Sparse Days

June 25-26th, 2012

CERFACS, Toulouse, France

Monday, June 25th 2012

9.30 - 10.00 Registration and welcome coffee

Session I

10.00 - 10.30 *Finding weighted matchings for singular symmetric matrices*
J. A. Scott (Rutherford Appleton Laboratory (RAL), UK)

10.30 - 11.00 *Multithreaded algorithms for maximum matching in bipartite graphs*
E. G. Boman (Sandia National Laboratories, USA)

11.00 - 11.15 Coffee break

11.15 - 11.45 *Partitioning, ordering, and load balancing in a hierarchically parallel hybrid* linear solver
X. S. Li (Lawrence Berkeley National Laboratory, USA)

11.45 - 12.15 *Towards the implementation of sparse, direct methods on GPU-based systems*
A. Decollas (INRIA Bordeaux Sud-Ouest, France) and
F. Lopez (IRIT, Toulouse, France)

Lunch

Session II

14.00 - 14.30 *Selected inversion with application to electronic structure calculation*
C. Yang (Lawrence Berkeley National Laboratory, USA)

14.30 - 15.00 *Randomized sparse direct solvers and applications*
J. Xia (Department of Mathematics, Purdue University, USA)

15.00 - 15.30 *Minimizing communication in sparse matrix-vector multiplication using a novel representation*
R. Kannan (School of Mathematics, University of Manchester, Arup/Oasys Limited, UK)

15.30 - 16.00 Coffee break

Session III

16.00 - 16.30 *Using overlapping and filtering techniques for parallel preconditioners*
L. Qu (Laboratoire de Recherche en Informatique, Université Paris-Sud XI, INRIA Saclay - Ile de France, France)

16.30 - 17.00 *FEAST eigenvalue algorithm and solver: review and perspectives*
E. Polizzi (Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, USA)

19.30 Banquet at the restaurant Le Pôvre Yves in Toulouse.

Tuesday, June 26th 2012

Session IV

9.30 - 10.00 *Vector orthogonalization algorithms*
C. Paige (Computer Science, McGill University, Montreal, Canada)

10.00 - 10.30 *Sparse linear algebra issues arising in the analysis of complex networks*
M. Benzi (Department of Mathematics and Computer Science
Emory University, Atlanta, USA)

10.30 - 11.00 Coffee break

Session V

11.00 - 11.30 *Iterative methods for symmetric quasi-definite linear systems*
M. Arioli (Rutherford Appleton Laboratory (RAL), UK)

11.30 - 12.00 *Spectral information and GMRES convergence*
G. Meurant (France)

Lunch

Session VI

14.00 - 14.30 *The bitter truth about interior-point methods*
D. Orban (Ecole Polytechnique de Montreal, Canada)

14.30 - 15.00 *Preconditioning of linear least-squares problems*
M. Tůma (Institute of Computer Science, Academy of Sciences
of the Czech Republic)

15.00 - 15.30 *New perspectives on computational mathematics: designing probabilistic-based algorithms suited for massively parallel computers*
A. Rodríguez-Rozas (Instituto Superior Técnico, Lisboa, Portugal)

Closure

June 25th - Session I

Finding weighted matchings for singular symmetric matrices

J. Scott (Rutherford Appleton Laboratory (RAL), UK).

The use of weighted matchings for the scaling of unsymmetric matrices is well-known and widely used. Duff and Pralet adapted the work of Duff and Koster on unsymmetric matrices to the symmetric case and their approach has been successfully used for computing scalings and orderings for sparse symmetric systems. However, in the singular case, the Duff and Pralet weighted matching is generally suboptimal and the computed ordering may still need to be substantially modified during the numerical factorization to maintain numerical stability. In this talk, modifications to the Duff and Koster algorithm for structurally singular symmetric matrices are proposed and their effectiveness illustrated using problems arising from practical applications.

This is joint work with Jonathan Hogg.

Multithreaded algorithms for maximum matching in bipartite graphs

E. G. Boman (Sandia National Laboratory, USA).

We design, implement, and evaluate algorithms for maximum cardinality matching in bipartite graphs. Such algorithms are used to obtain a zero-free diagonal in sparse matrices and to compute the block triangular form. Our parallel algorithms use shared-memory and we compare breadth-first-search and depth-first-search augmenting-path strategies on several architectures, including conventional multicore machines and a massively multithreaded machine (Cray XMT). We show that substantial speedup can be obtained over a serial implementation. We further observe that the best algorithm depends on the architecture.

This is joint work with Ariful Azad, Mahantesh Halapanavar, Siva Rajamanickam, Arif Khan, and Alex Pothen.

Partitioning, ordering, and load balancing in a hierarchically parallel hybrid linear solver

X.S. Li (Lawrence Berkeley National Laboratory, USA).

PDSLin is a general-purpose algebraic parallel hybrid (direct/iterative) linear solver based on the Schur complement method. The most challenging step of the solver is the computation of a preconditioner based on an approximate global Schur complement. We investigate two combinatorial problems to enhance PDSLin's performance at this step. The first is a multi-constraint partitioning problem to balance the workload while computing the preconditioner in parallel. For this, we describe and evaluate a number of graph and hypergraph partitioning algorithms to satisfy our particular objective and constraints. The second problem is to reorder the sparse right-hand side vectors to improve the data access locality during the parallel solution of a sparse triangular system with multiple right-hand sides. This is needed to eliminate the unknowns associated with the interface in PDSLin. We study two reordering techniques: one based on a postordering of the elimination tree and the other based on a hypergraph partitioning. To demonstrate the effect of these techniques on the performance of PDSLin, we present the numerical results of solving large-scale linear systems arising from numerical simulations of modeling accelerator cavities and of modeling fusion devices.

This is joint work with I. Yamazaki, F.-H. Rouet and B. Ucar.

Towards the implementation of sparse, direct methods on GPU-based systems

A. Decollas (INRIA Bordeaux Sud-Ouest, France) and F. Lopez (ENSEEIHT-IRIT, France).

Direct solvers for sparse linear systems are extremely rich in floating point computations and their intense use of Level-3 BLAS routines makes them ideal candidates for GPU acceleration. Nonetheless, these methods involve extremely complex computational patterns and heterogeneous kernel operations which may be both memory and computation intensive. Therefore, in order to achieve the porting of direct sparse methods on moderns multicore accelerated computers, programming and execution models which are capable of handling this double level of heterogeneity (i.e., architectural and algorithmic) are necessary. This talk presents preliminary results towards this goal. Specifically, it will present an implementation of the QR multifrontal method as well as of some elementary operations commonly occurring in sparse, direct methods based on the StarPU runtime system and programming model.

June 25th - Session II

Selected inversion with application to electronic structure calculation

C. Yang (Computational Research Division, Lawrence Berkeley National Laboratory, USA).

One of the most time consuming tasks in Kohn-Sham density functional theory based electronic structure calculation is the evaluation of the electron density. This calculation is commonly done by computing the lowest k eigenvalues and eigenvectors of a Kohn-Sham Hamiltonian. The complexity of this approach scales cubically with respect to the number of atoms. An alternative way to evaluate the electron density is to express it as the diagonal of the Fermi-Dirac function evaluated at a Kohn-Sham Hamiltonian. A rational approximation to the Fermi-Dirac function based on an accurate pole expansion technique can be used in practice. However, this approach requires one to compute the diagonal of the inverses of a few shifted Kohn-Sham Hamiltonians. We will discuss how this type of calculation can be performed by using sparse matrix techniques that do not require computing the full inverse of a sparse Hamiltonian. The complexity of this approach scales at most quadratically with respect to the number of atoms. Numerical examples will be given to demonstrate the efficiency and accuracy of this approach.

Randomized sparse direct solvers and applications

J. Xia (Department of Mathematics, Purdue University, USA).

We propose some new structured direct solvers for large linear systems, using randomization and other techniques. The work involves new flexible methods to exploit structures in large matrix computations. Our randomized structured techniques provide both higher efficiency and better applicability than some existing structured methods. New efficient ways are developed to conveniently perform various complex operations which are difficult in standard rank-structured solvers. Our new solvers have nearly linear complexity for various types of large linear systems.

We also study the following issues:

- Develop matrix-free structured solvers.
- Update a structured factorization when few matrix entries change.
- Relaxed rank requirements in structured solvers. We show the feasibility of our methods for solving various difficult problems, especially high dimensional ones.
- Develop effective preconditioners for problems without significant rank structures. We analyze the criterion for compressing off-diagonal blocks so as to achieve nearly optimal effectiveness and efficiency in our preconditioner.

Applications of the techniques to selected sparse inversion, eigenvalue problems, and PDE solutions will be shown.

Minimizing communication in sparse matrix-vector multiplication using a novel representation

R. Kannan (School of Mathematics, University of Manchester and Arup/Oasys Limited, UK).

We consider the problem of multiplying sparse matrix $A \in \mathbb{R}^{n\times n}$ with a dense matrix $X \in \mathbb{R}^{n\times \ell}$ with the aim of increasing register reuse. A new block compressed format that is shown to have less loads compared with compressed sparse row and block compressed sparse row formats, is proposed. The format stores blocks as bitmaps and is less sensitive to the non–zero structure (in comparison with blocked formats). A C++ based algorithm that decodes blocks in $\mathcal{O}(nz_b)$ time, where nz_b is the number of non–zeros in a block, is developed. Performance comparisons with Intel MKL sparse BLAS and with standard CSR and BSR implementations are presented.

June 25th - Session III

Using overlapping and filtering techniques for parallel preconditioners

L. Qu (Laboratoire de Recherche en Informatique, INRIA Saclay, France).

In this talk we discuss a parallel preconditioner for solving large sparse linear systems of equations. The preconditioner is based on nested dissection ordering and uses filtering techniques to alleviate the effect of low frequency modes on the convergence of the iterative method. We also discuss overlapping techniques to further increase the numerical efficiency of the preconditioner.

FEAST eigenvalue algorithm and solver: review and perspectives

E. Polizzi (Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, USA).

The eigenvalue problem is a central topic in Science and Engineering arising from a wide range of applications and posing major numerical challenges. The FEAST solver package is a new free high-performance numerical library for solving the standard or generalized eigenvalue problem, and obtaining all the eigenvalues and eigenvectors within a given search interval. It is based on an innovative fast and stable numerical algorithm presented in [1] – named the FEAST algorithm – which deviates fundamentally from traditional eigenvalue algorithms. The FEAST algorithm offers many important capabilities for achieving high performance, robustness, accuracy, and scalability on parallel architectures. The current version v2.0 of the FEAST package [2, 3] focuses on solving the symmetric eigenvalue problems (real symmetric or complex Hermitian systems) on both shared-memory and distributed-memory architectures (i.e contains both FEAST-SMP and FEAST-MPI libraries). The FEAST-MPI library, in particular, includes a three level parallelism capability (MPI-MPI-OpenMP).

After reviewing the main properties and performance scalability of the FEAST algorithm and solver, we propose to broaden the capabilities of FEAST (beyond the linear symmetric case) for solving the non-symmetric and the non-linear problem. For the electronic structure problem, in particular, we demonstrate the efficiency of an alternative approach deriving from a generalization of the FEAST algorithm for solving directly the full coupled non-linear Density-Functional Theory/Kohn-Sham equations.

References

[1] E. Polizzi, *Density-Matrix-Based Algorithms for Solving Eigenvalue Problems*, Phys. Rev. B. Vol. 79, 115112 (2009).

[2] E. Polizzi, *FEAST Solver v2.0 User's Guide*, arxiv.org/abs/1203.4031 (2012).

[3] http://www.ecs.umass.edu/∼polizzi/feast

June 26th - Session IV

Vector orthogonalization algorithms

Chris Paige (Computer Science, McGill University, Montreal, Canada).

Many iterative algorithms for large sparse matrix problems are based on orthogonality (or A-orthogonality, bi-orthogonality, etc.), but these properties can be lost very rapidly using vector orthogonalization (subtracting multiples of earlier supposedly orthogonal vectors from the latest vector to produce the next orthogonal vector). Yet many of these algorithms are some of the best we have. There are those based on orthogonalizing with respect to all previously produced vectors, such as MGS, Arnoldi's method, and MGS-GMRES, and those that orthogonalize against a very few recent vectors such as Conjugate Gradients, Lanczos' method for the eigenproblem of symmetric or unsymmetric matrices, and Golub and Kahan bidiagonalization. This latter class depends on mathematically correct implicit orthogonalization against the remaining vectors, but this implicit behaviour is quite weak numerically, and very difficult to understand.

The elegant concept of backward stability for matrix transformation algorithms that was so effectively developed and applied by J. H. Wilkinson (FRS) just does not apply to these vector orthogonalization algorithms, leaving a large gap in our understanding of their numerical behaviours that we need to fill in order to use them most effectively.

Here we describe an ideal form of orthogonal matrix that arises from any sequence of supposedly orthogonal vectors. We will indicate how this ideal orthogonal matrix leads to expressions for new concepts of stability of such vector orthogonalization algorithm, and use this to understand some properties of the very basic and remarkably numerically subtle symmetric matrix tridiagonalization algorithm suggested by Cornelius Lanczos.

If we can sort this one out, we can probably eventually handle most of the others.

This is joint work with Wolfgang Wülling.

Sparse linear algebra issues arising in the analysis of complex networks

M. Benzi (Department of Mathematics and Computer Science, Emory University, Atlanta, USA).

In recent years there has been a surge of interest in the mathematical and computational analysis of complex networks. Such networks arise in a wide variety of applications throughout the physical sciences, in biology, in the social sciences, in engineering, in finance, and so forth. Most real-world networks are large and sparse, but their structural and spectral properties tend to be drastically different from those of the highly regular graphs arising in more traditional areas of scientific computing, such as the finite element method for PDEs. In this introductory talk I will review some of the computational problems arising in network analysis, with an emphasis on the calculation of centrality rankings and communicability measures between pairs of nodes. These problems lead to the computation of selected entries of certain matrix functions, such as the exponential of the adjacency matrix associated with the network.

This is joint work with Ernesto Estrada (Strathclyde) and Christine Klymko (Emory).

June 26th - Session V

Iterative methods for symmetric quasi-definite linear systems

M. Arioli (Rutherford Appleton Laboratory (RAL), UK).

We propose generalized versions of LSQR, Craig and LSMR well suited to the solution of symmetric quasi-definite systems of equations such as those arising in regularized interior-point methods for convex optimization or in stabilized control problems. Those methods essentially operate on the normal equations. We establish a connection between the iterates that they generate and those generated by CG and Minres on the original system.

Spectral information and GMRES convergence

G. Meurant (France).

A. Greenbaum and Z. Strakoš have introduced the concept of GMRES(A, b)-equivalent matrices. They are matrices B which give the same residual norm convergence curve as (A, b) with the same right-hand side b. In this talk we will exhibit unitary matrices that are GMRES(A, b)-equivalent. Moreover, we will give expressions of the residual norms as functions of the eigenvalues (and some entries of the eigenvectors) of the unitary matrices. This shows why and when GMRES is converging fast. These unitary matrices are related to the orthogonal bases of the Krylov subspaces $\mathcal{K}_n(A, b)$ and $A\mathcal{K}_n(A, b)$. We will illustrate these results with some numerical experiments.

This is joint work with J. Duintjer Tebbens, H. Sadok and Z. Strakoš.

June 26th - Session VI

The bitter truth about interior-point methods

D. Orban (Ecole Polytechnique of Montreal, Canada).

Interior-point methods feature prominently in the solution of constrained optimization problems, and involve the need to solve a sequence of 3×3 block indefinite system which become increasingly ill-conditioned throughout the iteration. To solve these systems, it is common practice to perform a block Gaussian elimination, and then either solve the resulting reduced 2×2 block indefinite system that has a typical saddle-point form, or further reduce the system to the normal equations and apply a symmetric positive-definite solver. We explore whether the step of reducing the system from 3×3 block form to a 2×2 block form necessarily pays off. We use energy estimates to obtain bounds on the eigenvalues of the matrices in question, which indicate that in fact at least in terms of spectral structure, it may be better to keep the system in its original unreduced form rather than perform a partial elimination.

Preconditioning of linear least-squares problems

M. Tůma (Institute of Computer Science, Academy of Sciences of the Czech Republic).

In this contribution we consider iterative methods for solving large and sparse linear least squares problems. We are interested in algebraic techniques to precondition the CGLS method. In particular we describe two such preconditioning techniques. First of them is based on the LU factorization, solving strategy which goes back to sixties. Our approach includes a new reordering based on a specific weighted transversal problem. Direct preconditioning of the normal equations by the balanced symmetric and positive definite factorization is our second approach. Numerical experiments demonstrate effectiveness of the algorithmic and implementational features ofthe new approaches.

This work was supported by the project No. P108/11/0853 of the Grant Agency of the Czech Republic. This is joint work with Rafael Bru, José Mas and José Marín (Institut de Matemàtica Multidisciplinar, Universitat Politècnica de Valencia, Spain).

References

[1] R. Bru, J. Marín, J. Mas and M. Tůma, *Preconditioned iterative methods for solving linear least squares problems*, submitted to SIAM SISC (2012).

Designing probabilistic-based algorithms suited for massively parallel computers

Ángel Rodríguez-Rozas (Center for Mathematics and its Applications, Department of Mathematics, Instituto Superior Técnico, Lisboa, Portugal).

New parallel numerical algorithms based on generating suitable *generalized random trees* are developed for solving non-linear parabolic and hyperbolic partial differential equations. Different approaches are discussed, both requiring generating suitable random trees, governed by new probabilistic representations of the solution, combined with a Pade approximant for approximating accurately a given divergent series. Such series are obtained by summing the partial contribution to the solution coming from trees with arbitrary number of branches.

The new representations largely expand the class of problems amenable to be solved probabilistically, and is used successfully to develop a generalized probabilistic *domain decomposition* method. Such a method has been shown to be suited for massively parallel computers [1, 2, 3], enjoying full scalability and fault tolerance. In short, the idea consists of generating only few interfacial values using the probabilistic approach along a given possibly artificial interfaces inside the domain, obtaining approximate values interpolating on such interfaces, and then use such values as boundary data in order to split the original problem into fully decoupled sub-problems.

While classical techniques based on a deterministic domain decomposition exhibits strong limitations in terms of scalability, probabilistic methods are capable of exploit massively parallel architectures since the problem can be fully decoupled. Some important examples are given, such as the Vlasov-Poisson system in Fourier space, describing a practical implementation and showing a remarkable scalability and performance of the method on a massively parallel supercomputer.

This is joint work with Juan A. Acebrón.

References

[1] J. A. Acebrón and A. Rodríguez-Rozas and R. Spigler, *Domain decomposition solution of nonlinear two-dimensional parabolic problems by random trees*, J. Comput. Phys. **15** (2009), 5574-5591

[2] J. A. Acebrón and A. Rodríguez-Rozas and R. Spigler, *Efficient Parallel Solution of Nonlinear Parabolic Partial Differential Equations by a Probabilistic Domain Decomposition*, J. Sci. Comput. **43** (2010), 135-157.

[3] J. A. Acebrón and A. Rodríguez-Rozas, *A new parallel solver suited for arbitrary semilinear parabolic partial differential equations based on generalized random trees*, J. Comput. Phys. **230** (2011), 7891-7909.