Problem 1 Finding df, d^2f, ∇f and $\nabla^2 f$.
1(a) $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{2}\|xx^T - A\|_F^2$, $A \in \mathbb{S}^n$
Solution:

$$\begin{array}{rcl}
f(x) &=& \frac{1}{2}\|xx^T - A\|_F^2 \\
&=& \frac{1}{2}\mathrm{tr}\left(\left(xx^T - A\right)^T\left(xx^T - A\right)\right) \\
&=& \frac{1}{2}\mathrm{tr}(((xx^T)^T - A^T)(xx^T - A)) \\
&=& \frac{1}{2}\mathrm{tr}\left(\left(xx^T - A\right)\left(xx^T - A\right)\right) \text{ since, } A \in \mathbb{S}^n \\
&=& \frac{1}{2}\mathrm{tr}\left(xx^Txx^T - xx^TA - Axx^T + A^2\right) \\
&=& \frac{1}{2}\left(\mathrm{tr}\left(xx^Txx^T\right) - \mathrm{tr}\left(xx^TA\right) - \mathrm{tr}\left(Axx^T\right) + \mathrm{tr}\left(A^2\right)\right)
\end{array}$$

Applying cyclic properties of trace,

$$\begin{array}{rcl}
&=& \frac{1}{2}\left(\mathrm{tr}\left(x^Txx^Tx\right) - \mathrm{tr}\left(x^TAx\right) - \mathrm{tr}\left(x^TAx\right) + \mathrm{tr}\left(A^2\right)\right) \\
&=& \frac{1}{2}\left(\mathrm{tr}\left((x^Tx)^2\right) - 2\mathrm{tr}\left(x^TAx\right) + \mathrm{tr}\left(A^2\right)\right)
\end{array}$$

x^Tx and x^TAx are scalar quantities

$$\begin{array}{rcl}
&=& \frac{1}{2}\left((x^Tx)^2 - 2x^TAx + \mathrm{tr}\left(A^2\right)\right) \\
&=& \frac{1}{2}\left(\langle x, x\rangle^2 - 2\langle Ax, x\rangle + \mathrm{tr}\left(A^2\right)\right) \\
&=& \frac{1}{2}\langle x, x\rangle^2 - \langle Ax, x\rangle + \frac{1}{2}\mathrm{tr}\left(A^2\right)
\end{array}$$

Thus,

$$\begin{array}{rcl}
\Rightarrow df &=& \frac{1}{2}d(\langle x, x\rangle^2) - d\langle Ax, x\rangle + 0 \\
&=& \frac{1}{2}(d\langle x, x\rangle)\langle x, x\rangle + \frac{1}{2}\langle x, x\rangle(d\langle x, x\rangle) - \langle(A + A^T)x, dx\rangle \\
&=& (d\langle x, x\rangle)\langle x, x\rangle - \langle(A + A)x, dx\rangle \\
&=& \langle(I_n + I_n^T)x, dx\rangle\langle x, x\rangle - \langle(2A)x, dx\rangle \\
&=& \langle(2\langle x, x\rangle I_n)x, dx\rangle - \langle(2A)x, dx\rangle \\
&=& \langle(2\langle x, x\rangle I_n)x - (2A)x, dx\rangle \\
df &=& \langle 2(\langle x, x\rangle I_n - A)x, dx\rangle \\
\Rightarrow \nabla f &=& 2(\langle x, x\rangle I_n - A)x \\
d^2f &=& d\left(\langle 2(\langle x, x\rangle I_n - A)x, dx_1\rangle\right) \\
&=& \langle d\left(2(\langle x, x\rangle I_n - A)x\right), dx_1\rangle
\end{array}$$

Now, we will first resolve, $d\left(2(\langle x,x\rangle I_n - A)x\right)$

$$\begin{aligned}
d\left(2(\langle x,x\rangle I_n - A)x\right) &= 2d\left((\langle x,x\rangle I_n - A)x\right)\\
&= 2d(\langle x,x\rangle I_n x) - 2d(Ax)\\
&= 2d(\langle x,x\rangle x) - 2d(Ax)\\
&= 2(d\langle x,x\rangle)x + 2\langle x,x\rangle(dx) - 2Adx\\
&= 4\langle x,dx\rangle x + 2\langle x,x\rangle dx - 2Adx\\
&= 4x\langle x,dx\rangle + 2\langle x,x\rangle dx - 2Adx\\
&= 4xx^T dx + 2\langle x,x\rangle dx - 2Adx\\
d\left(2(\langle x,x\rangle I_n - A)x\right) &= (4xx^T + 2\langle x,x\rangle I_n - 2A)dx
\end{aligned}$$

Now, back to the second derivative

$$\begin{aligned}
d^2 f &= \langle d\left(2(\langle x,x\rangle I_n - A)x\right), dx_1\rangle\\
&= \langle (4xx^T + 2\langle x,x\rangle I_n - 2A)dx, dx_1\rangle\\
d^2 f &= \langle (4xx^T + 2\langle x,x\rangle I_n - 2A)dx_1, dx\rangle\\
\Rightarrow \nabla^2 f &= 4xx^T + 2\langle x,x\rangle I_n - 2A
\end{aligned}$$

1(b) $f : \mathbb{R}^n\backslash\{0\} \to \mathbb{R}$, $f(x) = \frac{\langle Ax,x\rangle}{\|x\|^2}$, $A \in \mathbb{S}^n$

$$\begin{aligned}
df &= d\left(\frac{\langle Ax,x\rangle}{\|x\|^2}\right)\\
&= \frac{\|x\|^2 d\langle Ax,x\rangle - \langle Ax,x\rangle d(\|x\|^2)}{\|x\|^4}\\
&= \frac{\|x\|^2\langle (A+A^T)x,dx\rangle - \langle Ax,x\rangle d\langle x,x\rangle}{\|x\|^4}\\
&= \frac{\|x\|^2\langle (A+A)x,dx\rangle - \langle Ax,x\rangle d\langle xI_n,x\rangle}{\|x\|^4}\\
&= \frac{\|x\|^2\langle 2Ax,dx\rangle - \langle Ax,x\rangle\langle (I_n+I_n^T)x,dx\rangle}{\|x\|^4}\\
&= \frac{\|x\|^2\langle 2Ax,dx\rangle - \langle Ax,x\rangle\langle 2x,dx\rangle}{\|x\|^4}\\
&= \left\langle \frac{2Ax}{\|x\|^2}, dx\right\rangle - \left\langle \frac{2\langle Ax,x\rangle x}{\|x\|^4}, dx\right\rangle\\
\Rightarrow df &= \left\langle \frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}, dx\right\rangle\\
\Rightarrow \nabla f &= \frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\\
d^2 f = d\left(\left\langle \frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}, dx_1\right\rangle\right) &= \left\langle d\left(\frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\right), dx_1\right\rangle
\end{aligned}$$

Now, we will first resolve $d\left(\frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\right)$

$$\begin{aligned}
d\left(\frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\right) &= d\left(\frac{2Ax}{\|x\|^2}\right) - d\left(\frac{2\langle Ax,x\rangle x}{\|x\|^4}\right)\\
&= \frac{\|x\|^2 d(2Ax) - d(\|x\|^2)2Ax}{\|x\|^4} - \frac{\|x\|^4 d(2\langle Ax,x\rangle x) - d(\|x\|^4)2\langle Ax,x\rangle x}{\|x\|^8}\\
d\left(\frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\right) &= \frac{2\|x\|^2 Adx - \langle 2x,dx\rangle 2Ax}{\|x\|^4} - \frac{2\|x\|^4 d(\langle Ax,x\rangle x) - 2d(\|x\|^4)\langle Ax,x\rangle x}{\|x\|^8}
\end{aligned}$$

We need to resolve $d(\langle Ax, x\rangle x)$ and $d(\|x\|^4)$

$$\begin{aligned}
d(\langle Ax, x\rangle x) &= (d\langle Ax, x\rangle)x + \langle Ax, x\rangle dx \\
&= \langle 2Ax, dx\rangle x + \langle Ax, x\rangle dx \\
&= 2x\langle Ax, dx\rangle + \langle Ax, x\rangle dx \\
&= 2xx^T Adx + \langle Ax, x\rangle dx \\
\Rightarrow d(\langle Ax, x\rangle x) &= (2xx^T A + \langle Ax, x\rangle I_n)dx \\
d(\|x\|^4) = d(\langle x, x\rangle^2) &= 2\langle x, x\rangle d(\langle x, x\rangle) \\
\Rightarrow d(\|x\|^4) &= 2\|x\|^2\langle 2x, dx\rangle \\
d\left(\frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\right) &= \frac{2\|x\|^2 Adx - 4(x^T dx)Ax}{\|x\|^4} - \frac{2\|x\|^4(2xx^T A + \langle Ax,x\rangle I_n)dx - 4\|x\|^2\langle 2x,dx\rangle\langle Ax,x\rangle x}{\|x\|^8} \\
&= \frac{2\|x\|^2 Adx - 4Ax(x^T dx)}{\|x\|^4} - \frac{2\|x\|^4(2xx^T A + \langle Ax,x\rangle I_n)dx - 8\|x\|^2(x^T dx)\langle Ax,x\rangle x}{\|x\|^8} \\
&= \frac{(2\|x\|^2 A - 4Axx^T)dx}{\|x\|^4} - \frac{2\|x\|^2(2xx^T A + \langle Ax,x\rangle I_n)dx - 8\langle Ax,x\rangle xx^T dx}{\|x\|^6} \\
&= \left(\frac{2\|x\|^2 A - 4Axx^T}{\|x\|^4} - \frac{2\|x\|^2(2xx^T A + \langle Ax,x\rangle I_n) - 8\langle Ax,x\rangle xx^T}{\|x\|^6}\right) dx \\
\Rightarrow d\left(\frac{2Ax}{\|x\|^2} - \frac{2\langle Ax,x\rangle x}{\|x\|^4}\right) &= \frac{1}{\|x\|^6}\left(2A\|x\|^4 - (4Axx^T + 4xx^T A + 2\langle Ax, x\rangle I_n)\|x\|^2 + 8\langle Ax, x\rangle xx^T\right) dx
\end{aligned}$$

$$\begin{aligned}
\Rightarrow d^2 f &= \left\langle \frac{1}{\|x\|^6}\left(2A\|x\|^4 - (4Axx^T + 4xx^T A + 2\langle Ax, x\rangle I_n)\|x\|^2 + 8\langle Ax, x\rangle xx^T\right) dx, dx_1 \right\rangle \\
\Rightarrow d^2 f &= \left\langle \frac{1}{\|x\|^6}\left(2A\|x\|^4 - (4Axx^T + 4xx^T A + 2\langle Ax, x\rangle I_n)\|x\|^2 + 8\langle Ax, x\rangle xx^T\right) dx_1, dx \right\rangle \\
\Rightarrow \nabla^2 f &= \frac{1}{\|x\|^6}\left(2A\|x\|^4 - (4Axx^T + 4xx^T A + 2\langle Ax, x\rangle I_n)\|x\|^2 + 8\langle Ax, x\rangle xx^T\right)
\end{aligned}$$

1(c) $f : \mathbb{R}^n \backslash \{0\} \to \mathbb{R},\ f(x) = \langle x, x\rangle^{\langle x,x\rangle}$
Note that

$$dy^y = y^y(\ln y + 1)dy$$

$$\begin{aligned}
df &= d\left(\langle x, x\rangle^{\langle x,x\rangle}\right) \\
&= \langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1)d\langle x, x\rangle \\
&= \langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1)\langle 2x, dx\rangle \\
\Rightarrow df &= \left\langle 2x\langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1), dx\right\rangle \\
\Rightarrow \nabla f &= 2x\langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1) \\
d^2 f = d\left(\left\langle 2x\langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1), dx_1\right\rangle\right) &= \left\langle d\left(2x\langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1)\right), dx_1\right\rangle
\end{aligned}$$

Now, we resolve

$$d\left(2x\langle x, x\rangle^{\langle x,x\rangle}(\ln\langle x, x\rangle + 1)\right)$$

$$
\begin{aligned}
d\left(2x\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1)\right) &= d\left(2x\langle x,x\rangle^{\langle x,x\rangle}\ln\langle x,x\rangle+2x\langle x,x\rangle^{\langle x,x\rangle}\right)\\
&= 2d\left(x\langle x,x\rangle^{\langle x,x\rangle}\ln\langle x,x\rangle\right)+2d\left(x\langle x,x\rangle^{\langle x,x\rangle}\right)\\
&= 2d\left(x\langle x,x\rangle^{\langle x,x\rangle}\right)\ln\langle x,x\rangle+2x\langle x,x\rangle^{\langle x,x\rangle}d\left(\ln\langle x,x\rangle\right)+2d\left(x\langle x,x\rangle^{\langle x,x\rangle}\right)\\
&= (2\ln\langle x,x\rangle+2)d\left(x\langle x,x\rangle^{\langle x,x\rangle}\right)+2x\langle x,x\rangle^{\langle x,x\rangle}\frac{d(\langle x,x\rangle)}{\langle x,x\rangle}\\
&= (2\ln\langle x,x\rangle+2)d\left(x\langle x,x\rangle^{\langle x,x\rangle}\right)+2x\langle x,x\rangle^{\langle x,x\rangle-1}\langle 2x,dx\rangle\\
&= (2\ln\langle x,x\rangle+2)d(x)\left(\langle x,x\rangle^{\langle x,x\rangle}\right)+(2\ln\langle x,x\rangle+2)xd\left(\langle x,x\rangle^{\langle x,x\rangle}\right)+2x\langle x,x\rangle^{\langle x,x\rangle-1}\langle 2x,dx\rangle
\end{aligned}
$$

$$
=(2\ln\langle x,x\rangle+2)\langle x,x\rangle^{\langle x,x\rangle}dx+(2\ln\langle x,x\rangle+2)x\left\langle 2x\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1),dx\right\rangle+2x\langle x,x\rangle^{\langle x,x\rangle-1}\langle 2x,dx\rangle
$$

$$
=(2\ln\langle x,x\rangle+2)\langle x,x\rangle^{\langle x,x\rangle}dx+4(\ln\langle x,x\rangle+1)x\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1)x^Tdx+4x\langle x,x\rangle^{\langle x,x\rangle-1}x^Tdx
$$

Thus,

$$
\begin{aligned}
\Rightarrow\ & d\left(2x\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1)\right)\\
=\ & \left(2(\ln\langle x,x\rangle+1)\langle x,x\rangle^{\langle x,x\rangle}I_n+4\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1)^2xx^T+4\langle x,x\rangle^{\langle x,x\rangle-1}xx^T\right)dx\\
\Rightarrow d^2f =\ & \left\langle\left(2(\ln\langle x,x\rangle+1)\langle x,x\rangle^{\langle x,x\rangle}I_n+4\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1)^2xx^T+4\langle x,x\rangle^{\langle x,x\rangle-1}xx^T\right)dx_1,dx\right\rangle\\
\Rightarrow \nabla^2 f =\ & 2(\ln\langle x,x\rangle+1)\langle x,x\rangle^{\langle x,x\rangle}I_n+4\langle x,x\rangle^{\langle x,x\rangle}(\ln\langle x,x\rangle+1)^2xx^T+4\langle x,x\rangle^{\langle x,x\rangle-1}xx^T
\end{aligned}
$$

1(d) $f:\mathbb{R}^n\to\mathbb{R}$,

$$
f(x)=\log\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)
$$

where $a_i\in\mathbb{R}^n$.

$$
\begin{aligned}
df &= d\left(\log\left(\textstyle\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)\right)\\
&= \frac{d\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\\
&= \frac{\sum_{i=1}^m d(e^{\langle a_i,x\rangle})}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\\
&= \frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}d(\langle a_i,x\rangle)}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\\
&= \frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}\langle a_i,dx\rangle}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\\
&= \frac{\sum_{i=1}^m \langle e^{\langle a_i,x\rangle}a_i,dx\rangle}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\\
\Rightarrow df &= \left\langle\frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}},dx\right\rangle\\
\Rightarrow \nabla f &= \frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\\
d^2f &= d\left(\left\langle\frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}},dx_1\right\rangle\right)\\
&= \left\langle d\left(\frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\right),dx_1\right\rangle
\end{aligned}
$$

Now, we resolve

$$d\left(\frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\right)$$

$$\begin{aligned}
d\left(\frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\right) & \\
&= \frac{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)d\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i\right)-d\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)\sum_{i=1}^m d\left(e^{\langle a_i,x\rangle}a_i\right)-\sum_{i=1}^m d\left(e^{\langle a_i,x\rangle}\right)\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)\sum_{i=1}^m a_i e^{\langle a_i,x\rangle}d\langle a_i,x\rangle-(\sum_{i=1}^m e^{\langle a_i,x\rangle}d\langle a_i,x\rangle)\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)\sum_{i=1}^m a_i e^{\langle a_i,x\rangle}a_i^T dx-(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i^T dx)\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i a_i^T dx-(\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i^T I_n\left(\sum_{j=1}^m e^{\langle a_j,x\rangle}a_j\right)dx)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}\left(\sum_{j=1}^m e^{\langle a_j,x\rangle}\right)a_i a_i^T dx-\sum_{i=1}^m\left(\sum_{j=1}^m e^{\langle a_i,x\rangle}e^{\langle a_j,x\rangle}a_i^T I_n a_j\right)dx}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\sum_{i=1}^m\left(\sum_{j=1}^m e^{\langle a_i,x\rangle}e^{\langle a_j,x\rangle}a_i a_i^T\right)dx-\sum_{i=1}^m\left(\sum_{j=1}^m e^{\langle a_i,x\rangle}e^{\langle a_j,x\rangle}a_i^T I_n a_j\right)dx}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\sum_{i=1}^m\left(\sum_{j=1}^m e^{\langle a_i,x\rangle}e^{\langle a_j,x\rangle}a_i a_i^T-e^{\langle a_i,x\rangle}e^{\langle a_j,x\rangle}a_i^T I_n a_j\right)dx}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
&= \frac{\left(\sum_{i=1}^m\sum_{j=1}^m e^{\langle a_i+a_j,x\rangle}a_i a_i^T-e^{\langle a_i+a_j,x\rangle}a_i^T I_n a_j\right)dx}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}\\
\Rightarrow d\left(\frac{\sum_{i=1}^m e^{\langle a_i,x\rangle}a_i}{\sum_{i=1}^m e^{\langle a_i,x\rangle}}\right) & \\
&= \frac{\left(\sum_{i=1}^m\sum_{j=1}^m e^{\langle a_i+a_j,x\rangle}\left(a_i a_i^T-a_i^T I_n a_j\right)\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}dx
\end{aligned}$$

Plugging in, we have,

$$\begin{aligned}
d^2 f &= \left\langle \frac{\left(\sum_{i=1}^m\sum_{j=1}^m e^{\langle a_i+a_j,x\rangle}\left(a_i a_i^T-a_i^T I_n a_j\right)\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}dx, dx_1\right\rangle\\
\Rightarrow d^2 f &= \left\langle \frac{\sum_{i=1}^m\sum_{j=1}^m e^{\langle a_i+a_j,x\rangle}\left(a_i a_i^T-a_i^T I_n a_j\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}dx_1, dx\right\rangle\\
\nabla^2 f &= \frac{\sum_{i=1}^m\sum_{j=1}^m e^{\langle a_i+a_j,x\rangle}\left(a_i a_i^T-a_i^T I_n a_j\right)}{\left(\sum_{i=1}^m e^{\langle a_i,x\rangle}\right)^2}
\end{aligned}$$

Problem 2 Computing f'

2(a) $f : E \to \mathbb{R}$, $f(t) = \det(A - tI_n), A \in \mathbb{R}^{n\times n}, E := \{t \in \mathbb{R} : \det(A - tI_n) \neq 0\}$ First, setting $X = A - tI_n$, we have $dX = dA - d(tI_n) \Rightarrow dX = -(dt)I_n$. Thus,

$$\begin{aligned} f' = d(\det(A - tI_n)) &= \det(A - tI_n)\mathrm{tr}((A - tI_n)^{-1}d(A - tI_n)) \\ \Rightarrow f' &= -\det(A - tI_n)\mathrm{tr}((A - tI_n)^{-1}d(t)I_n) \end{aligned}$$

2(b) $f : \mathbb{R}_{++} \to \mathbb{R}, f(t) = \|(A + tI_n)^{-1}b\|, A \in \mathbb{S}^n_+, b \in \mathbb{R}^n$
Set $X = (A + tI_n)$ and $x = X^{-1}b$. Now, applying the results from the lecture note.

$$\begin{aligned} f' = d(\|x\|) &= \left\langle \tfrac{x}{\|x\|}, dx \right\rangle \\ &= \left\langle \tfrac{(A+tI_n)^{-1}b}{\|(A+tI_n)^{-1}b\|}, d(X^{-1}b) \right\rangle \\ &= \left\langle \tfrac{(A+tI_n)^{-1}b}{\|(A+tI_n)^{-1}b\|}, d(X^{-1})b \right\rangle \\ &= \left\langle \tfrac{(A+tI_n)^{-1}b}{\|(A+tI_n)^{-1}b\|}, -X^{-1}d(X)X^{-1}b \right\rangle \\ &= \left\langle \tfrac{(A+tI_n)^{-1}b}{\|(A+tI_n)^{-1}b\|}, -(A + tI_n)^{-1}d(A + tI_n)(A + tI_n)^{-1}b \right\rangle \\ &= \left\langle \tfrac{(A+tI_n)^{-1}b}{\|(A+tI_n)^{-1}b\|}, -(A + tI_n)^{-1}d(t)(A + tI_n)^{-1}b \right\rangle \\ &= \left\langle \tfrac{(A+tI_n)^{-1}b}{\|(A+tI_n)^{-1}b\|}, -(A + tI_n)^{-2}bdt \right\rangle \\ &= -\tfrac{1}{\|(A+tI_n)^{-1}b\|}\left(((A + tI_n)^{-1}b)^T(A + tI_n)^{-2}b\right) dt \\ &= -\tfrac{1}{\|(A+tI_n)^{-1}b\|}\left(b^T(A^T + tI_n^T)^{-1}(A + tI_n)^{-2}b\right) dt \\ &= -\tfrac{1}{\|(A+tI_n)^{-1}b\|}\left(b^T(A + tI_n)^{-1}(A + tI_n)^{-2}b\right) dt \\ &= -\tfrac{b^T(A+tI_n)^{-3}b}{\|(A+tI_n)^{-1}b\|}dt \end{aligned}$$

3 Smoothness
3(a) $g : \mathbb{R} \to \mathbb{R}, |g''(x)| \leq L, \forall x \in \mathbb{R}.a \in \mathbb{R}^n, b \in \mathbb{R}, f(x) = g(\langle a, x\rangle + b)$
We first compute $\nabla^2 f$.

$$\begin{aligned} df = d(g(\langle a, x\rangle + b)) &= g'(\langle a, x\rangle + b)d(\langle a, x\rangle + b) \\ &= g'(\langle a, x\rangle + b)\langle a, dx\rangle \\ \Rightarrow df &= \langle ag'(\langle a, x\rangle + b), dx\rangle \\ d^2f = d(\langle ag'(\langle a, x\rangle + b), dx_1\rangle) &= \langle d(ag'(\langle a, x\rangle + b)), dx_1\rangle \\ &= \langle ad(g'(\langle a, x\rangle + b)), dx_1\rangle \\ &= \langle ag''(\langle a, x\rangle + b)d(\langle a, x\rangle + b), dx_1\rangle \\ &= \langle ag''(\langle a, x\rangle + b)\langle a, dx\rangle, dx_1\rangle \\ &= \langle ag''(\langle a, x\rangle + b)a^T dx, dx_1\rangle \\ \Rightarrow \nabla^2 f &= ag''(\langle a, x\rangle + b)a^T \end{aligned}$$

Next we compute $\|\nabla^2 f\|_{\mathrm{op}}$

$$\begin{aligned}
\|\nabla^2 f\|_{\text{op}} = \|ag''(\langle a,x\rangle + b)a^T\|_{\text{op}} &= |g''(\langle a,x\rangle + b)|\|aa^T\|_{\text{op}} \leq L|\|aa^T\|_{\text{op}} \\
\text{Now, by properties of 2-norm, } L|\|aa^T\|_{\text{op}} &= L|\|a\|_2^2 = L|\|a\|^2 \\
\Rightarrow \|\nabla^2 f\|_{\text{op}} &\leq L|\|a\|^2
\end{aligned}$$

We are done by Theorem 2.
3(b) $f : \mathbb{R}^n \to \mathbb{R}, a_i \in \mathbb{R}^n$,

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \ln\left(1 + e^{\langle a_i,x\rangle}\right)$$

$$\begin{aligned}
df &= d\left(\tfrac{1}{m}\textstyle\sum_{i=1}^m \ln\left(1 + e^{\langle a_i,x\rangle}\right)\right) \\
&= \tfrac{1}{m}\textstyle\sum_{i=1}^m d\left(\ln\left(1 + e^{\langle a_i,x\rangle}\right)\right) \\
&= \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{d\left(1+e^{\langle a_i,x\rangle}\right)}{1+e^{\langle a_i,x\rangle}} \\
&= \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{\left\langle e^{\langle a_i,x\rangle}a_i, dx\right\rangle}{1+e^{\langle a_i,x\rangle}} \\
\Rightarrow df &= \left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i}{1+e^{\langle a_i,x\rangle}}, dx\right\rangle \\
\Rightarrow d^2 f &= d\left(\left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i}{1+e^{\langle a_i,x\rangle}}, dx_1\right\rangle\right) \\
&= \left\langle d\left(\tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i}{1+e^{\langle a_i,x\rangle}}\right), dx_1\right\rangle \\
&= \left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m d\left(\frac{e^{\langle a_i,x\rangle}a_i}{1+e^{\langle a_i,x\rangle}}\right), dx_1\right\rangle \\
&= \left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{\left(1+e^{\langle a_i,x\rangle}\right)d\left(e^{\langle a_i,x\rangle}a_i\right) - d\left(1+e^{\langle a_i,x\rangle}\right)\left(e^{\langle a_i,x\rangle}a_i\right)}{\left(1+e^{\langle a_i,x\rangle}\right)^2}, dx_1\right\rangle \\
&= \left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{\left(1+e^{\langle a_i,x\rangle}\right)a_i e^{\langle a_i,x\rangle}a_i^T dx - \left\langle e^{\langle a_i,x\rangle}a_i, dx\right\rangle\left(e^{\langle a_i,x\rangle}a_i\right)}{\left(1+e^{\langle a_i,x\rangle}\right)^2}, dx_1\right\rangle \\
&= \left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i a_i^T dx + e^{2\langle a_i,x\rangle}a_i a_i^T dx - e^{2\langle a_i,x\rangle}a_i a_i^T dx}{\left(1+e^{\langle a_i,x\rangle}\right)^2}, dx_1\right\rangle \\
&= \left\langle \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i a_i^T}{\left(1+e^{\langle a_i,x\rangle}\right)^2} dx, dx_1\right\rangle \\
\Rightarrow \nabla^2 f &= \tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i a_i^T}{\left(1+e^{\langle a_i,x\rangle}\right)^2} \\
\|\nabla^2 f\| = \left\|\tfrac{1}{m}\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i a_i^T}{\left(1+e^{\langle a_i,x\rangle}\right)^2}\right\| &= \tfrac{1}{m}\left\|\textstyle\sum_{i=1}^m \frac{e^{\langle a_i,x\rangle}a_i a_i^T}{\left(1+e^{\langle a_i,x\rangle}\right)^2}\right\| \leq \tfrac{1}{m}\textstyle\sum_{i=1}^m \left\|\frac{e^{\langle a_i,x\rangle}a_i a_i^T}{\left(1+e^{\langle a_i,x\rangle}\right)^2}\right\|
\end{aligned}$$

From,

$$e^{\langle a_i,x\rangle} \geq 1 \Rightarrow \frac{1}{1 + e^{\langle a_i,x\rangle}} \leq \frac{1}{2}$$

we have that,

$$\begin{aligned}
\Rightarrow \|\nabla^2 f\| &\leq \tfrac{1}{m}\textstyle\sum_{i=1}^m \left\| \frac{e^{\langle a_i,x\rangle}a_i a_i^T}{(2)^2} \right\| = \tfrac{1}{4m}\textstyle\sum_{i=1}^m \left\| e^{\langle a_i,x\rangle}a_i a_i^T \right\| = \tfrac{1}{4m}\textstyle\sum_{i=1}^m e^{\langle a_i,x\rangle}\|a_i\|^2 \\
\Rightarrow \|\nabla^2 f\| &\leq \tfrac{1}{4m}\textstyle\sum_{i=1}^m e^{\langle a_i,x\rangle}\|a_i\|^2
\end{aligned}$$

We are done by Theorem 2.

4 Analytical solution of Linear Regression

Given $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n$,

$$\min_{x\in\mathbb{R}^d} \left\{ f(x) = \frac{1}{2n}\|Ax-b\|^2 \right\}$$

is such that

$$A^T A x^* = A^T b$$

We first find the gradient and equate it to zero. Next we show that the problem is convex.

$$\begin{aligned}
df &= d\left(\tfrac{1}{2n}\|Ax-b\|^2\right) \\
&= \tfrac{1}{2n} d(\|Ax-b\|^2) \\
&= \tfrac{1}{2n} d((Ax-b)^T(Ax-b)) = \tfrac{1}{2n} d((x^TA^T - b^T)(Ax-b)) \\
&= \tfrac{1}{2n} d(x^TA^TAx - x^TA^Tb - b^TAx + b^Tb) \\
&= \tfrac{1}{2n} d(x^TA^TAx) - \tfrac{1}{2n} d(x^TA^Tb) - \tfrac{1}{2n} d(b^TAx) + \tfrac{1}{2n} d(b^Tb) \\
&= \tfrac{1}{2n} d(\langle A^TAx, x\rangle) - \tfrac{1}{2n} d(\langle A^Tb, x\rangle) - \tfrac{1}{2n} d((A^Tb)^T x) \\
&= \tfrac{1}{2n}\langle (A^TA + (A^TA)^T)x, dx\rangle - \tfrac{1}{2n}\langle A^Tb, dx\rangle - \tfrac{1}{2n} d(\langle A^Tb, x\rangle) \\
&= \tfrac{1}{2n}\langle 2(A^TA)x, dx\rangle - \tfrac{1}{n}\langle A^Tb, dx\rangle \\
\Rightarrow df &= \tfrac{1}{n}\langle A^TAx - A^Tb, dx\rangle \\
\Rightarrow \nabla f &= \tfrac{1}{n}A^TAx - A^Tb \\
\nabla f(x^*) &= 0 \Rightarrow A^TAx^* = A^Tb \\
\Rightarrow d^2 f &= \tfrac{1}{n}\langle df(A^TAx - A^Tb), dx_1\rangle \\
&= \tfrac{1}{n}\langle A^TA dx, dx_1\rangle \\
\Rightarrow \nabla^2 f &= \tfrac{1}{n}A^TA
\end{aligned}$$

Let $y \in \mathbb{R}^n$ be an arbitrary vector.

$$\begin{aligned}
0 \leq \|Ay\|^2 = \langle Ay, Ay\rangle = (Ay)^T(Ay) &= y^TA^TAy = y^T(A^TA)y \\
\Rightarrow 0 \leq y^T(A^TA)y &\Rightarrow A^TA \succeq 0 \Rightarrow \nabla^2 f \succeq 0
\end{aligned}$$

Convexity shown. Hence, we conclude that

$$A^TAx^* = A^Tb$$

5 Oracle complexity

Given $L \geq 1, R \geq 1, 0 \leq \mu \leq 1$ and $0 < \epsilon \leq 1$, try to find the minimum function amongst $\sqrt{\frac{LR^2}{\epsilon}}, \frac{LR^2}{\epsilon}, \frac{L}{\mu}\log\frac{R^2}{\epsilon}$ and $\sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon}$ for different parameters.

$R \geq 1 \Rightarrow R^2 \geq 1$ and $L \geq 1 \Rightarrow LR^2 \geq 1$. Also, $\epsilon \leq 1 \Rightarrow \frac{1}{\epsilon} \geq 1$. Combining these gives $\frac{LR^2}{\epsilon} \geq 1 \Rightarrow \sqrt{\frac{LR^2}{\epsilon}} \geq 1$ since square-root is an increasing function. Now, multiplying both sides by left handside term gives

$$\frac{LR^2}{\epsilon} \geq \sqrt{\frac{LR^2}{\epsilon}}$$

Using similar arguments, we can derive that $L \geq 1$ and $\mu \leq 1 \Rightarrow \frac{1}{\mu} \geq 1 \Rightarrow \sqrt{\frac{L}{\mu}} \geq 1 \Rightarrow \frac{L}{\mu} \geq \sqrt{\frac{L}{\mu}}$. Hence,

$$\frac{L}{\mu}\log\frac{R^2}{\epsilon} \geq \sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon}$$

Now, we compare $\sqrt{\frac{LR^2}{\epsilon}}$ and $\sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon}$

$$\sqrt{\frac{LR^2}{\epsilon}} > \sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon} \Leftrightarrow \mu > \frac{\epsilon}{R^2}\left(\log\frac{R^2}{\epsilon}\right)^2$$

and

$$\sqrt{\frac{LR^2}{\epsilon}} \leq \sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon} \Leftrightarrow \mu \leq \frac{\epsilon}{R^2}\left(\log\frac{R^2}{\epsilon}\right)^2$$

So, if $\mu > \frac{\epsilon}{R^2}\left(\log\frac{R^2}{\epsilon}\right)^2$, we choose $\sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon}$ else we choose $\sqrt{\frac{LR^2}{\epsilon}}$.

We can conclude here.

For comparison that may not involve R

Let

$$f(R) = R - 2\sqrt{\frac{\epsilon}{\mu}}\log R + \sqrt{\frac{\epsilon}{\mu}}\log\epsilon$$

We consider this function because

$$\sqrt{\frac{LR^2}{\epsilon}} - \sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon} = \sqrt{\frac{L}{\epsilon}}f(R)$$

This gives,

$$f'(R) = 1 - \frac{2}{R}\sqrt{\frac{\epsilon}{\mu}} \Rightarrow f'(R^*) = 0 \Leftrightarrow R^* = 2\sqrt{\frac{\epsilon}{\mu}}$$

For second derivative,

$$f''(R) = \frac{2}{R^2}\sqrt{\frac{\epsilon}{\mu}} \geq 0$$

Hence, we have that $f(R)$ is convex. Hence, R^* gives the global minimum.

Now, $f(R^*) = R^* - 2\sqrt{\frac{\epsilon}{\mu}}\log R^* + \sqrt{\frac{\epsilon}{\mu}}\log\epsilon = 2\sqrt{\frac{\epsilon}{\mu}} - 2\sqrt{\frac{\epsilon}{\mu}}\log\left(2\sqrt{\frac{\epsilon}{\mu}}\right) + \sqrt{\frac{\epsilon}{\mu}}\log\epsilon = 2\sqrt{\frac{\epsilon}{\mu}} - \sqrt{\frac{\epsilon}{\mu}}\log\frac{4}{\mu}$.

Thus,

$$f_{\min}(R) = 2\left(\sqrt{\frac{\epsilon}{\mu}} - \log\frac{4}{\mu}\right)$$

.

$$f_{\min}(R) \geq 0 \Leftrightarrow \sqrt{\tfrac{\epsilon}{\mu}} \geq \log\tfrac{4}{\mu}$$

So, if we have $\sqrt{\frac{\epsilon}{\mu}} \geq \log\frac{4}{\mu}$, we could immediately choose $\sqrt{\frac{L}{\mu}}\log\frac{R^2}{\epsilon}$ else, we revert to previous method.