Preconditioning for parallel-in-time

Andy Wathen Oxford University, UK



joint work with Fede Danieli (Oxford), Ben Southworth (Los Alamos)

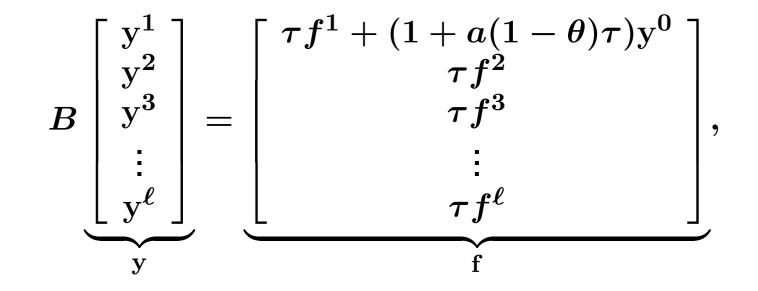
ODE IVP

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

discretise: e.g.

$$rac{\mathrm{y}^{k+1}-\mathrm{y}^k}{ au}= heta a\mathrm{y}^{k+1}+(1- heta)a\mathrm{y}^k+f^k, \hspace{1em}\mathrm{y}^0=y_0,$$

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



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

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

$$b=1-a heta au$$
 , $c=-1-a(1- heta) au$.

i.e. B is a bidiagonal Toeplitz (constant diagonal) matrix.

• forward substitution \rightarrow sequential—*causality*

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

From $x_0 = 0$ (typically) generate $\{x_1, x_2, \dots, x_k, \dots\}$ using one matrix × vector product at each iteration:

 $Bc, B(Bc), \ldots, B^kc, \ldots$ so that

 $x_1 \in \text{span}\{c\}, x_2 \in \text{span}\{c, Bc\}, \dots, x_k \in \text{span}\{c, Bc, \dots, B^{k-1}c\}, \dots$

 \Rightarrow Krylov subspace methods generally described by:

 $\mathsf{r}_{\mathsf{k}}=\mathsf{p}_{\mathsf{k}}(\mathsf{B})\mathsf{r}_{\mathsf{0}}, \qquad \mathsf{r}_{\mathsf{k}}=\mathsf{c}-\mathsf{B}\mathsf{x}_{\mathsf{k}}, \ \ \mathsf{p}_{\mathsf{k}}\in\Pi_{\mathsf{k}}, \mathsf{p}_{\mathsf{k}}(\mathsf{0})=1$

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

 $\|\mathbf{r}_{k}\| \leq \|\mathbf{X}\| \|\mathbf{p}_{k}(\mathbf{\Lambda})\| \|\mathbf{X}^{-1}\| \|\mathbf{r}_{0}\|$

and if $\mathbf{B} = \mathbf{B}^{\mathsf{T}}$ so that $\mathbf{X}^{-1} = \mathbf{X}^{\mathsf{T}}$ then this bound on convergence in $\|\cdot\|_2$ depends only on eigenvalues Well distributed (clustered) eigenvalues \Rightarrow fast convergence for symmetric matrices.

All-at-once system

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

$$B\mathbf{y} = egin{bmatrix} b & & & \ c & b & & \ & c & b & & \ & \ddots & \ddots & \ & & & c & b \end{bmatrix} egin{bmatrix} \mathbf{y}^1 \ \mathbf{y}^2 \ \mathbf{y}^3 \ \mathbf{y}^3 \ \mathbf{y}^3 \ \mathbf{y}^4 \end{bmatrix} = egin{bmatrix} au f^0 - c \mathbf{y}^0 \ au f^1 \ au f^2 \ \mathbf{y}^\ell \end{bmatrix} = \mathbf{f},$$

All-at-once system

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

i.e. f only its first entry non-zero \Rightarrow with zero initial guess

$$y_{1} \in \text{span}\left\{ \begin{bmatrix} \times \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}, y_{2} \in \text{span}\left\{ \begin{bmatrix} \times \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \times \\ \times \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}, \dots$$

$$\mathbf{y}_{k} \in \operatorname{span} \left\{ \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \times \\ \times \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} \times \\ \vdots \\ \times \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}, \quad k = 1, 2, \dots$$

but solution is an exponential (non-zero for every time step) \Rightarrow need ℓ iterations.

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 only at ℓ^{th} iteration

NOT a parallel-in-time method!!

Iterative methods for linear systems

This is true for any of the available iterative methods:

(For self-adjoint problems/symmetric matrices, iterative methods of choice exist: conjugate gradients for Symmetric Positive Definite matrices, MINRES otherwise)

and any of the many possible methods for non-self-adjoint problems/nonsymmetric matrices: GMRES, BICGSTAB, LSQR, QMR, IDR, ...

Iterative methods for linear systems

This is true for any of the available iterative methods:

(For self-adjoint problems/symmetric matrices, iterative methods of choice exist: conjugate gradients for Symmetric Positive Definite matrices, MINRES otherwise)

and any of the many possible methods for non-self-adjoint problems/nonsymmetric matrices: GMRES, BICGSTAB, LSQR, QMR, IDR, ...

But it is well know that for almost all systems we need *preconditioning*

Preconditioner **P** such that

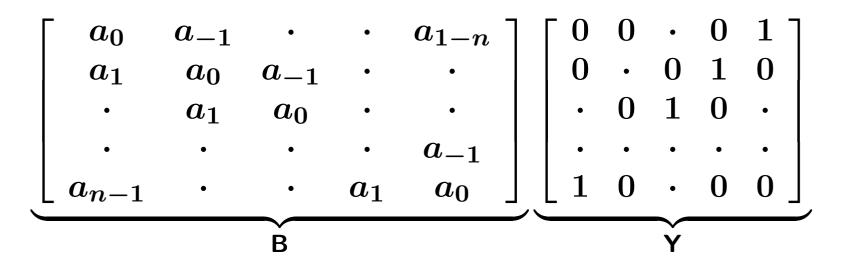
$$\mathbf{P}^{-1}\mathbf{B}\mathbf{x} = \mathbf{P}^{-1}\mathbf{b}^{\mathbf{n}}$$

has much faster convergence with the appropriate iterative method than Bx = b.

All-at-once system

A practical and guaranteed preconditioning approach: use *Pestana & W, 2015*:

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



is the real symmetric (Hankel) matrix

Thus MINRES can be robustly applied to **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, C (*Strang, 1986, Chan, 1988, Chan, 1989, Tyrtishnikov, 1996/7*) which are diagonalised by an FFT in $O(n \log n)$ work: $C = F^*\Lambda F$,

For many symmetric Toeplitz matrices we have that the Strang or Optimal (Chan) circulant **C** satisfy

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

where **R** is of small rank and **E** is of small norm

 \Rightarrow eigenvalues clustered around 1 except for a few outliers

For example, the Strang circulant for the standard Toeplitz matrix (as above) is

$$\begin{bmatrix} a_0 & a_{-1} & \dots & a_{-\lfloor \frac{n}{2} \rfloor} & a_{\lfloor \frac{n-1}{2} \rfloor} & \dots & a_2 & a_1 \\ a_1 & a_0 & a_{-1} & \dots & a_{-\lfloor \frac{n}{2} \rfloor} & a_{\lfloor \frac{n-1}{2} \rfloor} & \dots & a_2 \\ \dots & a_1 & a_0 & \ddots & \dots & a_{-\lfloor \frac{n}{2} \rfloor} & \ddots & \vdots \\ a_{\lfloor \frac{n}{2} \rfloor} & \dots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & a_{\lfloor \frac{n-1}{2} \rfloor} \\ a_{-\lfloor \frac{n-1}{2} \rfloor} & \ddots & \dots & \ddots & \ddots & \ddots & \ddots & a_{-\lfloor \frac{n}{2} \rfloor} \\ \vdots & \ddots & a_{\lfloor \frac{n}{2} \rfloor} & \dots & \ddots & a_0 & a_{-1} & \dots \\ a_{-2} & \dots & a_{-\lfloor \frac{n-1}{2} \rfloor} & a_{\lfloor \frac{n}{2} \rfloor} & \dots & a_1 & a_0 & a_1 \\ a_{-1} & a_{-2} & \dots & a_{-\lfloor \frac{n-1}{2} \rfloor} & a_{\lfloor \frac{n}{2} \rfloor} & \dots & a_1 & a_0 \end{bmatrix}$$

To ensure a symmetric and positive definite preconditioner for **BY** just use

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

which is real symmetric and positive definite

To ensure a symmetric and positive definite preconditioner for **BY** just use

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

which is real symmetric and positive definite

Theorem (*Pestana & W, 2015*)

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

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 **BY** just use

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

which is real symmetric and positive definite

Theorem (Pestana & W, 2015)

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

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

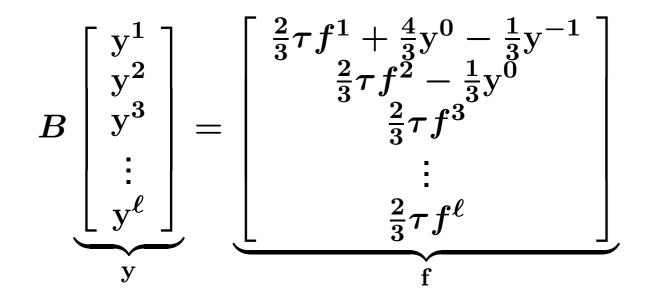
For the ODE problem (au=0.2, a=-0.3, heta=0.8):

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

Multistep method: BDF2

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

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



where the coefficient matrix B is

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

Same approach:

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

•

This gives a *parallel-in-time* method if all components are efficiently implemented in parallel.

Moreover we *observe* that GMRES with just **C** as preconditioner gives even better convergence (but no proof!)

Preconditioning

Note that the circulant preconditioner here simply represents preconditioning the Initial Value Problem

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

with the *nearby* periodic problem

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

for which Fourier technology gives rapid (and parallel) application

This approach also applies for systems of ODEs and for time-dependent PDEs \Rightarrow block Toeplitz/block circulant matrices for which standard parallel technologies can be applied in space and the periodic preconditioning only applied in time

PDEs: diffusion problem

 $egin{array}{rcl} u_t&=&\Delta u+f& ext{in }\Omega imes(0,T],&\Omega\subset\mathbb{R}^2 ext{ or }\mathbb{R}^3,\ u&=&g& ext{on }\partial\Omega,\ u(x,0)&=&u_0(x)& ext{ at }t=0 \end{array}$

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

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

or

$$\mathcal{A}_{BE} \mathrm{x} := \left[egin{array}{ccc} A_0 & & & \ A_1 & A_0 & & \ & \ddots & \ddots & \ & & A_1 & A_0 \end{array}
ight] \left[egin{array}{c} \mathrm{u}_1 \ \mathrm{u}_2 \ dots \ \mathrm{u}_\ell \end{array}
ight] = \left[egin{array}{c} M \mathrm{u}_0 + au \mathrm{f}_1 \ au \mathrm{f}_2 \ dots \ \ dots \ \ dots \ \ dots \ \ \ \$$

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

We use the block circulant preconditioner

$$\mathcal{P}_{BE} := \left[egin{array}{cccc} A_0 & & A_1 \ A_1 & A_0 & & \ & \ddots & \ddots & \ & & \ddots & \ddots & \ & & A_1 & A_0 \end{array}
ight]$$

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*.

 \Rightarrow fast and ℓ -independent convergence (with flip to be guaranteed) of the appropriate iterative methods

2nd part of talk: Incompressible Navier-Stokes

$$\begin{cases} \frac{\partial \mathbf{u}}{\partial t}(\mathbf{x},t) + (\mathbf{w}(\mathbf{x},t)\cdot\nabla)\mathbf{u}(\mathbf{x},t) \\ -\mu\nabla^2\mathbf{u}(\mathbf{x},t) + \nabla p(\mathbf{x},t) &= \mathbf{f}(\mathbf{x},t) \\ \nabla\cdot\mathbf{u}(\mathbf{x},t) &= \mathbf{0} \end{cases}$$

u: velocity, p: pressure, w: wind (= u for full N-S)

Linearisation is time-dependent Stokes equations:

$$\begin{cases} \frac{\partial \mathbf{u}}{\partial t}(\mathbf{x},t) - \mu \nabla^2 \mathbf{u}(\mathbf{x},t) + \nabla p(\mathbf{x},t) &= \mathbf{f}(\mathbf{x},t) \\ \nabla \cdot \mathbf{u}(\mathbf{x},t) &= \mathbf{0} \end{cases}$$

Discretize: Galerkin finite elements in space, Backwards Euler in time, constant time-step Δt

$$\left\{ egin{array}{l} rac{1}{\Delta t}\mathcal{M}_{\mathrm{u}}\left(\mathrm{u}^{k}-\mathrm{u}^{k-1}
ight)+\mu\mathcal{A}_{\mathrm{u}}\mathrm{u}^{k}+\mathcal{W}_{\mathrm{u},k}\mathrm{u}^{k}+\mathcal{B}^{T}\mathrm{p}^{k} &=\mathrm{f}^{k} \ \mathcal{B}\mathrm{u}^{k} &=0 \end{array}
ight.$$

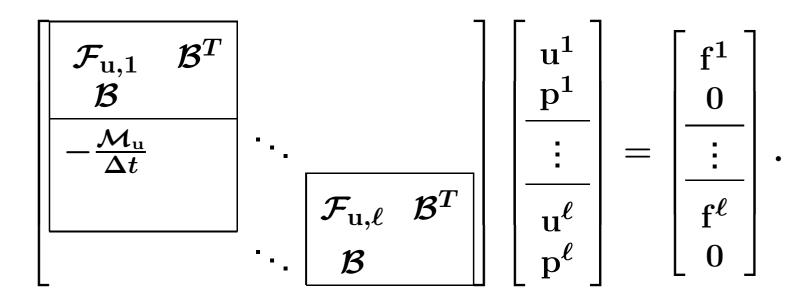
 $k=1,\ldots,\ell$ with $\ell\Delta t=T.$

To simplify notation define the time-dependent advection-diffusion operator

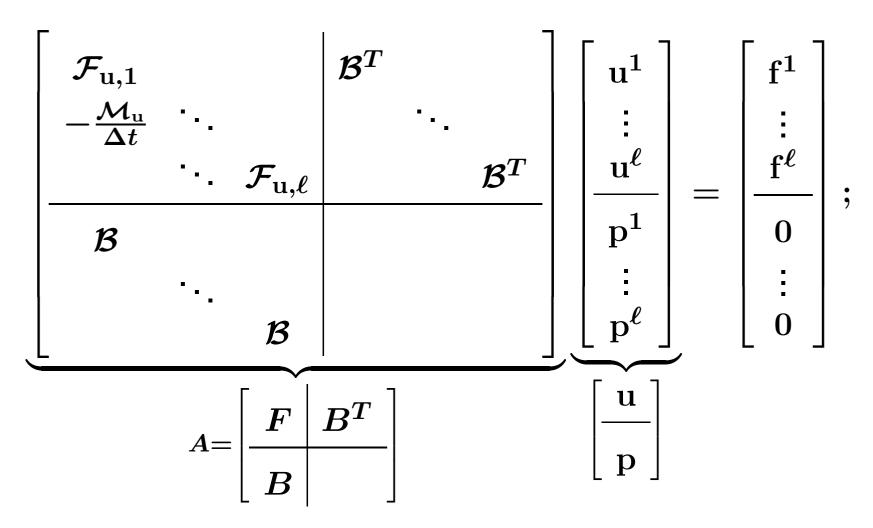
$$\mathcal{F}_{\mathrm{u},k} = rac{\mathcal{M}_{\mathrm{u}}}{\Delta t} + \mathcal{W}_{\mathrm{u},k} + \mu \mathcal{A}_{\mathrm{u}}, \qquad k = 1, \ldots \ell.$$

9

This is thus the system



Reorder to



Important point: Saddle-point system: *F* is now a time-dependent (advection-)diffusion operator; something we might have excellent PinT methods for!!

Block preconditioning for Saddle-point systems

based on the observation (Murphy, Golub, W (2000))

$$\left[egin{array}{cc} F & B^T \ B & 0 \end{array}
ight]$$

preconditioned by • $\begin{bmatrix} F & 0 \\ 0 & S \end{bmatrix}$ has 3 distinct eigenvalues • $\begin{bmatrix} F & B^T \\ 0 & S \end{bmatrix}$ has 2 distinct eigenvalues

where $S = BF^{-1}B^T$ (Schur Complement)

- \Rightarrow MINRES /GMRES terminates in 3 / 2 iterations
- \Rightarrow want approximations $\widehat{F}, \ \widehat{S} \Rightarrow 3 / 2$ clusters
- \Rightarrow fast convergence

Preconditioning the All-at-once system

use PinT method for *F* and Schur complement approximation

$$S^{-1} pprox M_p^{-1} F_p A_p^{-1}$$

where

- *M_p* is a block diagonal matrix of pressure mass matrices
- F_p represents time-integration on the pressure space analogous to F_u
- A_p represents a block diagonal matrix of discrete pressure Laplacians

which is essentially an identical approach to the highly successful PCD approach for the steady-state problem *Elman, Silvester, W (2014)*

Number of GMRES iterations: driven cavity problem for $t \in [0, 1]$ (thus T = 1)

$\Delta x^{\Delta t}$	2^{-3}		2^{-4}		2^{-5}		2^{-6}		2^{-7}	
2^{-4}	25	(25)	25	(26)	24	(25)	24	(31)	23	(28)
2^{-5}	23	(25)	23	(24)	22	(24)	23	(32)	22	(24)
2^{-6}	22	(26)	21	(26)	21	(26)	22	(28)	21	(28)
2^{-7}	21	(28)	20	(28)	21	(31)	21	(32)	20	(33)
2^{-8}	20	(28)	19	(30)	20	(32)	19	(32)	19	(34)

exact subsystem solves (iterative subsystem solves)

- F_u^{-1} : AIR
- M_p^{-1} : Chebyshev (semi-)iteration
- F_p : matrix multiply
- A_p^{-1} : AMG (BoomerAMG)



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.

Wathen, A.J., 2022,

'Some observations on preconditioning for non-self-adjoint and time-dependent problems',

Comput. Math. Appl. 116, pp. 176–180.

Danieli, F., Southworth, B. & Wathen, A.J., 2022, 'Space-time block preconditioning for incompressible flow', SIAM J. Sci. Comput. 44, pp. A337–A363.

Elman, H.C., Silvester, D.J. & Wathen, A.J., 2014, "Finite Elements and Fast Iterative Solvers" Second Edition, OUP, Oxford, 2014.

Wathen, A.J., 2015,

'Preconditioning',

Acta Numerica (2015) 24, pp. 329-376.