Parallel preconditioning for time-dependent PDEs and PDE control

Andy Wathen
Oxford University, UK

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

Systems/PDEs

Symmetric/self-adjoint

vs

Nonsymmetric/non-self-adjoint

Iterative methods for linear systems

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 , . . .

Iterative methods for linear systems

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

$$``\mathbf{P}^{-1}\mathbf{B}\mathbf{x} = \mathbf{P}^{-1}\mathbf{b}"$$

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

Iterative methods: $Bx = c, \quad B \in \mathbb{R}^{n \times n}$

From $\mathbf{x_0}$ generate $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots\}$ using one matrix×vector product at each iteration:

$$\mathbf{Br}, \mathbf{B}(\mathbf{Br}), \ldots, \mathbf{B}^k\mathbf{r}, \ldots$$

$\Rightarrow$ Krylov subspace methods:

$$\mathbf{r}_k = \mathbf{p}_k(\mathbf{B})\mathbf{r}_0, \qquad \mathbf{r}_k = \mathbf{c} - \mathbf{B}\mathbf{x}_k, \quad \mathbf{p}_k \in \Pi_k, \mathbf{p}_k(0) = 1$$

so if $\mathbf{B} = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{-1}$ then

$$\|\mathbf{r}_k\| \leq \|\mathbf{X}\| \; \|\mathbf{p}_k(\boldsymbol{\Lambda})\| \; \|\mathbf{X}^{-1}\| \; \|\mathbf{r}_0\|$$

and if $\mathbf{B} = \mathbf{B}^T$ so that $\mathbf{X}^{-1} = \mathbf{X}^T$ then this bound on convergence in $\|\cdot\|_2$ depends only eigenvalues

Well distributed (clustered) eigenvalues $\Rightarrow$ fast convergence for symmetric matrices.

Nonsymmetric matrices: GMRES

- Eigenvalue/eigenvector bound:

$$\|r_k\|/\|r_0\| \leq \|X\| \, \|X^{-1}\| \min_{p\in\Pi_k, p(0)=1} \max_j |p(\lambda_j)|$$

- Field of Values bound:

$$W(B) = \{x^*Bx/x^*x; x \neq 0\}$$

$$\|r_k\|/\|r_0\| \leq 2 \min_{p\in\Pi_k, p(0)=1} \max_{z\in W(B)} |p(z)|$$

- Pseudospectral bound:

$$\Lambda_\epsilon(B) = \{z \in \mathbb{C} : \|(zI - B)^{-1}\| > \epsilon^{-1}\}$$

$$\|r_k\|/\|r_0\| \leq \mathcal{L}(\partial\Lambda_\epsilon)/2\pi\epsilon \min_{p\in\Pi_k, p(0)=1} \max_{z\in\Lambda_\epsilon(B)} |p(z)|$$

Krylov subspace methods

So difference between symmetric and non-symmetric problems 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 *guarantees* 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!

2 Theorems: $Bx = c$

If $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ is self-adjoint in $\langle\cdot,\cdot\rangle_{\boldsymbol{H}}$, $\quad \boldsymbol{H} = \boldsymbol{H}^T$ SPD

—means $\langle \boldsymbol{y}, \boldsymbol{z}\rangle_{\boldsymbol{H}} = \boldsymbol{z}^T\boldsymbol{H}\boldsymbol{y}$ and $\forall \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$ that $\langle \boldsymbol{B}\boldsymbol{y}, \boldsymbol{z}\rangle_{\boldsymbol{H}} = \langle \boldsymbol{y}, \boldsymbol{B}\boldsymbol{z}\rangle_{\boldsymbol{H}}$ ($\Leftrightarrow$ $\boldsymbol{H}\boldsymbol{B} = \boldsymbol{B}^T\boldsymbol{H}$)—

then for minimum residual methods (MINRES), iterates $\boldsymbol{x}_k$,

$$\|\boldsymbol{c} - \boldsymbol{B}\boldsymbol{x}_k\|_{\boldsymbol{H}} \leq \min_{p\in\Pi_k, p(0)=1} \max_j |p(\boldsymbol{\lambda}_j)| \, \|\boldsymbol{c} - \boldsymbol{B}\boldsymbol{x}_0\|_{\boldsymbol{H}}$$

$\boldsymbol{\lambda}_j \in \mathbb{R}$: eigenvalues of $\boldsymbol{B}$ (not $\boldsymbol{H}\boldsymbol{B}$)

and if $\boldsymbol{B}$ is also positive definite in $\langle\cdot,\cdot\rangle_{\boldsymbol{H}}$ (Conjugate Gradients) then

$$\|\boldsymbol{x} - \boldsymbol{x}_k\|_{\boldsymbol{H}\boldsymbol{B}} \leq \min_{p\in\Pi_k, p(0)=1} \max_j |p(\boldsymbol{\lambda}_j)| \, \|\boldsymbol{x} - \boldsymbol{x}_0\|_{\boldsymbol{H}\boldsymbol{B}}$$

$\boldsymbol{\lambda}_j \in \mathbb{R}$: eigenvalues of $\boldsymbol{B}$ (not $\boldsymbol{H}\boldsymbol{B}$)

2 Theorems: $Bx = c$

If $B \in \mathbb{R}^{n\times n}$ is self-adjoint in $\langle \cdot, \cdot \rangle_H$, $\qquad H = H^T$ SPD

—means $\langle y, z\rangle_H = z^T H y$ and $\forall y, z \in \mathbb{R}^n$ that $\langle By, z\rangle_H = \langle y, Bz\rangle_H$ ($\Leftrightarrow HB = B^T H$)—

then for minimum residual methods (MINRES), iterates x_k,

$$\|c - Bx_k\|_H \leq \min_{p\in\Pi_k, p(0)=1} \max_j |p(\lambda_j)| \, \|c - Bx_0\|_H$$

$\lambda_j \in \mathbb{R}$: eigenvalues of B (not HB)

makes sense because there exists such and H if and only if B is diagonalisable and has real eigenvalues.

2 Theorems: $Bx = c$

Given any nonincreasing positive sequence $f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0$ there exists $B \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}^n$ with $\|c - Bx_0\| = f(0)$ such that $f(k) = \|c - Bx_k\|$, $k = 1, \cdots, n-1$, where x_k is the iterate at step k of the GMRES algorithm applied to $Bx = c$, with initial iterate x_0. Moreover, the matrix B can be chosen to have any eigenvalues.

Greenbaum, Ptak & Strakos (1996)

To summarise

- for a symmetric system convergence of iterative methods can be bounded only in terms of eigenvalues
- badly distributed eigenvalues can give you bad convergence of iterative methods, but for nonsymmetric systems even with good eigenvalues one can get arbitrarily bad convergence.

Consequence:

- preconditioning for symmetric systems is well-founded: clustering eigenvalues *guarantees* fast convergence.
- preconditioning for nonsymmetric systems must generally be heuristic unless the preconditioner endows symmetry

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 $\boldsymbol{B}$ is

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

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

i.e. $\boldsymbol{B}$ 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

Fortunately it is well known that many Toeplitz matrices are well approximated by related circulant matrices, $\mathbf{C}$ (*Strang, 1986, Chan, 1988, Chan, 1989, Tyrtishnikov, 1996/7*) which are diagonalised by an FFT in $O(n \log n)$ work: $\mathbf{C} = \mathbf{F}^{\star}\mathbf{\Lambda}\mathbf{F}$,

For many symmetric Toeplitz matrices we have that the Strang or Optimal (Chan) circulant $\mathbf{C}$ satisfy

$$\mathbf{C}^{-1}\mathbf{B} = \mathbf{I} + \mathbf{R} + \mathbf{E}$$

where $\mathbf{R}$ is of small rank and $\mathbf{E}$ is of small norm

$\Rightarrow$eigenvalues clustered around 1 except for a few outliers

To ensure a symmetric and positive definite preconditioner for $\mathbf{BY}$ just use

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

which is real symmetric and positive definite

To ensure a symmetric and positive definite preconditioner for $\mathbf{BY}$ just use

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

which is real symmetric and positive definite

Theorem (*Pestana & W, 2015*)

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

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!

To ensure a symmetric and positive definite preconditioner for $\mathbf{BY}$ just use

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

which is real symmetric and positive definite

Theorem (*Pestana & W, 2015*)

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

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

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

ℓ	$\kappa(B)$	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

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

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:

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

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.

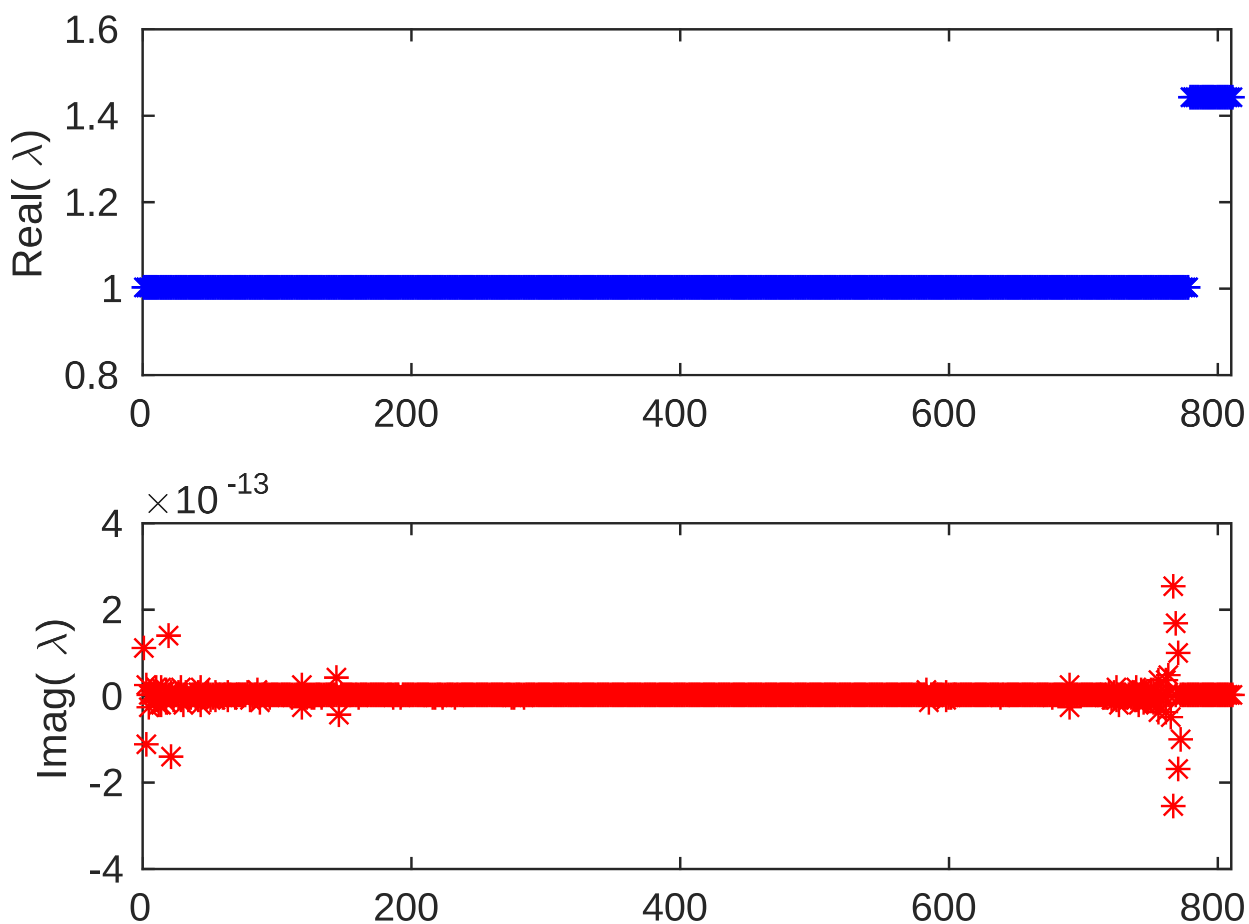

Figure 0: The eigenvalues of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$, $n = 81$, $\ell = 10$ and $\tau = 0.1$.

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

(63.30 $\times$ 32 = 2025.6)

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

(68.10 × 32 = 2179.2)

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

$$\mathcal{Y} := \begin{bmatrix} & & & \boldsymbol{I_n} \\ & & \cdot & \\ & \boldsymbol{I_n} & & \\ \boldsymbol{I_n} & & & \end{bmatrix} = \boldsymbol{Y} \otimes \boldsymbol{I_n}, \ \boldsymbol{Y} = \begin{bmatrix} & & \mathbf{1} \\ & \cdot & \\ \mathbf{1} & & \end{bmatrix} \text{ as before}$$

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

When $\boldsymbol{A_i}$ 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 itertions 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 $\lvert\mathcal{P}\rvert^{-1}\mathcal{Y}\mathcal{A}$	FGMRES $\mathcal{P}_{\mathrm{MG}}^{-1}\mathcal{A}$
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 $\|\mathcal{P}\|^{-1}\mathcal{Y}\mathcal{A}$	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

Control

Minimise

$$\frac{1}{2}\int_0^T\int_{\Omega_1}(\mathrm{y}(\mathrm{x},\mathrm{t})-\bar{\mathrm{y}}(\mathrm{x},\mathrm{t}))^2\,dxdt+\frac{\beta}{2}\int_0^T\int_{\Omega_2}(\mathrm{u}(\mathrm{x},\mathrm{t}))^2\,dxdt$$

subject to

$$y_t-\nabla^2 y=u \qquad \text{in } \Omega\times[0,T]$$

with boundary conditions $y=0$ on $\partial\Omega$ and initial condition $y(x,0)=y_0(x)$.

Adjoint PDE:

$$-p_t-\nabla^2 p=y-\bar{y}$$

with $p=0$ on $\partial\Omega, p(x,T)=y(x,T)-\bar{y}(x,T)$.

Backwards Euler in time, Galerkin finite elements in space

Discretise, Optimize $\Rightarrow$ saddle point system

$$\begin{bmatrix} M & 0 & K^T \\ 0 & \beta M & -M \\ K & -M & 0 \end{bmatrix} \begin{bmatrix} y \\ u \\ p \end{bmatrix} = r.h.s$$

where K is the discrete forward differential operator and K^T its adjoint

Main issue for time-dependent PDE is that K is monolithic as above, so *all* time steps values $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N$ are involved:

$$\underbrace{\begin{bmatrix} M+\tau K & & & & \\ -M & M+\tau K & & & \\ & -M & M+\tau K & & \\ & & \ddots & \ddots & \\ & & & -M & M+\tau K \end{bmatrix}}_{\underline{K}} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_\ell \end{bmatrix}$$

so

$$\begin{bmatrix} \underline{M} & 0 & -\underline{K}^T \\ 0 & \beta\underline{M} & \underline{M} \\ -\underline{K} & \underline{M} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \underline{M}\mathbf{y} \\ 0 \\ d \end{bmatrix}.$$

is a very large dimensional system, but matrix$\times$vector multiplication easy and block preconditioning ideas apply...

Main issue: approximation of the Schur complement

$$\underline{KM}^{-1}\underline{K}^T + \frac{1}{\beta}\underline{M}$$

via the Pearson matching idea

$$\frac{1}{\beta}\underline{M} \quad + \quad \underline{KM}^{-1}\underline{K}^T$$

$$= -\frac{2}{\sqrt{\beta}}\underline{K} + \left(\underline{K} + \frac{1}{\sqrt{\beta}}\underline{M}\right)\underline{M}^{-1}\left(\underline{K} + \frac{1}{\sqrt{\beta}}\underline{M}\right)^T$$

So we need to approximate $\underline{K} + \frac{1}{\sqrt{\beta}}\underline{M}$, (similarly its transpose) and we use the above monolithic ideas

(*Pearson & W (2010), Schoberl & Zulehner (2007)*)

Numerics: Heat Control

h	ℓ	DoF	GMRES iterations
2^{-4}	20	17340	16
2^{-5}	20	65340	17
2^{-6}	20	253500	17
2^{-7}	20	998460	18
2^{-8}	20	3962940	16
2^{-5}	20	65340	17
2^{-5}	40	130680	19
2^{-5}	60	196020	21
2^{-5}	80	261360	21
2^{-5}	100	326700	21

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',
SIAM J. Sci. Comput. **40**, pp. A1012–A1033.

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

Goddard, A. & Wathen, A.J., 2019,
'A note on parallel preconditioning for all-at-once evolutionary PDEs',
ETNA **51**, pp. 135–150.

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)