Homework 6. Due Wednesday, Oct. 9.

1. **(3 pts)**

   (a) Prove the cyclic property of the trace:

   $$\mathsf{trace}(ABC) = \mathsf{trace}(BCA) = \mathsf{trace}(CAB) \tag{1}$$

   for all A, B , C such that their product is defined and is a square matrix.

   (b) Prove that

   $$\|A\|_F^2 = \sum_{i=1}^{d} \sigma_i^2. \tag{2}$$

   Hint: use the full SVD of A and *the cyclic property of trace.*

   (c) Prove that

   $$\|A+B\|_F^2 = \|A\|_F^2 + \|B\|_F^2 + 2\langle A, B\rangle_F, \tag{3}$$

   where $\langle A, B\rangle_F$ is the Frobenius inner product. The Frobenius inner product is defined as

   $$\langle A, B\rangle_F := \sum_{i,j} a_{ij}b_{ij} = \mathsf{trace}(A^\top B) = \mathsf{trace}(B^\top A). \tag{4}$$

2. (**5 pts**) Prove the Eckart-Young-Mirsky theorem for any Ky-Fan norm.

   Theorem 1. *Let $A = U\Sigma V^\top$ be an SVD of A and M be any matrix of the size of A such that $\mathsf{rank}(M) \le k$. Then*

   $$\|A - M\| \ge \|A - U_k\Sigma_k V_k^\top\| \quad \textit{for any Ky-Fan norm} \quad \|\cdot\|,$$

   where U_k and V_k consist of the first k columns of U and V, respectively, and $\Sigma_k = \mathsf{diag}\{\sigma_1, \dots, \sigma_k\}$.

   You can use Lemma 1 in Section 4.3 in `LinearAlgebra.pdf`. In this case, write it in your proof and explain every nontrivial equality in it. this will help you to understand this Lemma.

3. (**6 pts**) The dataset for this problem is downloaded from this webpage: TechTC – Technion Repository of Text Categorization. The particular file that I used is Preprocessed feature vectors: techtc300_preprocessed.zip (117,951,459 bytes; approx. 301Mb uncompressed). Its description is available here.

   I extracted two data files from it: `vectors.txt` and `words_idx.txt`. Each line of `words_idx.txt` is of the form

```
<word index><word>
```

A total of 18446 words. The file `vectors.txt` contains 278 lines. Lines 1, 3, 5, etc, i.e. all odd lines, contain a single number, the index of a document. Lines 2,4,6, etc, i.e. all even lines, contain information about the content of the document whose index is in the line above. The information is given as follows. The first number is 1 or -1, a label of the document. Label -1 attributes the document to Florida, while label 1 corresponds to Indiana. The rest of the numbers are the word indices and the counts of the corresponding words in the document. For example, line 2 starts with

```
-1 1   54 2   11 3   53 4   22 5   44 6
```

This means that document 1 belongs to the category -1. Furthermore, word 1 is encountered 54 times, word 2 is encountered 11 times, word 3 is encountered 53 times, word 4 is encountered 22 times, etc.

My code `DocsLeeSeung.ipynb` reads the data from files, creates an $N_{\text{words}} \times N_{\text{docs}}$ matrix A such that $A_{i,j} = 1$ if word i is present in document j, and computes its factorization $A \approx WH$ where $W \in \mathbb{R}_+^{N_{\text{words}} \times k}$ and $H \in \mathbb{R}_+^{k \times N_{\text{docs}}}$ using the Lee-Seung algorithm. I set $k = 10$.

You can choose to do this problem in Matlab or in Python.

(a) Implement the projected gradient descent to factorize $A \approx WH$ where $W \in \mathbb{R}_+^{N_{\text{words}} \times k}$ and $H \in \mathbb{R}_+^{k \times N_{\text{docs}}}$, $k = 10$. Plot the Frobenius norm of $R := A - WH$ versus the number of iteration. Determine (approximately) the number of iterations sufficient to make the residual $\|R\|_F$ stop changing visibly. Check if the eventual norm of the residual depends on the initial approximation. Check which words correspond to relatively high numbers in the columns of W and, looking at them, hypothesize what is the common theme of this set of documents.

(b) Do the same task for the HALS algorithm (Section 5.3.2 in `LinearAlgebra.pdf`).

(c) Find the best approximation A_{10} of A by a matrix of rank ≤ 10 using SVD. Compute $\|A - A_{10}\|_F$. Compare it with the residuals