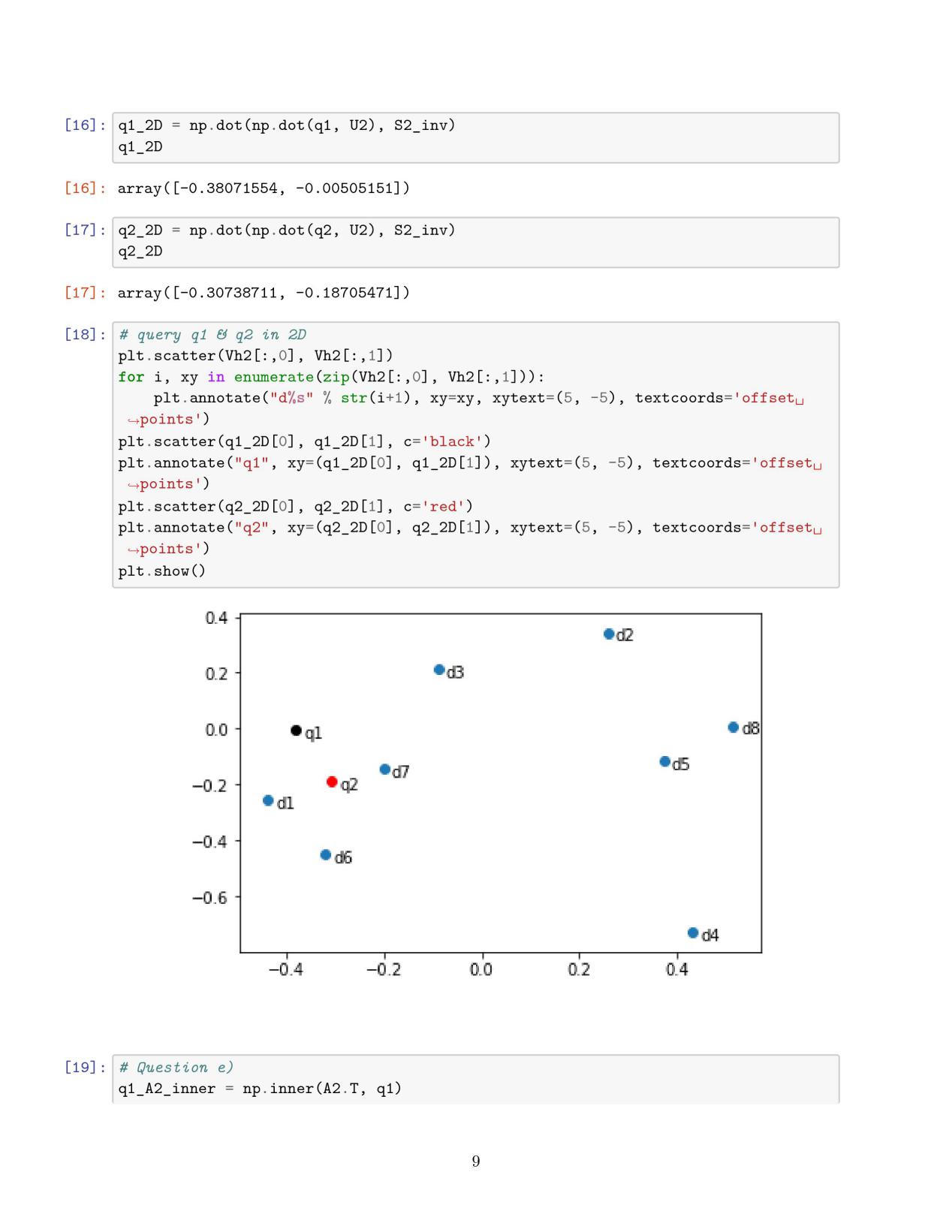
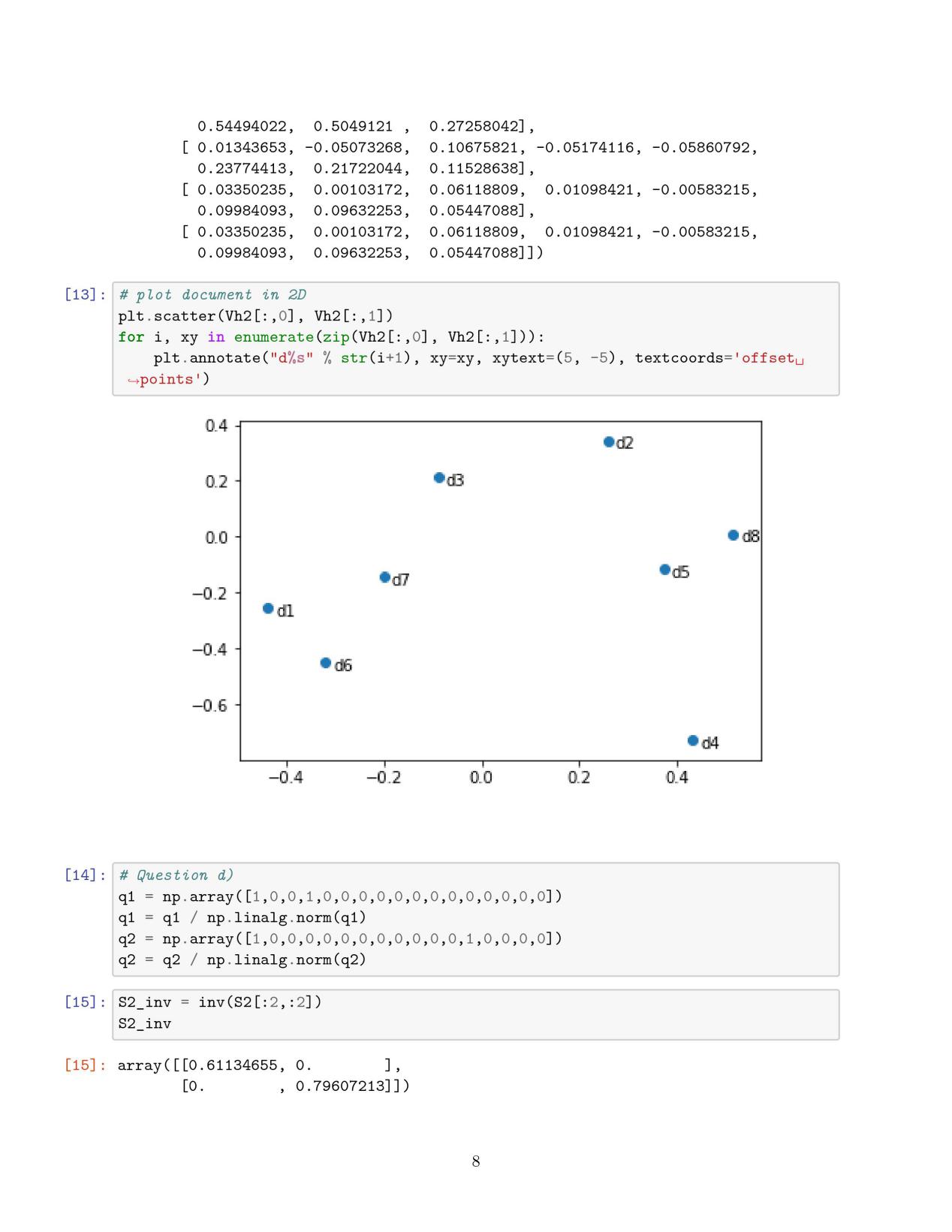
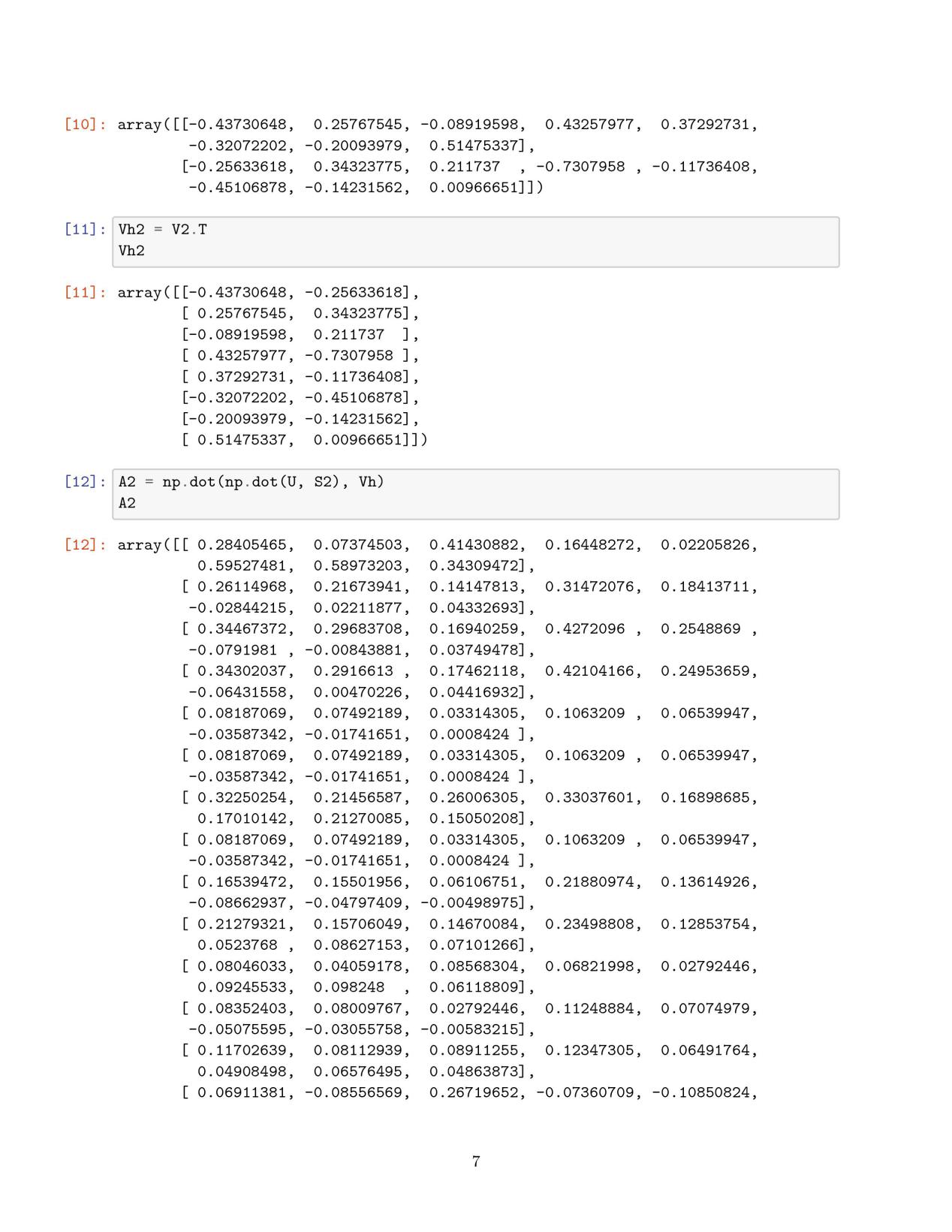
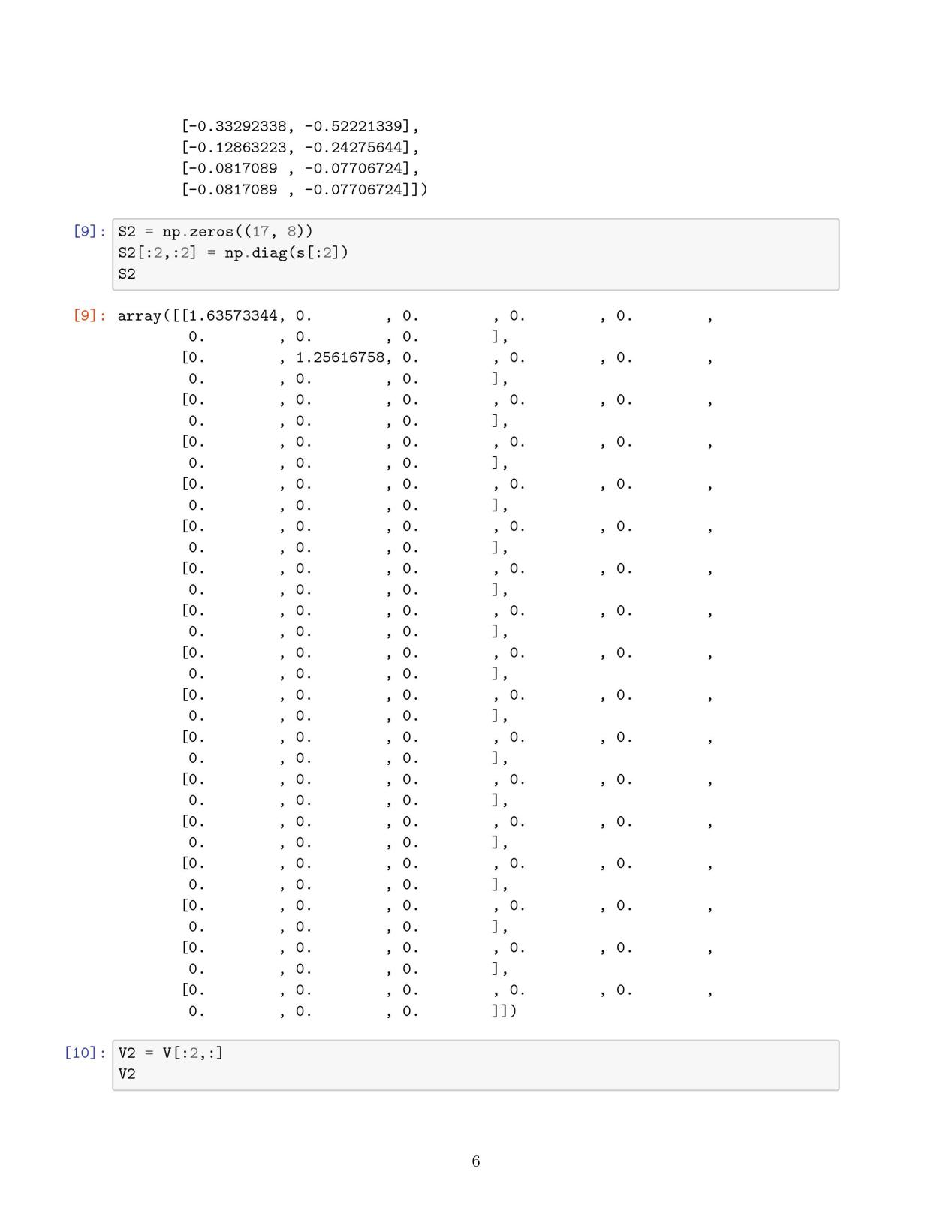
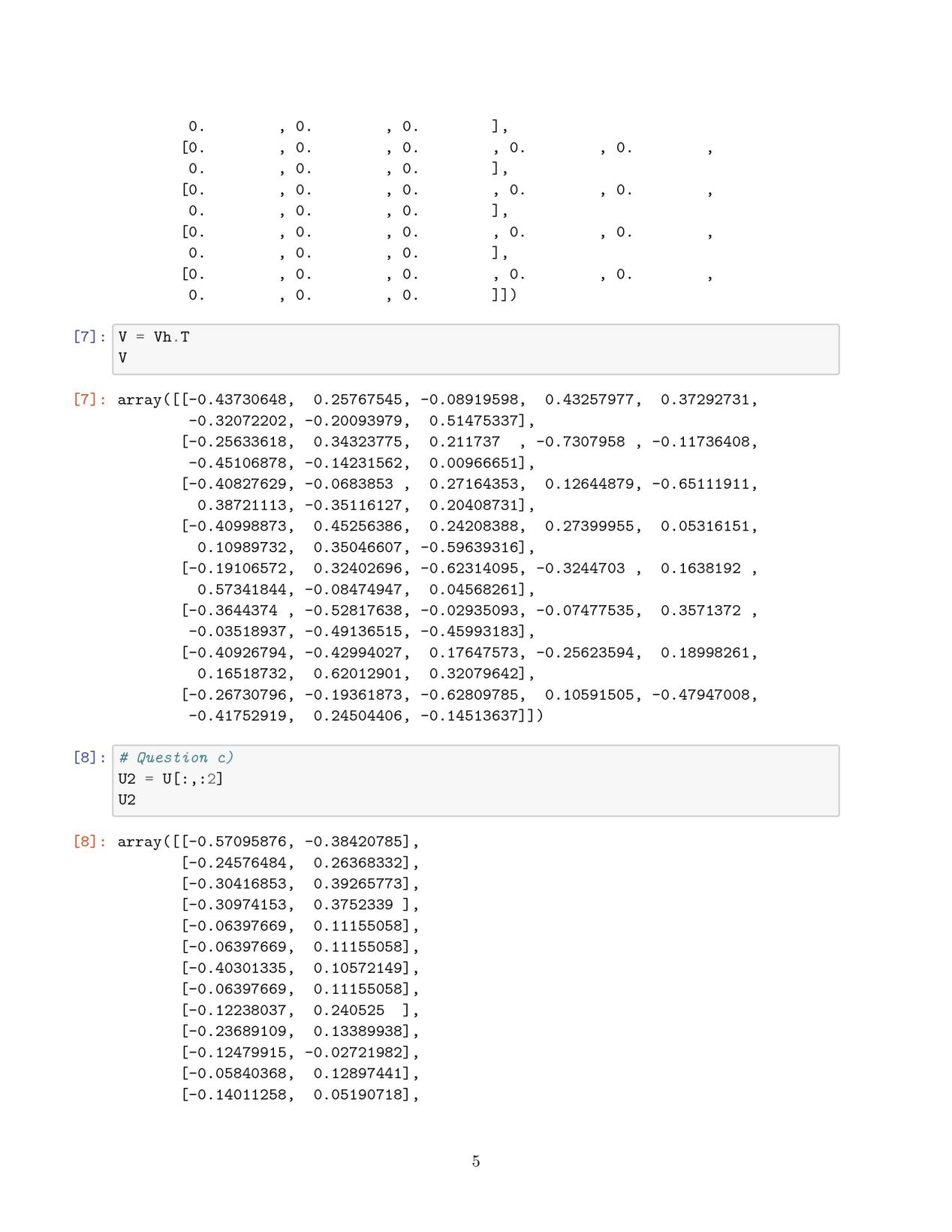
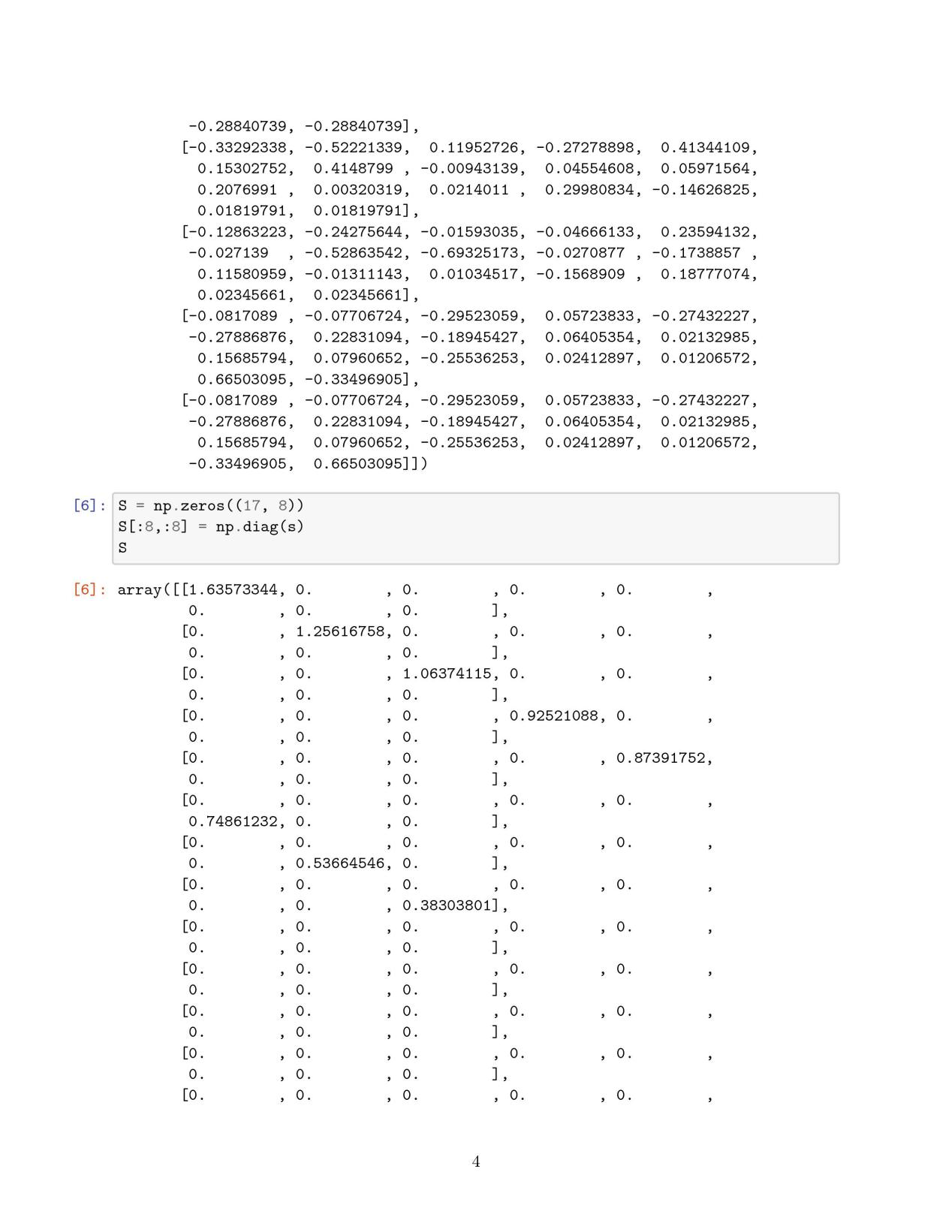
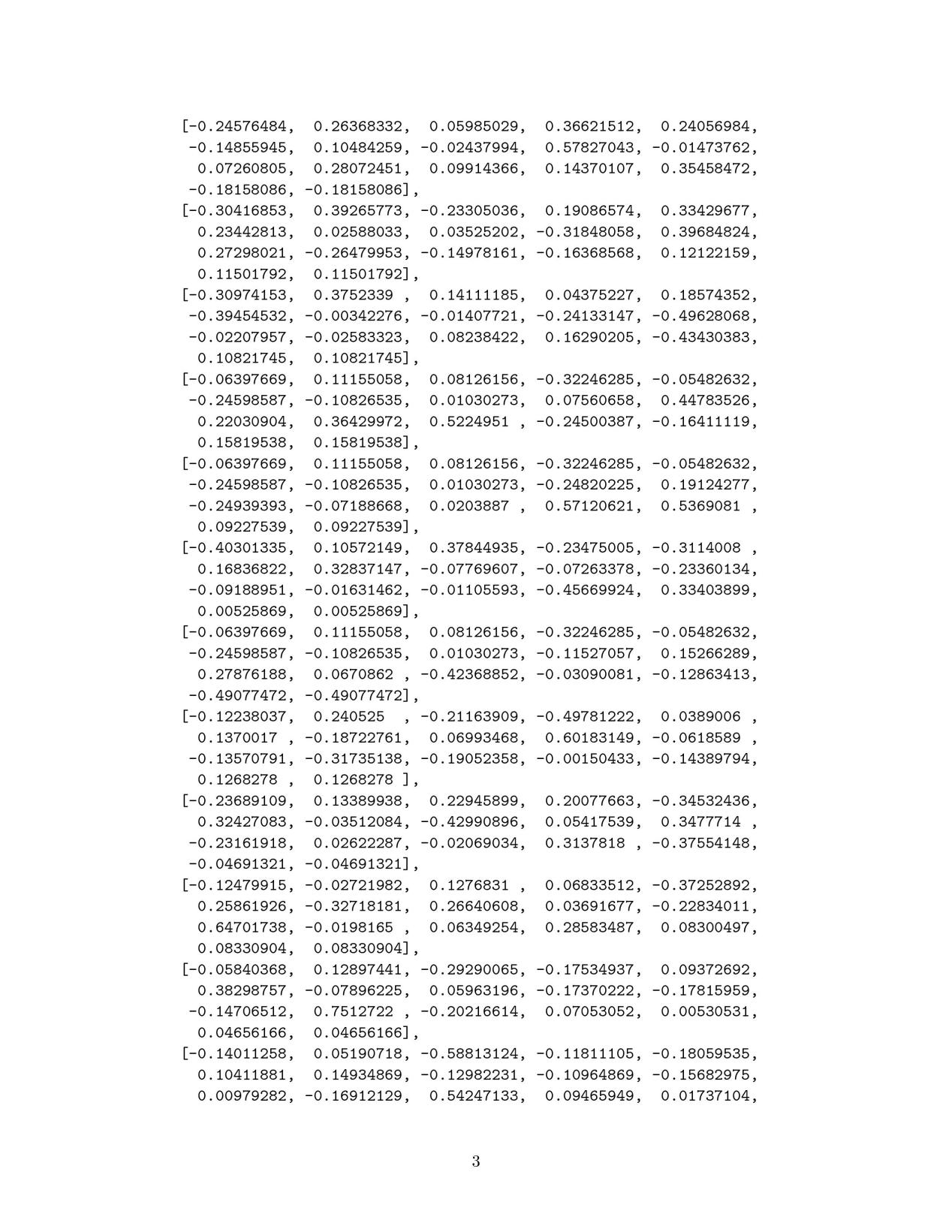
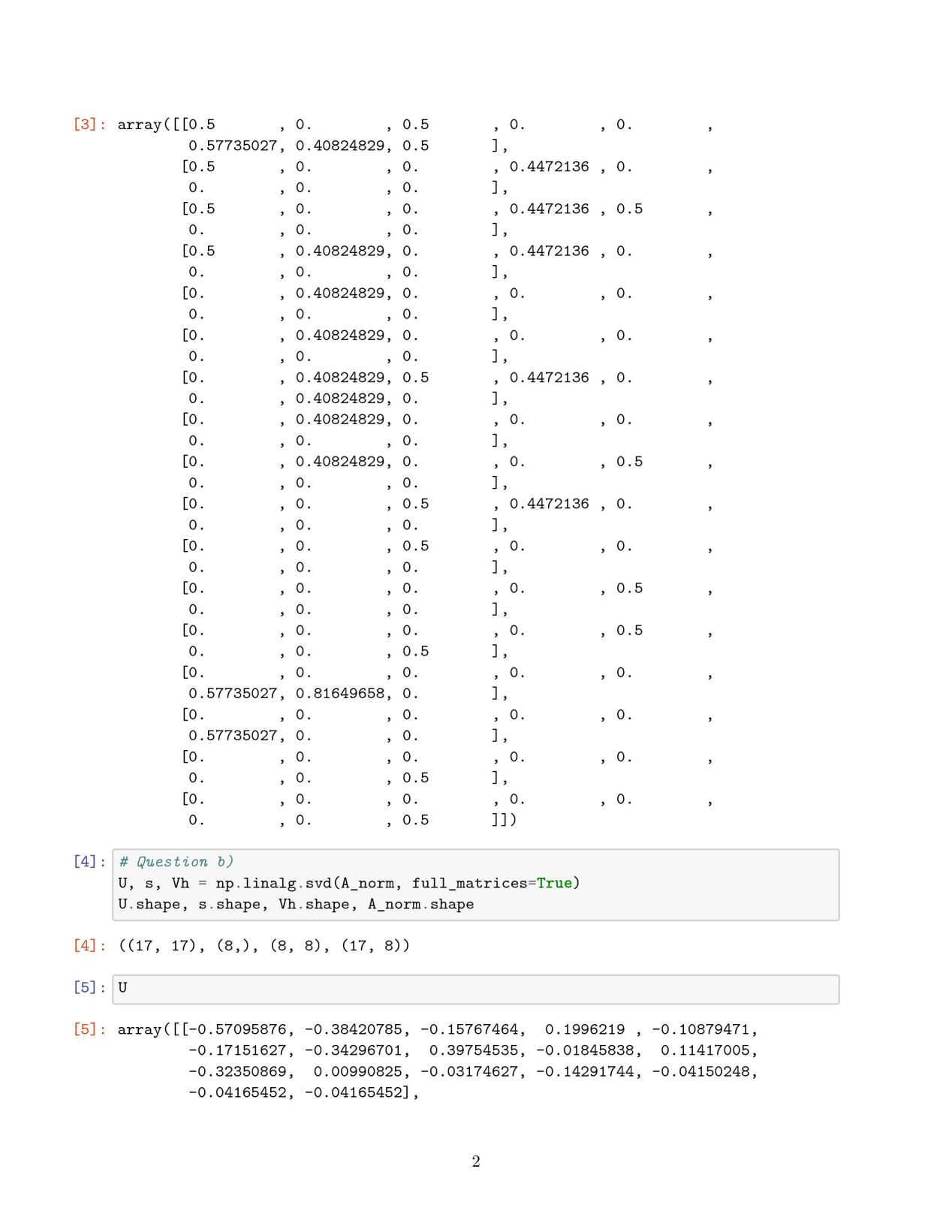
**Q1:**

|  |  |  |
| --- | --- | --- |
| TermID | term | DocID |
| T1 | windows | D1, D3, D6, D7, D8 |
| T2 | Azure | D1, D4 |
| T3 | Cloud | D1, D4, D5 |
| T4 | Microsoft | D1, D2, D4 |
| T5 | Chairman | D2 |
| T6 | Bill | D2 |
| T7 | Gates | D2, D3, D4, D7 |
| T8 | retirement | D2 |
| T9 | morning | D2, D5 |
| T10 | Success(successful) | D3, D4 |
| T11 | attributed | D3 |
| T12 | sunrise | D5 |
| T13 | garden | D5, D8 |
| T14 | wooden | D6, D7 |
| T15 | frames | D6 |
| T16 | trees | D8 |
| T17 | flowers | D8 |



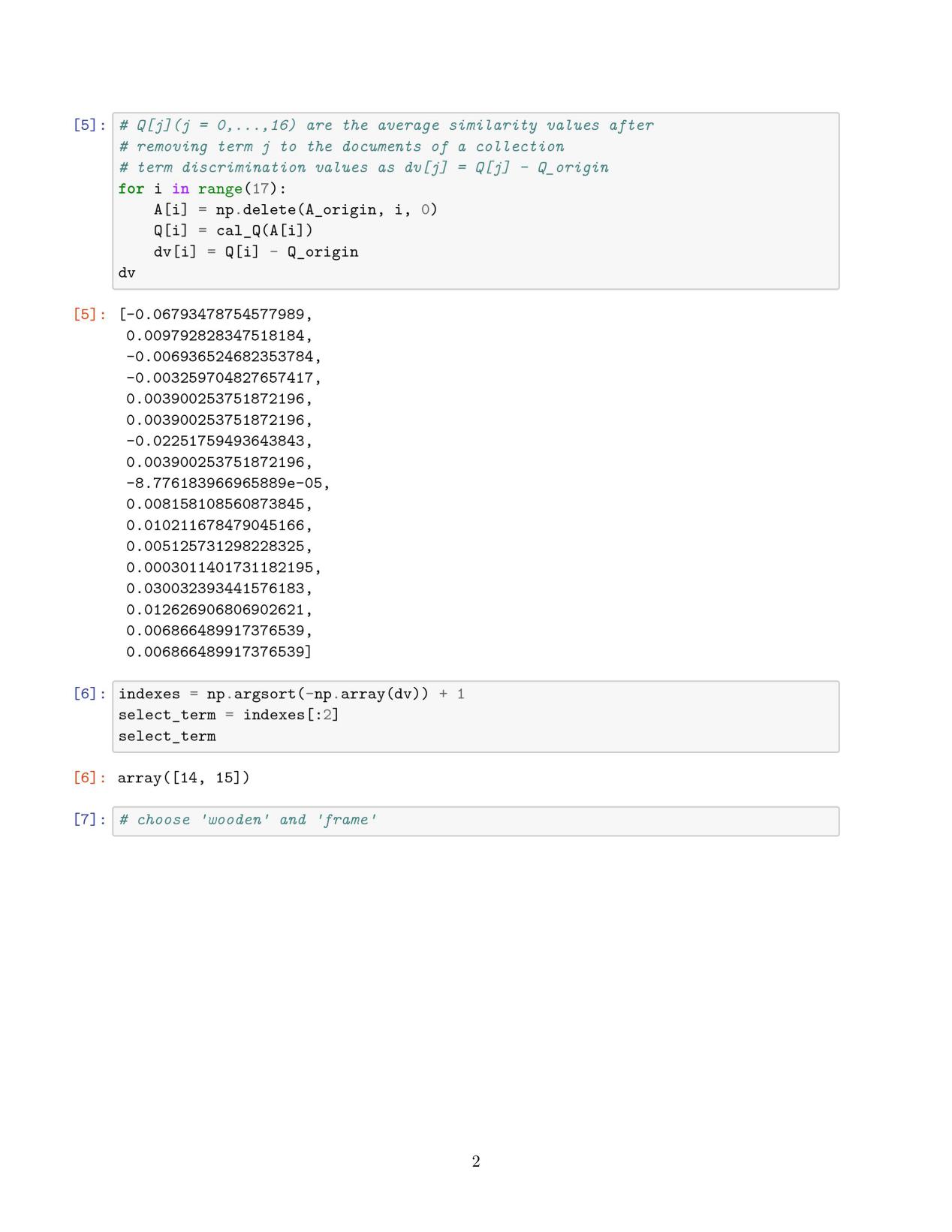
f)

‘success’ in D3 and ‘successful’ in D4 are synonymy, so put them in 1 term T10 ‘success’, this term occurs in D3 and D4.

‘windows’ has 2 meanings here: opening window in wall, a Microsoft system. In D1-D3, ‘windows’ means system, in D6-D8, ‘windows’ means window in the wall. The meaning of ‘windows’ in two queries is different, q1 has the same meaning with D1-D3, q2 has the same meaning with D6-D8. For q1 with LSI in this example, the top 3 documents in reduced or original vector space has similar result, D1 and D3 is correct, D6 and D7 are incorrect. For q2 with LSI, in reduced vector space it fails to get D8 for top 3 documents, but it gets D6 and D7, so Rank-2 approximation might reduce accuracy but increase the speed and reduce memory.

‘gates’ also have two different meanings, one is surname of Bill Gates, the other is the barrier.

**Q2:**



According to the result, choosing ‘wooden’ and ‘frames’ as index terms is better.

**Q3:**

Strategy 1: click > skip above

page 2 > page 1, page 8 > page 3, page 8 > page 4, page 8 > page 5, page 8 > page 6, page 8 > page 7

Strategy 2: last click > skip above

page 8 > page 1, page 8 > page 3, page 8 > page 4, page 8 > page 5, page 8 > page 6, page 8 > page 7

Strategy 3: click > earlier click

page 8 > page 2

Strategy 4: last click > skip previous

page 2 > page 1, page 8 > page 7

Strategy 5: click > no-click next

page 2 > page 3, page 8 > page 9

Above all:

page 8 > page 2 > (page 1, page 3)

page 8 > (page 4, page 5, page 6, page 7, page 9)

**Q4:**

1. Pages that cited from many places have a high score in Page Rank, when random selecting websites people have a high possibility to select such high score pages.
2. Pages that have a citation from important websites like Wikipedia have a high score in Page Rank, on the same way, they are worth looking at.
3. This pattern is similar to recommended system, people random surfer is actually based on Page Rank algorithm. People have small possibility to pick low Page Rank score pages even if they are also worth looking at.