



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

FINAL REPORT (MILESTONE DATE 9/30/11) FOR SUBCONTRACT NO. B594099 NUMERICAL METHODS FOR LARGE-SCALE DATA FACTORIZATION

H. De Sterck

October 21, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

**FINAL REPORT (MILESTONE DATE 9/30/11)
FOR SUBCONTRACT NO. B594099
NUMERICAL METHODS FOR LARGE-SCALE DATA FACTORIZATION**

Description of work performed for the subcontract

The following work has been performed by PI Hans De Sterck and graduate student Manda Winlaw for the required tasks 1-5 (as listed in the Statement of Work):

1. Graduate student Manda Winlaw has visited LLNL January 31-March 11, 2011 and May 23 – August 19, 2010, working with Van Henson and Mike O’Hara on non-negative matrix factorizations (NMF). She has investigated the dense subgraph clustering algorithm from “Finding Dense Subgraphs for Sparse Undirected, Directed, and Bipartite Graphs“ by Chen and Saad, testing this method on several term-document matrices and adapting it to cluster based on the rank of the subgraphs instead of the density. Manda Winlaw was awarded a first prize in the annual LLNL summer student poster competition for a poster on her NMF research.
2. PI Hans De Sterck has developed a new adaptive algebraic multigrid algorithm for computing a few dominant or minimal singular triplets of sparse rectangular matrices. This work builds on adaptive algebraic multigrid methods that were further developed by the PI and collaborators (including Sanders and Henson) for Markov chains. The method also combines and extends existing multigrid algorithms for the symmetric eigenproblem. The PI has visited LLNL February 22-25, 2011, and has given a CASC seminar “Algebraic Multigrid for the Singular Value Problem” on this work on February 23, 2011. During his visit, he has discussed this work and related topics with Van Henson, Geoffrey Sanders, Panayot Vassilevski, and others. He has tested the algorithm on PDE matrices and on a term-document matrix, with promising initial results.
3. Manda Winlaw has also started to work, with O’Hara, on estimating probability distributions over undirected graph edges. The goal is to estimate probabilistic models from sets of undirected graph edges for the purpose of prediction, anomaly detection and support to supervised learning. Graduate student Manda Winlaw is writing a paper on the results obtained with O’Hara which will be submitted some time later in 2011 to a data mining conference.
4. PI Hans De Sterck has developed a new optimization algorithm for canonical tensor approximation, formulating an extension of the nonlinear GMRES method to optimization problems. Numerical results for tensors with up to 8 modes show that this new method is efficient for sparse and dense tensors. He has written a paper on this which has been submitted to the SIAM Journal on Scientific Computing.
5. PI Hans De Sterck has further developed his new optimization algorithm for canonical tensor approximation, formulating an extension in terms of steepest-descent preconditioning, which makes the approach generally applicable for nonlinear optimization. He has written a paper on this extension which has been submitted to Numerical Linear Algebra with Applications.

Publications enabled by the subcontract:

1. Hans De Sterck, 'A Self-learning Algebraic Multigrid Method for Extremal Singular Triplets and Eigenpairs', submitted to SIAM J. Scientific Computing, February 2011 (arXiv:1102.0919)
2. Hans De Sterck, 'A Nonlinear GMRES Optimization Algorithm for Canonical Tensor Decomposition', submitted to SIAM J. Scientific Computing, May 2011 (arXiv:1105.5331)
3. Hans De Sterck, 'Steepest Descent Preconditioning for Nonlinear GMRES Optimization', submitted to Numerical Linear Algebra with Applications, June 2011 (arXiv:1106.4426)
4. Hans De Sterck, Killian Miller, Eran Treister and Irad Yavneh, 'Fast multilevel methods for Markov chains', Numerical Linear Algebra with Applications, accepted, 2011.
5. Hans De Sterck, Killian Miller, and Geoffrey Sanders, 'One-norm Top-level Acceleration of Algebraic Multigrid for Markov Chains via the Ellipsoid Method', Computing and Visualization in Science, accepted, 2011.
6. Hans De Sterck, Van Emden Henson, and Geoffrey Sanders, 'Multilevel Aggregation Methods for Small-World Graphs with Application to Random-Walk Ranking', Computing and Informatics 30, 1001-1022, 2011.
7. Graduate student Manda Winlaw is writing a paper on the recent results for estimating probability distributions over undirected graph edges with O'Hara which will be submitted in 2011 to a data mining conference.

Presentations enabled by the subcontract:

1. Hans De Sterck, Max Planck Institute for Informatics seminar, Saarbruecken, Germany, 16 August 2011. 'A residual-minimizing nonlinear optimization method applied to tensor approximation'.
2. Hans De Sterck, Workshop on Tensor Approximation in High Dimension, Bonn, Germany, 2 August 2011. 'Extending GMRES to Nonlinear Optimization: Application to Tensor Approximation'.
3. Hans De Sterck, Department of Computer Science seminar, Friedrich-Alexander-Universitaet Erlangen-Nuernberg, Germany, 7 July 2011. 'Extending GMRES to Nonlinear Optimization: Application to Tensor Approximation'.
4. Hans De Sterck, Department of Mathematics seminar, University of Trier, Germany, 5 July 2011. 'Extending GMRES to Nonlinear Optimization: Application to Tensor Approximation'.
5. Hans De Sterck, Applied Computer Science Group seminar, Department of Mathematics, University of Wuppertal, Germany, May 2011. 'Algebraic Multigrid for the Singular Value Problem'.
6. Hans De Sterck, 2011 SIAM Conference on Computational Science and Engineering Reno, NV, USA, March 1, 2011. 'Multilevel Aggregation of Small-World Graphs with Application to Random-Walk Ranking'.
7. Hans De Sterck, Center for Applied Scientific Computing Seminar. Lawrence Livermore National Laboratory, California, February 2011. 'Algebraic Multigrid for the Singular Value Problem'.