Technische Universität Bergakademie Freiberg
Institut für Numerische Mathematik und Optimierung
Prof. Dr. O. Rheinbach
Dipl. Math. S. Köhler

Wintersemester 2021/22
04. 01. 2022

Numerische Mathematik

Introduction to High Performance Computing and Optimization

Exercise 1

Google-Page-Rank We denote the importance (page-rank) of the i-th webpage by r_i, $i \in 1, \ldots, n$. We define the entries of the *link matrix* $L \in \mathbb{R}^{n \times n}$ by

$$l_{ij} = \begin{cases} 1, & \text{if a link from the } j\text{-th webpage to the } i\text{-th webpage exists} \\ 0, & \text{otherwise.} \end{cases}$$

For the *page-rank vector* $\mathbf{r} = (r_1, \ldots, r_n)^T$ the following condition are supposed to hold:

1. $r_i = \sum_{j=1}^{n} \frac{1}{n_j} l_{ij} r_j$
2. $\sum_{i=1}^{n} r_i = 1$
3. $r_i \geq 0 \;\; \forall i = 1, \ldots, n,$

where n_i denotes the number of links from the i-th webpage.
The matrix L represents the graph of the web. To be sure that a user of the web does not get stuck on a webpage which has no link to another webpage, the matrix L is modified as following: We define the matrix Q by $Q_{ij} = \frac{1}{n_j} l_{ij}$ and denote the j-th column of Q by Q_j. Furthermore, we define the vector $\mathbf{e} = (1, \ldots, 1)^T$, the vector $\mathbf{d}$ by

$$d_j = \begin{cases} 1 & Q_j = 0 \\ 0 & \text{otherwise,} \end{cases}$$

and the matrix $P = Q + \frac{1}{n}\mathbf{e}\mathbf{d}^T$. In P the columns of Q with zero entries are replaced by the vector $\frac{1}{n}\mathbf{e}$. This models the case that for a webpage with no links to another one, the user visits a random webpage with uniform distribution. P is a *left stochastic matrix*, which means the sum of each column is one. The eigenvalue problem

$$P\mathbf{r} = \lambda_{\max}\mathbf{r}$$

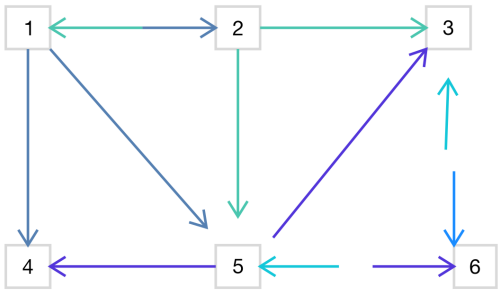
can be solved by the *Power Method*.

for *k = 1,2,... until convergence* **do**
  $\mathbf{q}^{(k)} = P\mathbf{r}^{(k-1)}$
  $\mathbf{r}^{(k)} = \mathbf{q}^{(k)} / \|\mathbf{q}^{(k)}\|_1$
end

The eigenvalue $\lambda_{\max}$ can be approximated by the *Rayleigh-Quotient*

$$\theta_k = \frac{\langle r_k, Ar_k \rangle}{\|r_k\|_2^2}.$$

Programming Project Implement a parallel dense matrix and vector format using MPI and only the C standard library or C++ standard template Library. Implement the power iteration in parallel to compute an approximation to the largest eigenvalue of the matrix and the corresponding eigenvector of a left stochastic matrix P induced by a link matrix L.
Test your implementation with the link matrix corresponding to the web



Implement also random webs with $n = 100, 1000, 10\,000$ webpages (**attention:** make sure your computation is reproducible by the use of seeds).
Perform strong and weak scaling results with 30 iteration of the power iteration.
Deadline: 21th March 2022.
Remark: Please send your project as an e-mail to stephan.koehler@math.tu-freiberg with the subject "HPC programming project 2021". Provide a compilable file (with mpicxx or mpicc) or an archive / directory with all your required files including a make or cmake file to us.