Block Preconditioning for Markov Chain Problems

2006 Markov Anniversary Meeting
Charleston, SC
13 June 2006

Michele Benzi
Department of Mathematics
and Computer Science
Emory University
Atlanta, GA

Acknowledgements

Thanks to:

- Bora Uçar (Emory)
- NSF (Computational Mathematics Program)

Preprint available at

http://www.mathcs.emory.edu/~benzi

Outline

- Problem description
- Block preconditioning
- Spectral analysis
- Matrix partitioning
- Experimental results using MARCA models
- Conclusions

Problem description

We consider the computation of the stationary probability distribution vector of ergodic, homogeneous, finite Markov chains.

Let N denote the number of states, and let $P = [p_{ij}] \in \mathbf{R}^{N \times N}$ denote the transition matrix of the chain. Then P is non-negative, row-stochastic and irreducible.

Let $A = I_N - P^T$ be the corresponding rate matrix. Note that A is a singular, irreducible M-matrix.

Perron–Frobenius Theorem $\Rightarrow$ rank$\,(A) = N - 1$.

The problem is then to find a nonzero vector $\mathbf{x} \in \mathbf{R}^N$ such that $A\mathbf{x} = \mathbf{0}$.

Problem description (cont.)

Here we assume that P (therefore, A) is large and sparse.

However, we make the assumption that the nonzero entries of A are explicitly available.

We consider preconditioned GMRES for solving the singular, homogeneous system $A\mathbf{x} = \mathbf{0}$.

We focus on developing effective preconditioners that can be efficiently implemented in parallel.

A main ingredient is the use of graph partitioning to reorder the matrix into a suitable block structure.

Block preconditioning

Assume that A is an irreducible M-matrix that has been partitioned in the following block 2×2 form:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is $n \times n$ and A_{22} is $m \times m$, with $n + m = N$. Typically, $n > m$.

Then it is well known that:

1. A_{11} is a nonsingular M-matrix.

2. The Schur complement $S := A_{22} - A_{21}A_{11}^{-1}A_{12}$ is an irreducible M-matrix. It is nonsingular iff A is. If A is singular, then S has rank $m - 1$.

Block preconditioning (cont.)

Assume now that A is nonsingular. Then A has the block LU factorization

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_n & O \\ A_{21}A_{11}^{-1} & I_m \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ O & S \end{bmatrix}.$$

Block preconditioning (cont.)

Assume now that A is nonsingular. Then A has the block LU factorization

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_n & O \\ A_{21}A_{11}^{-1} & I_m \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ O & S \end{bmatrix}.$$

Letting

$$P_T = \begin{bmatrix} A_{11} & A_{12} \\ O & S \end{bmatrix}$$

Block preconditioning (cont.)

Assume now that A is nonsingular. Then A has the block LU factorization

$$A = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right] = \left[\begin{array}{cc} I_n & O \\ A_{21}A_{11}^{-1} & I_m \end{array} \right] \left[\begin{array}{cc} A_{11} & A_{12} \\ O & S \end{array} \right].$$

Letting

$$P_T = \left[\begin{array}{cc} A_{11} & A_{12} \\ O & S \end{array} \right]$$

we have that $\sigma(AP_T^{-1}) = \sigma(P_T^{-1}A) = \{1\}$.

That is, the matrix AP_T^{-1} (and hence $P_T^{-1}A$) has only the eigenvalue $\lambda = 1$ (of multiplicity N).

Block preconditioning (cont.)

Furthermore, the preconditioned matrix $M = AP_T^{-1}$ satisfies $(M - I)^2 = O$. Hence, the minimum polynomial of M has degree 2, which implies that GMRES will converge to the solution of $A\mathbf{x} = \mathbf{0}$ in at most two steps (in exact arithmetic).

Block triangular preconditioners of the form P_T have been studied by Murphy, Golub and Wathen (SISC, 2000) and by others for solving saddle point problems where, typically,

$$A_{12} = A_{21}^T \quad \text{and} \quad A_{22} = O\,.$$

This approach has been extended to more general matrices by Ipsen (SISC, 2001). Here we study the application of block triangular preconditioning to Markov chains. A major difference is that now the 2×2 block structure is not given by the problem, but it must be imposed.

Block preconditioning (cont.)

The "ideal" block triangular preconditioner P_T is not practical. In practice one uses as preconditioner a block triangular matrix of the form

$$\mathcal{P}_T = \begin{bmatrix} \widehat{A}_{11} & A_{12} \\ O & \widehat{S} \end{bmatrix}$$

where $\widehat{A}_{11} \approx A_{11}$ and $\widehat{S} \approx S$ are invertible approximations to A_{11} and S.

Linear systems with $\widehat{A}_{11}$ and $\widehat{S}$ must be "easy" to solve. At the same time, the approximations must be good enough so as to retain fast convergence of GMRES.

Spectral analysis

For Markov chain problems, A (and hence S) is singular. The preconditioned matrix, therefore, will have have the simple eigenvalue $\lambda = 0$, and a cluster around $\lambda = 1$. The better the approximations $\widehat{A}_{11} \approx A_{11}$ and $\widehat{S} \approx S$, the tighter the cluster.

It is not easy to bound the eigenvalues of the preconditioned matrix, in general. Some simple results can be obtained by assuming that $\widehat{A}_{11} = A_{11}$ and that

$$\widehat{S} = A_{22} - A_{21} M_{11}^{-1} A_{21} \quad \text{where} \quad O \le M_{11}^{-1} \le A_{11}^{-1} \,.$$

Note that the last inequality cannot be an equality.

The conditions on $\widehat{S}$ are satisfied if M_{11} is obtained from A_{11} by deletion of off-diagonal entries. For instance, $M_{11} = \text{diag}(A_{11})$ will do, unless A_{11} is itself diagonal.

Spectral analysis (cont.)

Theorem: Let A be a singular, irreducible M-matrix partitioned in block 2×2 form. Let P_T be a block triangular preconditioner with

$$\widehat{A}_{11} = A_{11}, \quad \widehat{S} = A_{22} - A_{21} M_{11}^{-1} A_{12}$$

where $M_{11} \neq A_{11}$ satisfies $O \leq M_{11}^{-1} \leq A_{11}^{-1}$.

Then the spectrum of $P_T^{-1} A$ consists of:

- The simple eigenvalue $\lambda = 0$;
- The eigenvalue $\lambda = 1$ of multiplicity at least n;
- A cluster of at most $m - 1$ eigenvalues lying in the disk $D(1,1) = \{z \in \mathbf{C}\,;\, |z-1| < 1\}$. The diameter of this cluster goes to zero as $\|S - \widehat{S}\| \to 0$.

Moreover, the splitting $A = P_T - (P_T - A)$ is weak regular of the II kind; that is, $P_T^{-1} \geq O$ and $I - AP_T^{-1} \geq O$.

Matrix partitioning

The spectral analysis of the preconditioned matrix suggests that the choice of blocks should be such that n, the size of the (1,1) block A_{11}, is as large as possible, so as to maximize the number of eigenvalues at or near $\lambda = 1$.

Computational considerations, on the other hand, impose that the A_{11} block should be easy to (approximately) invert. This is because we need to solve linear systems with A_{11} and also because the inverse of A_{11} appears in the definition of the Schur complement.

Also, in view of a possible parallel implementation, A_{11} should be a block diagonal matrix.

Matrix partitioning (cont.)

Thus, for a given integer p, we would like to find a reordering (symmetric permutation) of A into the block form

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] = \left[\begin{array}{cccc|c} A_1 & & & & B_1 \\ & A_2 & & & B_2 \\ & & \ddots & & \vdots \\ & & & A_p & B_p \\ \hline C_1 & C_2 & \cdots & C_p & A_S \end{array}\right],$$

with the size of the block $A_{22} = A_S$ as small as possible.

With such a partitioning we have

$$A_{11} = \text{diag}\,(A_1, A_2, \ldots, A_p) \quad \text{and} \quad S = A_S - \sum_{i=1}^{p} C_i A_i^{-1} B_i\,.$$

Note that S is well-defined since each A_i is an invertible M-matrix.

Matrix partitioning (cont.)

Such a reordering of A can be obtained using graph partitioning by vertex separator (GPVS) techniques.

Given an undirected graph $\mathcal{G} = (V, E)$ and an integer p, the p-way GPVS problem consists of finding a set of vertices V_S of minimum size whose removal decomposes a graph into p disconnected subgraphs $V_1, V_2, \ldots, V_p$ with balanced sizes. The problem is NP-hard.

We use standard graph partitioning software (METIS) applied to the undirected graph $\mathcal{G}$ associated with the symmetrized matrix $A + A^T$. Note that

$$|V| = N \quad \text{and} \quad |E| = nnz(A + A^T)\,.$$

Matrix partitioning (cont.)

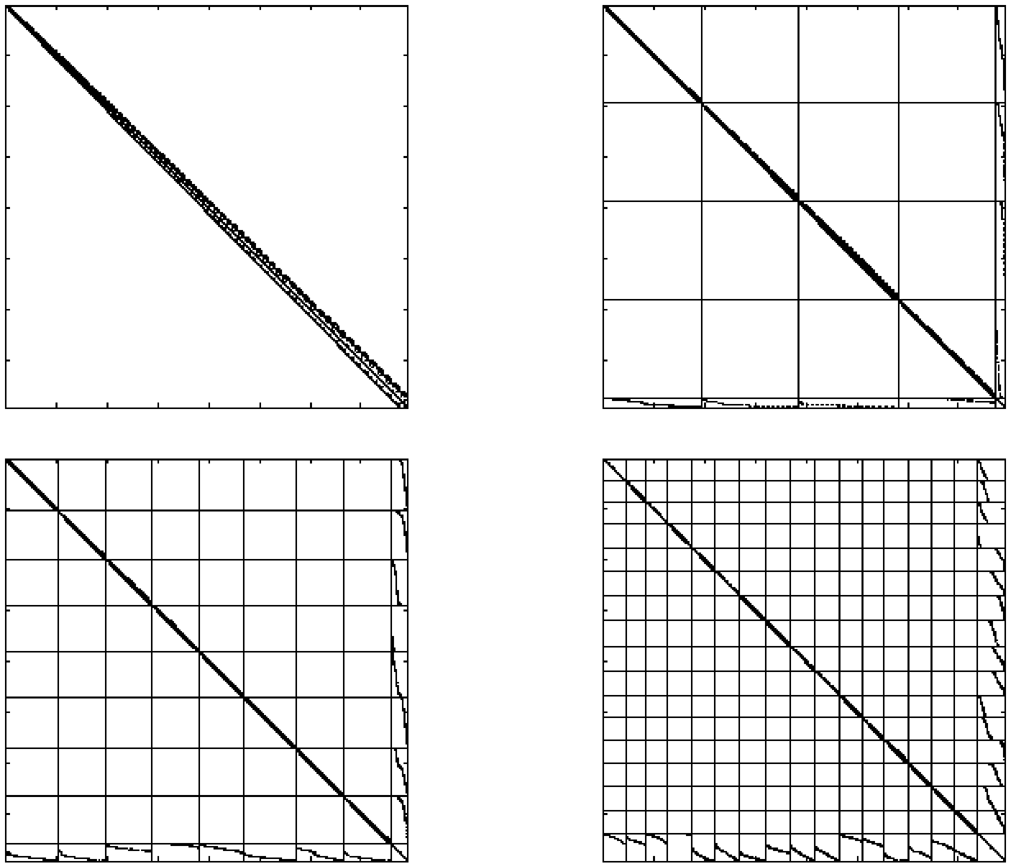
The matrix A can be put into the 2×2 block structure

$$A = \left[\begin{array}{cccc|c} A_1 & & & & B_1 \\ & A_2 & & & B_2 \\ & & \ddots & & \vdots \\ & & & A_p & B_p \\ \hline C_1 & C_2 & \cdots & C_p & A_S \end{array}\right]$$

by permuting the rows and columns associated with the vertices in $\bigcup_k V_k$ before the rows and columns associated with the vertices in V_S. That is, V_S defines the rows and columns of the (2,2) block $A_{22} \equiv A_S$.

Hence, the size of the separator set is $|V_S| = m$.

Example: MARCA matrix ‘qnatm06’



Matrix partitioning (cont.)

We use this block structure to define block triangular preconditioners of the form

$$P_T = \begin{bmatrix} \widehat{A}_1 & & & & B_1 \\ & \widehat{A}_2 & & & B_2 \\ & & \ddots & & \vdots \\ & & & \widehat{A}_p & B_p \\ & & & & \widehat{S} \end{bmatrix}$$

where $\widehat{A}_i \approx A_i$ for $i = 1 : p$ and $\widehat{S}$ is a sparse approximation to the Schur complement. Taking just $\widehat{S} \approx A_S$ leads to a block Gauss–Seidel preconditioner with inexact block solves.

Lower triangular and block diagonal variants could also be used.

Matrix partitioning (cont.)

In our code, we use ILUTH (threshold-based incomplete LU factorization) to build the diagonal blocks $\hat{A}_i$:

$$\hat{A}_i = L_i U_i \approx A_i, \quad i = 1 : p\,.$$

The approximate Schur complement $\hat{S}$ is obtained in two stages. First we compute

$$\bar{S} = A_S - \sum_{i=1}^{p} B_i M_i^{-1} C_i \quad \text{where} \quad M_i^{-1} \approx A_i^{-1} \quad \text{for} \quad i = 1 : p\,,$$

then we use ILUTH to approximate $\bar{S}$.

For the approximate inverses $M_i^{-1} \approx A_i^{-1}$ we found simple diagonal approximations to be sufficient. Note that each A_i is diagonally dominant (by columns).

Experimental results

- Solver: GMRES(50) with right preconditioning
- MATLAB 7.1.0 implementation
- 2.2 GHz dual core AMD Opteron (4GB main memory)
- Random $\mathbf{x}_0$, stopping criterion: $\|A\mathbf{x}_k\| < 10^{-10}$
- Test matrices from MARCA (W. J. Stewart)

Matrix	number of rows/cols	number of nonzeros					
		total	average	row		col	
	N		row/col	min	max	min	max
mutex09	65535	1114079	17.0	16	17	16	17
mutex12	263950	4031310	15.3	9	21	9	21
ncd07	62196	420036	6.8	2	7	2	7
ncd10	176851	1207051	6.8	2	7	2	7
qnatm06	79220	533120	6.7	3	9	4	7
qnatm07	130068	875896	6.7	3	9	4	7
tcomm16	13671	67381	4.9	2	5	2	5
tcomm20	17081	84211	4.9	2	5	2	5
twod08	66177	263425	4.0	2	4	2	4
twod10	263169	1050625	4.0	2	4	2	4

Experimental results (cont.)

- The 'mutex' matrices behave differently from the others:
  1. Preconditioning helps but is not essential;
  2. Separator set is huge ($m \approx \frac{N}{5}$ already for $p = 2$).
- Timings (secs.) for Block Gauss-Seidel, Product Splitting, and Block Triangular prec.
- Block Gauss-Seidel with $p = 8$ is best for this problem.

Matrix	GMRES	Preconditioned GMRES						
	Total time		Total time					
			Set-up			Solve		
	Solve	p	BGS	PS	BT	BGS	PS	BT
mutex09	7.0	2	3.60	4.53	4.30	1.82	1.96	1.42
		4	1.52	3.33	2.95	1.82	2.37	1.51
		8	**1.04**	3.67	3.16	**1.99**	2.86	1.72
		16	1.06	4.44	3.62	2.60	3.94	2.07
		32	1.45	5.47	4.16	3.63	5.54	2.82
mutex12	27.2	2	17.95	21.75	20.72	9.18	8.07	7.23
		4	5.49	12.63	11.10	8.05	8.29	6.21
		8	**3.37**	14.58	12.36	**7.79**	9.98	6.46
		16	3.42	16.11	12.69	9.47	12.13	7.47
		32	4.62	19.81	13.97	13.35	17.92	9.13

Experimental results (cont.)

- Experiments with 'mutex' matrices
- Average iteration counts for different values of p and different preconditioners
- BJ = Block Jacobi, BD = Block Diagonal, BGS = Block Gauss-Seidel, PS = Product Splitting, BT = Block Triangular

Matrix	GMRES	Preconditioned GMRES					
			Preconditioners				
		p	BJ	BD	BGS	PS	BT
mutex09	97	2	24	24	13	9	9
		4	27	27	14	9	9
		8	28	28	14	8	9
		16	29	29	15	9	9
		32	29	30	15	9	9
mutex12	91	2	26	27	14	8	10
		4	28	29	15	8	9
		8	28	29	15	8	9
		16	29	31	16	7	9
		32	30	31	16	7	8

Experimental results (cont.)

- Experiments with 'ncd' matrices
- BJ = Block Jacobi, BD = Block Diagonal, BGS = Block Gauss-Seidel, PS = Product Splitting, BT = Block Triangular
- Here '250' means no convergence within 250 iterations

Matrix	GMRES	Preconditioned GMRES					
			Preconditioners				
		p	BJ	BD	BGS	PS	BT
ncd07	250	2	40	68	23	12	12
		4	212	250	87	13	13
		8	222	250	101	15	15
		16	243	250	129	16	16
		32	250	250	192	18	18
ncd10	250	2	122	168	57	14	15
		4	160	249	47	15	15
		8	244	250	105	15	15
		16	250	250	145	18	17
		32	250	250	215	19	19

Experimental results (cont.)

- Iteration counts for ‘qnatm’ matrices
- BJ = Block Jacobi, BD = Block Diagonal, BGS = Block Gauss-Seidel, PS = Product Splitting, BT = Block Triangular
- Here ‘250’ means no convergence within 250 iterations

Matrix	GMRES	Preconditioned GMRES					
			Preconditioners				
		p	BJ	BD	BGS	PS	BT
qnatm06	250	2	60	60	38	36	36
		4	78	83	41	36	36
		8	104	19	46	39	39
		16	177	181	55	42	43
		32	223	250	83	46	47
qnatm07	250	2	51	56	40	39	39
		4	83	91	44	41	41
		8	110	138	51	44	44
		16	163	186	74	47	47
		32	232	246	92	55	56

Experimental results (cont.)

- Iteration counts for 'tcomm' matrices
- BJ = Block Jacobi, BD = Block Diagonal, BGS = Block Gauss-Seidel, PS = Product Splitting, BT = Block Triangular
- Here '250' means no convergence within 250 iterations

Matrix	GMRES	Preconditioned GMRES					
			Preconditioners				
		p	BJ	BD	BGS	PS	BT
tcomm16	250	2	30	30	18	16	16
		4	48	51	27	21	21
		8	112	131	39	29	29
		16	250	250	85	42	42
		32	250	250	105	46	46
tcomm20	250	2	33	34	20	17	18
		4	55	61	29	23	23
		8	132	157	43	32	32
		16	247	250	94	44	44
		32	250	250	221	94	98

Experimental results (cont.)

- Iteration counts for 'twod' matrices
- BJ = Block Jacobi, BD = Block Diagonal, BGS = Block Gauss-Seidel, PS = Product Splitting, BT = Block Triangular
- Here '250' means no convergence within 250 iterations

Matrix	GMRES	Preconditioned GMRES					
			Preconditioners				
		p	BJ	BD	BGS	PS	BT
twod08	250	2	28	27	15	10	10
		4	41	40	22	13	13
		8	49	48	25	18	18
		16	59	61	29	24	23
		32	92	93	36	29	29
twod10	250	2	38	38	21	18	18
		4	47	47	26	21	21
		8	58	58	30	24	24
		16	77	78	35	28	28
		32	94	94	39	32	32

Experimental results (cont.)

- Timings for larger matrix of each remaining type

Matrix		Total time					
		Set-up			Solve		
	p	BGS	PS	BT	BGS	PS	BT
ncd10	2	1.18	1.93	1.68	19.40	5.39	4.13
	4	1.04	1.99	1.59	15.85	6.16	4.01
	8	1.01	2.31	**1.58**	33.95	7.74	**3.85**
	16	1.01	3.04	1.67	45.84	12.34	4.30
	32	1.09	4.55	2.00	71.01	20.82	5.03
qnatm07	2	3.68	4.18	4.04	11.43	13.62	11.20
	4	2.34	2.98	2.71	11.69	15.12	11.02
	8	1.66	2.53	**2.04**	13.23	18.73	**11.23**
	16	1.34	2.75	1.84	18.07	26.62	12.13
	32	1.30	3.72	1.96	23.85	46.35	14.73
tcomm20	2	0.10	0.13	**0.11**	0.43	0.50	**0.40**
	4	0.10	0.10	0.10	0.63	0.74	0.50
	8	0.10	0.18	0.10	0.93	1.28	0.65
	16	0.10	0.20	0.10	1.98	2.47	0.97
	32	0.10	0.30	0.10	5.01	8.30	2.21
twod10	2	7.04	7.92	**7.64**	10.46	10.90	**8.89**
	4	6.65	7.55	7.15	12.85	14.32	10.43
	8	4.95	6.47	5.64	14.41	18.83	11.17
	16	4.09	6.43	4.89	16.78	30.60	13.00
	32	3.02	7.17	4.18	18.27	53.05	13.82

Conclusions

- Block triangular preconditioning is promising
- Results are fairly stable with respect to p
- Comparisons on MARCA models suggest the method is often superior to other techniques with similar complexity and storage requirements
- Future work:

  1. Larger problems
  2. Parallel implementation
  3. What to do when the separator set is huge?
  4. Multilevel version?