Zhao-Li Shen
**An efficient preconditioner based on a partial low-rank
factorization for PageRank problems**

Institute of Mathematics and Computing Science
University of Groningen
Nijenborgh 9
P O Box 407
9700 AK Groningen
The Netherlands
`szlxiaoyao@163.com`
Bruno Carpentieri

The PageRank problem used by Google for ranking web pages can be mathematically stated as the following linear system

$$(I - \alpha P)x = v, \tag{1}$$

where $0 < \alpha < 1$ is the damping factor, $P = \{p_{ij}\}$ is the transition probability matrix representing the web structure where entry p_{ij} is the probability of transition from page j to page i, v is a personalization probability vector, and finally x is the unknown PageRank vector. With the rapid development of the Internet, the size of system (1) is reaching order of millions or even billions unknowns. For solving problems of this size, iterative methods are mandatory to use. When the damping factor approaches 1, i.e., the model reflects more closely the web structure, iterative solvers tend to suffer from slow convergence due to the unfavourable spectral distribution and the high condition number of system (1) [1, 2]. Thus preconditioning is necessary. Some preconditioning techniques that have proved successful for general linear systems have been applied for solving the PageRank problem. However, general purpose methods do not exploit well any special property or structure of PageRank matrices.

In this study we exploit the similarity of pattern structures and the special distribution of the numerical values on these patterns in the transition matrix P during the construction of preconditioner. This similarity often appears due to the similar link distributions of pages from same hosts. Note that the problem matrix $A = I - \alpha P$ maintains this similarity structure, except at the diagonal indexes. We propose an economic compression algorithm that computes similarity patterns in PageRank matrices and compresses matrix blocks of A by a low-rank matrix factorization, saving considerably storage and computational cost. Accordingly, system matrix $A \in \mathbb{R}^{n \times n}$ is reformulated as

$$A = D + FH, \tag{2}$$

where $D \in \mathbb{R}^{n \times n}$ is more sparse than A, $F \in \mathbb{N}^{n \times m}$ and $H \in \mathbb{R}^{m \times n}$ are low-rank factors. We demonstrate that both the matrix D and the matrix

1

$I + HD^{-1}F$ are unconditionally invertible for PageRank problems. Then we propose a preconditioner M based on the Woodbury formula of the form

$$M \approx A^{-1} = D^{-1}(I - F(I + HD^{-1}F)^{-1}HD^{-1}) \tag{3}$$

that exploits effectively the computed structure (2).

Numerical experiments show that our algorithm can solve difficult PageRank problems, i.e. for $\alpha = 0.99, 0.999$, with over one million unknows efficiently, also in comparison to state-of-the-art solvers, and may yield a high compression rate sometimes up to 40%. In this talk we present the theoretical background of the proposed method and we illustrate numerical experiments of realistic PageRank simulations that support our findings. Further improvement is under investigation.

Bibliography

[1] S.D. Kamvar and H. Haveliwala, The Condition Number of the PageRank Problem, Technical report, Stanford University, Stanford, CA, 2003.

[2] G.M. Del Corso, A. Gulli, F. Romani, Comparison of Krylov subspace methods on the PageRank problem, *J. Comput. Appl. Math.*, 210 (2007), pp. 159-166.