This report summarizes the research activities at Boeing Computer Services on AFOSR Contract F49620-87-C-0037 from April 1, 1987 to March 31, 1988. We report significant progress in the two areas of research, we have developed methodologies that provide impressive improvements in the performance of sparse linear equations solver on vector supercomputers and we have made several advances toward the development of ordering methodologies for solving sparse linear equations on parallel computers. IN ADDITION, WE HAVE submitted for publication and are now formally distributing the Harwell-Boeing collection of sparse matrix test problems. In this report we present the status for each task in our original plan. We also list relevant reports and publications of project personnel and discuss related sparse matrix activities at Boeing.
ORDERING METHODS FOR SPARSE MATRICES AND VECTOR COMPUTERS

AFOSR Contract F49620-87-C-0037

Report No. 1

Final Report
April 1, 1987 - March 31, 1988

by

John G. Lewis

Boeing Computer Services Co.

March 31, 1988

KEY WORDS AND PHRASES: Linear equations, sparse matrices, reordering algorithms, Cholesky factorization, Gauss elimination, frontal methods, multifrontal methods, minimum degree ordering, nested dissection, vector computers, supercomputers, parallel computers

ABSTRACT: This report summarizes the research activities at Boeing Computer Services on AFOSR Contract F49620-87-C-0037 from April 1, 1987 until March 31, 1988. We report significant progress in two of our areas of research: we have developed methodologies that provide impressive improvements in the performance of sparse linear equations solvers on vector supercomputers and we have made several advances toward the development of ordering methodologies for solving sparse linear equations on parallel computers. In addition, we have submitted for publication and are now formally distributing the Harwell-Boeing collection of sparse matrix test problems. In this report we present the status for each task in our original plan. We also list relevant reports and publications of project personnel and discuss related sparse matrix activities at Boeing.
INTRODUCTION

Direct factorization methods for solving large sparse linear equations are used as fundamental building blocks for the numerical solution of many scientific and computational problems. It is well known that reordering the variables and equations is crucial in reducing the cost of performing direct solution techniques. The problem of finding the optimal reordering is known to be an NP-complete problem. As a result, practical reordering algorithms are heuristic, and their behavior is usually only known empirically. Different reordering heuristics have been developed in a number of different disciplines, reflecting the different types of sparse linear systems and different views of the cost of computing.

This research has been concerned with furthering our understanding of how ordering heuristics and their companion numerical solution routines behave on high performance computers. The availability of such computers has led to a dramatic increase in the size and complexity of scientific computations. This is the arena in which better heuristics have the largest effect on the cost of scientific computing, but it is also an arena in which architectural constraints chosen for high speed often appear to conflict with sparsity. Our research indicates that this conflict is only superficial and that, with minor modifications, the methods that have proved best on ordinary scalar sequential computers continue to hold their advantages on both vector and parallel supercomputers.

This report describes the final status of the project. The original research objectives, the results of the research effort, relevant publications by project personnel, current makeup of the project team and related sparse matrix activities at Boeing Computer Services are discussed.

RESEARCH OBJECTIVES

In previous reports and in the Technical Proposal Modifications the research objectives were given as five separate tasks:

1. analysis of multifrontal factorization
2. creation of a symmetric indefinite out-of-core sparse column Cholesky factorization algorithm
3. analysis of an outer-product sparse Cholesky algorithm
4. analysis of quotient tree orderings
5. publication of the Harwell-Boeing sparse matrix collection.

The status of each of these tasks is discussed in turn in the following section. We should note, however, that as a result of major successes in Task 1, our emphasis was shifted to continue work on Task 1.

RESEARCH STATUS

Task 1: Analysis of Multifrontal Orderings

Under the technical proposal modifications the research in this area was to be directed toward two subtasks: possible modifications of multifrontal factorizations and orderings, and development of better orderings for factorization on parallel computers. We have found significant success in both of these -- as a result most of our research effort has been addressed to this task. We first present our results on vectorizing sparse factorizations.

We have been able to characterize many of the relationships between sparse column Cholesky factorization methods and multifrontal factorization methods. The first of our successful modifications of the multifrontal factorization method is a result of the previous work on tasks 1
and 3. The analysis of an outer-product version of a sparse Cholesky factorization produces a
obvious bottleneck in the creation of row indices for the sparse vector operations. A solution
that partially removes the bottleneck is the use of relative row indices as developed by Schreiber
[4]. This same solution can be applied to the multifrontal method, where the effect is to remove
the one formidable obstacle to using the hardware gather-scatter feature that is now standard on
vector supercomputers.

An experimental code was developed that uses relative row indices to eliminate the major non-
vector computations in the multifrontal factorization. This was combined with the use of higher
level ‘kernels’ in the dense portion of this sparse factorization. The result is an algorithm that
makes extremely good use of vector hardware in the factorization of sparse matrices. The
experimental code, running on a CRAY X-MP vector supercomputer, achieves over half the
maximum rated speed of this machine on a problem that has been thought to be inherently non-
vectorizable. Further, the methods are not machine-specific, and should apply to all vector
computers. In particular, the results should be usable on the important class of MIMD parallel
computers where each node is a vector processor, the class of machines that occur most
frequently in commercial efforts toward the next generation of supercomputers. These results
have been published in [BCS1, BCS3, BCS4].

The relationships between multifrontal and sparse column Cholesky factorizations allowed the
development of similar techniques for improvements in the column Cholesky factorization.
Higher level sparse matrix kernels were developed for this algorithm; these result in similar
speedups and again the methodology is quite general. The final performance for this sparse
supernodal Cholesky factorization is slightly less than for the multifrontal method, but
limitations in compiler techniques make a final comparison impossible at this time. We have
also included these techniques in [BCS3, BCS4]. Finer details will be given in [BCS5].

The cross-fertilization between the two factorization methodologies went in both directions. The
analysis of the supernode structure of a sparse factorization provided the tools for yet another
technique, relaxing the supernodal partition for improved vectorization of the multifrontal
factorization. This is based on identifying a natural structure where a limited amount of
additional work can be performed in exchange for a reduction in the amount of sparse memory
traffic. This is often proposed as an approach to vectorizing sparse computations, but it has
rarely been successful. In this case success is a result of making only small, local, modifications
to a structure that is already good for both sparsity and vectorization. These results have been
submitted for publication [BCS2]. This idea could also be applied in a straightforward manner
to the column Cholesky factorization, a programming exercise we did not carry out.

The second research topic under Task 1 was the analysis of orderings for parallel sparse
factorization. This is a very active topic today, with much interest being generated by the
proliferation of parallel computing hardware. Previous sparse matrix tools, specifically the
notion of the elimination tree, are already at hand to provide an analysis of the parallelism in a
sparse factorization, once given the ordering. These tools have already shown that the best
sequential orderings allow a significant degree of parallelism. At issue are whether the measures
of parallelism are correct, whether the tools are efficient and whether these standard orderings
are sufficiently close to optimal. Our research has been directed to all three of these topics, with
significant results on the latter two.

One of the standard measures of parallelism is the height of the elimination tree. Tall
elimination trees are clearly worse than short trees. Several years ago an algorithm was
developed by Jess and Kees [1] that could be used, in theory, to minimize the height of the
elimination tree for an ordered matrix. The algorithm finds an ordering that has exactly the same
sparsity as the original and, in addition, has the least elimination tree height of all such
equivalent orderings.
Such an algorithm is a useful tool, both for gaining additional parallelism, and for use in evaluating different ordering heuristics. In the latter use, it enables us to compare the best possible of various families of orderings, rather than simply comparing arbitrary members. Unfortunately, the original presentation was only a theoretical characterization. Liu and Mirzaian [2] recently developed an implementation of the Jess and Kees algorithm, whose complexity and running time are relatively large. Liu [3] separately developed a heuristic that approximates the optimal solution. The complexity of finding this approximation was substantially reduced.

We have applied the notion of clique trees to develop a characterization and implementation of the Jess and Kees algorithm whose complexity and development are much closer to the usual sparse ordering problem than the Liu and Mirzaian implementation. Although the complexity of our algorithm and Liu's heuristic are incommensurate, making a theoretical comparison impossible, the complexities are similar and in practice the implementations run in essentially the same time. This development means that the best equivalent ordering can be found for a relatively small additional cost. We are preparing a paper presenting these results [BCS9].

The algorithm described above finds the most parallel ordering equivalent to some specified ordering. It cannot find parallelism if the original ordering is poorly chosen. Our second project on the parallel ordering problem was to pursue orderings that transparently exhibit parallelism. The approach is now becoming standard, in part because of our success. The nested dissection algorithm is the classical example of the divide and conquer paradigm applied to the ordering problem. This paradigm clearly develops a structure amenable to parallel computation. However, earlier attempts to apply the nested dissection algorithm to general graphs were somewhat unsuccessful, in that the resulting orderings were usually worse, sometimes much worse, than the standard (minimum degree) orderings.

We have applied techniques for bisecting graphs to provide a basis for finding dissectors in graphs. The graph bisection problem is important in VLSI design, and a number of heuristic algorithms have been developed for approximating its solution. We developed a framework for performing nested dissection upon being given a bisection of the graph, and used one of the standard algorithms to solve the graph bisection problem. Our nested dissection orderings have elimination trees that are, on average, only 74% as high as those produced by the standard approach. This represents a considerable decrease in the tree height, which should be reflected by a corresponding decrease in the parallel execution time for the sparse factorization. In addition, these orderings are only very slightly worse, on average, than the accepted ordering for sequential computers.

These results have been submitted for publication [BCS8]. They represent the first successful application of the nested dissection paradigm to general sparse matrices. Unfortunately the VLSI graph bisection heuristics are not efficient enough to compete with the standard minimum degree ordering with respect to the time requirements for the ordering itself. However, in demonstrating that better orderings can be found with the nested dissection heuristic, this work has rekindled interest in nested dissection in the sparse matrix community. Work following on this success is planned at Boeing, Yale, Penn State and York University.

Task 2: Symmetric Indefinite Sparse Column Cholesky Factorization

This task called for investigation of a symmetric indefinite factorization algorithm based on Liu's out of core Cholesky factorization algorithm. A preliminary design for the necessary data structures was completed under the previous contract, which provides one model for such an algorithm. Joseph Liu (York University) published an alternative model, with simpler data structures, but potentially larger storage requirements. Due to our shifting resources to extend
the successes in Task 1, no further work was performed on this task. Further, it was concluded that a choice between the more elaborate model developed herein and Liu's model could most easily be made by using the multifrontal algorithm of Task 1 as a test bed.

Task 3: Outer Product Sparse Cholesky Factorization

The success in incorporating relative row indices and higher level vectorization kernels into the multifrontal algorithm in Task 1 made it clear that an (undistributed) outer product sparse factorization algorithm would not be competitive in speed with the multifrontal factorization, a distributed outer product algorithm. The analysis of the storage requirements carried out in the previous contract indicated little difference between the two approaches. Therefore, we concluded that it would unfruitful to continue pursuing this task, and resources were redirected to further work in Task 1.

Task 4: Analysis of Quotient Tree Algorithms

The primary motivation for further work on this task was to use the tree structure of a quotient tree ordering to support parallel factorization. In all of our preliminary investigations, the elimination tree structure proved to be a richer source of parallelism than the quotient tree structure. Due of a lack of confidence of success, the resources of this task were redirected to more fertile areas.

Task 5: Harwell-Boeing Sparse Matrix Collection

The Harwell-Boeing sparse matrix collection was expanded considerably in scope during this contract. A formal announcement of its structure and general availability was made in [BCS6]. More detailed documentation of the contents of the collection has been prepared as [BCS7]. Final discussions between the Boeing authors and our British colleague is the final Boeing activity under this contract: we expect this document to be released formally next month. With the release of the expanded collection, we anticipate further use of this already popular research benchmark.

REFERENCES


J.W.H. Liu, "Reordering sparse matrices for parallel elimination", TR CS-87-01, Dept. of Computer Science, York University, 1987

RECENT RELEVANT PUBLICATIONS AND PRESENTATIONS OF THE PROJECT PERSONNEL

Publications relevant to research on this contract


Presentations relevant to research on this contract


Other Related Publications


C.C. Ashcraft and D.J. Pierce, "Domain Decoupled Incomplete Factorizations", to appear in Parallel Computing


R.G. Grimes, D.J. Pierce and H.D. Simon, "A new algorithm for finding a pseudo-peripheral node in a graph", submitted for publication


**Other Related Presentations**


J.G. Lewis, with R.G. Grimes and H.D. Simon, "Industrial Strength Lanczos", University of Illinois, Center for Supercomputing Research & Development, Jan. 1987; Massachusetts Institute of Technology, Mathematics Department, Jan. 1987; Rensselaer Polytechnic Institute, May 1987

J.G. Lewis, "Numerical Computation on a Massively Parallel Computer", University of Washington, Applied Mathematics Department, Nov. 1987

**PROFESSIONAL PERSONNEL ASSOCIATED WITH THE PROJECT**

The project team consisted of C. Cleveland Ashcraft, Roger G. Grimes, John G. Lewis, Barry W. Peyton and Horst D. Simon, with Horst Simon serving as project manager. When Simon stepped down as project manager to take a Boeing position in Santa Clara, California, in support of NASA Ames, Lewis assumed the role of project manager. Both Ashcraft and Lewis took academic leaves during part of the contract period. Ashcraft went on academic leave in August 1987 to pursue a PhD in Computer Science at Yale University. Lewis returned in August from a year as a Boeing Fellow at M.I.T. Task 5 was carried out in collaboration with Iain S. Duff of AERE, Harwell, England. Lewis’s work on parallel orderings was carried out partly in collaboration with Charles E. Leiserson of M.I.T.

**RELATED SPARSE MATRIX ACTIVITIES AT BOEING COMPUTER SERVICES**

The project personnel are active in other projects at Boeing Computer Services that involve sparse matrix computations. This section briefly describes some of the most recent activities. These projects are not funded by this AFOSR contract, but they indicate the level of importance of sparse matrix research at Boeing.

**Iterative Methods and Preconditioners on Vector and Parallel Computers**

As a Boeing internal research project, C. Ashcraft and R. Grimes considered the problem of vectorizing the recursive calculations found in modified incomplete factorizations and SSOR preconditioners for the conjugate gradient method. For matrix problems derived from the discretization of partial differential equations on regular 2 and 3 dimensional grids, they developed vectorized implementation of three modified incomplete factorizations as well as the SSOR preconditioner. All four preconditioners achieve overall computational rates near 100 megaflops on a Cray X-MP/24, thus providing very fast implementations of good preconditioners for the conjugate gradient method.
D. Pierce is currently investigating use of a parallel incomplete Cholesky factorization based on Schur complements. The approach is novel in using hyperbolic transformations to form the incomplete factorization. Preliminary results on a 20 processor Sequent Balance 21000 indicate a 20% decrease in the required number of iterations compared to a similar preconditioner due to H. C. Elman [1987], with the same amount of work required per iteration.

Out-of-core Nested Dissection Code

J. Lewis continues to work on a production sparse matrix program for a major industrial customer. This code is used to solve systems of millions of sparse linear equations, using the nested dissection technique for ordering the problem and managing the required data transfers. This code is currently in production use on a Cray 2 computer, and is being extended to increase the size capabilities.

Sparse Matrix Computations on a Massively Parallel Computer

J. Lewis was supported by Boeing to spend the 1986-87 academic year as a Fellow in the Center for Advanced Engineering Study at M.I.T. His primary objective at M.I.T. was the study of algorithms and languages for parallel computation. As a part of this study, he spent two months as a visiting researcher at Thinking Machines Corporation, analyzing the requirements for implementing dense and sparse linear algebra algorithms on the Connection Machine. The need to ordering algorithms for this machine led to the collaboration between Lewis and C. Leiserson on algorithms for parallel orderings.

Sparse Matrix Code for CRI

R. Grimes and B. Peyton developed a supermodal general sparse factorization capability under Cray Research, Inc. funding for inclusion in their scientific library. This software is based on results from the successful research performed under this AFOSR contract and internally funded research projects, and provides users of Cray computers with software that is 4 to 5 times faster than previously available general sparse matrix codes.

Development of In-core and Out-of-core Multifrontal Sparse Linear Equation Solvers

As a internally-funded project C. Ashcraft and R. Grimes developed both an in-core and out-of-core prototype implementation of the multifrontal algorithm. This software demonstrates the results of the AFOSR research and provides a capability for solving sparse equations on Cray computers which is challenged in efficiency only by the supermodal general sparse algorithm. It also provides the capabilities of easily solving problems on the order of 40,000 to 50,000 variables on an Cray X-MP/24 with SSD.

Dynamic Analysis for Structural Engineering

R. Grimes incorporated the multifrontal codes discussed above into a program for solving the real sparse symmetric generalized eigenproblem using our block Lanczos method. This software has been made available to the structural engineering staff at Boeing and is currently being tested. It is expected that the software will provide an efficient means for solving dynamic analysis problems for structures with tens of thousands of degrees of freedom.

J. Lewis will investigate the solution of the sparse damped vibration problems in a separate internally funded project. This work will consider variations on the
unsymmetric Lanczos algorithm as the fundamental tool for reducing the dimensionality of the problem. Lewis also supports a research project for Boeing Aerospace in which damped vibration problems are attacked using the symmetric Lanczos algorithm.

Parallel Multifrontal Factorization

As an internally funded project, R. Grimes is developing a parallel implementation of the multifrontal factorization. The research tools developed under this AFOSR contract have been applied to the load balance problem for this computation. A new strategy for distributing the computation based on the sparsity structure is being tested on a 20 processor Sequent computer.

Improved Minimum Degree Orderings for Sparse Factorization

B. Peyton is exploring, under Boeing IMPD funding, alternative tie-breaking strategies for the minimum degree ordering heuristic. The minimum degree ordering is the standard against which other sparse orders are measured. However, 'arbitrary' tie-breaking within the heuristic leads occasionally to poor results. A lack of understanding as to the cause of the poor results is an obstacle to developing variants that could exploit parallelism; during the ordering heuristic itself, or for use 'out-of-core' for very large problems. Peyton is pursuing two approaches to the tie-breaking problem, one of which also shows promise for improving the performance and reducing the cost of the heuristic itself.

Orderings for Parallel Sparse Factorization

J. Lewis will consider parallel implementations of simulated annealing as a mechanism for finding nested dissection orderings.

Sparse Matrix Methods for Computational Fluid Dynamics

B. Peyton participated in the development of a special purpose sparse linear equation solver that is used in solving very large systems of linear equations arising in computational fluid dynamics.

Sparse Matrix Workshop and Conference

H. Simon served as co-organizer of the workshop on Supercomputer Applications of Sparse Matrix Algorithms, cosponsored by Boeing Computer Services and Cray Research, Inc. This workshop was held in Santa Cruz, California on March 27th-30th, 1988, and attracted 73 participants from academics and industry.

J. Lewis is chairman of the organizing committee for the SIAM Activity Group on Linear Algebra’s meeting on sparse matrix methods, to be held at Glenenden Beach, Oregon, in May of 1989. H. Simon is a member of the organizing committee. This is expected to be an international meeting attracting around 150 participants.