max planck institut
informatik

You can discuss these problems with other students, but everybody must hand in their own answers. You can use computers etc. to perform the algebraic operations, but you must show the intermediate steps (and "computer said so" is never a valid answer). You can return either computer-typeset solutions by email (but no scanned or photographed solutions are accepted), or legibly hand-written or computer-typeset solutions personally to the lecture. Notice that the DL is strict. *Remember to write your name and matriculation number to every answer sheet!*

Problem 1 (Householder transformations). Consider the following matrix:

$$\boldsymbol{A} = \begin{pmatrix} 8 & 5 & 0 & 2 & 3 \\ 7 & 4 & 2 & 7 & 1 \\ 0 & 5 & 7 & 2 & 6 \\ 6 & 8 & 5 & 9 & 7 \\ 4 & 3 & 2 & 3 & 5 \\ 9 & 8 & 3 & 8 & 4 \\ 0 & 5 & 6 & 2 & 6 \end{pmatrix} .$$

Use your favourite program (e.g. R) to compute the Householder bidiagonalization of $\boldsymbol{A}$. For your answer, report β and $\boldsymbol{v}$ (required to compute the Householder matrix $\boldsymbol{P}$) and matrix $\boldsymbol{A}$ after every reflection.

Problem 2 (KLT). Let $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ and let $\boldsymbol{U\Sigma V}^T$ be the SVD of it. In the lecture it was said that, assuming $\boldsymbol{A}$ is normalized, its k-dimensional Karhunen–Lòeve transform can be computed as

$$\tilde{\boldsymbol{A}}_k = \boldsymbol{A}\boldsymbol{V}_k \; , \tag{2.1}$$

where $\boldsymbol{V}_k$ contains the first k columns of $\boldsymbol{V}_k$. Show that we can equivalently compute it as

$$\tilde{\boldsymbol{A}}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_k \; , \tag{2.2}$$

where $\boldsymbol{U}_k$ contains the first k columns of $\boldsymbol{U}$ and $\boldsymbol{\Sigma}_k$ is the principal k-by-k submatrix of $\boldsymbol{\Sigma}$.

Problem 3 (Computing eigendecompositions). *Fully solving this problem earns you one extra point, i.e. you can earn two points from this problem.*

Our goal is to compute the *eigendecomposition* of n-by-n symmetric, full-rank matrix $\boldsymbol{A}$. The eigendecomposition is of form $\boldsymbol{A} = \boldsymbol{Q\Lambda Q}^T$, where $\boldsymbol{Q}$ is orthogonal and $\boldsymbol{\Lambda}$ is diagonal (for simplicity, we assume $\boldsymbol{A}$ admits such decomposition). We do this by applying the Givens rotations to $\boldsymbol{A}$ (from both sides) in such a way that every rotation minimizes the Frobenius norm of the off-diagonal entries,

$$\operatorname{off}(\boldsymbol{A}) = \left(\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}} a_{ij}^2 \right)^2 . \tag{3.1}$$

Let $c = \cos(\theta)$ and $s = \sin(\theta)$ for some θ and consider the Givens rotations applied to the 2-by-2 submatrix (using R's notation) $\boldsymbol{A}[c(p,q), c(p,q)]$:

$$\begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}^T \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} . \tag{3.2}$$

Assume that θ has been selected so that the left-hand-side matrix $\left(\begin{smallmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{smallmatrix}\right)$ is diagonal. Notice that if we let $\boldsymbol{B} = \boldsymbol{G}(p,q,\theta)^T \boldsymbol{A} \boldsymbol{G}(p,q,\theta)$, then $\boldsymbol{B}$ differs from $\boldsymbol{A}$ exactly at that 2-by-2 submatrix. Use this fact, the fact that $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric, and the fact that the Frobenius norm is orthogonaly invariant to prove that

$$\operatorname{off}(\boldsymbol{B})^2 = \operatorname{off}(\boldsymbol{A})^2 - 2a_{pq}^2 . \tag{3.3}$$

Develop an algorithm that uses Givens rotations to find the eigendecomposition of $\boldsymbol{A}$. Use (3.3) to minimize the number of rotations you will need. What can you say about the speed of convergence of the algorithm (i.e. can you bound the off-diagonal Frobenius with a function of the number of iterations, the original off-diagonal Frobenius, and the size of the matrix)?

Problem 4 (Nonnegative rank). In the lectures it was claimed that for all $\boldsymbol{A} \in \mathbb{R}_+^{n \times m}$

$$\operatorname{rank}(\boldsymbol{A}) \le \operatorname{rank}_+(\boldsymbol{A}) \le \min\{n, m\} . \tag{4.1}$$

Prove the claim.

Problem 5 (Biconvexity of NMF). Prove that when one of the factor matrices is fixed, the NMF loss function

$$f(\boldsymbol{H}) = \frac{1}{2} \|\boldsymbol{A} - \boldsymbol{H}\boldsymbol{W}\|_F^2 \tag{5.1}$$

is convex.

Problem 6 (Multiplicative rules as gradient descent). Lee and Seung's multiplicative updates for NMF would update $\boldsymbol{H}$ as

$$\boldsymbol{H}_{ij} \leftarrow \boldsymbol{H}_{ij} \frac{(\boldsymbol{W}^T \boldsymbol{A})_{ij}}{(\boldsymbol{W}^T \boldsymbol{W} \boldsymbol{H})_{ij}} . \tag{6.1}$$

Show that this can be considered as a gradient descent approach with gradient updates

$$\boldsymbol{H}_{ij} \leftarrow \boldsymbol{H}_{ij} + \varepsilon_{ij} \Big((\boldsymbol{W}^T \boldsymbol{A})_{ij} - (\boldsymbol{W}^T \boldsymbol{W} \boldsymbol{H})_{ij} \Big) , \tag{6.2}$$

where the step size ε_{ij} is set separately for every element.

Hint: Find ε_{ij} such that you can transform (6.2) to (6.1).