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.

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

$k = 0, 1, \ldots, \ell$ 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 (constant diagonal) matrix.

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

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

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

$$\mathbf{Bc}, \mathbf{B}(\mathbf{Bc}), \ldots, \mathbf{B^k c}, \ldots \qquad \text{so that}$$

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

$\Rightarrow$ Krylov subspace methods generally described by:

$$\mathbf{r_k} = \mathbf{p_k}(\mathbf{B})\mathbf{r_0}, \qquad \mathbf{r_k} = \mathbf{c} - \mathbf{Bx_k}, \quad \mathbf{p_k} \in \Pi_\mathbf{k}, \mathbf{p_k}(0) = 1$$

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

$$\|\mathbf{r_k}\| \le \|\mathbf{X}\| \; \|\mathbf{p_k}(\mathbf{\Lambda})\| \; \|\mathbf{X^{-1}}\| \; \|\mathbf{r_0}\|$$

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 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

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

All-at-once system

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

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

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

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

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

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 $\mathbf{P}$ 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 $\mathbf{B}\mathbf{x} = \mathbf{b}$.

All-at-once system

A practical and guaranteed preconditioning approach: 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

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

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

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

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

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

$\Rightarrow$ 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 ($\boldsymbol{\tau = 0.2}$, $\boldsymbol{a = -0.3}$,$\boldsymbol{\theta = 0.8}$):

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

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

Moreover we *observe* that GMRES with just $\mathbf{C}$ 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

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

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

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

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

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

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.

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.

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

2nd part of talk: Incompressible Navier-Stokes

$$\left\{ \begin{array}{rcl} \dfrac{\partial \mathbf{u}}{\partial t}(\mathbf{x},t) + (\mathbf{w}(\mathbf{x},t)\cdot\nabla)\mathbf{u}(\mathbf{x},t) & & \\ -\mu\nabla^2\mathbf{u}(\mathbf{x},t) + \nabla p(\mathbf{x},t) & = & \mathbf{f}(\mathbf{x},t) \\ \nabla\cdot\mathbf{u}(\mathbf{x},t) & = & 0 \end{array} \right.$$

$\mathbf{u}$: velocity, p: pressure, $\mathbf{w}$: wind ($= \mathbf{u}$ for full N-S)

Linearisation is time-dependent Stokes equations:

$$\left\{ \begin{array}{rcl} \dfrac{\partial \mathbf{u}}{\partial t}(\mathbf{x},t) - \mu\nabla^2\mathbf{u}(\mathbf{x},t) + \nabla p(\mathbf{x},t) & = & \mathbf{f}(\mathbf{x},t) \\ \nabla\cdot\mathbf{u}(\mathbf{x},t) & = & 0 \end{array} \right.$$

Discretize: Galerkin finite elements in space, Backwards Euler in time, constant time-step $\boldsymbol{\Delta t}$

$$\begin{cases} \frac{1}{\Delta t}\mathcal{M}_{\mathrm{u}}\left(\mathrm{u}^k - \mathrm{u}^{k-1}\right) + \mu\mathcal{A}_{\mathrm{u}}\mathrm{u}^k + \mathcal{W}_{\mathrm{u},k}\mathrm{u}^k + \mathcal{B}^T\mathrm{p}^k & = \mathrm{f}^k \\ \mathcal{B}\mathrm{u}^k & = 0 \end{cases},$$

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

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

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

This is thus the system

$$
\left[\begin{array}{c|c|c}
\begin{array}{cc} \boldsymbol{\mathcal{F}}_{\mathbf{u},1} & \boldsymbol{\mathcal{B}}^T \\ \boldsymbol{\mathcal{B}} & \end{array} & & \\
\hline
-\frac{\boldsymbol{\mathcal{M}}_{\mathbf{u}}}{\Delta t} & \ddots & \\
 & \ddots & \begin{array}{cc} \boldsymbol{\mathcal{F}}_{\mathbf{u},\ell} & \boldsymbol{\mathcal{B}}^T \\ \boldsymbol{\mathcal{B}} & \end{array}
\end{array}\right]
\left[\begin{array}{c}
\mathbf{u}^1 \\ \mathbf{p}^1 \\ \hline \vdots \\ \hline \mathbf{u}^\ell \\ \mathbf{p}^\ell
\end{array}\right]
=
\left[\begin{array}{c}
\mathbf{f}^1 \\ \mathbf{0} \\ \hline \vdots \\ \hline \mathbf{f}^\ell \\ \mathbf{0}
\end{array}\right].
$$

Reorder to

$$
\underbrace{\left[\begin{array}{ccc|ccc}
\mathcal{F}_{\mathbf{u},1} & & & \mathcal{B}^T & & \\
-\frac{\mathcal{M}_{\mathrm{u}}}{\Delta t} & \ddots & & & \ddots & \\
& \ddots & \mathcal{F}_{\mathbf{u},\ell} & & & \mathcal{B}^T \\
\hline
\mathcal{B} & & & & & \\
& \ddots & & & & \\
& & \mathcal{B} & & &
\end{array}\right]}_{A=\left[\begin{array}{c|c} F & B^T \\ \hline B & \end{array}\right]}
\underbrace{\left[\begin{array}{c}
\mathbf{u}^1 \\ \vdots \\ \mathbf{u}^\ell \\ \hline \mathbf{p}^1 \\ \vdots \\ \mathbf{p}^\ell
\end{array}\right]}_{\left[\begin{array}{c} \mathbf{u} \\ \hline \mathbf{p} \end{array}\right]}
=
\left[\begin{array}{c}
\mathbf{f}^1 \\ \vdots \\ \mathbf{f}^\ell \\ \hline \mathbf{0} \\ \vdots \\ \mathbf{0}
\end{array}\right];
$$

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)*)

$$\begin{bmatrix} F & B^T \\ B & 0 \end{bmatrix}$$

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 $\boldsymbol{F}$ and Schur complement approximation

$$\boldsymbol{S}^{-1} \approx \boldsymbol{M}_p^{-1} \boldsymbol{F}_p \boldsymbol{A}_p^{-1}$$

where

- $\boldsymbol{M}_p$ is a block diagonal matrix of pressure mass matrices
- $\boldsymbol{F}_p$ represents time-integration on the pressure space analogous to $\boldsymbol{F}_u$
- $\boldsymbol{A}_p$ 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 $T = 1$)

$\Delta x \backslash \Delta t$	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)

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.

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.