Parallel preconditioning for time-dependent PDEs

Andy Wathen
Oxford University, UK

Thanks:
Elle McDonald (CSIRO, Australia),
Jen Pestana (University of Strathclyde, UK),
Anthony Goddard (Durham University, UK)
Fede Danieli (Oxford University, UK)

Time-dependence: ODE IVP

$$y' = ay + f, \qquad y(0) = y_0$$

discretise: θ-method, integrate $0 \to T$

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

$k = 1, 2, \ldots, \ell$ with $\ell\tau = T$, τ being the time-step.

Time-dependence: ODE IVP

$$y' = ay + f, \qquad y(0) = y_0$$

discretise: θ-method, integrate $0 \to T$

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

$k = 1, 2, \ldots, \ell$ with $\ell\tau = T$, τ being the time-step.

As a linear all-at-once system this is

$$\underbrace{\begin{bmatrix} b & & & & \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^\ell \end{bmatrix}}_{\mathrm{y}} = \underbrace{\begin{bmatrix} \tau f^0 - c\mathrm{y}^0 \\ \tau f^1 \\ \tau f^2 \\ \vdots \\ \tau f^{\ell-1} \end{bmatrix}}_{\mathrm{f}},$$

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

Solve

$$\boldsymbol{B}\mathbf{y} = \begin{bmatrix} \boldsymbol{b} & & & & \\ \boldsymbol{c} & \boldsymbol{b} & & & \\ & \boldsymbol{c} & \boldsymbol{b} & & \\ & & \ddots & \ddots & \\ & & & \boldsymbol{c} & \boldsymbol{b} \end{bmatrix} \begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ \mathbf{y}^3 \\ \vdots \\ \mathbf{y}^\ell \end{bmatrix} = \begin{bmatrix} \boldsymbol{\tau f^0 - c}\mathbf{y}^0 \\ \boldsymbol{\tau f^1} \\ \boldsymbol{\tau f^2} \\ \vdots \\ \boldsymbol{\tau f^{\ell-1}} \end{bmatrix} = \mathbf{f},$$

- forward substitution $\rightarrow$ sequential—*causality*

Solve

$$\boldsymbol{B}\mathbf{y} = \begin{bmatrix} b & & & & \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix} \begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ \mathbf{y}^3 \\ \vdots \\ \mathbf{y}^\ell \end{bmatrix} = \begin{bmatrix} \tau f^0 - c\mathbf{y}^0 \\ \tau f^1 \\ \tau f^2 \\ \vdots \\ \tau f^{\ell-1} \end{bmatrix} = \mathbf{f},$$

- forward substitution $\rightarrow$ sequential—*causality*
- other (iterative) linear solver? Krylov subspace methods (Conjugate-Gradient-like methods) ie. methods that rely on a *matrix* $\times$ *vector* multiply

matrix $\times$vector multiply (`matvec`) is easily *parallelised*

Krylov subspace methods

need 1 `matvec` at each iteration $+$ vector operations

to compute iterate vectors

$$\mathbf{y}_k \in \text{span}\left\{\mathbf{f}, B\mathbf{f}, B^2\mathbf{f}, \ldots, B^{k-1}\mathbf{f}\right\}, \quad k = 1, 2, \ldots$$

when $\mathbf{y}_0 = \mathbf{0}$

$\Big[$ else

$$\mathbf{y}_k - \mathbf{y}_0 \in \text{span}\left\{\mathbf{r}_0, B\mathbf{r}_0, B^2\mathbf{r}_0, \ldots, B^{k-1}\mathbf{r}_0\right\}, \mathbf{r}_0 = \mathbf{f} - B\mathbf{y}_0\Big]$$

and an iterative method such as GMRES will compute the *optimal* vector in this Krylov subspace (the one that gives the smallest residual $\|\mathbf{f} - B\mathbf{y}_k\|_2$)

Thus an iterative solution method that converges in a number of iterations independent of ℓ can give an effective *parallel-in-time* (PinT) method for the time dependent problem.

With enough processors, time to solution (at *all* time-steps) is independent of the number of time-steps ℓ.

But consider $y' = ay, \;\; y(0) = y_0$ i.e. $f = 0 \Rightarrow$ all-at-once system

$$By = \begin{bmatrix} b & & & & \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix} \begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^\ell \end{bmatrix} = \begin{bmatrix} \tau f^0 - c\mathrm{y}^0 \\ \tau f^1 \\ \tau f^2 \\ \vdots \\ \tau f^{\ell-1} \end{bmatrix} = \mathrm{f},$$

But consider $y' = ay, \;\; y(0) = y_0$ i.e. $f = 0 \Rightarrow$ all-at-once system

$$\mathcal{B}\mathrm{y} = \begin{bmatrix} b & & & & \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix} \begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^\ell \end{bmatrix} = \begin{bmatrix} 0 - c\mathrm{y}^0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathrm{f},$$

i.e. f has only its first entry non-zero $\Rightarrow$

$$\mathrm{y}_1 \in \text{span}\left\{ \begin{bmatrix} \times \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}, \mathrm{y}_2 \in \text{span}\left\{ \begin{bmatrix} \times \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \times \\ \times \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}, \ldots$$

$$\mathbf{y}_k \in \text{span}\left\{ \begin{bmatrix} \times \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \times \\ \times \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \ldots, \begin{bmatrix} \times \\ \vdots \\ \times \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \right\}, \quad \boldsymbol{k = 1, 2, \ldots}$$

but solution is an exponential (non-zero for every time step)$\Rightarrow$ need ℓ iterations before all-at-once solution can be obtained.

Precisely: exact solution up to $k\tau$ at k^{th} iteration, zero for all other time steps: causality

thus solution for all ℓ time-steps at ℓ^{th} iteration

NOT a parallel-in-time method!!

Preconditioning

Instead of GMRES for $B\mathrm{y} = \mathrm{f}$, solve

$$P^{-1}B\mathrm{y} = P^{-1}\mathrm{f} \quad \text{or} \quad BP^{-1}\mathrm{x} = \mathrm{f}, \;\; \mathrm{y} = P^{-1}\mathrm{x}$$

where

- it is easy to solve $Pz = r$ for z given r: preconditioner
- GMRES converges in number of iterations independent of ℓ

Preconditioning

Instead of GMRES for $B\mathrm{y} = \mathrm{f}$, solve

$$P^{-1}B\mathrm{y} = P^{-1}\mathrm{f} \quad \text{or} \quad BP^{-1}\mathrm{x} = \mathrm{f}, \;\; \mathrm{y} = P^{-1}\mathrm{x}$$

where

- it is easy to solve $Pz = r$ for z given r: preconditioner
- GMRES converges in number of iterations independent of ℓ

One idea: precondition with a periodic solution

same ODE, same time-step, but initial condition replaced by $\mathrm{y}^\ell = \mathrm{y}^0$

Periodic preconditioning

$$y' = ay + f, \qquad y(0) = y(T)$$

leads to the *circulant* all-at-once system

$$\underbrace{\begin{bmatrix} b & & & & c \\ c & b & & & \\ & c & b & & \\ & & \ddots & \ddots & \\ & & & c & b \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} \mathrm{y}^1 \\ \mathrm{y}^2 \\ \mathrm{y}^3 \\ \vdots \\ \mathrm{y}^\ell \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \tau f^0 \\ \tau f^1 \\ \tau f^2 \\ \vdots \\ \tau f^\ell \end{bmatrix}}_{\mathbf{f}},$$

with b, c as before.
Such periodic systems can be efficiently solved with (parallel) FFT ($O(\ell \log \ell)$ operations in serial):

in fact P is diagonalised—$U \Lambda U^{-1}$—by FFT

Toeplitz/Circulant

Theory: B is a Toeplitz(constant-diagonal) matrix

If $B = B^T$ then the related circulant P is an excellent preconditioner for B: eigenvalues of $P^{-1}B$ all clustered at 1 except for a handful of outliers (*Strang (1986), Chan (1989)*).

If $B \neq B^T$ then can pre- or post-multiply by

$$Y = \begin{bmatrix} 0 & 0 & \cdot & 0 & 1 \\ 0 & \cdot & 0 & 1 & 0 \\ \cdot & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & 0 & 0 \end{bmatrix} \Rightarrow BY = \begin{bmatrix} 0 & \cdot & \cdot & 0 & b \\ \cdot & \cdot & 0 & b & c \\ \cdot & \cdot & b & c & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ b & c & \cdot & \cdot & 0 \end{bmatrix}$$

which is real symmetric, and related theory applies (*Pestana and W (2015)*) so a related (in fact an absolute value) circulant is also an excellent preconditioner in this case

Crucially,

- MINRES converges in a number of iterations independent of ℓ
- multiplication of a vector by B takes time independent of ℓ on a parallel machine
- solution of system with P is efficient (indep of ℓ?) in parallel

So circulant (periodic) preconditioned all-at-once system is a parallel-in-time method

What we can not prove, (but seems to be true!): GMRES with the relevant circulant preconditioner converges fast also.

Since there is currently no generally descriptive convergence bound for GMRES (or any other non-symmetric Krylov subspace iterative method) and no more specific results that cover this case!

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

ℓ	$\kappa(B)$	MINRES Iterations
10	10.474	4
100	30.852	4
1000	33.887	4

Limitations:

- fixed time-step, τ
- constant coefficient, a
- linear

but practical computation shows that variation can be tolerated without losing all useful properties: averaging along diagonals is a simple expedient, for example, for allowing circulant preconditioners (approximations) to be defined

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 $\boldsymbol{B}$ 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

But PDEs: we concentrate on

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

but have worked on similar ideas for simple hyperbolic wave equations

$$\begin{aligned} u_{tt} &= u_{xx} + f \quad \text{in } [a,b] \times (0,T], \\ u(a,t) &= \alpha \quad \text{and} \qquad u(b,t) = \beta, \\ u(x,0) &= u_0(x) \quad \text{and} \quad u_t(x,0) = u_0'(x) \qquad \text{at } t=0, \end{aligned}$$

(*Danieli & W (2021)*)

Galerkin finite elements/central finite differences in space, Backwards Euler in time:

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

or

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

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

$M \in \mathbb{R}^{n\times n}$: the mass matrix
$K \in \mathbb{R}^{n\times n}$: negative discrete Laplacian (stiffness matrix)

both are symmetric and positive definite.

Block lower triangular: block forwards substitution $\Rightarrow$ sequential solution (causality)

but precondition with a block circulant:

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

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.

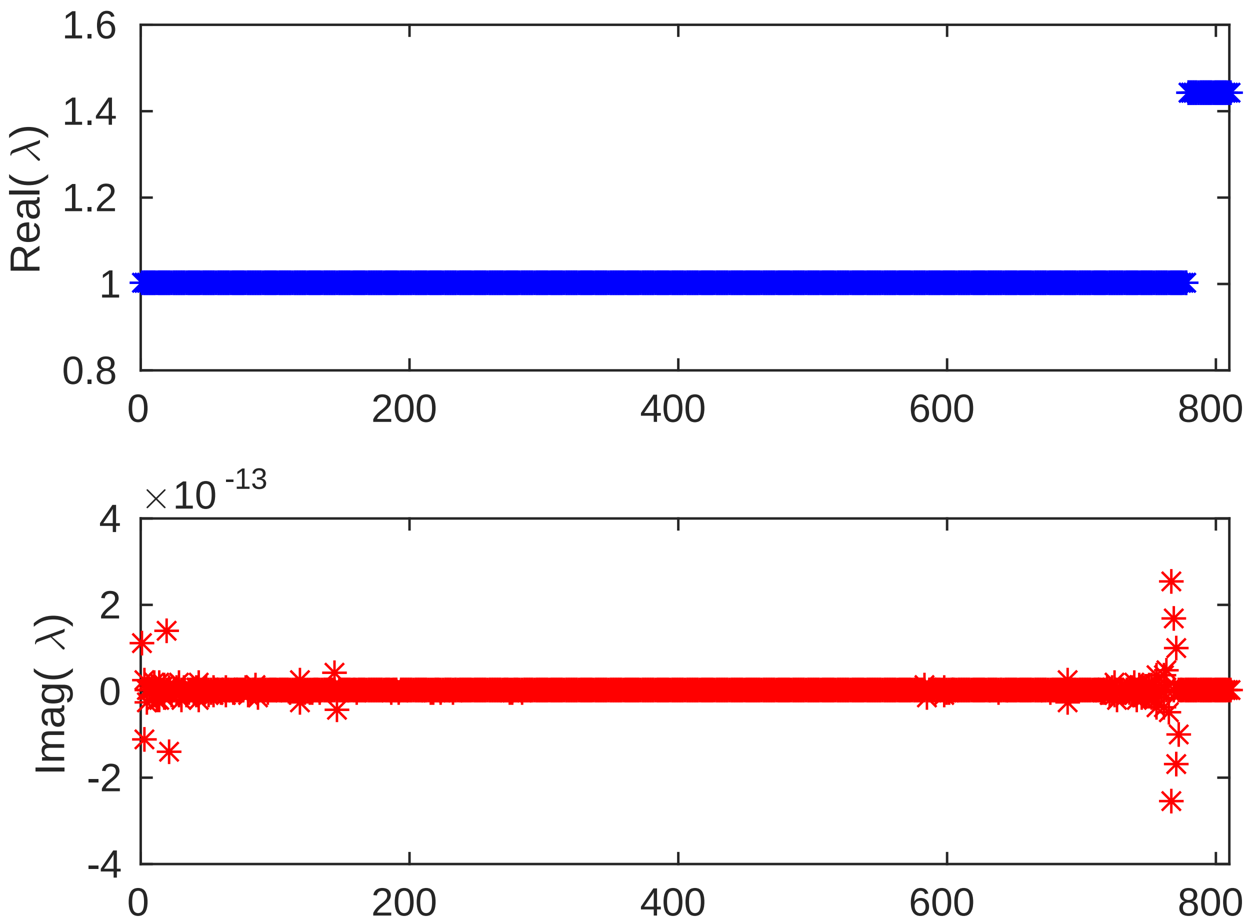

Figure 1: The eigenvalues of $\mathcal{P}_{BE}^{-1}\mathcal{A}_{BE}$, $n = 81$, $\ell = 10$ and $\tau = 0.1$.

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

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

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

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

- multiplication of $\boldsymbol{r}$ by $\boldsymbol{U \otimes I_n}$ and $\boldsymbol{U^* \otimes I_n}$; parallel over $\boldsymbol{n}$ processors?
- inversion of the block diagonal matrix $\boldsymbol{I_\ell \otimes A_0 + \Lambda \otimes A_1}$: easier if $\boldsymbol{A_i}$ are symmetric as for diffusion, but could use Algebraic Multigrid (AMG) when $\boldsymbol{A_i}$ 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	$p=1$	$p=2$	$p=4$	$p=8$	$p=16$	$p=32$
768	320	77.72	29.26	15.32	8.95	5.11	3.34
	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
1024	320	146.67	54.68	28.40	17.059	10.35	6.07
	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
1440	320	325.14	124.67	63.64	39.78	22.74	13.06
	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

Time: Störmer-Verlet - Space: Central-Differences

$\ell \rightarrow$ / $n \downarrow$	120	240	480	960	1920	3840	theoretical bound
80	15	17	26	39	70	136	161
160	–	16	19	30	53	95	321
320	–	–	18	21	33	76	641
640	–	–	–	19	22	35	1281
1280	–	–	–	–	21	23	2561
2560	–	–	–	–	–	24	5121

GMRES iterations (single processor)

Can use

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

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 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 $\lvert\mathcal{P}\rvert^{-1}\mathcal{Y}\mathcal{A}$	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 $\mathcal{P}_{\mathrm{MG}}^{-1}\mathcal{A}$
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

Summary

- Parallel computing over time is possible
- Causality can be circumvented
- simple initial results for linear constant coefficients, constant time-steps, but some variation allowable

References

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.

Danieli, F. & Wathen, A.J., 2021,
'All-at-once solution of linear wave equations',
Numer. Linear Algebra Appl. 2021;28:e2386.

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