Fall 2017: Numerical Methods I
Assignment 5 (due Nov 16, 2017)

1. **[Properties of the power method, 2+2+1pt]** Let $A \in \mathbb{R}^{n\times n}$ be symmetric with eigenvalues
$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|,$$
and denote the corresponding eigenvectors by $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n$. We consider the power method for finding the eigenvector corresponding to the dominant (i.e., largest in absolute value) eigenvalue. We consider the power method to find the eigenvector, i.e., given an initialization $\boldsymbol{x}^0$ not orthogonal to $\boldsymbol{\eta}_1$, we compute $\boldsymbol{x}^{k+1} := A\boldsymbol{x}^k$ for $k = 0, 1, \ldots$. The Rayleigh quotient, an approximation to the corresponding eigenvalue, is given by
$$r_k := \frac{(\boldsymbol{x}^k)^T A \boldsymbol{x}^k}{(\boldsymbol{x}^k)^T \boldsymbol{x}^k} \tag{1}$$

   (a) Show that
   $$r_k = \lambda_1 \left[1 + \mathcal{O}\left((\lambda_2/\lambda_1)^{2k}\right)\right].$$

   (b) Consider a symmetric matrix $A \in \mathbb{R}^{5\times 5}$
   $$A = A^T = \begin{bmatrix} -9 & * & * & * & * \\ * & 0 & * & * & * \\ * & * & 1 & * & * \\ * & * & * & 4 & * \\ * & * & * & * & 21 \end{bmatrix},$$
   where $*$ represents elements of absolute values $\leq 1/4$. Suppose the power method is applied with A and the initial vector $\boldsymbol{x}^0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$. Show that $\boldsymbol{x}^0$ is an "appropriate" initial vector, i.e., that the sequence $\boldsymbol{y}^k = \boldsymbol{x}^k / \|\boldsymbol{x}^k\|$ does converge toward the eigenvector belonging to the dominant eigenvalue of A.[1]

   (c) Estimate how many correct digits r_{k+5} gains compared to r_k, with r_k as defined in (1).

The next problems discuss eigenvalues and eigenvectors of stochastic matrices. Stochastic matrices are very useful for ranking sport teams or the importance of web pages. In fact, Google's famous PageRank algorithm[2] is based on eigenvectors of stochastic matrices and is (part of) the reason why Google's search was/is superior to other search engines. The basic algorithm is described in a paper[3] from 1998 by Sergey Brin and Larry Page, the founders of Google, who were

[1] This and the problem below require basic estimation of the eigenvalues of A, for which you can use Gershgorin circle theorem. If you are not familiar with this theorem, that allows to estimate eigenvalues of a matrix using its entries, take a look at any linear algebra textbook, or the Wikipedia page.

[2] The PageRank algorithm is named after Larry Page, one of the founders of Google.

[3] http://infolab.stanford.edu/pub/papers/google.pdf

at that point both students at Stanford. This original paper has been cited more than $10,000$ times, and Google is worth more than 500,000,000,000 USD today[4].

The basic idea is to give each web page a non-negative score describing its importance. This score is derived from links pointing to that page from other web pages. Links from more important web pages are more valuable as the score of each page is distributed amongst the pages it links to. Let us consider an example with 4 web pages, where page 1 links to all other pages, page 2 links to pages 3 and 4, page 3 links to page 1, and page 4 links to pages 1 and 3. Denoting the scores for the ith page by x_i, this mini-web has the following conditions for its scores:

$$x_1 = x_3/1 + x_4/2,\ x_2 = x_1/3,\ x_3 = x_1/3 + x_2/2 + x_4/2,\ x_4 = x_1/3 + x_2/2,$$

or, equivalently, the eigenvalue equation $L\boldsymbol{x} = \boldsymbol{x}$ (i.e., the eigenvalue is 1), where $\boldsymbol{x} \in \mathbb{R}^4$, and

$$L = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}. \tag{2}$$

Thus, the solution of the eigenvalue problem $L\boldsymbol{x} = \boldsymbol{x}$ provides the importance score for our mini-web. The matrix L has a special structure, it is a column-stochastic matrix. In general, a column-stochastic matrix $L \in \mathbb{R}^{n\times n}$ is a matrix with all non-negative entries, such that each column sum of L is equal to 1, i.e., $\sum_{j=1}^{n} l_{jk} = 1$ for all $k = 1, \ldots, n$. In the following problems, we study properties of these matrices.

2. **[Stochastic matrices, 1+1pt]** Let $L \in \mathbb{R}^{n\times n}$ be a column-stochastic matrix.

   (a) Show that the column vector $\boldsymbol{e}$ of all ones is an eigenvector of L^T. What's the corresponding eigenvalue?

   (b) Argue that L has an eigenvector corresponding to the eigenvalue 1.

3. **[Positive stochastic matrices, 2+2+1pt]** Assume that all entries of a column-stochastic matrix L are positive. We will prove that eigenvectors $\boldsymbol{v}$ corresponding to the eigenvalue 1 of L either have all positive or all negative values.[5] We prove this by contradiction.

   (a) Suppose an eigenvalue $\boldsymbol{v} \in \mathbb{R}^n$ corresponding to the eigenvalue 1 has negative and positive components. Show that for every i we have

   $$|v_i| < \sum_{j=1}^{n} l_{ij}|v_j|.$$

   Where in this argument is the positivity of L used?

   (b) By summing over all i and using properties of L, show that this results in a contradiction.

[4] A good read is also the 2006 SIREV paper *The 25,000,000,000 eigenvector. The linear algebra behind Google* by Kurt Bryan and Tanya Leise. It's easy to find—just google it!

[5] These entries are used to rank the pages–the higher, the more important a page is.

(c) Thus, either all $v_i \geq 0$ or all $v_i \leq 0$. Argue that either all $v_i > 0$ or all $v_i < 0$.

4. **[Positive stochastic matrices, eigenspace dimension, 2+2pt]** Under the same assumptions as in the previous problem, we argue that the dimension of the eigenspace corresponding to the eigenvalue 1 is one[6].

   (a) Assume given two linearly independent vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. Then there exists a vector $\boldsymbol{w} \in \text{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ such that $\boldsymbol{w}$ has both, positive and negative components. *Hint:* Try to find a nonzero vector $\boldsymbol{w}$ such that $\sum_j w_j = 0$.

   (b) Use the above result to argue that the eigenspace corresponding to the eigenvalue 1 of a column-stochastic matrix L is of dimension 1.

5. **[Application of power method, 2+2pt]** In many applications[7], the column-stochastic matrix L is very large and solving an eigenvalue problem with L is difficult. A simple possibility to compute the largest eigenvalue and the corresponding eigenvector is the power method. Let us assume a column-stochastic matrix L and assume that the eigenvalue 1 has an algebraic multiplicity of 1[8]. We assume that the starting vector $\boldsymbol{v}^0$ has a nonzero component in the direction of the eigenvector for the eigenvalue 1.

   (a) Generalize the result on the convergence of the power method for symmetric matrices to matrices that are diagonalizable over $\mathbb{C}$.

   (b) Use power iterations to compute (in MATLAB or Python) the dominant eigenvalue for the matrix L defined in (2).

6. **[Stability of eigenvalues for non-symmetric matrices, 2+1pt]** Let A_ε be a family of matrices given by

$$\begin{bmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \varepsilon & & & \lambda \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

   Obviously, A_0 has λ as its only eigenvalue with multiplicity n.

   (a) Show that for $\varepsilon > 0$, A_ε has n different eigenvalues given by

$$\lambda_{\varepsilon,k} = \lambda + \varepsilon^{1/n}\exp(2\pi i k/n), \quad k = 0, \ldots, n-1,$$

   and thus that $|\lambda - \lambda_{\varepsilon,k}| = \varepsilon^{1/n}$.

   (b) Based on the above result, what accuracy can be expected for the eigenvalues of A_0 when the machine epsilon is 10^{-16}?

[6] Thus, the importance score vector is unique.

[7] The number of internet web pages is around 15 billion.

[8] For positive column-stochastic matrices, it can be shown that the algebraic multiplicity of the eigenvalue 1 is one.

7. **[Finding all roots of a polynomial, 2+1pt]** An efficient way to find individual roots of a polynomial is to use Newton's method. However, as we have seen, Newton's method requires an initialization close to the root one wants to find, and it can be difficult to find *all* roots of a polynomial. Luckily, one can use the relation between eigenvalues and polynomial roots to find all roots of a give polynomial. Let us consider a polynomial of degree n with leading coefficient 1:

$$p(x) = a_0 + a_1 x + \ldots + a_{n-1}x^{n-1} + x^n \quad \text{with } a_i \in \mathbb{R}.$$

   (a) Show that $p(x)$ is the characteristic polynomial of the matrix (sometimes called a companion matrix for p)

$$A_p := \begin{bmatrix} 0 & & & -a_0 \\ 1 & & & -a_1 \\ & \ddots & & \vdots \\ & & 1 & -a_{n-1} \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

   Thus, the roots of $p(x)$ can be computed as the eigenvalues of A_p using the QR algorithm (as implemented, e.g., in MATLAB's `eig` function).

   (b) Let us consider Wilkinson's polynomial $p_w(x)$ of order 25, i.e., a polynomial with the roots $1, 2, \ldots, 25$:

$$p_w(x) = (x-1)\cdot(x-2)\cdot\ldots\cdot(x-25).$$

   The corresponding coefficients can be found using the `poly()` function. Use these coefficients in the matrix A_p to find the original roots again as eigenvalues of A_p. Use your own implementation of the QR algorithm for that purpose, and hand in your code.[9]

8. **[Properties of the SVD, 1+2+1+1+1pt]** Consider the singular value decomposition of $A \in \mathbb{C}^{m\times n}$, i.e.,

$$A = U\Sigma V^*, \tag{3}$$

   where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary, $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal and V^* denotes the adjoint of V (i.e., the complex conjugate matrix). The diagonal entries in Σ are real and nonnegative, and we assume them to be ordered in non-increasing order, i.e., $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$, $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_p \geq 0$, where $p = \min(m, n)$. If A is a real matrix, it has a real SVD, i.e., U, V can be chosen as real orthonormal matrices. The singular values are uniquely determined. If all the singular values are different, the columns of U and V (which are called the left and right singular vectors), are uniquely determined up to multiplication with scalars of absolute value 1.

[9] If you use MATLAB, you can compare the method developed here with the build-in method (called `roots()`) for finding the roots of a polynomial. For many MATLAB functions that do not use external libraries, you can see how they are implemented by typing `edit name_of_function`. Doing that for the `roots` function will show you how MATLAB computes roots of polynomials—basically, it is using the method outlined above.

(a) Show that the columns of U are the eigenvectors of $AA^* \in \mathbb{C}^{m\times m}$, and that the columns of V are the eigenvectors of $A^*A \in \mathbb{C}^{n\times n}$. What are the corresponding eigenvales?

(b) Use the previous two properties to compute a (real) SVD for the matrix

$$A = \begin{pmatrix} 1 & 0 & 3 \\ -3 & 0 & -1 \end{pmatrix}.$$

(c) Let $m = n$ and A be invertible. Using the SVD of A, give an expression for A^{-1}.

(d) Let $m < n$, and rank$(A) = m$, i.e., A has full rank. Use the SVD of A to compute the pseudoinverse $A^\dagger := A^*(AA^*)^{-1}$.

(e) The Frobenius norm[10] of a matrix $A \in \mathbb{R}^{m\times n}$ is given by

$$\|A\|_F = \sqrt{\sum\nolimits_{i,j} a_{ij}^2}.$$

Given the SVD of A, compute $\|A\|_F$.

[10]Note that we have discussed matrix norms, that were induced by norms of vectors. The Frobenius norm is not naturally induced by a vector norm, since, for instance, an induced norm of the identity matrix should be 1.