IRis

Northeastern University

Research Thrust R3 Presentations

CenSSIS Research Thrust R3: Image and Data Information Management

January 01, 2007

Characterizing the Relationship Between ILU-Type Preconditioners and the Storage Hierarchy

Diego Rivera
Northeastern University

David Kaeli
Northeastern University

Misha Kilmer
Tufts University

Recommended Citation

Rivera, Diego; Kaeli, David; and Kilmer, Misha, "Characterizing the Relationship Between ILU-Type Preconditioners and the Storage Hierarchy" (2007). *Research Thrust R3 Presentations.* Paper 3. http://hdl.handle.net/2047/d10009127

This work is available open access, hosted by Northeastern University.

"Characterizing the Relationship between ILU-type Preconditioners and the Storage Hierarchy"

Diego Rivera[1] , David Kaeli[1] and Misha Kilmer[2]

[1] Department of Electrical and Computer Engineering
Northeastern University, Boston, MA
{drivera, kaeli}@ece.neu.edu

[2] Department of Mathematics
Tufts University, Medford, MA
misha.kilmer@tufts.edu

www.ece.neu.edu/students/drivera/tlg/tunlib.html

Objective

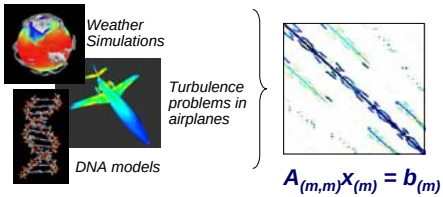
- To improve the performance of preconditioners targeting sparse matrices
- To accelerate the memory accesses associated with these codes

Motivation

- Prior work targeted Krylov subspace methods
- However, little has been done in the case of preconditioners

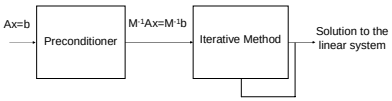
"Nothing will be more central to computational science in the next century than the art of transforming a problem that appears intractable into another whose solution can be approximated rapidly. For Krylov subspace matrix iterations, this is preconditioning" from Numerical Linear Algebra by Trefethen and Bau (1997).

Common target applications



$$A_{(m,m)}x_{(m)} = b_{(m)}$$

- Incomplete LU factorization type Preconditioners are used to accelerate the convergence of Krylov subspace methods



- A drawback of these approaches is that it is difficult to choose the best values for their tuning parameters
- Choosing good values depends heavily on the structure of non-zero elements of the coefficient matrix
- In our work we have found that it depends also on the memory hierarchy of the machine used to carry out the computation
- The parameter values used to obtain the fastest execution time, given an acceptable final error, may be different for different memory hierarchies

Target preconditioners

Preconditioner	Parameters	Description parameters
ILU(κ)	κ	level-of-fill
ILUT	ρ,τ	level-of-fill, drop tolerance
ILUTP	ρ,τ, permtol	level-of-fill, drop tolerance and tolerance ratio
ILUD	τ, α	drop tolerance, diagonal compensation parameter
ILUDP	τ, α , permtol	drop tolerance, diagonal compensation parameter and tolerance ratio

- Multilevel preconditioners based on ILU factorization

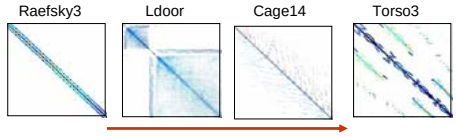
Evaluation environment

	Intel XEON 3.06 GHz	Ultra Sparc-III 750 MHz
Level 1	8KB 4-way for data	64KB 4-way for data
Level 2	512 KB 8-way	8MB 2-way
Level 3	1 MB 8-way	N/A
RAM	2 GB RAM	1 GB RAM

Matrices

Name	Non-zero elements	Rows	NS	B	NS/B
Raefsky3	1,488,768	21,200	48%	0.0596	8.05
Ldoor	42,493,817	952,203	100%	0.7215	1.39
Cage14	27,130,349	1,505,785	21%	0.4490	0.47
Torso3	4,429,042	259,156	0%	0.8181	0

- NS (Numerical Symmetry)
- B (matrix's Bandwidth)



- The difference in performance on different memory hierarchies becomes significant when the problem's conditions make it more difficult to solve
- These conditions are related to the dropping strategy adopted in the preconditioner algorithm

Error norm vs. 13 first duple sorted in increasing order for execution time of ILUT and GMRES

Same duple (level of fill-in, drop tol) in both machines

Different duple (level of fill-in, drop tol) in both machines

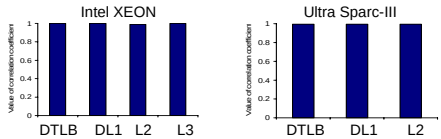
RAEFSKY3							
Xeon				Ultra			
level of fill-in	drop tol.	iterations	Residual error	level of fill-in	drop tol.	iterations	Residual error
30	1.0E-03	23	7.5754E-07	30	1.0E-03	23	7.5754E-07
30	8.0E-04	23	5.8466E-07	30	8.0E-04	23	5.8466E-07
32	1.0E-03	23	5.3237E-07	32	1.0E-03	23	5.3237E-07
34	1.0E-03	22	6.5689E-07	34	1.0E-03	22	6.5689E-07
32	8.0E-04	22	4.8701E-07	32	8.0E-04	22	4.8701E-07
30	6.0E-04	23	7.5087E-07	30	6.0E-04	23	7.5087E-07
34	8.0E-04	22	6.3722E-07	34	8.0E-04	22	6.3722E-07
36	1.0E-03	22	6.9058E-07	36	1.0E-03	22	6.9058E-07
32	6.0E-04	22	6.7871E-07	32	6.0E-04	22	6.7871E-07
30	4.0E-04	23	6.6286E-07	30	4.0E-04	23	6.6286E-07
38	1.0E-03	22	3.7978E-07	38	1.0E-03	22	3.7978E-07
34	6.0E-04	22	5.0248E-07	34	6.0E-04	22	5.0248E-07
36	8.0E-04	22	4.4528E-07	36	8.0E-04	22	4.4528E-07

LDOOR							
Xeon				Ultra			
level of fill-in	drop tol.	iterations	Residual error	level of fill-in	drop tol.	iterations	Residual error
50	1.0E-02	3	4.5742E-02	50	1.0E-02	3	4.5742E-02
50	1.0E-03	3	4.5558E-02	50	1.0E-03	3	4.5558E-02
50	1.0E-04	3	4.5558E-02	50	1.0E-04	3	4.5558E-02
50	1.0E-07	3	4.5558E-02	50	1.0E-07	3	4.5558E-02
50	1.0E-06	3	4.5558E-02	50	1.0E-06	3	4.5558E-02
50	1.0E-10	3	4.5558E-02	50	1.0E-10	3	4.5558E-02
50	1.0E-05	3	4.5558E-02	50	1.0E-05	3	4.5558E-02
50	1.0E-01	4	3.7216E-03	50	1.0E-01	4	3.7216E-03
40	1.0E-01	4	1.0742E-02	50	5.0E-02	4	5.0180E-04
50	5.0E-02	4	5.0180E-04	50	2.5E-02	4	2.4878E-04
50	2.5E-02	4	2.4878E-04	40	1.0E-01	4	1.0742E-02
50	2.5E-01	5	3.8002E-03	40	5.0E-02	4	6.2204E-03
40	5.0E-02	4	6.2204E-03	40	2.5E-02	4	5.7514E-03

CAGE14							
Xeon				Ultra			
level of fill-in	drop tol.	iterations	Residual error	level of fill-in	drop tol.	iterations	Residual error
1	5.0E-01	8	1.4892E-02	13	2.5E-01	7	2.6528E-02
40	5.0E-01	8	2.3926E-02	15	2.5E-01	7	2.8517E-02
2	5.0E-01	8	1.5111E-02	1	2.5E-01	8	1.4892E-02
20	5.0E-01	8	2.3926E-02	2	5.0E-01	8	1.5111E-02
17	5.0E-01	8	2.3882E-02	1	1.0E-01	8	2.6387E-02
3	5.0E-01	8	1.7360E-02	1	5.0E-01	8	1.4892E-02
15	5.0E-01	8	2.3695E-02	30	5.0E-01	8	2.3926E-02
5	5.0E-01	8	1.7308E-02	40	5.0E-01	8	2.3926E-02
30	5.0E-01	8	2.3926E-02	50	5.0E-01	8	2.3926E-02
9	5.0E-01	8	2.2126E-02	20	5.0E-01	8	2.3926E-02
50	5.0E-01	8	2.3926E-02	13	5.0E-01	8	2.3546E-02
11	5.0E-01	8	2.3420E-02	3	5.0E-01	8	1.7360E-02
13	5.0E-01	8	2.3546E-02	11	5.0E-01	8	2.3420E-02

TORSO							
Xeon				Ultra			
level of fill-in	drop tol.	iterations	Residual error	level of fill-in	drop tol.	iterations	Residual error
20	4.0E-02	10	2.1274E-08	30	1.0E-02	7	2.0350E-08
17	4.0E-02	10	2.1105E-08	30	2.5E-02	9	2.1215E-08
13	4.0E-02	10	4.1135E-08	30	1.5E-02	8	8.3086E-09
15	4.0E-02	10	2.9366E-08	13	4.0E-02	10	4.1135E-08
30	4.0E-02	10	2.1951E-08	30	4.0E-02	10	2.1951E-08
30	2.5E-02	9	2.1215E-08	17	4.0E-02	10	2.1105E-08
17	3.5E-02	10	1.4079E-08	15	4.0E-02	10	2.9366E-08
30	3.5E-02	10	1.1416E-08	20	4.0E-02	10	2.1274E-08
20	3.5E-02	10	1.1082E-08	30	3.5E-02	10	1.1416E-08
20	6.0E-02	11	3.1833E-08	20	3.5E-02	10	1.1082E-08
13	6.0E-02	11	3.7707E-08	17	3.5E-02	10	1.4079E-08
15	6.0E-02	11	3.3234E-08	30	2.0E-02	9	1.0323E-08
30	6.0E-02	11	3.1831E-08	20	3.0E-02	10	7.4593E-09

Correlation of load accesses and execution time



- We use the PIN tool to capture cache events
- Our results show a high correlation between the execution time, memory accesses and cache misses

Our Algorithm Approach to:

1) Extract the problem's conditions related to the dropping strategies adopted in the preconditioner

2) Detect if the computation of a solution depends upon the relationship between the preconditioner's parameters and the memory hierarchy of the machine used

3) Suggest values for the preconditioner's parameters which can help to reduce the time required to compute the preconditioner and the solution for matrices with similar characteristics

- Our experimental results show that 78.4% of the time, the suggested values of the preconditioner's parameters were appropriate in reducing the overall execution time

Plans and future work

- Explore more sophisticated heuristics for our algorithmic approach
  - Increase the percentage of suggested values appropriated in reducing the overall execution time.
- Extend our study to multilevel preconditioners based on ILU factorization

Acknowledgement

This project is supported by a grant from the NSF Advanced Computational Research Division, award No. CCF-0342555 and the Institute for Complex Scientific Software at Northeastern University.