Sparse Hypermatrix Cholesky: Customization for High Performance

IMECS’06

Josep R. Herrero, Juan J. Navarro
{josepr,juanjo}@ac.upc.es

Computer Architecture Department
Universitat Politècnica de Catalunya
Barcelona
Spain
Hypermatrix Structure

Can store 0’s within data submatrices
- Storage
- Computation

Trade-off in data submatrix size
- BLAS3 efficiency
- (Useless) operation on 0’s
Overview

Goals:

- Efficient implementation of sparse Cholesky factorization
  - Focus on matrices arising in Interior Point Methods (IPMs)

Facts:

- Efficient execution requires adaptation of the code to:
  - Problem
  - Platform

Approach:

- Based on previous work
  - Small Matrix Library (SML)
  - Windows within data sub-matrices
  - Intra-Block amalgamation

- New
  - New values for Intra-Block amalgamation on Itanium2
  - Sparse matrix reordering for IPMs
  - Submatrix size & storage
Windows within data submatrices

Windows within data submatrices:

Intra-block amalgamation:
Intra-Block amalgamation on Itanium2: submatrices of size $4 \times 32$

\[
\text{Effective Mflops} = \frac{\# \text{flops (excluding operations on zeros)}}{\text{Time (including operations on zeros)}} \cdot 10^{-6}
\]
Ordering matrices for IPM

METIS parameters used: 1, 3, 1, 1, 0, 3, 60, and 5.

- Larger ordering time

# iterations of IPM necessary to amortize cost of improved ordering
Data submatrix compression

Data submatrices before compression:

Data submatrices after compression:
Impact of submatrix compression on HM space

Reduction in HM space after submatrix compression
Sparse HM Cholesky: performance for several submatrix sizes

Performance depends on

- Overhead due to 0’s within submatrices
- Efficiency of the operations on submatrices

<table>
<thead>
<tr>
<th>Submatrix Size</th>
<th>Performance (in ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 × 32</td>
<td>4005</td>
</tr>
<tr>
<td>8 × 32</td>
<td>4080</td>
</tr>
<tr>
<td>16 × 32</td>
<td>4488</td>
</tr>
<tr>
<td>32 × 32</td>
<td>4401</td>
</tr>
</tbody>
</table>

Performance of the $C = C - A \times B^T$ matrix multiplication routine for each submatrix size
Sparse HM Cholesky: performance for several submatrix sizes

Variation in execution time for each submatrix size relative to $4 \times 32$