k^T B_k g_k) = \phi(\tau)$$

$$\frac{\partial \phi(\tau)}{\partial \tau} = 0 = -\|g_k\|^2 + \tau (g_k^T B_k g_k) \Rightarrow \tau = \frac{\|g_k\|^2}{g_k^T B_k g_k}$$

اینجا به بعدش رو دیگه نمی‌تونم جلو ببرم و باید بپرسم که دقیقاً باید چیکار کرد . . .

Nocedal

4.10 show that if B is any symmetric matrix, then there exists $\lambda \geq 0$ such that $B + \lambda I$ is positive definite.

B is symmetric $\Rightarrow B = B^T \Rightarrow B = Q \Lambda Q^T$

$\Rightarrow x^T B x = x^T (Q \Lambda Q^T) x = (Q^T x)^T \Lambda (Q^T x)$, $Q^T x = y \Rightarrow x^T B x = y^T B y$

$$y^T B y = \sum_{i=1}^{n} \lambda_i y_i^2$$ $\Rightarrow$ درصورتی که تمام λ_i ها مثبت باشند آنگاه B مثبت معین است

$B + \lambda I = Q \Lambda Q^T + \lambda I \Rightarrow B + \lambda I = Q \Lambda Q^T + \lambda Q I Q^T = Q \underline{(\Lambda + \lambda I)} Q^T$

$\underline{\lambda + \lambda_i}$ که مقادیر ویژه‌ی این ماتریس چون قطریه درایه‌های قطر اصلیش

$$\begin{cases} \lambda_{min} \leq 0 \longrightarrow \lambda_{min} - \lambda_i = \lambda + \lambda_i' \\ \lambda_{min} > 0 \quad \boxtimes \end{cases}$$

یک مقدار ϵ به λ_{min} اضافه کنیم اون وقت B مثبت معین میشه

n-point

مسئله‌ی رگرسیون n نقطه با چند بعدای d را به صورت یک مسئله‌ی کمترین مربعات بنویسید.

a) شرط لازم و کافی برای نقطه‌ی بهینه را به دست آورید.

b) برای بهبود کیفیت جواب تنظیم کننده (regularization) را نیز اضافه کنید و شرط لازم و کافی را در این حالت بنویسید.

$$J(w) = \frac{1}{2N} \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \frac{\lambda}{2} \|w\|^2$$

$x_i \in \mathbb{R}^d$

$y_i \in \mathbb{R}$

$w \in \mathbb{R}^d$

λ : Regularization پارامتر

$\|w\|^2 = w^T w$: norm-2 w

a) $$\frac{\partial J(w)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - w^T x_i)^2\right) = \frac{1}{2N} \sum_{i=1}^{N} \frac{\partial}{\partial w}(y_i - w^T x_i)^2 = \frac{1}{2N} \sum_{i=1}^{N} -2(y_i - w^T x_i) x_i$$

$$= -\frac{1}{N} \sum_{i=1}^{N} (y_i - w^T x_i) x_i$$

$+$

$$\frac{\partial}{\partial w}\left(\frac{\lambda}{2}\|w\|^2\right) = \frac{\partial}{\partial w}(w^T w) = 2w$$

$$\Rightarrow \frac{\partial J(w)}{\partial w} = -\frac{1}{N} \sum_{i=1}^{N} (y_i - w^T x_i) x_i + \lambda w = 0 \Rightarrow -\frac{1}{N} \sum_{i=1}^{N} (y_i - w^T x_i) x_i = \lambda w$$

$$= \underbrace{-\frac{1}{N} \sum_{i=1}^{N} y_i x_i}_{\text{ثابت}} - \underbrace{\frac{1}{N} \sum_{i=1}^{N} (w^T x_i) x_i} = \lambda w$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i (x_i^T w) = \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T\right) w$$

$$\Rightarrow w = (X^T X + N \lambda I)^{-1} X^T y \quad \text{s.t.} \begin{cases} \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T = \frac{1}{N} X^T X \\ \frac{1}{N} \sum_{i=1}^{N} y_i x_i = \frac{1}{N} X^T y \end{cases}$$

b) $\min_{w} \frac{1}{2N} \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \frac{\lambda}{2} \|w\|^2 \equiv \begin{cases} \min_{w} \frac{1}{2N} \sum_{i=1}^{N} (y_i - w^T x_i)^2 \\ \|w\|^2 \leq C \end{cases}$