Problem 1

(a)假设

$$\begin{aligned} \Delta J &= J_k - J_{k-1} \\ \Delta \vec{x} &= \vec{x}_k - \vec{x}_{k-1} \\ \vec{d} &= f(\vec{x}_k) - f(\vec{x}_{k-1}) - J_{k-1}\Delta \vec{x} \end{aligned}$$

那么原问题可以化为如下形式：

$$\min_{\Delta J} \|\Delta J\|_{\text{Fro}}^2$$
$$\text{such that } \Delta J. \Delta \vec{x} = \vec{d}$$

构造拉格朗日乘子：

$$\Lambda = \|\Delta J\|_{\text{Fro}}^2 + \vec{\lambda}^T(\Delta J. \Delta \vec{x} - \vec{d})$$

关于$(\Delta J)_{ij}$求偏导并令其为0得到

$$0 = \frac{\partial \Lambda}{(\Delta J)_{ij}} = 2(\Delta J)_{ij} + \lambda_i (\Delta \vec{x})_j$$

所以

$$\Delta J = -\frac{1}{2}\vec{\lambda}(\Delta \vec{x})^T$$

带回$\Delta J. \Delta \vec{x} = \vec{d}$可得

$$\vec{\lambda}(\Delta \vec{x})^T(\Delta \vec{x}) = -2\vec{d} \Rightarrow \vec{\lambda} = -\frac{2\vec{d}}{\|\Delta \vec{x}\|^2}$$

因此

$$\Delta J = -\frac{1}{2}\vec{\lambda}(\Delta \vec{x})^T = \frac{\vec{d}(\Delta \vec{x})^T}{\|\Delta \vec{x}\|^2}$$

回顾各项的定义，我们得到

$$\begin{aligned} J_k &= J_{k-1} + \Delta J \\ &= J_{k-1} + \frac{\vec{d}(\Delta \vec{x})^T}{\|\Delta \vec{x}\|^2} \\ &= J_{k-1} + \frac{(f(\vec{x}_k) - f(\vec{x}_{k-1}) - J_{k-1}\Delta \vec{x})}{\|\vec{x}_k - \vec{x}_{k-1}\|^2}(x_k - \vec{x}_{k-1})^T \end{aligned}$$

(b)带入验证即可：

$$
\begin{aligned}
\left(A+\vec{u}\vec{v}^T\right)\left(A^{-1}-\frac{A^{-1}\vec{u}\vec{v}^T A^{-1}}{1+\vec{v}^T A^{-1}\vec{u}}\right) &= I+\vec{u}\vec{v}^T A^{-1}-\frac{\vec{u}\vec{v}^T A^{-1}}{1+\vec{v}^T A^{-1}\vec{u}}-\frac{\vec{u}\vec{v}^T A^{-1}\vec{u}\vec{v}^T A^{-1}}{1+\vec{v}^T A^{-1}\vec{u}} \\
&= I+\vec{u}\vec{v}^T A^{-1}-\frac{\vec{u}\vec{v}^T A^{-1}}{1+\vec{v}^T A^{-1}\vec{u}}-\frac{\vec{u}(\vec{v}^T A^{-1}\vec{u})\vec{v}^T A^{-1}}{1+\vec{v}^T A^{-1}\vec{u}} \\
&= I+\vec{u}\vec{v}^T A^{-1}-(\vec{u}\vec{v}^T A^{-1})\frac{1+\vec{v}^T A^{-1}\vec{u}}{1+\vec{v}^T A^{-1}\vec{u}} \\
&= I+\vec{u}\vec{v}^T A^{-1}-\vec{u}\vec{v}^T A^{-1} \\
&= I
\end{aligned}
$$

(c)原始的迭代形式为：

$$
J_k = J_{k-1}+\vec{u}_k\vec{v}_k^T
$$

由(b)可得

$$
J_k^{-1} = J_{k-1}^{-1}-\frac{J_{k-1}^{-1}\vec{u}_k\vec{v}_k^T J_{k-1}^{-1}}{1+\vec{v}_k^T J_{k-1}^{-1}\vec{u}_k}
$$

其中

$$
\begin{aligned}
\vec{u}_k &= \frac{(f(\vec{x}_k)-f(\vec{x}_{k-1})-J_{k-1}\Delta\vec{x})}{\|\vec{x}_k-\vec{x}_{k-1}\|^2} \\
\vec{v}_k &= x_k-\vec{x}_{k-1}
\end{aligned}
$$

Problem 2

(a)使用奇异值分解：

$$
A = U\Sigma V^T
$$

因为$m=n$，所以Σ为对角阵，即

$$
\Sigma^T = \Sigma
$$

因此

$$
\begin{aligned}
A^T A &= V\Sigma^T U^T U\Sigma V^T \\
&= V\Sigma^2 V^T \\
\sqrt{A^T A} &= V\Sigma V^T
\end{aligned}
$$

计算trace可得

$$
\begin{aligned}
\operatorname{trace}(\sqrt{A^TA}) &= \operatorname{trace}(V\Sigma V^T) \\
&= \operatorname{trace}(V^TV\Sigma) \qquad \operatorname{trace}(AB)=\operatorname{trace}(BA) \\
&= \operatorname{trace}(\Sigma) \\
&= \sum_{i=1}^n \sigma_i(A) \\
&= \|A\|_*
\end{aligned}
$$

(b)证明一般情形，如果$A\in\mathbb{R}^{n\times m},B\in\mathbb{R}^{m\times n}$，那么

$$
\operatorname{trace}(AB)=\operatorname{trace}(BA)
$$

注意到

$$
(AB)_{ii}=\sum_{s=1}^m A_{is}B_{si},(BA)_{ss}=\sum_{i=1}^n B_{si}A_{is}
$$

注意到

$$
AB\in\mathbb{R}^{n\times n},BA\in\mathbb{R}^{m\times m}
$$

所以

$$
\begin{aligned}
\operatorname{trace}(AB) &= \sum_{i=1}^n (AB)_{ii} \\
&= \sum_{i=1}^n\sum_{s=1}^m A_{is}B_{si} \\
&= \sum_{s=1}^m\sum_{i=1}^n B_{si}A_{is} \\
&= \sum_{s=1}^m (BA)_{ss} \\
&= \operatorname{trace}(BA)
\end{aligned}
$$

(c)由SVD分解可得

$$
AC=U\Sigma V^TC
$$

记

$$
(V')^T=V^TC
$$

那么

$$
AC=U\Sigma(V')^T
$$

并且

$$
(V')(V')^T=C^TVV^TC=C^TC=I
$$

利用定义可得

$$
\begin{aligned}
\operatorname{trace}(AC) &= \operatorname{trace}(U\Sigma(V')^T) \\
&= \sum_{i=1}^n \sigma_i(A) u_i^T v_i' \\
&\le \sum_{i=1}^n \sigma_i(A) \|u_i\|. \|v_i'\| \\
&= \sum_{i=1}^n \sigma_i(A) \\
&= \|A\|_*
\end{aligned}
$$

当且仅当

$$
v_i' = u_i
$$

时等号成立，即

$$
V' = C^T V = U, C^T = UV^T, C = VU^T
$$

(d)对于满足条件$C^TC=I$的C，我们有

$$
\begin{aligned}
\operatorname{trace}\Big((A+B)C\Big) &= \operatorname{trace}(AC) + \operatorname{trace}(BC) \\
&\le \|A\|_* + \|B\|_*
\end{aligned}
$$

所以

$$
\begin{aligned}
\|A+B\|_* &= \max_{C^TC=I} \operatorname{trace}\Big((A+B)C\Big) \\
&\le \|A\|_* + \|B\|_*
\end{aligned}
$$

(e)令

$$
A' = (\sigma_1(A), \ldots, \sigma_n(A))^T
$$

那么我们需要最小化

$$
\|A - A_0\|_{\text{Fro}}^2 + \|A'\|_1
$$

由L_1正则化的特性，我们的结果会使得A'某些项为0，不妨设非零项的下标为

$$
k_1, \ldots, k_m
$$

由SVD可得

$$
A = \sum_{\sigma_i(A)\ne 0} \sigma_i(A) u_i v_i^T = \sum_{j=1}^m \sigma_{k_j}(A) u_{k_j} v_{k_j}^T
$$

所以得到A_0的低秩近似。

Problem 3

(a)回顾割线法的定义：

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

注意到

$$f(x')=0$$

如果$x_k=x'$，那么

$$f(x_k)=0$$

即

$$\begin{aligned}x_{k+1}&=x_k-\frac{f(x_k)(x_k-x_{k-1})}{f(x_k)-f(x_{k-1})}\\&=x'\end{aligned}$$

如果$x_{k-1}=x'$，那么

$$f(x_{k-1})=0$$

即

$$\begin{aligned}x_{k+1}&=x_k-\frac{f(x_k)(x_k-x')}{f(x_k)}\\&=x_k-(x_k-x')\\&=x'\end{aligned}$$

(b)假设A的奇异值为

$$\sigma_1(A)\ge\ldots\ge\sigma_k(A)$$

回顾SVD的推导，我们知道

$$R_A(\vec{x})=\frac{\|A\vec{x}\|}{\|\vec{x}\|}\in[\sigma_k(A)，\ \sigma_1(A)]$$

现在假设增加一行$\vec{\alpha}^T$，那么

$$\tilde{A}=\begin{bmatrix}A\\\vec{\alpha}^T\end{bmatrix},\tilde{A}\vec{x}=\begin{bmatrix}A\vec{x}\\\vec{\alpha}^T\vec{x}\end{bmatrix}$$

因此

$$
\begin{aligned}
R_{\tilde{A}}(\vec{x}) &= \frac{\|\tilde{A}\vec{x}\|}{\|\vec{x}\|} \\
&= \frac{\left\| \begin{bmatrix} A\vec{x} \\ \vec{\alpha}^T\vec{x} \end{bmatrix} \right\|}{\|\vec{x}\|} \\
&\geq \frac{\|A\vec{x}\|}{\|\vec{x}\|} \\
&= R_A(\vec{x})
\end{aligned}
$$

因此$\tilde{A}$的最小奇异值和最大奇异值均不小于A的最小奇异值和最大奇异值。

Problem 4

(a)将$\frac{1}{a}$视为

$$
f(x) = \frac{1}{x} - a
$$

的零点，然后利用牛顿迭代法迭代即可，注意

$$
f'(x) = -\frac{1}{x^2}
$$

所以

$$
x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k + x_k^2(\frac{1}{x_k} - a) = 2x_k - ax_k^2
$$

(b)

$$
\begin{aligned}
\epsilon_{k+1} &= ax_{k+1} - 1 \\
&= 2ax_k - a^2x_k^2 - 1 \\
&= -(ax_k - 1)^2 \\
&= -\epsilon_k^2
\end{aligned}
$$

(c)由(b)可得

$$
|\epsilon_{k+1}| = |\epsilon_k^2|, |\epsilon_k| = |\epsilon_0|^{2^k}
$$

要使得计算结果达到d位2进制小数，我们有

$$
\begin{aligned}
|\epsilon_0|^{2^k} &= 2^{-d} \\
2^k \ln|\epsilon_0| &= -d\ln 2 \\
2^k &= -\frac{d\ln 2}{\ln|\epsilon_0|} \\
k &= \log_2(-\frac{d\ln 2}{\ln|\epsilon_0|})
\end{aligned}
$$