MA576 HW1

Aidan Fischer

September 2023

I pledge my honor that I have abided by the Stevens Honor System

Q1

a)

We want to show that $x^T LL^T x \geq 0$

$A = LL^T$ is an $n \times n$ matrix with the $(j, k)^{th}$ element being a dot product of the j^{th} and k^{th} rows of L

The elements of $x^T A$ are

$$(x^T A)_n = \sum_{k=1}^{n} x_k * A_{k,n} = \sum_{k=1}^{n} x_k * L_{n,:} \cdot L_{k,:}$$

$$x^T Ax = \sum_{i=1}^{n} \sum_{k=1}^{n} x_i * x_k * L_{i,:} \cdot L_{k,:}$$

$$= \sum_{i=1}^{n} x_i * x_i * L_{i,:} \cdot L_{i,:} + \sum_{i=1}^{n} \sum_{k=1,k\neq i}^{n} x_i * x_k * L_{i,:} \cdot L_{k,:}$$

In the second sum, we can consider only the lower diagonal elements twice because the summation body is equal for (i, k) and (k, i)

$$= \sum_{i=1}^{n} x_i * x_i * L_{i,:} \cdot L_{i,:} + 2 \sum_{i=1}^{n} \sum_{k=1}^{i-1} x_i * x_k * L_{i,:} \cdot L_{k,:}$$

This seems like a familiar form.

$$||L^T x||^2 = \sum_{i=1}^{n} (L^T x)_i^2$$

where

$$(L^T x)_i = \sum_{j=1}^{n} x_j L_{j,i}$$

So

$$||L^T x||^2 = \sum_{i=1}^{n} (\sum_{j=1}^{n} x_j L_{j,i})^2$$

Consider the elements of this summation. The expansion of the square term will have 1 $x_j^2 L_{j,i}^2$ term per j, and 2 $x_j x_k L_{j,i} L_{k,i}$ terms per j.

Considering all i's here, we are left with 1 $x_j^2 L_{j,i}^2$ term per j per i, and 2 $x_j x_k L_{j,i} L_{k,i}$ terms per (j,k) $j \neq k$ per i. This representation doesn't group terms, but we can group by rows (dot product). In this case, we get exactly the prior sum for $x^T A x$, since both summations have the same terms.

Therefore, $x^T A x = ||L^T x||^2$. Since this is a norm, this is always ≥ 0. Therefore, LL^T is psd.

b)

Taking the conclusion from part a, we can show that L being full row rank means that A is pd.

$L^T x$ is a linear combination of the columns of L^T, which are the rows of L.

By definition, L having full row rank means all of its rows are linearly independent. If L does not have full row rank, then at least 2 of its rows are linearly dependent.

This means that if L is not full row rank, you can construct an $x \neq 0$ that will make $L^T x = 0$ by combining the linearly dependent rows and therefore $||L^T x||^2 = 0$, meaning A is not pd.

If L is full row rank, then by definition, no linear combination of its rows can result in the zero vector. Because a norm is always greater than 0 with a vector that is not the 0 vector, this means for L with full row rank, $||L^T x|| > 0$ and therefore A is pd.

If A is pd, that means $||L^T x|| > 0$ for all x. That means no linear combination of L's rows results in the 0 vector. By definition, that means L's rows are linearly independent, and therefore L is full row rank.

Therefore, A is pd iff L is full row rank.

Q2

a)

If $Q \in \mathbb{R}^{n \times n}$ is pd, then $x^T Q x > 0$ for all x.

$$(x^T Q)_i = \sum_{j=1}^{n} x_j Q_{i,j}$$

$$x^T Q x = \sum_{j=1}^{n} (x^T Q)_j x_j = \sum_{j=1}^{n} \sum_{k=1}^{n} x_j x_k Q_{j,k}$$

$$\sum_{j=1}^{n} \sum_{k=1}^{n} x_j x_k Q_{j,k} = \sum_{i=1}^{n} x_i^2 Q_{i,i} + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} x_i x_j Q_{i,j}$$

The diagonal elements of Q must be positive because we can construct an x such that $x^T Q x \leq 0$ in the other cases.

Say there is a Q such that some diagonal element $Q_{i,i} \leq 0$. Choose x such that $x_j = 0$ for all $j \neq i$. Then set $x_i \neq 0$. The first sum will be ≤ 0 because most terms will be 0 and $x_i^2 Q_{i,i} \leq 0$. The second sum will be 0 because either x_i or x_j will be 0 since only one element of x is nonzero. Therefore, $x^T Q x \leq 0$ and Q is not pd.

b)

If there are positive and negative elements in the diagonal of Q, then we can construct an x_n such that $x_n^T Q x_n < 0$ and x_p such that $x_p^T Q x_p > 0$

Say there is a Q such that $Q_{i,i} > 0$ and $Q_{j,j} < 0$.

Construct x_n such that all elements are 0 except for the j^{th} element. By similar logic to Q2a, $x_n^T Q x_n < 0$.

Construct x_p such that all elements are 0 except for the i^{th} element. By similar logic to Q2a, $x_p^T Q x_p > 0$.

Therefore, Q is indefinite because there exists x such that $x^T Q x > 0$ and there exists x such that $x^T Q x < 0$.

Q3

```
import numpy as np
import math
from tqdm import trange
```

```python
# Yields a randomly generated positive definite matrix
# by generating a randomly sized symmetric matrix with
# positive diagonal elements
def pd_generator(seed=None, minsize=2, maxsize=30):
    r = np.random.default_rng(seed)
    while True:
        # Choose a random size between minsize and maxsize
        s = math.floor((maxsize - minsize) * r.random() + minsize)
        # Generate a fully random array
        res = r.standard_normal(size=(s, s))
        # Convert to symmetric matrix
        res = (res + res.T)/2
        # Ensure diagonal dominance
        # Reference: https://en.wikipedia.org/wiki/Diagonally_dominant_matrix
        for i in range(s):
            res[i, i] = np.sum(np.abs(res[i, np.arange(s) != i])) + r.random()
        yield res


# Sylvester's Criterion
def sylvesters(A):
    s = A.shape[0]
    for i in range(1, s+1, 1):
        if np.linalg.det(A[:i, :i]) <= 0:
            return False
    return True


# Run 1 million tests
def main():
    g = pd_generator()
    for _ in trange(1000000):
        A = next(g)
        assert sylvesters(A)


if __name__ == "__main__":
    main()
```

Q4

We have

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$a + b = c + d$

To get the eigenvalues, we use

$$\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}$$

and set the determinant to 0.

We obtain the equation $ad - \lambda d - \lambda a + \lambda^2 - bc = 0$

$=> \quad \lambda^2 - (a+d)\lambda = bc - ad$

$=> \quad \lambda^2 - (a+d)\lambda + \frac{a^2+2ad+d^2}{4} = bc - ad + \frac{a^2+2ad+d^2}{4}$

$=> \quad (\lambda - \frac{a+d}{2})^2 = bc + \frac{a^2-2ad+d^2}{4}$

$=> \quad (\lambda - \frac{a+d}{2})^2 = bc + (\frac{a-d}{2})^2$

Manipulating the 2nd given equation, we can obtain $a - d = c - b$

$=> \quad (\lambda - \frac{a+d}{2})^2 = bc + (\frac{c-b}{2})^2$

$=> \quad (\lambda - \frac{a+d}{2})^2 = bc + \frac{c^2-2bc+b^2}{4}$

$=> \quad (\lambda - \frac{a+d}{2})^2 = \frac{c^2+2bc+b^2}{4}$

$=> \quad (\lambda - \frac{a+d}{2})^2 = (\frac{b+c}{2})^2$

$=> \quad \lambda - \frac{a+d}{2} = \pm\frac{b+c}{2}$

$=> \quad 2\lambda - (a+d) = \pm(b+c)$

$=> \quad \lambda = \frac{(a+d)\pm(b+c)}{2}$

$a + b = c + d => a = c + d - b => a - c = d - b$

$\lambda_1 = \frac{(a+d)+(b+c)}{2} = \frac{a+b+c+d}{2} = \frac{a+b+a+b}{2} = a + b$

and

$\lambda_2 = \frac{(a+d)-(b+c)}{2} = \frac{a+d-b-c}{2} = \frac{a+a-c-c}{2} = a - c$

In which case we have

$$\lambda_1 -> \begin{pmatrix} a-(a+b) & b \\ c & d-(a+b) \end{pmatrix} = \begin{pmatrix} -b & b \\ c & d-b-a \end{pmatrix}$$

$$= \begin{pmatrix} -b & b \\ c & a-c-a \end{pmatrix} = \begin{pmatrix} -b & b \\ c & -c \end{pmatrix}$$

and

$$\lambda_2 -> \begin{pmatrix} a-(a-c) & b \\ c & d-(a-c) \end{pmatrix} = \begin{pmatrix} c & b \\ c & d-a+c \end{pmatrix}$$

$$= \begin{pmatrix} c & b \\ c & a+b-a \end{pmatrix} = \begin{pmatrix} c & b \\ c & b \end{pmatrix}$$

The eigenvectors are obained through

$\begin{pmatrix} -b & b \\ c & -c \end{pmatrix} * v_1 = 0$

$\begin{pmatrix} c & b \\ c & b \end{pmatrix} * v_2 = 0$

The first gives $v_1 = k_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

The second gives $v_2 = k_2 \begin{pmatrix} b \\ -c \end{pmatrix}$

Therefore $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of A

Q5

a)

$f(x) = e^{||x||_2^2}$

We know that $||x||_2^2 = \sum_j x_j^2$

The gradient of $\sum_j x_j^2$ is $2x$ because the partial derivative of the sum with respect to each x_j is $2x_j$

Using the chain rule, then, the gradient of $f(x)$ is $2e^{||x||_2^2}x$

$\nabla f(x)_i = 2e^{||x||_2^2}x_i$

$\nabla^2 f(x)_{ij} = \frac{\partial}{\partial x_j}(2e^{||x||_2^2}x_i)$

For the non diagonal entries, the x_i is a constant and we can use the chain rule again to get

$\nabla^2 f(x)_{ij} = 4e^{||x||_2^2}x_i x_j$

For the diagonal entries, since we are taking the derivative with respect to x_i again, we need to use the product rule.

$\nabla^2 f(x)_{ii} = 2e^{||x||_2^2} + 4e^{||x||_2^2}x_i^2$

Therefore,

$\nabla^2 f(x) = 2e^{||x||_2^2}(2xx^T + I)$

b)

The gradient elements are

$\nabla g(x)_i = \prod_{j=1,j\neq i}^{n} x_j$

Because x_i is a constant when taking a derivative with respect to x_j when $j \neq i$

Because of this, the diagonal entries of the Hessian are 0 and the nondiagonal entries $\nabla^2 g(x)_{ij}$ are $\prod_{k=1,k\neq i,j}^{n} x_k$

Q6

$$\lim_{(x,y)\to(0,0)} \frac{xy - sin(x)sin(y)}{x^2 + y^2}$$

The gradient of the numerator is $[y - cos(x)sin(y), x - sin(x)cos(y)]^T$

The Hessian of the numerator is $\begin{pmatrix} sin(x)sin(y) & 1 - cos(x)cos(y) \\ 1 - cos(x)cos(y) & sin(x)sin(y) \end{pmatrix}$

The gradient of the denominator is $[2x, 2y]^T$

The Hessian of the denominator is $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

The Taylor expansion of a function $f(x_0+h)$ is $f(x_0)+\nabla f(x_0)^T h+\frac{1}{2}h^T\nabla^2 f(x_0)h+r(h)$

For the numerator, $f((0,0)) = 0$, and the gradient and Hessian are also 0s. Therefore we are left with just $r_f(h)$ in the numerator.

For the denominator, $f(x_0)$ and the gradient are 0, but the Hessian isn't. The denominator is $h_x^2 + h_y^2 + r_g(h)$

Leaving us with

$$\lim_{h->(0,0)} \frac{r_f(h)}{h_x^2 + h_y^2 + r_g(h)}$$

Notice that $h_x^2 + h_y^2 = ||h||^2$

Dividing the numerator and denominator by $||h||^2$, we get

$$\lim_{h->(0,0)} \frac{\frac{r_f(h)}{||h||^2}}{1 + \frac{r_g(h)}{||h||^2}}$$

By Taylor's theorem, because $x_0 = (0,0)$, the two $\frac{r(h)}{||h||^2}$ terms tend towards 0 as h approaches $(0,0)$. Therefore, the limit is 0/1=0.

Q7

For each problem, I am plugging in the gradients and Hessians into Taylor's approximation formula.

a)

$g(x,y,z) = e^{xe^y} + e^{ze^y} + y^2(1+x+z)$

$$\nabla g(x,y,z) = \begin{pmatrix} e^{xe^y}e^y + y^2 \\ e^{xe^y}xe^y + e^{ze^y}ze^y + 2y(1+x+z) \\ e^{ze^y}e^y + y^2 \end{pmatrix}$$

$$\nabla^2 g(x,y,z) = \begin{pmatrix} e^{2y}e^{xe^y} & e^{xe^y}xe^{2y} + e^{xe^y}e^y + 2y & 0 \\ e^{xe^y}e^y + xe^{xe^y}e^{2y} + 2y & x^2e^{xe^y}e^{2y} + xe^{xe^y}e^y & e^{ze^y}e^y + ze^{ze^y}e^{2y} + 2y \\ & +z^2e^{ze^t}e^{2y} + ze^{ze^y}e^y + 2(x+z+1) & \\ 0 & e^{ze^y}ze^{2y} + e^{ze^y}e^y + 2y & e^{ze^y}e^{2y} \end{pmatrix}$$

$p_0 = (0,1,0)$

$g(p_0) = 1+1+1 = 3$

$$\nabla g(p_0) = \begin{pmatrix} e+1 \\ 2 \\ e+1 \end{pmatrix}$$

$$\nabla^2 g(p_0) = \begin{pmatrix} e^2 & e+2 & 0 \\ e+2 & 2 & e+2 \\ 0 & e+2 & e^2 \end{pmatrix}$$

$g(x,y,z) \approx g(p_0) + \nabla g(p_0)^T((x,y,z)-p_0) + \frac{1}{2}((x,y,z)-p_0)^T\nabla^2 g(p_0)((x,y,z)-p_0)$

$$= 3 + \begin{pmatrix} e+1 & 2 & e+1 \end{pmatrix} \begin{pmatrix} x \\ y-1 \\ z \end{pmatrix} + \frac{1}{2}\begin{pmatrix} x & y-1 & z \end{pmatrix} \begin{pmatrix} e^2 & e+2 & 0 \\ e+2 & 2 & e+2 \\ 0 & e+2 & e^2 \end{pmatrix} \begin{pmatrix} x \\ y-1 \\ z \end{pmatrix}$$

$g(x,y,z) \approx 3 + xe + x + 2y - 2 + ze + z + 2y^2 + 2exy + 4xy + 2eyz + 4yz - 4y + e^2x^2 - 2ex - 4x + e^2z^2 - 2ez - 4z + 2$

b)

$h(x) = \langle x, Ax\rangle + e^{\langle c,x\rangle}$

Derivative of $\langle x, Ax\rangle = 2Ax$

$\nabla h(x) = 2Ax + ce^{\langle c,x\rangle}$

$\frac{\partial x^T Ax}{\partial x_i x_j} = 2A_{i,j}$

$\nabla c_i e^{\langle c,x\rangle} = c_i * c * e^{\langle c,x\rangle}$

$\nabla^2 h(x) = 2A + cc^T e^{\langle c,x\rangle}$

$x_0 = (1,1,1,...,1)^T$

$h(x) \approx h(x_0) + \nabla h(x_0)^T(x-x_0) + \frac{1}{2}(x-x_0)^T\nabla^2 h(x_0)(x-x_0)$

$$h(x) \approx \sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j} + e^{\sum_{i=1}^{n} c_i} + (2x^T A^T + c^T e^{\sum_{i=1}^{n} c_i})(x - x_0)$$

$$+\frac{1}{2}(x - x_0)^T (2A^T + cc^T e^{\sum_{i=1}^{n} c_i})(x - x_0)$$

Q8

a)

$\mathcal{A}$ is open because despite being a union of closed sets, for any member at the boundary of those closed sets, I can choose $B(c, r) \subseteq \mathcal{A}$ with c a boundary point of that single closed set with $r > 0$ by considering the next set in the given sequence. This is because $\frac{1}{n+1} < \frac{1}{n}$ and $\frac{n+1}{n+2} > \frac{n}{n+1}$.

$\mathcal{B}$ is open because given any point $x \in \mathcal{B}$, there is a $B(x, r) \subseteq \mathcal{B}$ with $r > 0$ because no matter how close the maximum of the absolute values of x_1, x_2, and x_3 are to 1, they cannot be equal to one and therefore we can choose $r = 1 - max(|x_1|, |x_2|, |x_3|) > 0$.

$\mathcal{C}$ is neither open nor closed. It is not open because it contains boundary points where no $B(c, r) \subseteq \mathcal{C}$ exists with $r > 0$. $x = [0, 1]^T$ is an example of this. No radius will work because x_1 cannot be less than 0. It is not closed because the point $y = [0, 0]^T \not\subseteq \mathcal{C}$, yet it is a limiting sequence of points in the set with $y_1 = 0$ and $y_2 > 0$. '

$\mathcal{D}$ is closed because it contains its limiting points.

b)

As open sets, the interiors of $\mathcal{A}$ and $\mathcal{B}$ are themselves.

The interior of $\mathcal{C}$ replaces the $\leq$ in its definition with a $<$

The interior of $\mathcal{D}$ replaces the $\geq$ with a $>$

The closure of $\mathcal{A}$ contains 0 and 1 (since those numbers are the limiting numbers of the set boundaries of the closed sets). Therefore the closure of $\mathcal{A}$ is $[0, 1]$

The closure of $\mathcal{B}$ replaces the $<$ in its definition with a $\leq$

The closure of $\mathcal{C}$ replaces the $<$ in its definition with a $\leq$

As a closed set, the closure of $\mathcal{D}$ is itself.

c)

$\mathcal{A}$ is bounded by $B(0, 1)$

$\mathcal{B}$ is bounded by $B([0,0,0]^T, \sqrt{3})$ This is because each of the elements of x can be up to 1. The magnitude of [1,1,1] is $\sqrt{3}$

$\mathcal{C}$ is not bounded because x_1 and x_2 are unrestricted except by themselves. They can tend towards infinity.

$\mathcal{D}$ is bounded by $B([0,0,0]^T, 1)$ because the sum of squares of x_1, x_2, and x_3 cannot be greater than 1 if each of them is at most 1, because between 0 and 1, the square function decreases the number, and if one of them is equal to 1, the restraints mean the other 2 must be 0.

References

matrixcookbook.pdf fundamentals.pdf