Inversion Homework #1

Professor H. Yao & H. Zhang

Jintao Li

SA20007037

E-mail: lijintao@mail.ustc.edu.cn

October 11, 2020

Appendix A

Exercise 4

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 1 & 3 \\ 4 & 6 & 7 & 11 \end{bmatrix} \tag{1}$$

Find bases for $\mathcal{N}(\mathbf{A})$, $\mathcal{R}(\mathbf{A})$, $\mathcal{N}(\mathbf{A}^T)$, and $\mathcal{R}(\mathbf{A}^T)$. What are the dimensions of the four subspaces?

Solution:

To find $\mathcal{N}(\mathbf{A})$, we solve the system of equations $\mathbf{Ax} = 0$,

$$\mathbf{Ax} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 1 & 3 \\ 4 & 6 & 7 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

Put the system of equations into an augmented matrix,

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 0 \\ 2 & 2 & 1 & 3 & 0 \\ 4 & 6 & 7 & 11 & 0 \end{array}\right] \tag{3}$$

and then find the reduced row echelon form (**RREF**),

$$\left[\begin{array}{cccc|c} 1 & 0 & -2 & -1 & 0 \\ 0 & 1 & \frac{5}{2} & \frac{5}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{4}$$

We can find that

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2x_3 + x_4 \\ -\frac{5}{2}x_3 - \frac{5}{2}x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} 1 \\ -\frac{5}{2} \\ 0 \\ 1 \end{bmatrix} x_4 \tag{5}$$

So,

$$\mathcal{N}(\mathbf{A}) = \textbf{space}\left(\begin{bmatrix} 2 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{5}{2} \\ 0 \\ 1 \end{bmatrix} \right), \tag{6}$$

and the dimension of $\mathcal{N}(\mathbf{A})$ is **2**.

To find $\mathcal{R}(\mathbf{A})$, the equations becomes $\mathbf{Ax} = \mathbf{b}$,

$$\mathbf{Ax} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 1 & 3 \\ 4 & 6 & 7 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b} \tag{7}$$

Because the **RREF** of **A** is

$$\begin{bmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & \frac{5}{2} & \frac{5}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

We can write $\mathbf{b}$ as a linear combination of the first two columns of $\mathbf{A}$:

$$\mathbf{b} = x_1 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix}, \tag{9}$$

i.e.

$$\mathcal{R}(\mathbf{A}) = \textbf{space} \left(\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix} \right), \tag{10}$$

and the dimension of $\mathcal{R}(\mathbf{A})$ is **2**.

To find $\mathcal{N}(\mathbf{A}^T)$ and $\mathcal{R}(\mathbf{A}^T)$, we first calculate the **RREF** of $\mathbf{A}^T$.

$$\mathbf{A}^T = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 2 & 6 \\ 3 & 1 & 7 \\ 4 & 3 & 11 \end{bmatrix}, \tag{11}$$

$$\textbf{RREF of A is } \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{12}$$

So,

$$\mathbf{x} = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} x_3, \tag{13}$$

i.e.

$$\begin{aligned} \mathcal{N}(\mathbf{A}^T) &= \textbf{space} \left(\begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} \right), \\ \mathcal{R}(\mathbf{A}^T) &= \textbf{space} \left(\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1 \\ 3 \end{bmatrix} \right). \end{aligned} \tag{14}$$

And the dimension of $\mathcal{N}(\mathbf{A}^T)$ is **1**, the dimension of $\mathcal{R}(\mathbf{A}^T)$ is **2**.

Exercise 10

Show that if $\mathbf{x} \perp \mathbf{y}$, then

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 . \tag{15}$$

Solution:

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + 2\|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos\theta \tag{16}$$

Because $\mathbf{x} \perp \mathbf{y}$, so $\cos\theta = 0$, i.e.

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 \tag{17}$$

We can consider another solution,

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \sum_{i=1}^{n}(x_i+y_i)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 + 2\sum_{i=1}^{n} x_i^2 y_i^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 \tag{18}$$

because $\mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i^2 y_i^2 = 0$ when $\mathbf{x} \perp \mathbf{y}$.

Exercise 11

In this exercise, we will derive the formula (A.88) for the 1-norm of a matrix. Begin with the optimization problem

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1 . \tag{19}$$

(a) Show that if $\|\mathbf{x}\|_1 = 1$, then

$$\|\mathbf{A}\mathbf{x}\|_1 \le \max_j \sum_{i=1}^{m} |A_{i,j}| . \tag{20}$$

Solution:

We define that $\mathbf{A}$ is a $m \times n$ matrix and $\mathbf{x}$ is a $m \times 1$ vector. Then we can write that:

$$\begin{aligned}
\|\mathbf{A}\mathbf{x}\|_1 &= \sum_{i=1}^{m} |\mathbf{A}\mathbf{x}|_i = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{i,j} x_j \right| \\
&\le \sum_{i=1}^{m}\sum_{j=1}^{n} |A_{i,j}|\,|x_j| = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} |A_{i,j}|\right)|x_j| \\
&\le \left(\max_j \sum_{i=1}^{m} |A_{i,j}|\right)\sum_{j=1}^{n} |x_j| = \max_j \sum_{i=1}^{m} |A_{i,j}| ,
\end{aligned} \tag{21}$$

i.e.

$$\|\mathbf{A}\mathbf{x}\|_1 \le \max_j \sum_{i=1}^{m} |A_{i,j}| . \tag{22}$$

Note that $\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j| = 1$.

(b) Find a vector $\mathbf{x}$ such that $\|\mathbf{x}\|_1 = 1$, and

$$\|\mathbf{A}\mathbf{x}\|_1 = \max_j \sum_{i=1}^{m} |A_{i,j}| \, . \tag{23}$$

Solution:

To find such $\mathbf{x}$, we assume that the maximum is arrived in the k^{th} column, i.e.

$$\max_j \sum_{i=1}^{m} |A_{i,j}| = \sum_{i=1}^{m} |A_{i,k}| \, . \tag{24}$$

So, we can write that:

$$\begin{aligned}
\|\mathbf{A}\mathbf{x}\|_1 &= \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{i,j} x_j \right| = \sum_{i=1}^{m} |A_{i,k}| \, , \\
\Longrightarrow & \left| \sum_{j=1}^{n} A_{i,j} x_j \right| = |A_{i,k}| \, , \\
\Longrightarrow & \, x_j = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} .
\end{aligned} \tag{25}$$

Thus, the vector $\mathbf{x} = (0, 0, ..., 1, 0, ..., 0)^T$, i.e. the k^{th} component of the vector is $\mathbf{1}$, otherwise is $\mathbf{0}$.

(c) Conclude that

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{x}\|_1 = 1} \|\mathbf{A}\mathbf{x}\|_1 = \max_j \sum_{i=1}^{m} |A_{i,j}| \, . \tag{26}$$

Solution:

$$\begin{aligned}
&\|\mathbf{A}\|_1 = \|\mathbf{A}\|_1 \|\mathbf{x}\|_1 \geq \|\mathbf{A}\mathbf{x}\|_1, \\
\Longrightarrow &\|\mathbf{A}\|_1 = \max_{\|\mathbf{x}\|_1 = 1} \|\mathbf{A}\mathbf{x}\|_1 = \max_j \sum_{i=1}^{m} |A_{i,j}| \, .
\end{aligned} \tag{27}$$

Appendix B

Exercise 6

Suppose that $\mathbf{x} = (X_1, X_2)^T$ is a vector composed of two random variables with a multivariate normal distribution with expected value μ and covariance matrix $\mathbf{C}$, and that $\mathbf{A}$ is a 2 by 2 matrix. Use properties of expected value and covariance to show that $\mathbf{y} = \mathbf{A}\mathbf{x}$ has expected value $\mathbf{A}\mu$ and covariance $\mathbf{A}\mathbf{C}\mathbf{A}^T$

Solution:

According to the question, we know that the JDF of this multivariate normal distribution (the dimension is 2) is:

$$f(\mathbf{x}) = |2\pi\mathbf{C}|^{-\frac{1}{2}} \,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{C}^{-1}(\mathbf{x}-\mu)\right), \tag{28}$$

and we can simply define $D = |2\pi\mathbf{C}|^{-\frac{1}{2}}$ as D is a constant. Thus,

$$f(\mathbf{x}) = D\,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{C}^{-1}(\mathbf{x}-\mu)\right). \tag{29}$$

So,

$$E[\mathbf{y}] = E[\mathbf{A}\mathbf{x}] = \int_{-\infty}^{\infty} \mathbf{A}\mathbf{x} f(\mathbf{x})\, d\mathbf{x} = \mathbf{A}\int_{-\infty}^{\infty} \mathbf{x} f(\mathbf{x})\, d\mathbf{x} = \mathbf{A}E[\mathbf{x}] = \mathbf{A}\mu. \tag{30}$$

To covariance matrix of $\mathbf{y}$, we consider that:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}, \tag{31}$$

and then, we replace $\mathbf{x}$ with $\mathbf{A}^{-1}\mathbf{y}$ in $f(\mathbf{x})$:

$$\begin{aligned}
f &= D\,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{C}^{-1}(\mathbf{x}-\mu)\right) \\
&= D\,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{A}^{-1}\mathbf{y}-\mu)^T\mathbf{C}^{-1}(\mathbf{A}^{-1}\mathbf{y}-\mu)\right) \\
&= D\,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{A}^{-1}\mathbf{y}-\mathbf{A}^{-1}\mathbf{A}\mu)^T\mathbf{C}^{-1}(\mathbf{A}^{-1}\mathbf{y}-\mathbf{A}^{-1}\mathbf{A}\mu)\right). \\
&= D\,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{y}-\mathbf{A}\mu)^T(\mathbf{A}^T)^{-1}\mathbf{C}^{-1}\mathbf{A}^{-1}(\mathbf{y}-\mathbf{A}\mu)\right) \\
&= D\,\mathbf{exp}\left(-\frac{1}{2}(\mathbf{y}-\mathbf{A}\mu)^T(\mathbf{A}\mathbf{C}\mathbf{A}^T)^{-1}(\mathbf{y}-\mathbf{A}\mu)\right)
\end{aligned} \tag{32}$$

It's obvious that the covariance matrix of $\mathbf{y}$ is $\mathbf{A}\mathbf{C}\mathbf{A}^T$

Exercise 9

Using MATLAB, repeat the following experiment 1000 times. Generate five exponentially distributed random numbers from the exponential probability density function (B.10) with means $\mu = 1/\lambda = 10$. You may find the library function **exprnd** to be useful here. Use (B.74) to calculate a 95% confidence interval for the 1000 mean determinations. How many times out of the 1000 experiments did the 95% confidence interval cover the expected value of 10? What happens if you instead generate 50 exponentially distributed random numbers at time? Discuss your results.

Solution:

Matlab code

```matlab
clear all;

NumE = 1000;
countIn = 0;
testOfNum = 10;

mu = 10;
nNum = 5;
lmd = 0.1;
tConst = 2.262;

saveRes = zeros(NumE,nNum);

for ind=1:NumE
    R = exprnd(mu,1,nNum);
    saveRes(ind,:) = R;
    meanR = mean(R);
    sR = sqrt(sum((R-meanR).^2)/(nNum-1));
    sec2 = tConst*sR/sqrt(nNum);
    interL = meanR-sec2;
    interR = meanR+sec2;
    disp(["Epoch ",int2str(ind),": [",num2str(interL),", ",num2str(interR),"]"]);
    if testOfNum>=interL & testOfNum<=interR
        countIn = countIn+1;
    end
end

% accuracy
in2all = countIn/NumE;
countIn
NumE
in2all
disp(str)
```

(1) About 847 times.
(2) 959 times out of the 1000 experiments did the 95% confidence interval cover the expected value of 10 when I instead generate 50 exponentially distributed random numbers at time.
(3) The random numbers under 50 samplings is closer to the exponential distribution than that under 5 samplings. **Long live understanding!**

Appendix C

Exercise 1

Let

$$f(\mathbf{x}) = x_1^2 x_2^2 - 2x_1 x_2^2 + x_2^2 - 3x_1^2 x_2 + 12x_1 x_2 - 12x_2 + 6. \tag{33}$$

Find the gradient, $\nabla f(\mathbf{x})$, and Hessian, $\mathbf{H}(f(\mathbf{x}))$. What are the critical points of f? Which of these are minima and maxima of f?

Solution:

$$\begin{aligned}\nabla f(\mathbf{x}) &= \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 x_2^2 - 2x_2^2 - 6x_1 x_2 + 12x_2 \\ 2x_1^2 x_2 - 4x_1 x_2 + 2x_2 - 3x_1^2 + 12x_1 - 12 \end{bmatrix}.\end{aligned} \tag{34}$$

And

$$\begin{aligned}\mathbf{H}(f(\mathbf{x})) &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} \\ &= \begin{bmatrix} 2x_2^2 - 6x_2 & 4x_1 x_2 - 4x_2 - 6x_1 + 12 \\ 4x_1 x_2 - 4x_2 - 6x_1 + 12 & 2x_1^2 - 4x_1 + 2 \end{bmatrix}.\end{aligned} \tag{35}$$

A point $\mathbf{x}^*$ is a critical point when $\nabla f(\mathbf{x}^*) = \mathbf{0}$,

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 x_2^2 - 2x_2^2 - 6x_1 x_2 + 12x_2 \\ 2x_1^2 x_2 - 4x_1 x_2 + 2x_2 - 3x_1^2 + 12x_1 - 12 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{36}$$

Solving the function, we can get **2** critical points:

$$\mathbf{x}^* = (x_1, x_2) : (0, 6), (2, 0). \tag{37}$$

As the bottom figure shown (fig 1), $(0, 6)$ is minima of f, $f(0, 6) = -30$, and f doesn't exist maxima or its maxima is a infinite value.

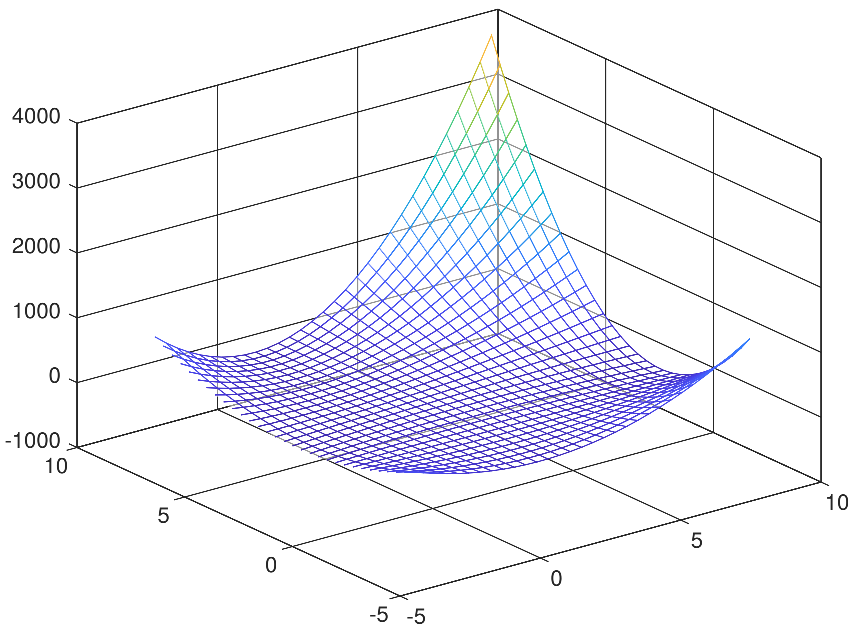

Figure 1: a function f graph of Appendix C exercise 1

Exercise 2

Find a Taylor's series approximation for $f(\mathbf{x} + \Delta\mathbf{x})$, where

$$f(\mathbf{x}) = e^{-(x_1+x_2)^2} \tag{38}$$

is near the point

$$\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}. \tag{39}$$

Solution:

$$\begin{aligned}
\nabla f(\mathbf{x}) &= \begin{bmatrix} -2(x_1+x_2)e^{-(x_1+x_2)^2} \\ -2(x_1+x_2)e^{-(x_1+x_2)^2} \end{bmatrix}, \\
\mathbf{H}(f(\mathbf{x})) &= \begin{bmatrix} -2(1-2(x_1+x_2)^2)e^{-(x_1+x_2)^2} & -2(1-2(x_1+x_2)^2)e^{-(x_1+x_2)^2} \\ -2(1-2(x_1+x_2)^2)e^{-(x_1+x_2)^2} & -2(1-2(x_1+x_2)^2)e^{-(x_1+x_2)^2} \end{bmatrix}, \\
f(\mathbf{x}+\Delta\mathbf{x}) &= f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \mathbf{H}(f(\mathbf{x}))\Delta\mathbf{x} + o^n.
\end{aligned} \tag{40}$$

Because $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, so:

$$f(\mathbf{x}+\Delta\mathbf{x}) = e^{-25} + \begin{bmatrix} -10e^{-25}, -10e^{-25} \end{bmatrix} \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \begin{bmatrix} 98e^{-25} & 98e^{-25} \\ 98e^{-25} & 98e^{-25} \end{bmatrix} \Delta\mathbf{x} + o^n \tag{41}$$