CIS4930 - Math for Machine Learning

Homework 3

Fernando Sckaff

February 21^{th}, 2023

Question 1

Use the Lagrange multiplier (parameter) approach to solve the following problems:

1. $f(x) = \|x\|_2^2, h(x) = \sum_{i=1}^{n} x_i - 1$.
2. $f(x) = \sum_{i=1}^{n} x_i, h(x) = \|x\|_2^2 - 1$.
3. $f(x) = \|x\|_2^2, h(x) = x^T Q x - 1$, where Q is positive definite (symmetric with all eigenvalues greater than zero).

Solution:

1. Solve for:

$$\begin{aligned}
L(x, \lambda) &= \|x\|_2^2 + \lambda(\sum_{i=1}^{n} x_i - 1) \\
&= \|x\|_2^2 + \lambda(1^T x - 1) \\
&= x^T x + \lambda 1^T x - \lambda
\end{aligned}$$

Apply theorem, $\nabla_x L = O$

$$\begin{aligned}
\nabla_x L &= \nabla_x (x^T x + \lambda 1^T x - \lambda) = O \\
O &= 2x + \lambda 1^T I \\
&= 2x + \lambda
\end{aligned}$$

Therefore, $x = -\lambda/2$. Going back to the constraint:

$$\begin{aligned}
h(x) &= \sum_{i=1}^{n} x_i - 1 \\
0 &= 1^T x - 1 \\
0 &= -\lambda/2 - 1
\end{aligned}$$

Therefore,

$$\lambda = 2 \quad 😄$$

2. Solve for:

$$\begin{aligned}
L(x, \lambda) &= \sum_{i=1}^{n} x_i + \lambda(\|x\|_2^2 - 1) \\
&= 1^T x + \lambda(x^T x - 1)
\end{aligned}$$

Apply theorem, $\nabla_x L = O$

$$\begin{aligned}
\nabla_x L &= \nabla_x (1^T x + \lambda(x^T x - 1)) = O \\
O &= 2x\lambda
\end{aligned}$$

Therefore,

$$\lambda = 0 \quad 😄$$

3. Solve for:

$$\begin{aligned} L(x,\lambda) &= \|x\|_2^2 + \lambda(x^TQx - 1) \\ &= x^Tx + \lambda x^TQx - \lambda \end{aligned}$$

Apply theorem, $\nabla_x L = O$

$$\begin{aligned} \nabla_x L = \nabla_x(x^Tx + \lambda x^TQx - \lambda) = O \\ O = 2x + \lambda 2Qx \end{aligned}$$

Isolate for λ:

$$\begin{aligned} 0 &= 2x + \lambda 2Qx \\ -2x &= 2x * \lambda Q \\ -1 &= \lambda Q \\ -Q^{-1} &= \lambda \end{aligned}$$

Therefore,

$$\lambda = -Q^{-1} \quad 😄$$

Question 2

Arithmetic-Geometric Mean Inequality:
Let $\alpha_1, \ldots, \alpha_n$ be positive scalars with $\sum_{i=1}^{n} \alpha_i = 1$. Use a Lagrange multiplier to solve the problem

$$\min_x \sum_{i=1}^{n} \alpha_i x_i$$

subject to

$$\prod_{i=1}^{n} x_i^{\alpha_i} = 1, x_i > 0 \forall i.$$

Establish the arithmetic-geometric mean inequality

$$\prod_{i=1}^{n} x_i^{\alpha_i} \leqslant \sum_{i=1}^{n} \alpha_i x_i$$

for a set of positive numbers $x_i . \forall i \in \{1, \ldots, n\}$. Hint: Use a change of variables $y_i = \log(x_i)$ where log is ln.

Solution:

$$\begin{aligned}
\prod_{i=1}^{n} x_i^{\alpha_i} &= 1 \\
x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_n^{\alpha_n} &= 1 \\
\alpha_1 ln(x_1) + \alpha_2 ln(x_2) + \ldots + \alpha_n ln(x_n) &= 0 && \text{Taking the natural logarithm on both sides} \\
\sum_{i=1}^{n} \alpha_i ln(x_i) &= 0 \\
\sum_{i=1}^{n} \alpha_i y_i &= 0
\end{aligned}$$

Therefore,

$$\begin{aligned}
L(y, \lambda) &= \sum_{i=1}^{n} \alpha_i x_i + \lambda \sum_{i=1}^{n} \alpha_i y_i \\
0 &= \sum_{i=1}^{n} \alpha_i e^{y_i} + \lambda \sum_{i=1}^{n} \alpha_i y_i \\
0 &= \alpha_i e^{y_i} + \lambda \alpha_i && \text{Derivative on } y_i \\
0 &= e^{y_i} + \lambda \\
e^{y_i} &= -\lambda \\
y_i &= ln| - \lambda| \\
y_i &= ln(\lambda)
\end{aligned}$$

Additionally,

$$\begin{aligned}
\sum_{i=1}^{n} \alpha_i y_i &= 0 \\
\sum_{i=1}^{n} \alpha_i ln(\lambda) &= 0 \\
ln(\lambda) &= 0 && \text{for all } i.
\end{aligned}$$

In case $y_1 = y_2 = ... = y_n = 0$,

$$\alpha_i e^{y_i} + \alpha_i e^{y_i} + ... + \alpha_i e^{y_i} = 1$$

This can only mean,

$$\begin{aligned}
\alpha_i e^{y_i} + \alpha_i e^{y_i} + ... + \alpha_i e^{y_i} &\geqslant 1 \\
\alpha_i e^{y_i} + \alpha_i e^{y_i} + ... + \alpha_i e^{y_i} &\geqslant x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot ... \cdot x_n^{\alpha_n} \\
\sum_{i=1}^{n} \alpha_i x_i &\geqslant \prod_{i=1}^{n} x_i^{\alpha_i} \quad 😄
\end{aligned}$$

Question 3

Consider the optimization problem

$$\min_{x,y} f(x,y) = \frac{1}{2}\left((x-x_0)^2 + (y-y_0)^2\right)$$

with (x_0, y_0) fixed and subject to

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

- Write out the Lagrangian for this problem.
- Solve for x, y in terms of the Lagrange parameter.
- Rewrite the ellipse constraint in terms of the Lagrange parameter.
- Show that this leads to a quartic. Can a general quartic be solved using radicals?

Solution:
Lagrangian:

$$L(x,y,\lambda) = \frac{1}{2}(x^2 - 2xx_o + x_o^2) + \frac{1}{2}(y^2 - 2yy_o + y_o^2) + \lambda(\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1)$$

$$\nabla_x L = x - x_o + \lambda\frac{2x}{a^2} = 0 \rightarrow x = \frac{a^2 x_o}{a^2 + 2\lambda} \rightarrow \lambda = \frac{a^2(x_o - x)}{2x} \tag{1}$$

$$\nabla_y L = y - y_o + \lambda\frac{2y}{b^2} = 0 \rightarrow y = \frac{b^2 y_o}{b^2 + 2\lambda} \rightarrow \lambda = \frac{b^2(y_o - y)}{2y} \tag{2}$$

Rewriting the ellipse function:

$$\begin{aligned}
\frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1 \\
\frac{(\frac{a^2 x_o}{a^2+2\lambda})^2}{a^2} + \frac{(\frac{b^2 y_o}{b^2+2\lambda})^2}{b^2} &= 1 \\
\frac{a^2 x_o^2}{(a^2+2\lambda)^2} + \frac{b^2 y_o^2}{(b^2+2\lambda)^2} &= 1 \\
a^2 x_o^2 (b^2+2\lambda)^2 + b^2 y_o^2 (a^2+2\lambda)^2 &= (a^2+2\lambda)^2(b^2+2\lambda)^2 \\
(a^2+2\lambda)^2(b^2+2\lambda)^2 - a^2 x_o^2(b^2+2\lambda)^2 - b^2 y_o^2 (a^2+2\lambda)^2 &= 0
\end{aligned}$$

Conclusion: $(a^2 + 2\lambda)^2(b^2 + 2\lambda)^2$ will lead to a λ^4. A quartic. They can be solved using radicals, but it is extremely difficult to solve by hand. 😃

Question 4

You are given the least squares problem

$$E(\boldsymbol{w}) = \left\| y - \sum_{k=1}^{K} w_k x_k \right\|_2^2$$

where $\boldsymbol{w} \in \mathbb{R}^K$ and $\boldsymbol{x}_i, \boldsymbol{y} \in \mathbb{R}^D$.

- Write the Lagrangian for this problem when supplied with the constraint $\sum_{k=1}^{K} w_k = 1$.
- Solve for the Lagrange parameter explaining all steps.
- Solve for $\boldsymbol{w}$ while satisfying the constraint.
- Describe the criteria for solutions to exist for this problem.

Solution:

Lagrangian:

$$\begin{aligned}
L(w, \lambda) &= \left\| y - \sum_{k=1}^{K} w_k x_k \right\|_2^2 + \lambda\left(\sum_{k=1}^{K} w_k - 1\right) \\
&= \|y - wx\|_2^2 + \lambda(1^T w - 1) && \text{- Acknowledging dot product and sum of elements of w} \\
&= (y - wx)^T (y - wx) + \lambda(1^T w - 1) && \text{- } \|y - wx\|_2^2 \text{ is the euclidean norm of the residual. } \|r\|_2^2 = r^T r \\
&= y^T y - 2x^T w^T y - x^T w^T wx + \lambda 1^T w - \lambda && \text{- Expand} \\
\nabla_w L &= -2x^T y - 2x^T wx + \lambda = 0 && \text{- Take derivative w.r.t. } w \\
0 &= -2x^T y - 2x^T wx + \lambda
\end{aligned}$$

From that, you can derive:

$$w = \frac{\lambda - 2x^T y}{2x^T x} \tag{3}$$

$$\lambda = 2x^T (y - xw) \tag{4}$$

· Observe on (4), $y - xw$ is the residual.
· In order for solutions to exist, K must be equivalent to D.

Question 5

Write a program to roughly carry out the optimization problem inherent in $\|A\|_2$, namely

$$\max_{x \neq O_n} \frac{\|Ax\|_2}{\|x\|_2}$$

. The program should take an arbitrary $m \times n$ matrix A and sample over 100000 random vectors $x \in \mathbb{R}^n$ (and use Gaussian random numbers) to approximately pick the maximum. Compute and empirically show that the analytic solution is an upper bound for the search. For demonstration purposes, use $m = 9$ and $n = 6$. The matrix entries can also be randomly chosen. Your program should contrast the empirical maximum with the analytically derived maximum to the optimization problem.

Solution: on Google Colab, Link:

$$https://colab.research.google.com/drive/1aJ9PQJdT2-6eDrASEOfCzj7Bq_o oZXaA \quad ☺$$

Question 6: BONUS

Consider the L_2 norm of a matrix defined as $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$. Is $\|A\|_2 \geqslant \|A\|_F$ or is $\|A\|_2 \leqslant \|A\|_F$ or is it neither where $\|A\|_F$ is the Frobenius norm of the matrix? Give a clear mathematical explanation and justification for your choice (out of the three possible choices).

Solution:
We have $\|A\|_F = \sqrt{\sum_{i,j} |a_{i,j}|^2}$, where $a_{i,j}$ is the entry in the ith row and jth column of A.
Let x be a unit vector such that $\|Ax\|_2 = \|A\|_2$. Then we have:

$$\begin{aligned}
\|A\|_2 &= \|Ax\|_2 \\
&= \sqrt{\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{i,j} x_j \right)^2} && \text{Definition of the 2-norm of a vector} \\
&\leqslant \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^2 \sum_{j=1}^{n} x_j^2} && \text{by Cauchy-Schwarz inequality} \\
&= \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^2} && \text{since } \|x\|_2 = 1.
\end{aligned}$$

Therefore, we have $\|A\|_2 \leqslant \|A\|_F$. ☺

F.S.

Note:-

- Question 1
  - I. 2/5 Incomplete
  - II. 2/5 Incomplete
  - III. 5/10 Incomplete
- Question 2: 17/20 Not a general solution for the AM, GM Inequality
- Question 3: Correct
- Question 4: 10/20
- Question 5: Correct, but the proof is very clever and even though it needs a few more things to firm it up, the basic idea is great.