CS 111 Homework 6

Xiao Zhou

TOTAL POINTS

23.5 / 25

QUESTION 1

SPD matrices 8 pts

1.1 Non SPD matrix and nonzero vector x test 4 / 4

✓ - **0 pts** Correct

 - **2 pts** $$x$$ needs to be a vector, so that $$x^\top A x$$ is a scalar.

 - **2.5 pts** $$A$$ is not symmetric.

 - **2 pts** Need to find the vector $$x$$ asked by the question.

 - **4 pts** See comment below.

1.2 Proof: B.T@B is SPD 2.5 / 4

 - **0 pts** Correct

✓ - **1.5 pts** The fact that $$B$$ has rank $$n$$ has 1) not been used (which is crucial for knowing that $$x^\top B^\top B x>0$$ instead of $$x^\top B^\top B x\geq 0$$), or 2) has been mentioned without its proper use or explanation of use.

 - **4 pts** Didn't do it.

 - **4 pts** This is not a mathematical proof; please read what the question is asking for.

 - **4 pts** What has been written is an incorrect analysis that doesn't prove what the question asks.

 - **0.5 pts** There are some typos or confusing parts in the proof, as indicated on the document.

 - **2 pts** See comment below.

 - **4 pts** See comment below.

 1 $$B$$ is not invertible in general: it can be rectangular!

QUESTION 2

Digraph analysis 15 pts

2.1 Code for digraph analysis plot and Gettysburg text 6 / 6

✓ - **0 pts** Correct

 - **2 pts** with $$X = USV^T$$, used the first column of $$V^T$$ instead of the first column of $$V$$

 - **4.5 pts** No code included

 - **6 pts** No answer

 - **2 pts** No Plot

2.2 Plot and analysis for Alice in Wonderland text 3 / 3

✓ - **0 pts** Correct

 - **1 pts** with $$X = USV^T$$, used the first column of $$V^T$$ instead of the first column of $$V$$

 - **2 pts** No code included

 - **3 pts** No answer

 - **2 pts** No Plot

 - **1 pts** No analysis

2.3 Code for digraph analysis plot in different language and test 6 / 6

✓ - **0 pts** Correct

 - **1 pts** Did not say which language you use

 - **1.5 pts** Missing/Mistakes in code

 - **3 pts** Missing/Mistakes in graph

 - **2 pts** Missing comments

 - **6 pts** No answer provided

QUESTION 3

3 Correct LaTeX usage 1 / 1

✓ - **0 pts** Correct

QUESTION 4

4 Correct answer paging 1 / 1

✓ - **0 pts** Correct

- 1 pts Incorrect

CS 111. Homework 6
Xiao Zhou

1. Not SPD

```
A = np.array([[1,2],[2,3]])
print(A)
```

```
[[1 2]
 [2 3]]
```

```
evals, W = spla.eigh( A )
print(evals)
print(W)
```

```
[-0.2361  4.2361]
[[-0.8507  0.5257]
 [ 0.5257  0.8507]]
```

```
B = np.array([[-0.8507],[0.5257]])
B
```

```
array([[-0.8507],
       [ 0.5257]])
```

```
B.T@A@B
```

```
array([[-0.2361]])
```

1.2. $B^T B$ is SPD which means $\mathbf{x}^T A \mathbf{x} > 0$

$B^T B$ is symmetric because $(B^T B)^T = B^T (B^T)^T = B^T B$.

Then we have $x^T B^T B x = (Bx)^T (Bx) = ||Bx||^2 \geqslant 0$

Since B is a m by n matrix with rank n, B is not singular, B^T is also not singular because $(B^T)^{-1} = (B^{-1})^T$, so for any non-zero x, $B^T x \neq 0$, so $||Bx||^2 > 0$. Thus $\mathbf{x}^T B^T B \mathbf{x} > 0$

①

1.1 Non SPD matrix and nonzero vector x test **4 / 4**

✓ **- 0 pts** Correct

    - 2 pts x needs to be a vector, so that $x^\top A x$ is a scalar.

    - 2.5 pts A is not symmetric.

    - 2 pts Need to find the vector x asked by the question.

    - 4 pts See comment below.

ıl gradescope

CS 111. Homework 6
Xiao Zhou

1. Not SPD

```
A = np.array([[1,2],[2,3]])
print(A)
```

```
[[1 2]
 [2 3]]
```

```
evals, W = spla.eigh( A )
print(evals)
print(W)
```

```
[-0.2361  4.2361]
[[-0.8507  0.5257]
 [ 0.5257  0.8507]]
```

```
B = np.array([[-0.8507],[0.5257]])
B
```

```
array([[-0.8507],
       [ 0.5257]])
```

```
B.T@A@B
```

```
array([[-0.2361]])
```

1.2. $B^T B$ is SPD which means $\mathbf{x}^T A \mathbf{x} > 0$

$B^T B$ is symmetric because $(B^T B)^T = B^T (B^T)^T = B^T B$.

Then we have $x^T B^T B x = (Bx)^T (Bx) = ||Bx||^2 \geqslant 0$

①

Since B is a m by n matrix with rank n, B is not singular, B^T is also not singular because $(B^T)^{-1} = (B^{-1})^T$, so for any non-zero x, $B^T x \neq 0$, so $||Bx||^2 > 0$. Thus $\mathbf{x}^T B^T B \mathbf{x} > 0$

- **0 pts** Correct

✓ - **1.5 pts** The fact that $$B$$ has rank $$n$$ has 1) not been used (which is crucial for knowing that $$x^\top B^\top B x>0$$ instead of $$x^\top B^\top B x\geq 0$$), or 2) has been mentioned without its proper use or explanation of use.

- **4 pts** Didn't do it.

- **4 pts** This is not a mathematical proof; please read what the question is asking for.

- **4 pts** What has been written is an incorrect analysis that doesn't prove what the question asks.

- **0.5 pts** There are some typos or confusing parts in the proof, as indicated on the document.

- **2 pts** See comment below.

- **4 pts** See comment below.

1 $$B$$ is not invertible in general: it can be rectangular!

ılı gradescope

2. 2.1. gettysburg.txt

```python
with open('gettysburg.txt') as file:
    raw_text = file.read()
letters = []
for c in raw_text.upper():
    if c.isalpha():
        letters.append(c)
for i in range(len(letters)):
    letters[i] = ord(letters[i]) - ord('A')
A = np.zeros((26,26))
for i in range(len(letters)-1):
    A[letters[i]][letters[i+1]] += 1
A[letters[-1]][letters[0]] += 1
U,S,V = np.linalg.svd(A)
V = V.T
plt.figure(figsize = (8, 8))
plt.axis('square')
for i in range(26):
    plt.text(U[i][1], V[i][1], chr(i + ord('A')))
limits = (-0.6, 0.6)
plt.xlim(limits)
plt.ylim(limits)
plt.plot(limits, (0, 0), 'r')
plt.plot((0, 0), limits, 'r')
plt.title('The second principal component of a digraph matrix')
plt.xlabel('values from U_1')
plt.ylabel('values from V_1')
```


The second principal component of a digraph matrix
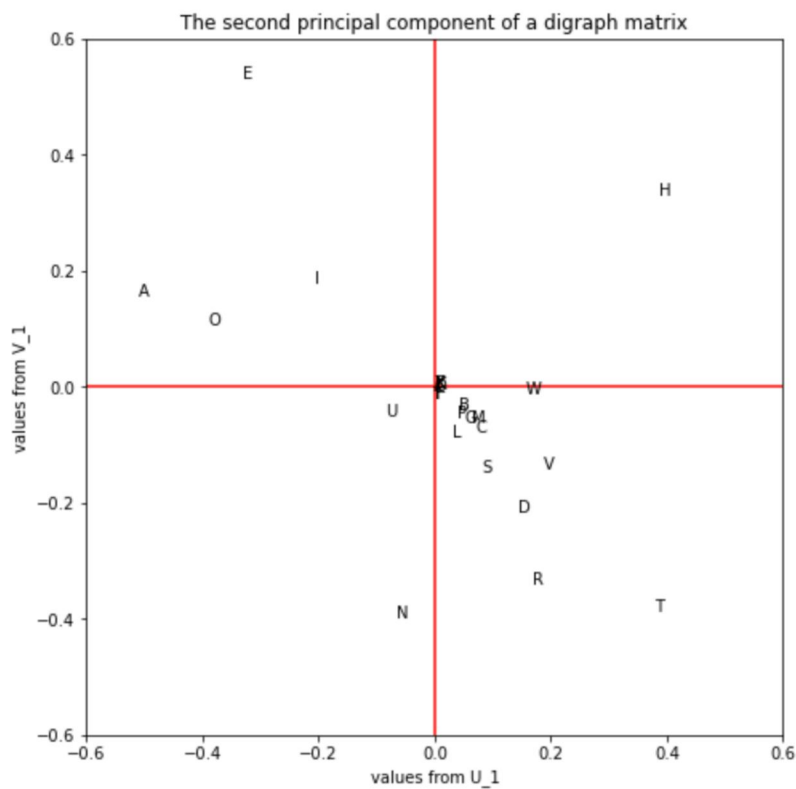
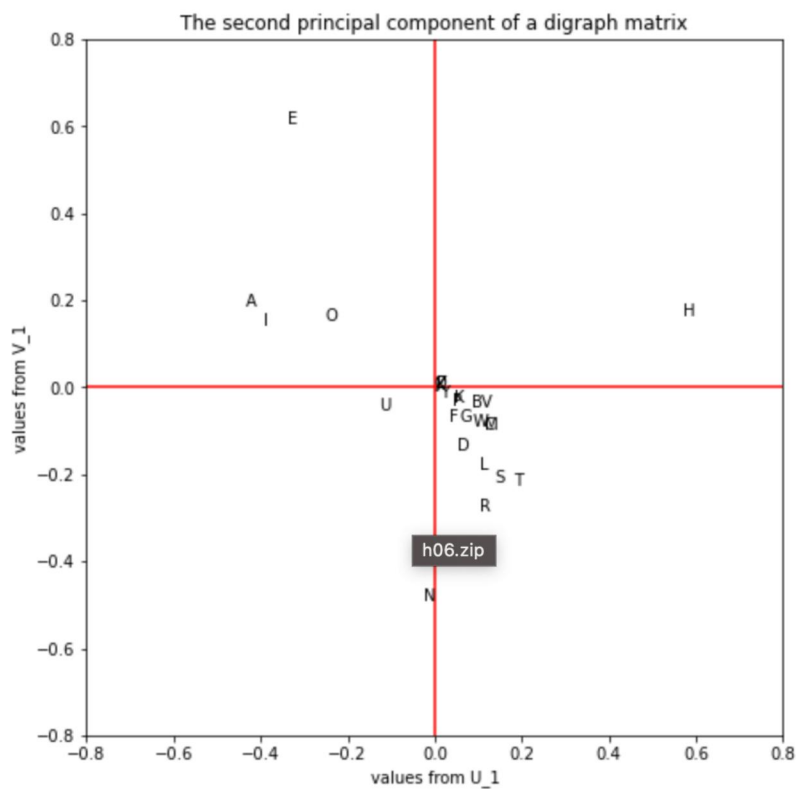**2.1** Code for digraph analysis plot and Gettysburg text **6 / 6**

✓ **- 0 pts** Correct

- 2 pts with $X = USV^T$, used the first column of V^T instead of the first column of V

- 4.5 pts No code included

- 6 pts No answer

- 2 pts No Plot

2.2. alice.txt

```python
with open('alice.txt') as file:
    raw_text = file.read()
letters = []
for c in raw_text.upper():
    if c.isalpha():
        letters.append(c)
for i in range(len(letters)):
    letters[i] = ord(letters[i]) - ord('A')
A = np.zeros((26,26))
for i in range(len(letters)-1):
    A[letters[i]][letters[i+1]] += 1
A[letters[-1]][letters[0]] += 1
U,S,V = np.linalg.svd(A)
V = V.T
plt.figure(figsize = (8, 8))
plt.axis('square')
for i in range(26):
    plt.text(U[i][1], V[i][1], chr(i + ord('A')))
limits = (-0.8, 0.8)
plt.xlim(limits)
plt.ylim(limits)
plt.plot(limits, (0, 0), 'r')
plt.plot((0, 0), limits, 'r')
plt.title('The second principal component of a digraph matrix')
plt.xlabel('values from U_1')
plt.ylabel('values from V_1')
```


The distribution of the vowels and consonants are similar that most of the vowels (A, E, I, O) show up in the upper left quadrant, most of the consonants show up in the lower right quadrant, H as a neuter shows up in the upper right quadrant, U and N are in the lower left quadrant since they can be followed by both vowels and consonants. The 'Alice.txt' graph's points is more well distributed and 'Gettyburg.txt' graph is more centered in the middle. The shorter the length the more dense the graph will be.
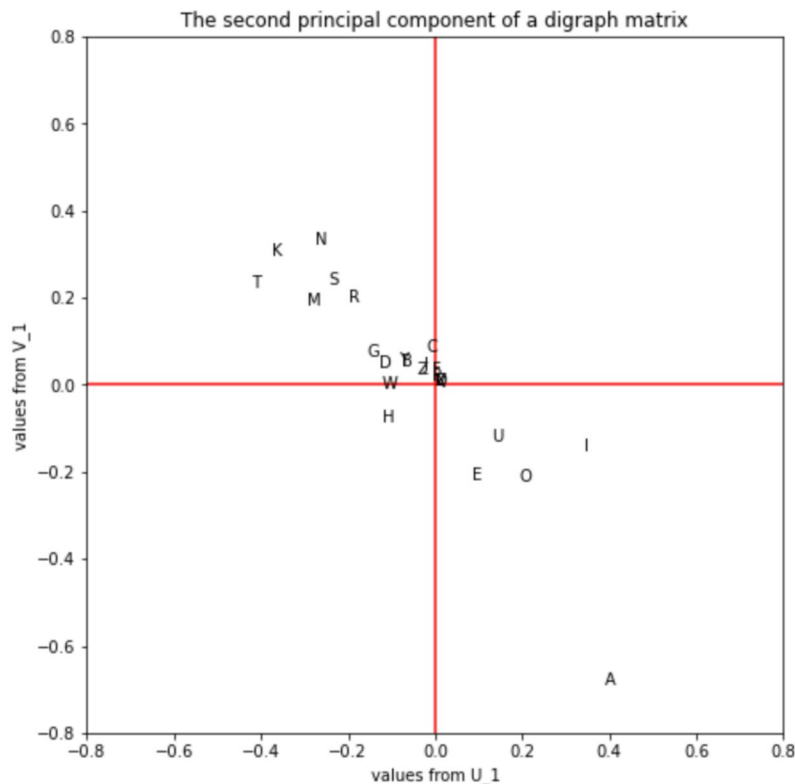
2.2 Plot and analysis for Alice in Wonderland text **3 / 3**

✓ **- 0 pts** Correct

    - 1 pts with $X = USV^T$, used the first column of V^T instead of the first column of V

    - 2 pts No code included

    - 3 pts No answer

    - 2 pts No Plot

    - 1 pts No analysis

ıl gradescope

2.3. Japanese text

```python
with open('jap.txt') as file:
    raw_text = file.read()
letters = []
for c in raw_text.upper():
    if c.isalpha():
        letters.append(c)
for i in range(len(letters)):
    letters[i] = ord(letters[i]) - ord('A')
A = np.zeros((26,26))
for i in range(len(letters)-1):
    A[letters[i]][letters[i+1]] += 1
A[letters[-1]][letters[0]] += 1
U,S,V = np.linalg.svd(A)
V = V.T
plt.figure(figsize = (8, 8))
plt.axis('square')
for i in range(26):
    plt.text(U[i][1], V[i][1], chr(i + ord('A')))
limits = (-0.8, 0.8)
plt.xlim(limits)
plt.ylim(limits)
plt.plot(limits, (0, 0), 'r')
plt.plot((0, 0), limits, 'r')
plt.title('The second principal component of a digraph matrix')
plt.xlabel('values from U_1')
plt.ylabel('values from V_1')
```


I chose Japanese and convert it into Romajin, the result is quite the opposite: all of the vowels (A, E, I, O, U) show up in the lower right quadrant, most of the consonants show up in the upper left quadrant. The graph is like a first reflect along y = -x and then y = x. N shows up in the upper left and H shows up in lower left. Japanese is different with English behavior
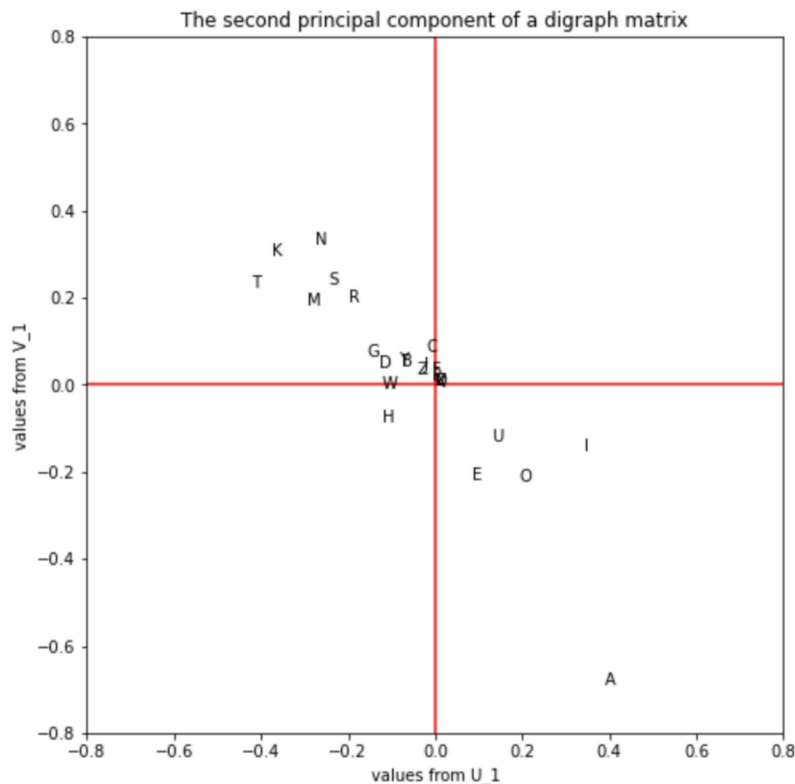
4

2.3 Code for digraph analysis plot in different language and test 6 / 6

✓ **- 0 pts** Correct

  - 1 pts Did not say which language you use

  - 1.5 pts Missing/Mistakes in code

  - 3 pts Missing/Mistakes in graph

  - 2 pts Missing comments

  - 6 pts No answer provided

2.3. Japanese text

```python
with open('jap.txt') as file:
    raw_text = file.read()
letters = []
for c in raw_text.upper():
    if c.isalpha():
        letters.append(c)
for i in range(len(letters)):
    letters[i] = ord(letters[i]) - ord('A')
A = np.zeros((26,26))
for i in range(len(letters)-1):
    A[letters[i]][letters[i+1]] += 1
A[letters[-1]][letters[0]] += 1
U,S,V = np.linalg.svd(A)
V = V.T
plt.figure(figsize = (8, 8))
plt.axis('square')
for i in range(26):
    plt.text(U[i][1], V[i][1], chr(i + ord('A')))
limits = (-0.8, 0.8)
plt.xlim(limits)
plt.ylim(limits)
plt.plot(limits, (0, 0), 'r')
plt.plot((0, 0), limits, 'r')
plt.title('The second principal component of a digraph matrix')
plt.xlabel('values from U_1')
plt.ylabel('values from V_1')
```
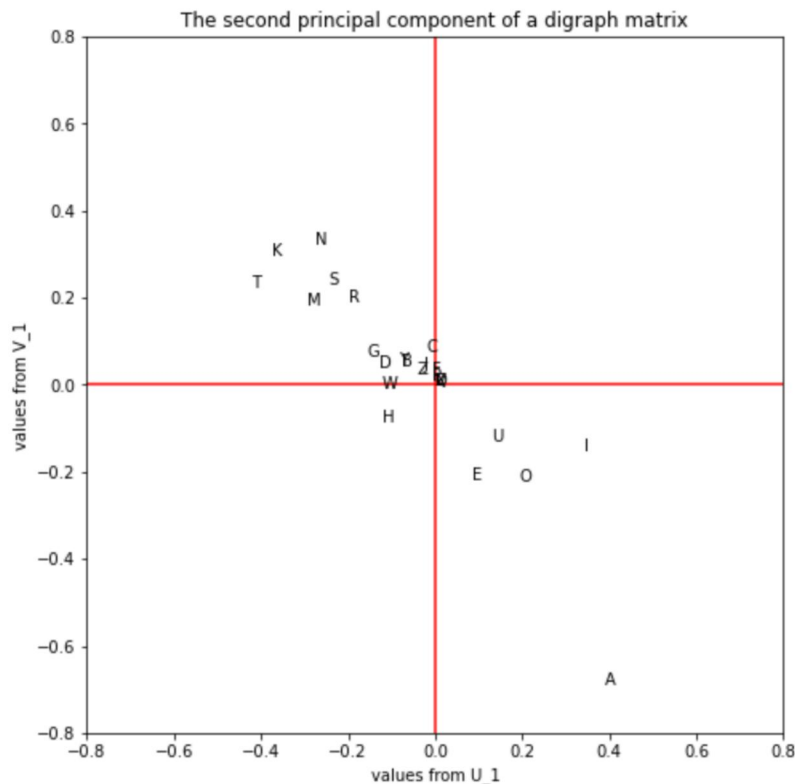

I chose Japanese and convert it into Romajin, the result is quite the opposite: all of the vowels (A, E, I, O, U) show up in the lower right quadrant, most of the consonants show up in the upper left quadrant. The graph is like a first reflect along y = -x and then y = x. N shows up in the upper left and H shows up in lower left. Japanese is different with English behavior

4

3 Correct LaTeX usage **1 / 1**

✓ **- 0 pts** Correct

2.3. Japanese text

```python
with open('jap.txt') as file:
    raw_text = file.read()
letters = []
for c in raw_text.upper():
    if c.isalpha():
        letters.append(c)
for i in range(len(letters)):
    letters[i] = ord(letters[i]) - ord('A')
A = np.zeros((26,26))
for i in range(len(letters)-1):
    A[letters[i]][letters[i+1]] += 1
A[letters[-1]][letters[0]] += 1
U,S,V = np.linalg.svd(A)
V = V.T
plt.figure(figsize = (8, 8))
plt.axis('square')
for i in range(26):
    plt.text(U[i][1], V[i][1], chr(i + ord('A')))
limits = (-0.8, 0.8)
plt.xlim(limits)
plt.ylim(limits)
plt.plot(limits, (0, 0), 'r')
plt.plot((0, 0), limits, 'r')
plt.title('The second principal component of a digraph matrix')
plt.xlabel('values from U_1')
plt.ylabel('values from V_1')
```


I chose Japanese and convert it into Romajin, the result is quite the opposite: all of the vowels (A, E, I, O, U) show up in the lower right quadrant, most of the consonants show up in the upper left quadrant. The graph is like a first reflect along y = -x and then y = x. N shows up in the upper left and H shows up in lower left. Japanese is different with English behavior

4

4 Correct answer paging **1 / 1**

✓ **- 0 pts** Correct

- 1 pts Incorrect

gradescope