Parallel preconditioning for time-dependent PDE problems

Andy Wathen
Oxford University, UK

joint work with
Elle McDonald (CSIRO, Australia),
Jen Pestana (University of Strathclyde)
Anthony Goddard (Oxford University)

Krylov subspace methods

For self-adjoint problems/symmetric matrices, iterative methods of choice exist: conjugate gradients for SPD, MINRES otherwise

but many possible methods for non-self-adjoint problems/nonsymmetric matrices: GMRES , BICGSTAB , QMR , IDR , . . .

Krylov subspace methods

For self-adjoint problems/symmetric matrices, iterative methods of choice exist: conjugate gradients for SPD, MINRES otherwise

but many possible methods for non-self-adjoint problems/nonsymmetric matrices: GMRES , BICGSTAB , QMR , IDR , . . .

For almost all need *preconditioning*

Preconditioner $\mathbf{P}$ such that

$$\text{“}\mathbf{P}^{-1}\mathbf{B}\mathbf{x} = \mathbf{P}^{-1}\mathbf{b}\text{”}$$

has much faster convergence with the appropriate iterative method than $\mathbf{B}\mathbf{x} = \mathbf{b}$.

Krylov subspace methods

Arises because of convergence guarantees:

- for symmetric matrices: descriptive convergence bounds based on eigenvalues $\Rightarrow$ a priori estimates of iterations for acceptable convergence; good preconditioning ensures fast convergence.
- for nonsymmetric matrices: by contrast, to date there are no generally applicable *and descriptive* convergence bounds even for GMRES ; for any of the other nonsymmetric methods without a minimisation property, convergence theory is extremely limited $\Rightarrow$ no good a priori way to identify what are the desired qualities of a preconditioner

A major theoretical difficulty, but heuristic ideas abound!

Nonsymmetric problems

$$\mathcal{L}u = f$$

where

$$\langle \mathcal{L}u, v \rangle \neq \langle u, \mathcal{L}v \rangle$$

$\langle \cdot, \cdot \rangle$ is any inner product

Classic examples:

- convection-diffusion (or most odd/even derivative problems)
- time-dependent problems

  (since $\langle u_t, v \rangle = -\langle u, v_t \rangle$.)

Time-dependent problems: ODEs

$$y' = ay + f, \qquad y(t_0) = y_0$$

discretise: e.g.

$$\frac{\mathrm{y}^{k+1} - \mathrm{y}^k}{\tau} = \theta a \mathrm{y}^{k+1} + (1-\theta) a \mathrm{y}^k + f^k, \quad \mathrm{y}^0 = y_0,$$

$k = 0, 1, \ldots, \ell$ with $\ell\tau = T$ gives

$$B \underbrace{\begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^\ell \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \tau f^1 + (1 + a(1-\theta)\tau)\mathrm{y}^0 \\ \tau f^2 \\ \tau f^3 \\ \vdots \\ \tau f^\ell \end{bmatrix}}_{\mathbf{f}},$$

where the $\ell \times \ell$ coefficient matrix B is

$$\begin{bmatrix} b & & & & \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix},$$

$b = 1 - a\theta\tau$, $c = -1 - a(1-\theta)\tau$.

i.e. B is a bidiagonal Toeplitz matrix.

Now use *Pestana & W, 2015*:

If **B** is a real Toeplitz matrix then

$$\underbrace{\begin{bmatrix} a_0 & a_{-1} & \cdot & \cdot & a_{1-n} \\ a_1 & a_0 & a_{-1} & \cdot & \cdot \\ \cdot & a_1 & a_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & a_{-1} \\ a_{n-1} & \cdot & \cdot & a_1 & a_0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} 0 & 0 & \cdot & 0 & 1 \\ 0 & \cdot & 0 & 1 & 0 \\ \cdot & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & 0 & 0 \end{bmatrix}}_{\mathbf{Y}}$$

is the real *symmetric* (Hankel) matrix

$$\begin{bmatrix} a_{1-n} & \cdot & \cdot & a_{-1} & a_0 \\ \cdot & \cdot & a_{-1} & a_0 & a_1 \\ \cdot & \cdot & a_0 & a_1 & \cdot \\ a_{-1} & \cdot & \cdot & \cdot & \cdot \\ a_0 & a_1 & \cdot & \cdot & a_{n-1} \end{bmatrix}$$

Thus MINRES can be robustly applied to $\mathbf{BY}$ — it is symmetric but generally indefinite — and its convergence will depend only on eigenvalues.

BUT preconditioning? – needs to be symmetric and positive definite for MINRES

Based on well-know Circulant approximations, $\mathbf{C}$, which are diagonalised by an FFT in $O(n \log n)$ work: $\mathbf{C} = \mathbf{F}^{\star}\mathbf{\Lambda}\mathbf{F}$, use

$$|\mathbf{C}| = \mathbf{F}^{\star}|\mathbf{\Lambda}|\mathbf{F}$$

which is real symmetric and positive definite

Theorem (*Pestana & W, 2015*)

$$|\mathbf{C}|^{-1}\mathbf{B}\mathbf{Y} = \mathbf{J} + \mathbf{R} + \mathbf{E}$$

where $\mathbf{J}$ is real symmetric and orthogonal with eigenvalues ± 1, $\mathbf{R}$ is of small rank and $\mathbf{E}$ is of small norm

$\Rightarrow$ guaranteed fast convergence because MINRES convergence only depends on eigenvalues which are clustered around ± 1 except for few outliers!

Theorem (*Pestana & W, 2015*)

$$|\mathsf{C}|^{-1}\mathsf{BY} = \mathsf{J} + \mathsf{R} + \mathsf{E}$$

where **J** is real symmetric and orthogonal with eigenvalues ± 1, **R** is of small rank and **E** is of small norm

$\Rightarrow$ guaranteed fast convergence because MINRES convergence only depends on eigenvalues which are clustered around ± 1 except for few outliers!

For the ODE problem ($\boldsymbol{\tau = 0.2}$, $\boldsymbol{a = -0.3}$,$\boldsymbol{\theta = 0.8}$):

$\boldsymbol{\ell}$	$\boldsymbol{\kappa(B)}$	Iterations
10	10.474	4
100	30.852	4
1000	33.887	4

Multistep method: BDF2

$$\frac{\mathrm{y}^{k+1} - \frac{4}{3}\mathrm{y}^{k} + \frac{1}{3}\mathrm{y}^{k-1}}{\tau} = \frac{2}{3} a \mathrm{y}^{k+1} + \frac{2}{3} f^{k+1},$$

with $\mathrm{y}^0 = y_0$ and $\mathrm{y}^{-1} = y_{-1}$ leads to the monolithic or all-at-once system

$$B \underbrace{\begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^\ell \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \frac{2}{3}\tau f^1 + \frac{4}{3}\mathrm{y}^0 - \frac{1}{3}\mathrm{y}^{-1} \\ \frac{2}{3}\tau f^2 - \frac{1}{3}\mathrm{y}^0 \\ \frac{2}{3}\tau f^3 \\ \vdots \\ \frac{2}{3}\tau f^\ell \end{bmatrix}}_{\mathbf{f}}$$

where the coefficient matrix B is

$$\begin{bmatrix} 1-\frac{2}{3}a\tau & & & & \\ -\frac{4}{3} & 1-\frac{2}{3}a\tau & & & \\ \frac{1}{3} & -\frac{4}{3} & 1-\frac{2}{3}a\tau & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{3} & -\frac{4}{3} & 1-\frac{2}{3}a\tau \end{bmatrix}.$$

Same approach:

ℓ	$\kappa(B)$	Iterations
10	29.33	6
100	67.49	6
1000	67.67	6

Thus we have *proved* and observed fast convergence for

$$|\mathbf{C}|^{-1}\mathbf{BY}\mathbf{y} = |\mathbf{C}|^{-1}\mathbf{f}.$$

To observe if it is more generally

Krylov friendly (Tim Kelley),

we try GMRES simply for

$$\mathbf{C}^{-1}\mathbf{B}\mathbf{y} = \mathbf{C}^{-1}\mathbf{f}$$

and find it is *quicker.*

PDEs: diffusion problem

$$\begin{aligned} u_t &= \Delta u + f && \text{in } \Omega \times (0,T], \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3, \\ u &= g && \text{on } \partial\Omega, \\ u(x,0) &= u_0(x) && \text{at } t = 0 \end{aligned}$$

Discretize - finite elements, mesh size h, and n spatial dofs:

$$M \frac{\mathrm{u}_k - \mathrm{u}_{k-1}}{\tau} + K\mathrm{u}_k = \mathrm{f}_k, \quad k = 1, \ldots, \ell,$$

(M: mass matrix, K: stiffness matrix) or

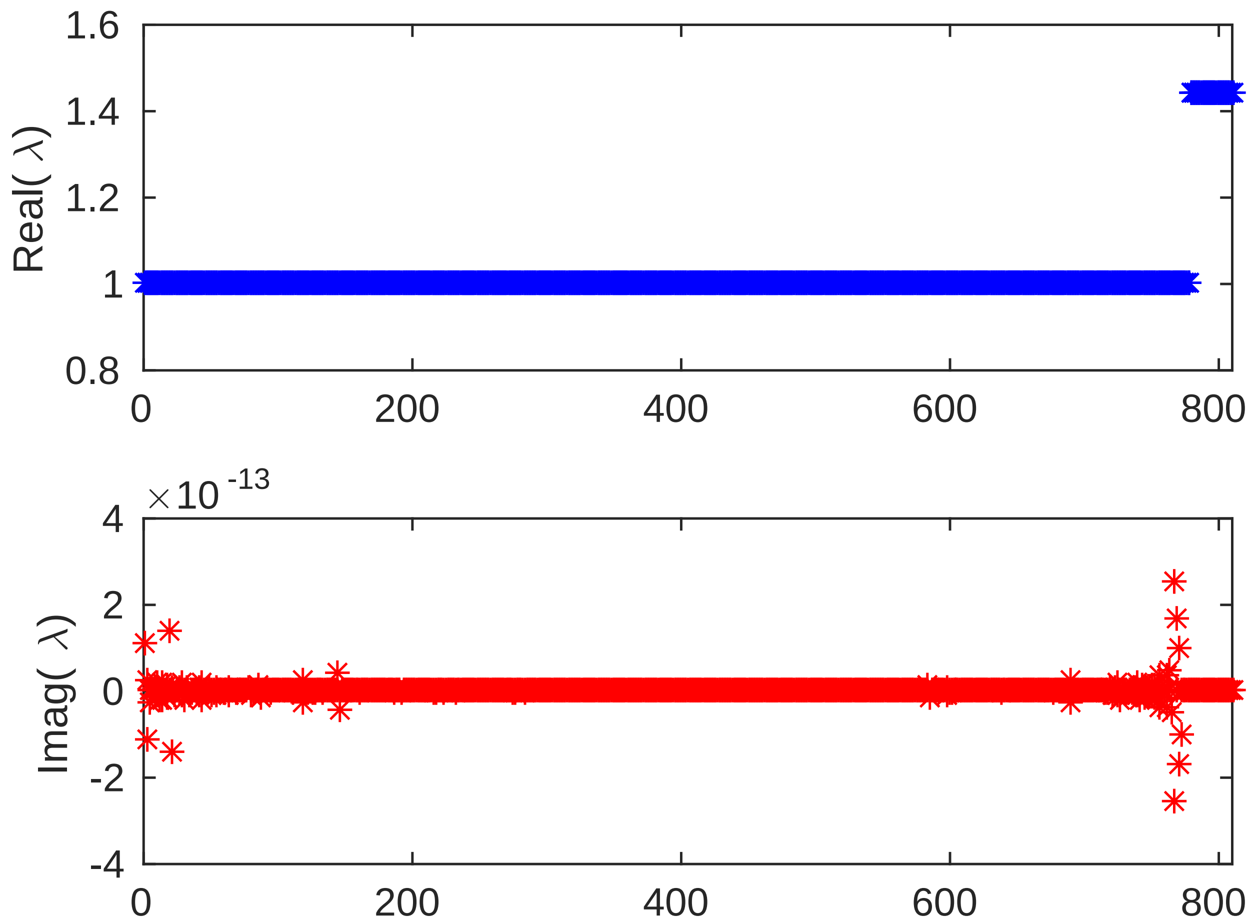
$$\mathcal{A}_{BE}\mathrm{x} := \begin{bmatrix} A_0 & & & \\ A_1 & A_0 & & \\ & \ddots & \ddots & \\ & & A_1 & A_0 \end{bmatrix} \begin{bmatrix} \mathrm{u}_1 \\ \mathrm{u}_2 \\ \vdots \\ \mathrm{u}_\ell \end{bmatrix} = \begin{bmatrix} M\mathrm{u}_0 + \tau\mathrm{f}_1 \\ \tau\mathrm{f}_2 \\ \vdots \\ \tau\mathrm{f}_\ell \end{bmatrix},$$

where $A_0 = M + \tau K$ is symmetric positive definite and $A_1 = -M$ is symmetric.

We use the block circulant preconditioner

$$\mathcal{P}_{BE} := \begin{bmatrix} A_0 & & & A_1 \\ A_1 & A_0 & & \\ & \ddots & \ddots & \\ & & A_1 & A_0 \end{bmatrix}.$$

Theorem (*McDonald, Pestana & W, 2018*)
$\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$ is diagonalisable, has $(\ell-1)n$ eigenvalues of 1 and n eigenvalues which cluster around 1 for small h.



The eigenvalues of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$, $n = 81$, $\ell = 10$ and $\tau = 0.1$.

Kronecker Product form

If

$$\Sigma = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & & & 1 \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} = U\Lambda U^*,$$

then

$$\begin{aligned} \mathcal{A}_{BE} &= I_\ell \otimes A_0 + \Sigma \otimes A_1, \\ \mathcal{P}_{BE} &= I_\ell \otimes A_0 + C \otimes A_1, \end{aligned}$$

and using $(W \otimes X)(Y \otimes Z) = (WY \otimes XZ)$

$$\mathcal{P}_{BE} = I_\ell \otimes A_0 + C \otimes A_1 = (U \otimes I_n)[I_\ell \otimes A_0 + \Lambda \otimes A_1](U^* \otimes I_n)$$

so

$$\mathcal{P}_{BE}^{-1} = (U \otimes I_n)[I_\ell \otimes A_0 + \Lambda \otimes A_1]^{-1}(U^* \otimes I_n)$$

$$\mathcal{P}_{BE}^{-1} = (U \otimes I_n)[I_\ell \otimes A_0 + \Lambda \otimes A_1]^{-1}(U^* \otimes I_n)$$

so $\mathcal{P}_{BE}^{-1} r$ requires

- multiplication of r by $U \otimes I_n$ and $U^* \otimes I_n$; parallel over n processors?
- inversion of the block diagonal matrix $I_\ell \otimes A_0 + \Lambda \otimes A_1$: easier if A_i are symmetric as for diffusion, but could use AMG when A_i are non symmetric.

lead to fast parallel execution?

Heat Equation

Times (seconds) for solving $\mathcal{P}^{-1}\mathcal{A}\mathrm{U} = \mathcal{P}^{-1}\mathrm{b}$ with GMRES (tol $= 10^{-5}$). p is the number of processors.

ℓ	n	$p=1$	$p=2$	$p=4$	$p=8$	$p=16$	$p=32$
	320	77.72	29.26	15.32	8.95	5.11	3.34
768	**512**	152.64	57.54	32.71	17.52	11.54	6.69
	768	245.47	97.77	50.81	30.71	16.66	9.65
	320	146.67	54.68	28.40	17.059	10.35	6.07
1024	**512**	265.22	107.07	60.86	34.13	20.40	11.75
	768	459.12	198.94	101.23	55.85	28.55	16.12
	320	325.14	124.67	63.64	39.78	22.74	13.06
1440	**512**	646.81	239.65	123.44	72.44	40.95	22.50
	768	979.85	432.46	215.77	114.99	59.80	32.41
1440	**1568**	2119.91	815.93	431.13	218.24	118.62	63.30

Wave Equation

Times (seconds) for solving $\mathcal{R}_{BD2}^{-1}\mathcal{C}_{BD2}\mathrm{U} = \mathcal{R}_{BD2}^{-1}\mathrm{b}_{BD2}$ with GMRES (tol $= 10^{-5}$). p is the number of processors.

ℓ	n	$p=1$	$p=2$	$p=4$	$p=8$	$p=16$	$p=32$
	320	79.07	31.29	16.20	9.53	6.11	4.36
768	**512**	163.68	61.33	34.33	19.76	11.54	7.09
	768	251.37	100.99	53.14	27.31	14.64	11.09
	320	153.37	53.39	30.47	20.99	12.38	7.39
1024	**512**	287.23	119.72	65.84	39.81	23.23	12.08
	768	497.12	222.93	115.24	60.84	32.46	18.01
	320	328.17	125.71	65.64	41.70	23.32	14.21
1440	**512**	680.15	243.53	124.65	73.92	41.42	24.51
	768	960.33	434.46	211.01	115.95	60.54	35.12
1440	**1568**	2211.97	820.21	444.31	230.42	122.63	68.10

In a similar manner for multistep methods:

$$\mathcal{A} := \begin{bmatrix} A_0 & & & & & \\ A_1 & A_0 & & & & \\ \vdots & \ddots & \ddots & & & \\ A_p & & \ddots & \ddots & & \\ & \ddots & & A_1 & A_0 & \\ & & A_p & \cdots & A_1 & A_0 \end{bmatrix},$$

$$\mathcal{P} = (U \otimes I_n)\mathcal{G}(U^* \otimes I_n),$$

where $\mathcal{G} = \mathrm{diag}(G_1, \ldots, G_\ell)$ and $G_j = \sum_{i=0}^{p} \lambda_j^i A_i$.

Can use

$$\mathcal{Y} := \begin{bmatrix} & & I_n \\ & \cdot^{\cdot} & \\ & I_n & \\ I_n & & \end{bmatrix} = Y \otimes I_n, \; Y = \begin{bmatrix} & & 1 \\ & \cdot & \\ 1 & & \end{bmatrix}$$ as before

to symmetrize any block Toeplitz matrix with symmetric blocks and use a SPD absolute value preconditioner as before.

When A_i are not symmetric (e.g. convection-diffusion problems) GMRES/FGMRES are necessary

Theory: MINRES for $|\mathcal{P}|^{-1}\mathcal{Y}\mathcal{A}$ guaranteed to converge in a number of iterations independent of ℓ

Practice:

- very few MINRES iterations required
- GMRES with $\mathcal{P}$ does better, but no guarantee!
- AMG (AGMG - Y. Notay) also few iterations

Numerics: Heat Eqn, Backwards Euler

n	ℓ	DoF	GMRES $\mathcal{P}^{-1}\mathcal{A}$	MINRES $\vert\mathcal{P}\vert^{-1}\mathcal{Y}\mathcal{A}$	FGMRES $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$
289	2^4	4624	3	11	8
	2^6	18496	3	13	8
	2^8	73984	3	15	8
	2^{10}	295936	3	19	8
	2^{12}	1183744	3	18	7
	2^{14}	4734976	3	16	7
1089	2^4	17424	3	10	8
	2^6	69696	3	13	8
	2^8	278784	3	14	8
	2^{10}	1115136	3	18	8
	2^{12}	4460544	3	20	7
	2^{14}	17842176	3	19	6
4225	2^4	67600	3	10	15
	2^6	270400	3	11	16
	2^8	1081600	3	13	16
	2^{10}	4326400	3	18	16
	2^{12}	17305600	3	20	17
	2^{14}	69222400	2	19	16

Numerics: Heat Eqn, BDF2

n	ℓ	DoF	GMRES $\mathcal{P}^{-1}\mathcal{A}$	MINRES $\lvert\mathcal{P}\rvert^{-1}\mathcal{Y}\mathcal{A}$	FGMRES $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$
289	2^4	4624	3	13	7
	2^6	18496	3	16	8
	2^8	73984	3	19	8
	2^{10}	295936	3	21	7
	2^{12}	1183744	3	24	7
	2^{14}	4734976	3	22	6
1089	2^4	17424	3	13	8
	2^6	69696	3	15	8
	2^8	278784	3	18	8
	2^{10}	1115136	3	22	7
	2^{12}	4460544	3	24	7
	2^{14}	17842176	3	25	6
4225	2^4	67600	3	11	15
	2^6	270400	3	13	16
	2^8	1081600	3	18	16
	2^{10}	4326400	3	21	17
	2^{12}	17305600	3	24	17
	2^{14}	69222400	3	25	16

Numerics: Convection-diffusion, BE

n	ℓ	DoF	GMRES $\mathcal{P}^{-1}\mathcal{A}$	FGMRES $\mathcal{P}_{\text{MG}}^{-1}\mathcal{A}$
289	2^4	4624	13	12
	2^6	18496	13	12
	2^8	73984	13	12
	2^{10}	295936	13	12
	2^{12}	1183744	13	12
	2^{14}	4734976	13	12
1089	2^4	17424	12	12
	2^6	69696	13	12
	2^8	278784	13	12
	2^{10}	1115136	13	12
	2^{12}	4460544	13	12
	2^{14}	17842176	13	12
4225	2^4	67600	12	22
	2^6	270400	12	22
	2^8	1081600	12	23
	2^{10}	4326400	12	23
	2^{12}	17305600	12	23
	2^{14}	69222400	12	23

References and Acknowledgement

McDonald, E, Pestana, J. & Wathen, A.J., 2018,
'Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations',
accepted for publication in SIAM J. Sci. Comput.

Goddard, A & Wathen, A.J., 'A note on parallel preconditioning for all-at-once evolutionary PDEs',
submitted to ETNA, 2018.

Pestana, J. & Wathen, A.J., 2015,
'A preconditioned MINRES method for nonsymmetric Toeplitz matrices',
SIAM J. Matrix Anal. Appl. **36**, pp. 273–288.

Wathen, A.J., 2015,
'Preconditioning',
Acta Numerica (2015) **24**, pp. 329–376.

This work is partially supported by Award No. KUK-C1-013-04 made by King Abdullah University of Science and Technology (KAUST).