Optimization Theory and Algorithm	**Homework 3 - 30/05/2021**

Homework 3

Lecturer: Xiangyu Chang *Scribe:* 赵敬业

1 HW 1

1.1 问题重述

(1) 已知 $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{bmatrix}$, 请计算列空间 $C(A)$ 和零空间 $N(A)$.

(2) 证明对任意 $A \in \mathbb{R}^{m\times n}$, 有

$$C(A) \perp N\left(A^{\top}\right)$$

1.2 解 (1)

$$A = \begin{bmatrix} 1,1,2 \\ 2,1,3 \\ 3,1,4 \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} 1,0,0 \\ 2,-1,0 \\ 3,-2,0 \end{bmatrix} = U.$$

则显然

$$C(A) = \{y : y = Ax, \forall x \in R^n\}$$
$$\Rightarrow C(A) = span\{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}\}$$
$$N(A) = \{x \in R^n : Ax = 0\} = span\{\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\}$$

1.3 证明 (2)

$$\forall \overrightarrow{y_1} \in C(A),$$
$$\forall \overrightarrow{x_2} \in N(A^{\mathrm{T}})$$

下面分情况讨论：
情况一 $r(A) = 0$:

$$r(y) = 0, then$$
$$\langle y_1, x_2 \rangle = 0$$

显然成立，
情况二：$r(A) > 0$:

$$\langle A, x_2 \rangle = 0, means, A^{\mathrm{T}} x_2 = 0$$
$$\Leftrightarrow x_1^{\mathrm{T}} A^{\mathrm{T}} x_2 = 0$$
$$\Leftrightarrow (Ax_1)^{\mathrm{T}} x_2 = 0$$
$$\Leftrightarrow \langle y_1, x_2 \rangle = 0$$

综合以上两种情况，得证

2 HW 2

2.1 问题重述

(1) 证明 ℓ_0 范数不是向量范数
(2) 证明 $\sigma_1 = \sup_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2$, 其中 σ_1 是 A. 的最大奇异值
(3) $\|AB\|_F \leq \|A\|_2 \|B\|_F$ 对任何 $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$ 成立

2.2 证明 (1)

$$||2x_0|| \neq 2 \times ||x_0||$$

不满足齐性要求，得证

2.3 证明 (2)

$$Lemma: \forall U \in Orthogonal \quad matrix \Rightarrow ||Ux||_2 = ||x||_2$$
$$Proof: \forall U \in Orthogonal \quad matrix:$$
$$||Ux||_2|| = ||\sum_j u_j x_j||_2 = \left(\sum_i \left(\sum_j u_{ij}x_j\right)^2\right)^{\frac{1}{2}} = \left(\sum_i \left(\langle u_i, x\rangle\right)^2\right)^{\frac{1}{2}}$$

因为这是正交矩阵, 故可以取 $k_i, i = 1, 2, \ldots, n$, 使得 $x = \sum_{j=1}^n k_j u_j$, 从而原式有：

$$\begin{aligned}
&= \left(\sum_i \left(\left\langle u_i, \sum_{j=1}^n k_j u_j \right\rangle\right)^2\right)^{\frac{1}{2}} \\
&= \left(\sum_i \left(\langle u_i, k_i u_i\rangle\right)^2\right)^{\frac{1}{2}} \\
&= \left(\sum_i \left(k_i \langle u_i, u_i\rangle\right)^2\right)^{\frac{1}{2}} \\
&= \left(\sum_i \left(k_i ||u_i||_2^2\right)^2\right)^{\frac{1}{2}} = \left(\sum_i (k_i)^2\right)^{\frac{1}{2}} = ||x||_2
\end{aligned}$$
$$for: ||x||_2^2 = ||\left(\sum_{j=1}^n k_j u_j\right)||_2^2 = \left(\sum_{i=1}^n k_i^2\right)$$

lemma 得证
对 A 进行奇异值分解：$A = U\Sigma V$

$$||Ax||_2 = ||U\Sigma V x||_2$$
$$(lemma) = ||\Sigma V x||_2$$

令：$Vx = y$

$$= ||\Sigma y||_2 \leq \left(\sum_i \left(\sum_j \sigma_{ij}^2 y_j^2\right)\right)^{\frac{1}{2}}$$
$$= \left(\sum_j \left(\sum_i sigma_{ij}^2 y_j^2\right)\right)^{\frac{1}{2}} \leq \sigma_1 \left(\sum_j y_j^2\right)^{\frac{1}{2}} = \sigma_1 ||y||_2$$
$$= \sigma_1 ||x||_2 = \sigma_1$$

当 $x=V_1$ 时，‘=’ 成立
QED

2.4 证明 (3)

$$Lemma: \forall V \in Orthogonal \qquad matrix \Rightarrow ||VB||_F = ||B||_F \quad for: V \in R^{n\times n} B \in R^{n\times p}$$

$$Proof: ||VB||_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\sum_{k=1}^{n} v_{ik}b_{kj}\right)^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\langle v_i, b_j\rangle\right)^2\right)^{\frac{1}{2}}$$

$$= \left(\sum_{j=1}^{p}||b_j||_2^2\right)^{\frac{1}{2}} = ||B||_F$$

Lemma QED
分解奇异值 $A=U\Sigma V$

$$||AB||_F = ||U\Sigma VB||_F = ||\Sigma VB||_F$$

$$Suppose: VB = C \qquad \Sigma \in R^{m\times n} \qquad C \in R^{n\times p} \qquad \sigma_i, i = 1, 2, \ldots, rank(\Sigma)$$

$$= ||\Sigma C||_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{n}\left(\sigma_{ik}C_{kj}\right)^2\right)^{\frac{1}{2}}$$

$$= \left(\sum_{i=1}^{m}\sum_{j=1}^{p}\left(\sigma_{ii}^2C_{ij}^2\right)^2\right)^{\frac{1}{2}} \le \left(\sum_{i=1}^{m}\sum_{j=1}^{p}\left(\sigma_1^2C_{ij}^2\right)^2\right)^{\frac{1}{2}}$$

$$= \sigma_1\left(\sum_{i=1}^{m}\sum_{j=1}^{p}c_{ij}^2\right)^{\frac{1}{2}}$$

$$= ||A||_2||C||_F = ||A||_2||B||_F$$

3 HW 3

3.1 问题重述

已知

$$f(\mathbf{x}) = \sum_{i=1}^{m}\log\left(1+\exp\left(\langle \mathbf{a}_i, \mathbf{x}\rangle\right)\right) - \langle \mathbf{b}, A\mathbf{x}\rangle$$

是 Logistic 回归的目标函数，那么：
(1) 请证明：$f(\mathbf{x})$ 是 β 是平滑的;
(2) 写下 Logistic 回归的梯度下降算法的迭代公式.

3.2 证明（1）

$$\nabla f(x)=\sum_{i=1}^{m}\left(\frac{\exp\langle a_i,x\rangle}{1+\exp\langle a_i,x\rangle}-b_i\right)a_i^{\mathrm{T}}$$

假设：

$$\begin{gathered}
z_i=\langle \mathbf{a}_i,\mathbf{x}\rangle\,, g(z_i)=\sum_{i=1}^{m}\log\left(1+\exp\left(z_i\right)\right)-b_iz_i\\
G(z)=\sum_{i=1}^{m}g(z_i)\\
\nabla G(z)_i=\left(\frac{\exp\left(z_i\right)}{1+\exp\left(z_i\right)}-b_i\right)
\end{gathered}$$

由通用公式:

$$\nabla f(x)=A^{\top}\nabla G(z)=(a_1,a_2,\ldots,a_m)\begin{pmatrix}\frac{\exp(\langle \mathbf{a}_1,\mathbf{x}\rangle)}{1+\exp(\langle \mathbf{a}_1,\mathbf{x}\rangle)}-b_1\\ \frac{\exp(\langle \mathbf{a}_2,\mathbf{x}\rangle)}{1+\exp(\langle \mathbf{a}_2,\mathbf{x}\rangle)}-b_2\\ \vdots\\ \frac{\exp(\langle \mathbf{a}_m,\mathbf{x}\rangle)}{1+\exp(\langle \mathbf{a}_m,\mathbf{x}\rangle)}-b_m\end{pmatrix}.$$

其中第 i 个

$$\frac{\partial f(x)}{\partial x}=\sum_{j=1}^{m}a_{ij}\left(\frac{\exp\left(\langle \mathbf{a}_j,\mathbf{x}\rangle\right)}{1+\exp\left(\langle \mathbf{a}_j,\mathbf{x}\rangle\right)}-b_j\right)=\sum_{j=1}^{m}a_{ij}\frac{\exp\left(\langle \mathbf{a}_j,\mathbf{x}\rangle\right)}{1+\exp\left(\langle \mathbf{a}_j,\mathbf{x}\rangle\right)}-\sum_{j=1}^{m}a_{ij}b_j.$$

以上就是其梯度，下面计算海森矩阵：
假设：

$$\begin{gathered}
t_i=\langle \mathbf{a}_i,\mathbf{x}\rangle\,\phi(t_i)=\sum_{j=1}^{m}a_{ij}\left(\frac{\exp\left(t_j\right)}{1+\exp\left(t_j\right)}-b_j\right)\\
\Phi(t)=\sum_{j=1}^{m}\phi(t_j).
\end{gathered}$$

由通用公式，得：

$$\nabla^2 f(x)_i = A^\top \nabla\Phi(t) = (a_1, a_2, \dots, a_m)\begin{pmatrix} a_1^{\mathrm{T}}\frac{\exp\langle \mathbf{a}_1,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_1,\mathbf{x}\rangle)^2} \\ a_2^{\mathrm{T}}\frac{\exp\langle \mathbf{a}_2,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_2,\mathbf{x}\rangle)^2} \\ \vdots \\ a_m^{\mathrm{T}}\frac{\exp\langle \mathbf{a}_m,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_m,\mathbf{x}\rangle)^2} \end{pmatrix}.$$

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\exp\langle \mathbf{a}_1,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_1,\mathbf{x}\rangle)^2}, 0, \dots, 0 \\ 0, \frac{\exp\langle \mathbf{a}_2,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_2,\mathbf{x}\rangle)^2}, \dots, 0 \\ \vdots, \quad \ddots, \quad \vdots \\ 0, 0, \dots, \frac{\exp\langle \mathbf{a}_m,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_m,\mathbf{x}\rangle)^2} \end{pmatrix} A^{\mathrm{T}}A.$$

特别地：

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \sum_{j=1}^{m} a_{ij}^2 \left(\frac{\exp\left(\langle \mathbf{a}_j, \mathbf{x}\rangle\right)}{\left(1+\exp\left(\langle \mathbf{a}_j, \mathbf{x}\rangle\right)\right)^2}\right)$$

.

下面计算 $\nabla^2 f(x)$ 的二范数：
首先注意到 $\nabla^2 f(x)$ 最左边的矩阵可以表示为 I 中每个矩阵元素乘以一个小于一的数的初等矩阵 E

$$||\nabla^2 f(x)||_2 = ||EA^{\mathrm{T}}A||_2 \le ||E||_2||A^2A||_2 \le ||A^{\mathrm{T}}A||_2 = \sigma_{max}^2.$$

因此，$f(x)$ 是 $\sigma_{max}^2 - smooth$ 函数，令 $\sigma_{max}^2 = \sigma_1^2$，
写作：$\sigma_1^2 - smooth$

3.3 解 (2)

1: 输入: 给出初值点 $x^0 \in dom(f)$, 容忍度 ϵ 和 $t = 0$
2: while$f(x^t) \ge \epsilon$do
3: $x^{t+1} = x^t - \frac{1}{\sigma_1^2}(a_1, a_2, \dots, a_m)\begin{pmatrix} \frac{\exp(\langle \mathbf{a}_1,\mathbf{x}\rangle)}{1+\exp(\langle \mathbf{a}_1,\mathbf{x}\rangle)} - b_1 \\ \frac{\exp(\langle \mathbf{a}_2,\mathbf{x}\rangle)}{1+\exp(\langle \mathbf{a}_2,\mathbf{x}\rangle)} - b_2 \\ \vdots \\ \frac{\exp(\langle \mathbf{a}_m,\mathbf{x}\rangle)}{1+\exp(\langle \mathbf{a}_m,\mathbf{x}\rangle)} - b_m \end{pmatrix}.,t = t + 1$
4: end while
5: 输出: x^T , 即循环最后的 x^{t+1}.

4 HW 4

4.1 问题重述

(1) $C \subseteq \mathbb{R}^n$ 是一个凸集,$\mathbf{x}_1, \dots, \mathbf{x}_k \in C$, 取 $\theta_1, \theta_2, \dots, \theta_k \in \mathbb{R}$ 满足 $\theta_i \ge 0, \sum_i \theta_i = 1$.

试证明 $\theta_1 \mathrm{x}_1+\theta_2 \mathrm{x}_2+\cdots+\theta_k \mathbf{x}_k \in C$.
(2) 证明椭球体 $E\left(\mathbf{x}_c\right)=\left\{\mathbf{x} \mid\left(\mathbf{x}-\mathbf{x}_c\right)^{\top} A\left(\mathbf{x}-\mathbf{x}_c\right) \leq 1, A \in \mathcal{S}_{++}^n\right\}$ 是一个凸集.
(3) 设 f 是一个凸函数，并且表示该集合包含所有可以达到 f 的全局最小值的点，即 G
请证明 G 是凸的.

4.2 证明 (1)

C 为凸集则 $\forall x_i, x_j \in C$ 满足：

$$
\begin{gathered}
\theta_i x_i+(1-\theta_i) x_j \in C \\
Suppose: x_i=x_{k-1}, x_j=x_k \\
\Rightarrow x_{k-1}^{\star}=\frac{\theta_{k-1}}{1-\theta_1-\theta_2-\ldots-\theta_{k-2}} x_{k-1}+\frac{\theta_k}{1-\theta_1-\theta_2-\ldots-\theta_{k-2}} x_k \in C \\
\Rightarrow x_{k-2}^{\star}=\frac{\theta_{k-2}}{1-\theta_1-\theta_2-\ldots-\theta_{k-3}} x_{k-2}+\frac{\theta_{k-1}+\theta_k}{1-\theta_1-\theta_2-\ldots-\theta_{k-3}} x_{k-1}^{\star} \in C \\
\ldots \\
\Rightarrow x_2^{\star}=\frac{\theta_2}{1-\theta_1} x_2+\frac{1-\theta_1-\theta_2}{1-\theta_1} x_k^{\star} \in C \Rightarrow \theta_1 x_1+(1-\theta_1) x_2^{\star} \in C
\end{gathered}
$$

其中 $\theta_1 x_1+(1-\theta_1) x_2^{\star} \in C$ 代入 $x_2^{\star}, x_3^{\star}, \ldots, x_{k-1}^{\star}$ 可以得到：$\theta_1 \mathrm{x}_1+\theta_2 \mathrm{x}_2+\cdots+\theta_k \mathbf{x}_k \in C$.

4.3 证明 (2)

$\forall \theta_i \neq 0$

$$
\forall x_1, x_2 \in E(x_c) \quad, \forall \beta \in(0,1)
$$

也就是：$(x_1-x_c)^{\mathrm{T}} A x_1-x_c \leq 1$;$(x_2-x_c)^{\mathrm{T}} A x_2-x_c \leq 1$

$$
\begin{gathered}
(\theta x_1+(1-\theta) x_2-x_c)^{\mathrm{T}} A(\theta x_1+(1-\theta) x_2-x_c) \\
=(\theta(x_1-x_c)+(1-\theta)(x_2-x_c))^{\mathrm{T}} A(\theta(x_1-x_c)+(1-\theta)(x_2-x_c)) \\
=\theta^2(x_1-x_c)^{\mathrm{T}} A(x_1-x_c)+\theta(1-\theta)(x_1-x_c)^{\mathrm{T}} A(x_2-x_c)+ \\
\theta(1-\theta)(x_2-x_c)^{\mathrm{T}} A(x_1-x_c)+(1-\theta)^2(x_2-x_c)^{\mathrm{T}} A(x_2-x_c) \\
<\theta^2+(1-\theta)^2+\theta(1-\theta)(x_1-x_c)^{\mathrm{T}} A(x_2-x_c)+(1-\theta) \theta(x_2-x_c)^{\mathrm{T}} A(x_1-x_c) \\
=\theta^2+(1-\theta)^2+ \\
\theta(1-\theta)\left((x_1-x_2)^{\mathrm{T}} A(x_2-x_c)+(x_2-x_1)^{\mathrm{T}} A(x_1-x_2)+(x_1-x_2)^{\mathrm{T}} A(x_1-x_c)+(x_2-x_c)^{\mathrm{T}} A(x_2-x_c)\right) \\
\leq \theta^2+(1-\theta)^2+2 \theta(1-\theta)=1
\end{gathered}
$$

QED

4.4 证明 (3)

$$\forall x_1 \in G, x_2 \in G which means : \forall x \in dom(f), f(x) \geq f(x_1), f(x) \geq f(x_2) \quad and \quad f(x_1) = f(x_2)$$
$$f(x_1) \leq f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta)f(x_2) = f(x_1)$$

则：$f(\theta x_1 + (1-\theta)x_2) = f(x_1)$，　$\theta x_1 + (1-\theta)x_2 \in G$
QED

5 HW5

5.1 问题重述

(1) 设 a 和 **b** 是 $\mathbb{R}^n$ 中的不同点. 表示比 **b** 更接近的所有点的集合是半空间.
(2) 两个平行超平面 $\{\mathrm{x} \mid \mathbf{a}^\top \mathrm{x} = b_1\}$ 和 $\{\mathrm{x} \mid \mathbf{a}^\top \mathbf{x} = b_2\}$ 之间的距离是多少?

5.2 证明 (1)

首先：$\exists x_0 = \frac{1}{2}a + \frac{1}{2}b$ 满足 $||x_0 - a||_2 = ||x_0 - b||_2$
且：$\forall x$ 满足

$$\begin{aligned}\langle x - x_0, b - a\rangle = 0 \Leftrightarrow x^{\mathrm{T}}b - x^{\mathrm{T}}a - \frac{1}{2}b^{\mathrm{T}}b + \frac{1}{2}a^{\mathrm{T}}a = 0 \\ \Leftrightarrow \frac{1}{2}b^{\mathrm{T}}b - x^{\mathrm{T}}b + \frac{1}{2}x^{\mathrm{T}}x = \frac{1}{2}a^{\mathrm{T}}a - x^{\mathrm{T}}a + \frac{1}{2}x^{\mathrm{T}}x \\ \Leftrightarrow \langle b - x, b - x\rangle = \langle a - x, a - x\rangle \Leftrightarrow d_1 = d_2\end{aligned}$$

$$Suppose : d_1 = ||\vec{x} - \vec{a}||_2 = \sqrt{\langle \vec{x}, \vec{a}\rangle} \quad d_2 = ||\vec{x} - \vec{a}||_2 = \sqrt{\langle \vec{x}, \vec{a}\rangle}$$
$$d_1 < d_2 \Leftrightarrow \langle x - a, x - a\rangle < \langle x - b, x - b\rangle \quad \Leftrightarrow \frac{1}{2}a^{\mathrm{T}}a - x^{\mathrm{T}}a + \frac{1}{2}x^{\mathrm{T}}x < \frac{1}{2}b^{\mathrm{T}}b - x^{\mathrm{T}}b + \frac{1}{2}x^{\mathrm{T}}x$$
$$\Leftrightarrow \frac{1}{2}a^{\mathrm{T}}a - x^{\mathrm{T}}a < \frac{1}{2}b^{\mathrm{T}}b - x^{\mathrm{T}}b$$
$$\Leftrightarrow \frac{1}{2}a^{\mathrm{T}}a - x^{\mathrm{T}}a - \frac{1}{2}b^{\mathrm{T}}b + x^{\mathrm{T}}b < 0$$
$$\Leftrightarrow \frac{1}{2}x^{\mathrm{T}}b - x^{\mathrm{T}}a - \frac{1}{2}x_0^{\mathrm{T}}b + x_0^{\mathrm{T}}a < 0$$
$$\Leftrightarrow \langle x, b - a\rangle - \langle x_0, b - a\rangle < 0 \Leftrightarrow \langle x - x_0, b - a\rangle < 0$$

是一个半空间，QED

5.3 解 (2)

$$\forall x_1 \in \{x \| a^{\mathrm{T}} x = b_1\}, x_2 \in \{x | a^{\mathrm{T}} = b_2\}$$
$$|b_1 - b_2| = |\langle x_2 - x_1, a\rangle| \leq ||a||_2 ||x_2 - x_1||_2$$
$$\Leftrightarrow ||x_2 - x_1||_2 \geq \frac{|b_2 - b_1|}{||a_2||_2}$$

距离是：$\frac{|b_2 - b_1|}{||a_2||_2}$

6 HW 6

6.1 问题重述

证明下列函数是凸的。

(1) 负熵：$f(x) = x\log(x), x > 0$.

(2) 二次超线性函数：$f(x, y) = \frac{x^2}{y}$ 有 $\mathrm{dom}(f) = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$.

(3) $f(\mathbf{x}) = \|A\mathbf{x} - b\|$.

(4) 支持函数：$S_C(\mathbf{x}) = \sup_{\mathbf{b}\in C} \mathbf{b}^\top \mathbf{x}$.

(5) 点 x 到集合 $S \subset \mathbb{R}^n$ 的距离 $d(\mathrm{x}, S) = \inf_{\mathbf{b}\in S} \|\mathrm{x} - \mathrm{b}\|$.

6.2 证明 (1)

$$\nabla f(x) = log(x) + 1 \tag{1}$$
$$\nabla^2 f(x) = \frac{1}{x} \geq 0 \tag{2}$$

则：：$f(x) = x\log(x), x > 0$ 是凸函数

6.3 证明 (2)

$$\nabla f(x,y)=\begin{bmatrix}\frac{2x}{y}\\ \frac{-x^2}{y^2}\end{bmatrix}\forall(x_1,y_1),(x_2,y_2)\in dom(f)$$

$$x_1^2y_2^2+x_2^2y_1^2-2x_1x_2y_1y_2\geq 0$$

$$\Leftrightarrow\frac{x_1y_2}{x_2^2y_1^2}\left(2x_2y_1-x_1y_2\right)\leq 1$$

$$\Leftrightarrow\frac{x_1}{y_1}\left(\frac{2x_2y_1-x_1y_2}{y_1}\right)\leq\frac{x_2^2}{y_2}$$

$$\Leftrightarrow\frac{x_1^2}{y_1}+\frac{x_1}{y_1}\left(2x_2-2x_1-\frac{x_1}{y_1(y_2-y-1)}\right)\leq f(x_2,y_2)$$

$$\Leftrightarrow f(x_1,y_2)+\left(\frac{2x_1}{y_1},\frac{-x_1^2}{y_1}\right)\begin{pmatrix}x_2-x_1\\ y_2-y_1\end{pmatrix}\leq f(x_2,y_2)$$

$$\Leftrightarrow f(x_2,y_2)\geq f(x_1,y_1)+\langle\nabla f(x),(x_2,y_2)-(x_1,y_1)\rangle$$

因此，$f(x,y)$ 是凸函数得证

6.4 证明 (3)

由通用公式，令 $z=Ax$, 则 $G(z)=||z-b||_2,\nabla G(z)=\frac{z-b}{||z-b||_2},\nabla f(x)=A^{\mathrm{T}}\frac{Ax-b}{||Ax-b||_2}$
则 Hessian 矩阵：

$$\frac{\partial\nabla f(x)}{\partial x}=A^{\mathrm{T}}\frac{\partial}{\partial x}\frac{Ax-b}{||Ax-b||_2}$$

$$=A^{\mathrm{T}}diag\left(\frac{||Ax-b||_2^2-(a_1^{\mathrm{T}}x-b)(a_1^{\mathrm{T}}x-b)^{\mathrm{T}}}{||Ax-b||_2^2},\ldots,\frac{||Ax-b||_2^2-(a_n^{\mathrm{T}}x-b)(a_n^{\mathrm{T}}x-b)^{\mathrm{T}}}{||Ax-b||_2^2}\right)A$$

显然 $\nabla^2f(x)$ 是半正定阵，则 $||\nabla^2f(x)||_2\geq 0$ 成立，得证

6.5 证明 (4)

设函数 $f(x,b_i)=b_i^{\mathrm{T}}x,b(i)\in C$. 显然:$f(x,b_i)$ 是凸函数.
则 $epi\left(f(x,b_i)\right)$ 是凸集合。
则：$\bigcap_{b_i\in C}epi\left(f(x,b_i)\right)=epi\left(f(x)\right)=epi\left(\max_{b_i\in C}f(x,bi)\right)$ 是凸集合。
其中：$f(x)=\sup_{\mathbf{b}\in C}b^{\mathrm{T}}x=S_c(x)$
可知：$f(x)$ 是凸函数，也就是 $S_c(x)$ 是凸函数

6.6 证明 (5)

设：$f(x,b)=||x-b||$
对 $\forall\epsilon\geq 0,\forall x_i\in dom(f(x,b)),i=1,2,\exists b_i,i=1,2$ 满足：

$f(x_i, b_i) \leq d(x_i, s) + \epsilon$

$d(\theta x_1 + (1-\theta)x_2, s) = \inf_{\mathbf{b} \in S} f(\theta x_1 + (1-\theta)x_2, b) \leq f(\theta x_1 + (1-\theta)x_2, \theta b_1 + (1-\theta)b_2) \quad \leq \theta f(x_1, b_1) + (1-\theta) f(x_2,$

其中令：$\epsilon = max(\theta\epsilon_1, (1-\theta)\epsilon_2) \to 0$

7 HW 7

7.1 问题重述

已知 $A = U\Sigma V = \begin{bmatrix} -0.91 & 0.37 & -0.18 \\ 0.19 & 0.78 & 0.59 \\ 0.36 & 0.50 & -0.78 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.25 & -0.73 & -0.62 \\ -0.74 & -0.56 & 0.35 \\ -0.61 & 0.37 & -0.69 \end{bmatrix}$

且 $\mathbf{b} = (-0.29, -2.09, -0.98)^\top$ 然后求解 LS 问题：

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2$$

* 通过回溯线搜索实现梯度下降算法，以解决 LS 问题。

* 为 $beta$-平滑函数实现梯度下降，以解决 LS 问题。

* 比较不同 A 的收敛速度。

$$A_1 = \begin{bmatrix} -0.91 & 0.37 & -0.18 \\ 0.19 & 0.78 & 0.59 \\ 0.36 & 0.50 & -0.78 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix} \begin{bmatrix} 0.25 & -0.73 & -0.62 \\ -0.74 & -0.56 & 0.35 \\ -0.61 & 0.37 & -0.69 \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} -0.91 & 0.37 & -0.18 \\ 0.19 & 0.78 & 0.59 \\ 0.36 & 0.50 & -0.78 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 10^{-6} \end{bmatrix} \begin{bmatrix} 0.25 & -0.73 & -0.62 \\ -0.74 & -0.56 & 0.35 \\ -0.61 & 0.37 & -0.69 \end{bmatrix}$$

7.2 求解过程

代码及完整运行结果见附页，运行结果如下：

第一种情况下，两种算法在迭代次数、时间和运行结果上相差不大, 相对残差都比较小

```
结果是:
[[ 1.04442925]
 [ 0.61473373]
 [-0.01459383]]
回退法相对残差、迭代次数及时间: {'迭代次数': 16, '相对残差:': 0.01300824454094233
8}
Wall time: 168 ms
这是使用β法求解第一种情况:
设定可以容忍的最大迭代次数: 1000000000000000000000000000000
默认s0是: 100
结果是:
[[ 1.05195036]
 [ 0.59173225]
 [-0.01125154]]
β法结相对残差、迭代次数及时间: {'迭代次数': 47, '相对残差:': 0.01182450914789473
5}
Wall time: 24 ms
```

Figure 1: 情况一运行结果

```
结果是:
[[ 0.79076079]
 [ 0.74915293]
 [-0.28764644]]
回退法相对残差、迭代次数及时间: {'迭代次数': 5, '相对残差:': 0.17811617184547832}
Wall time: 24 ms
这是使用β法求解第二种情况:
设定可以容忍的最大迭代次数: 1000000000000000000000000000000
默认s0是: 100
结果是:
[[ 20.65956274]
 [-11.48775518]
 [ 22.11766537]]
β法结相对残差、迭代次数及时间: {'迭代次数': 131138, '相对残差:': 0.0420197791869
6347}
Wall time: 29.4 s
```

Figure 2: 情况二运行结果

第二种情况下：最终回退法没有收敛到 x* 附近，相对残差比较大，运行速度可以接受 β 方法收敛到了 x* 附近，相对残差较小，但运行速度较慢，在 30 秒左右

第三种情况下：两种算法在运行结果都差不多，相对残差都较大。

```
结果是:
[[ 0.78902485]
 [ 0.750218  ]
 [-0.28960461]]
回退法相对残差、迭代次数及时间: {'迭代次数': 5, '相对残差:': 0.17814488504840534}
Wall time: 10 ms
这是使用β法求解第三种情况:
设定可以容忍的最大迭代次数: 1000000000000000000000000000000
默认s0是: 100
结果是:
[[ 0.79765114]
 [ 0.74838993]
 [-0.29824491]]
```

```
β法相对残差、迭代次数及时间: {'迭代次数': 21, '相对残差:': 0.1779828155123108}
Wall time: 9.98 ms
```

Figure 3: 情况三运行结果

参考文献