A Priori Ordering Strategies for High Performance Preconditioners

David Fritzsche Andreas Frommer

Bergische Universität Wuppertal

2nd June 2004

1. PABLO, TPABLO and XPABLO
   - Fullness Criterion and Connectivity Criterion
   - Threshold Criteria
   - Combining Criteria
2. Spanning Tree Preconditioners
   - Matrix Splitting Preconditioners
   - Spanning Trees
3. Preprocessing
   - Scaling and Transversals
4. Examples
   - Choosing the Parameters
   - Numerical Results

Graph Theory

Definition: Let $A \in \mathbb{R}^{n \times n}$. Then $G(A) = (V, E)$ is the *directed* graph of A, i.e.

- vertices $V = \{1, \ldots, n\}$
- edges $E = \{(i,j) : a_{ij} \neq 0\}$
- **weights** $\boldsymbol{w((i,j)) := |a_{ij}|}$

Definition:
Vertices v, w are *adjacent* $\Leftrightarrow$ (v, w) or (w, v) is edge in graph (direction of edges not important)

Overview of PABLO

Idea:

- Permute $A \to PAP^T$ to get dense "blocks" on the diagonal
- Find disjoint "clusters" in $G(A)$ which are "well connected"

Overview of PABLO

Idea:

- Permute $A \rightarrow PAP^T$ to get dense "blocks" on the diagonal
- Find disjoint "clusters" in $G(A)$ which are "well connected"

PABLO's Way: (Szyld, Neil 1990)

- Build one block at a time
- Check only adjacent vertices for possible inclusion
- Include an adjacent vertex if it fulfills either the **Fullness Criterion** or the **Connectivity Criterion**

Overview of PABLO

Idea:

- Permute $A \rightarrow PAP^T$ to get dense “blocks” on the diagonal
- Find disjoint “clusters” in $G(A)$ which are “well connected”

PABLO’s Way: (Szyld, Neil 1990)

- Build one block at a time
- Check only adjacent vertices for possible inclusion
- Include an adjacent vertex if it fulfills either the **Fullness Criterion** or the **Connectivity Criterion**

Other Features:

- Time complexity: $\mathcal{O}(n + \text{nnz}(A))$
- Block sizes are not known a priori.

Fullness Criterion and Connectivity Criterion

Let P be the vertices currently in the block, v the eligible vertex

Fullness Criterion:

The "fullness" of $P \cup \{v\}$ is at least α times the fullness of P

$$\text{fullness} := \frac{\text{number of edges}}{\text{number of all possible edges}}$$

Fullness Criterion and Connectivity Criterion

Let P be the vertices currently in the block, v the eligible vertex

Fullness Criterion:
The "fullness" of $P \cup \{v\}$ is at least α times the fullness of P

$$\text{fullness} := \frac{\text{number of edges}}{\text{number of all possible edges}}$$

Connectivity Criterion:
A fraction of β or more of all edges of v goes into P

Threshold Criteria

Observation: PABLO only looks at the structure
$\rightsquigarrow$ XPABLO which also cares about values

Threshold Criteria

Observation: PABLO only looks at the structure
$\rightsquigarrow$ XPABLO which also cares about values

New: Fix *threshold parameter* γ.
In addition to $G(A)$ we look at

$$G_\gamma := \text{remove edges with weight } \leq \gamma$$

Threshold Criteria

Observation: PABLO only looks at the structure
$\rightsquigarrow$ XPABLO which also cares about values

New: Fix *threshold parameter* γ.
In addition to $G(A)$ we look at

$$G_\gamma := \text{remove edges with weight } \leq \gamma$$

Threshold Fullness Criterion:
The fullness of $P \cup \{v\}$ in G_γ is at least ϑ.

Threshold Criteria

Observation: PABLO only looks at the structure
$\leadsto$ XPABLO which also cares about values

New: Fix *threshold parameter* γ.
In addition to $G(A)$ we look at

$$G_\gamma := \text{remove edges with weight } \leq \gamma$$

Threshold Fullness Criterion:
The fullness of $P \cup \{v\}$ in G_γ is at least ϑ.

Threshold Connectivity Criterion: A fraction of at least ζ of the edges of v into P in $G(A)$ are also present in G_γ.

Combining Criteria

The *flag parameter* F controls the combination of the criteria:

F	v fulfills the XPABLO criterion $\Leftrightarrow \ldots$
0	(fulln. **or** conn.) **or** (thr. fulln. **or** thr. conn.)
1	(fulln. **or** conn.) **and** (thr. fulln. **or** thr. conn.)
2	(fulln. **and** conn.) **or** (thr. fulln. **or** thr. conn.)
4	(fulln. **or** conn.) **or** (thr. fulln. **and** thr. conn.)
$3 = 1 + 2$	(fulln. **and** conn.) **and** (thr. fulln. **or** thr. conn.)
$5 = 1 + 4$	(fulln. **or** conn.) **and** (thr. fulln. **and** thr. conn.)
$6 = 2 + 4$	(fulln. **and** conn.) **or** (thr. fulln. **and** thr. conn.)
$7 = 1 + 2 + 4$	(fulln. **and** conn.) **and** (thr. fulln. **and** thr. conn.)

Combining Criteria

The *flag parameter* F controls the combination of the criteria:

F	v fulfills the XPABLO criterion $\Leftrightarrow \dots$
0	(fulln. **or** conn.) **or** (thr. fulln. **or** thr. conn.)
1	(fulln. **or** conn.) **and** (thr. fulln. **or** thr. conn.)
2	(fulln. **and** conn.) **or** (thr. fulln. **or** thr. conn.)
4	(fulln. **or** conn.) **or** (thr. fulln. **and** thr. conn.)
$3 = 1 + 2$	(fulln. **and** conn.) **and** (thr. fulln. **or** thr. conn.)
$5 = 1 + 4$	(fulln. **or** conn.) **and** (thr. fulln. **and** thr. conn.)
$6 = 2 + 4$	(fulln. **and** conn.) **or** (thr. fulln. **and** thr. conn.)
$7 = 1 + 2 + 4$	(fulln. **and** conn.) **and** (thr. fulln. **and** thr. conn.)

TPABLO: (Szyld, Choi 1996)

- TPABLO1: $F = 1$, $\vartheta = 2$, $\zeta = 1/2n$
- TPABLO2: $F = 5$, $\vartheta = 1$, $\zeta = 1$

XPABLO default: $F = 0$, $\vartheta = 2$

Matrix Splitting Preconditioners

Idea: Find a part of A that can easily be factorized and use it as a preconditioner (idea known from Jacobi, Gauss-Seidel etc.)

Matrix Splitting Preconditioners

Idea: Find a part of A that can easily be factorized and use it as a preconditioner (idea known from Jacobi, Gauss-Seidel etc.)

In the XPABLO framework:

1. Find splitting $PAP^T = M - N$
2. LU-factorize $M = Q_1 LU Q_2^T$
3. Use M as preconditioner

Matrix Splitting Preconditioners

Idea: Find a part of A that can easily be factorized and use it as a preconditioner (idea known from Jacobi, Gauss-Seidel etc.)

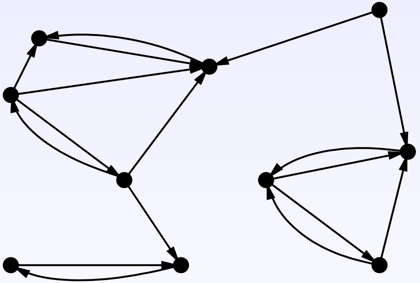
In the XPABLO framework:

1. Find splitting $PAP^T = M - N$
2. LU-factorize $M = Q_1 LU Q_2^T$
3. Use M as preconditioner

How to Choose M:

- Diagonal Blocks
- Tridiagonal Blocks
- Maximum Spanning Tree in Quotient Graph

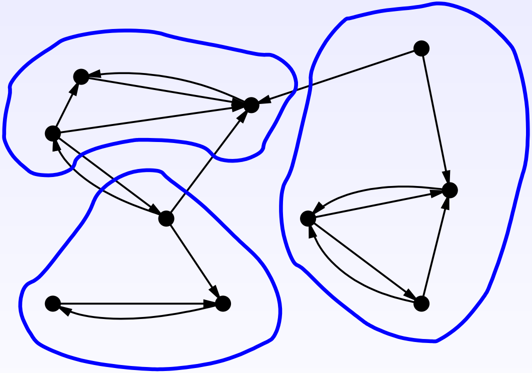
Graph Theory II - Quotient Graphs



- directed graph $G(A)$
- clusters found by (T/X)PABLO (blue)
- connections between clusters (green) $\rightsquigarrow$ edges in quotient graph
- quotient graph (green)

weights := sum of squares of all edge weights between clusters

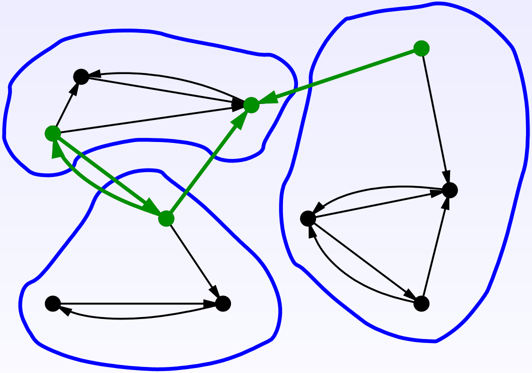
Graph Theory II - Quotient Graphs



- directed graph $G(A)$
- clusters found by (T/X)PABLO (blue)
- connections between clusters (green) $\rightsquigarrow$ edges in quotient graph
- quotient graph (green)

weights := sum of squares of all edge weights between clusters

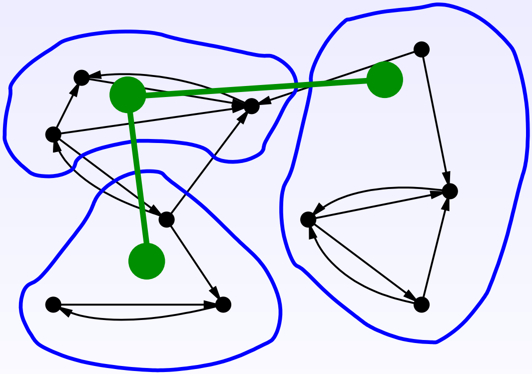
Graph Theory II - Quotient Graphs



- directed graph $G(A)$
- clusters found by (T/X)PABLO (blue)
- connections between clusters (green) $\rightsquigarrow$ edges in quotient graph
- quotient graph (green)

weights := sum of squares of all edge weights between clusters

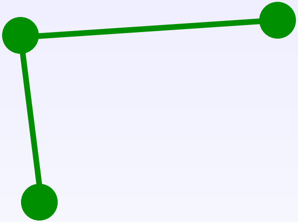
Graph Theory II - Quotient Graphs



- directed graph $G(A)$
- clusters found by (T/X)PABLO (blue)
- connections between clusters (green) $\rightsquigarrow$ edges in quotient graph
- quotient graph (green)

weights := sum of squares of all edge weights between clusters

Graph Theory II - Quotient Graphs



- directed graph $G(A)$
- clusters found by (T/X)PABLO (blue)
- connections between clusters (green) $\rightsquigarrow$ edges in quotient graph
- quotient graph (green)

weights := sum of squares of all edge weights between clusters

Block Tree Factorization

Situation: Find max. spanning tree in quotient graph $\rightsquigarrow$ prec. M

Proposition: Let $M \in \mathbb{R}^{n \times n}$. If

- Quotient graph of M is a tree
- blocks are numbered appropriately
- all $D_i := M_{ii} - \sum_{k=1}^{i-1} M_{ik} D_k^{-1} M_{ki}$ with $i = 1, \dots, q$ are non-singular

then we can factorize M as

$$M = \begin{pmatrix} D_1 & & & 0 \\ M_{21} & \ddots & & \\ \vdots & & \ddots & \\ M_{q1} & \cdots & \cdots & D_q \end{pmatrix} \begin{pmatrix} I & D_1^{-1} M_{12} & \cdots & D_1^{-1} M_{1q} \\ & \ddots & & D_2^{-1} M_{2q} \\ & & \ddots & \vdots \\ 0 & & & I \end{pmatrix}$$

Preprocessing: Scaling and Transversals

Scaling:

- Row-Column scaling $A \to RAC$ s.t.

$$\max_j \left\{ \left| (RAC)_{ij} \right| \right\} = 1 = \max_i \left\{ \left| (RAC)_{ij} \right| \right\}$$

Preprocessing: Scaling and Transversals

Scaling:

- Row-Column scaling $A \rightarrow RAC$ s.t.

$$\max_j \left\{ \left| (RAC)_{ij} \right| \right\} = 1 = \max_i \left\{ \left| (RAC)_{ij} \right| \right\}$$

Transversals:

- **Maximum Transversal (MT):** Find permutation $A \rightarrow PA$ s.t. $(PA)_{ii} \neq 0$ for all i.
- **Bottleneck Transversal (BT):** (Duff, Koster 1999) Find a maximum transversal which maximizes

$$\min_{i=1}^{n} |(PA)_{ii}|$$

Choosing the Parameters

Determined by Tests: $F = 0$, $\alpha = 1.1$, $\beta = 0.6$ and $\zeta = 1/2n$

Choosing the Parameters

Determined by Tests: $F = 0$, $\alpha = 1.1$, $\beta = 0.6$ and $\zeta = 1/2n$

“Seemed reasonable”: $\vartheta = 2$, $minbs = 2$, $maxbs = 100$

Choosing the Parameters

Determined by Tests: $F = 0$, $\alpha = 1.1$, $\beta = 0.6$ and $\zeta = 1/2n$

"Seemed reasonable": $\vartheta = 2$, $minbs = 2$, $maxbs = 100$

Difficult: Threshold parameter γ

- Fix γ s.t. a fraction of γ' entries are $< \gamma$
  Used in tests: $\gamma' = \{0.5, 0.6, 0.7, 0.8, 0.9\}$
  Best start: $\gamma' = 0.7$
- mean value of the absolute values of the non-zeros

Choosing the Parameters

Determined by Tests: $F = 0$, $\alpha = 1.1$, $\beta = 0.6$ and $\zeta = 1/2n$

"Seemed reasonable": $\vartheta = 2$, $minbs = 2$, $maxbs = 100$

Difficult: Threshold parameter γ

- Fix γ s.t. a fraction of γ' entries are $< \gamma$
  Used in tests: $\gamma' = \{0.5, 0.6, 0.7, 0.8, 0.9\}$
  Best start: $\gamma' = 0.7$
- mean value of the absolute values of the non-zeros

Preconditioners: Block Diagonal (D), Block Tridiagonal (T), Spanning Tree (S)

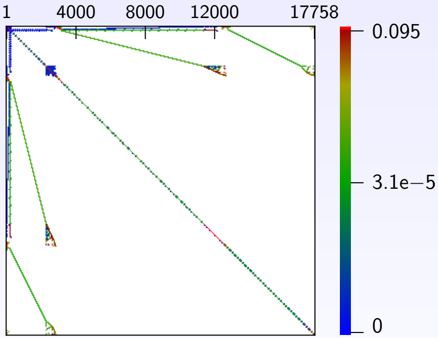
Solvers: BiCGSTAB, QMR, GMRES(20)

Numerical Results I

matrix	n	nnz	source	application
BCIRCUIT	68902	375558	UF/Hamm	circuit design
CIRCUIT_4	80209	307604	UF/Bomhof	circuit simulation
ECL32	51993	380415	UF/Sanghavi	semiconductor device simulat
HCIRCUIT	105676	513072	UF/Hamm	circuit design
IGBT3	10938	130500	UF/Shenk	semiconductor device simulat
LANGUAGE	399130	1216334	UF/Tromble	natural-language processing
MEMPLUS	17758	99147	MM/Hamm	circuit design
MULT_DCOP_01	25187	193276	UF/Sandia	circuit simulation
MULT_DCOP_02	25187	193276	UF/Sandia	circuit simulation
MULT_DCOP_03	25187	193216	UF/Sandia	circuit simulation
NMOS3	18588	237130	UF/Shenk	semiconductor device simulat
PESA	11738	79566	UF/Gaertner	??
SCIRCUIT	170998	958936	UF/Hamm	circuit design
WANG3	26064	177168	UF/Wang	semiconductor device simulat
ZHAO1	33861	166453	UF/Zhao	electromagnetic

Numerical Results II: MEMPLUS

The matrix MEMPLUS is part of the HAMM collection.



Matrix MEMPLUS real unsymmetric	
size	17758 × 17758
non-zero entries	99147
diagonal entries	17758
bandwidth	17751

	min	max	mean	median
M	1.4e−26	1.4	8.2e−3	4.0e−5
D	7.6e−3	1.4	0.026	0.018

Figure: Spy plot of matrix MEMPLUS

Numerical Results III: MEMPLUS (cont'd)

MEMPLUS	(X)PABLO num	(X)PABLO mean	BiCGSTAB flag	BiCGSTAB iter	time in seconds P	LU	solve	total
without preconditioner	–	–	0	**463.5**	0.01	–	21.97	**21.98**
diagonal preconditioner	–	–	0	771.5	0.01	–	36.50	36.51
XPABLO ($F = 0$, $\alpha = 1.1$, $\beta = 0.6$, $\zeta = 1/2n$, $\vartheta = 2$, $minbs = 2$, $maxbs = 100$)								
S $\gamma = 0.54$ ($\gamma' = 0.7$)	1614	11.00	0	9	0.40	0.22	0.49	1.11
T $\gamma = 0.54$ ($\gamma' = 0.7$)	1614	11.00	0	170.5	0.32	0.17	8.77	9.26
D $\gamma = 0.54$ ($\gamma' = 0.7$)	1614	11.00	0	218.5	0.33	0.08	11.18	11.59
S $\gamma = 0.0075$ ($\gamma' = 0.6$)	1533	11.58	0	**5.5**	0.42	0.23	0.30	0.95
T $\gamma = 0.0075$ ($\gamma' = 0.6$)	1533	11.58	0	381.5	0.33	0.18	19.37	19.88
D $\gamma = 0.0075$ ($\gamma' = 0.6$)	1533	11.58	0	289.5	0.34	0.08	14.72	15.14
S $\gamma = 0.31$	1614	11.00	0	9	0.40	0.23	0.49	1.12
T $\gamma = 0.31$	1614	11.00	0	170.5	0.31	0.18	8.91	9.40
D $\gamma = 0.31$	1614	11.00	0	218.5	0.31	0.08	11.13	11.52

Only subset of all results printed!!

Numerical Results IV

matrix	time in seconds direct	w/o	diag	XPAB
BCIRCUIT	3.72	–	–	316.15
IGBT3	2.58	–	–	5.22
MULT_DCOP_01	2.15	–	–	4.34
MULT_DCOP_02	1.67	–	–	2.56
PESA	1.93	–	–	5.58
SCIRCUIT	45.8	–	–	363.60
CIRCUIT_4	1282	–	–	70.02
ECL32	1770	–	78	30.98
HCIRCUIT	10.2	107	113	7.19
LANGUAGE	–	33	36	22.39
MEMPLUS	17.1	8.42	8.45	0.85
MULT_DCOP_03	7.53	7.67	3.44	1.47
NMOS3	8.64	–	–	5.00
WANG3	206	11.5	11.5	8.19
ZHAO1	77.6	4.11	4.09	3.40

XPAB = best of all XPABLO results

Conclusion

Conclusion:

- Flexible high performance framework
- Good for circuit design and similar applications
- Still work in progress (e.g. good values of ϑ, *minbs*, *maxbs*)

Outlook:

- Better implementation (block tree factorization, solving)
- Transversals: **MPS** (Duff, Koster 2001) Find $A \rightarrow PRAC$ s.t.

$$(PRAC)_{ii} = 1 \qquad \text{for } i = 1, \ldots, n$$
$$(PRAC)_{ij} \leq 1 \qquad \text{for } i \neq j$$

$\rightsquigarrow$ More systems can be solves iteratively