SVD preconditioners and Parallel-in-Time Solver for the All-at-Once Runge–Kutta Discretization

Á. Martínez[♭, 1], L. Bergamaschi[♮], S. Leveque[*] and J.W. Pearson[♯]

(♭) Department of Mathematics, Informatics and Geosciences, University of Trieste, Via A. Valerio 12/1, 34127, Trieste, Italy.

(♮) Dept. of Civil Environmental and Architectural Engineering, University of Padova, Via Marzolo 9, 35100, Padova, Italy

(∗) CRM Ennio De Giorgi, Scuola Normale Superiore, Piazza dei Cavalieri 3, 56126, Pisa, Italy

(♯) School of Mathematics, The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom

1 Introduction

Time-dependent PDEs arise quite often in many scientific areas, such as mechanics, biology, economics, and chemistry, just to name a few. Of late, researchers have devoted their effort to devising parallel-in-time methods for the numerical solution of time-dependent PDEs [1, 5]. As opposed to the classical approach, in which an approximation of the solution at a time t is computed after solving for all the previous times, parallel-in-time methods approximate the solution of the problem at all times concurrently. This in turn adds a new dimension of parallelism and allows to speed up the numerical solution on modern supercomputers.

In this work, we consider a fully parallelizable preconditioner for the all-at-once linear system arising when employing a Runge–Kutta method in time. Fully implicit Runge–Kutta methods offer the possibility to use high-order accurate time discretization to match space discretization accuracy, an issue of significant importance for many large-scale problems of current interest, where we may have fine space resolution with many millions of spatial degrees of freedom and long time intervals. In this work, we consider strongly A-stable implicit Runge–Kutta methods of arbitrary order of accuracy. For the arising huge algebraic systems we introduce an efficient parallel preconditioner that uses only real arithmetic. The proposed preconditioner results in a block-diagonal solve for all the stages at all the time steps, and a Schur complement obtained by solving again systems for the stages. To solve the system for the stages, we employ a new block preconditioner based on the SVD of the Runge–Kutta coefficient matrix. Parallel results on the Stokes equations show the robustness of the preconditioner with respect to the discretization parameters and the number of stages, as well as satisfactory scalability and parallel efficiency indices.

[1] amartinez@units.it

2 All-at-once Runge-Kutta discretization

Given a subset $\Omega \subset \mathbb{R}^d$, with $d = 1, 2, 3$, and a final time $t_f > 0$, we consider the following discrete algebraic system (DAE):

$$\begin{cases} \frac{\partial v}{\partial t} + \mathcal{D}_1 v = f(\mathbf{x}, t) & \text{in } \Omega \times (0, t_f), \\ \mathcal{D}_2 v = g(\mathbf{x}, t) & \text{in } \Omega \times (0, t_f), \end{cases}$$

given some suitable initial and boundary conditions. Here, $\mathcal{D}_1$ and $\mathcal{D}_2$ are differential operators (only) in space. In addition, the variable v may be a vector and contains all the physical variables described by the DAE (e.g., the *temperature* for the heat equation, or the *velocity* of the fluid and the *kinematic pressure* for the Stokes equations). In what follows, we will suppose that the differential operators $\mathcal{D}_1$ and $\mathcal{D}_2$ are linear and time-independent.

Given suitable discretizations $\mathbf{D}_1$ and $\mathbf{D}_2$ of $\mathcal{D}_1$ and $\mathcal{D}_2$ respectively, after dividing the time interval $[0, t_f]$ into n_t subintervals with constant time-step τ, a Runge–Kutta discretization in compact form reads as follows:

$$\begin{cases} \mathbf{M}\mathbf{v}_{n+1} = \mathbf{M}\mathbf{v}_n + \tau \mathbf{M}(\mathbf{b}_{\mathrm{RK}}^\top \mathbf{k}_n), \\ (I_s \otimes \mathbf{M})\mathbf{k}_n + (\mathbf{e} \otimes \mathbf{D}_1)\mathbf{v}_n + \tau(A_{\mathrm{RK}} \otimes \mathbf{D}_1)\mathbf{k}_n = \mathbf{f}_n, \\ (\mathbf{e} \otimes \mathbf{D}_2)\mathbf{v}_n + \tau(A_{\mathrm{RK}} \otimes \mathbf{D}_2)\mathbf{k}_n = \mathbf{g}_n, \end{cases}$$

where $\mathbf{e} \in \mathbb{R}^s$ is the column vector of all ones. Here, we set $\mathbf{k}_n = [\mathbf{k}_{1,n}, \dots, \mathbf{k}_{s,n}]^\top$, $\mathbf{f}_n = [\mathbf{f}_{1,n}, \dots, \mathbf{f}_{s,n}]^\top$, and $\mathbf{g}_n = [\mathbf{g}_{1,n}, \dots, \mathbf{g}_{s,n}]^\top$.
Setting $\mathbf{v} = [\mathbf{v}_0, \dots, \mathbf{v}_{n_t}]^\top$ and $\mathbf{k} = [\mathbf{k}_0, \dots, \mathbf{k}_{n_t-1}]^\top$, the all-at-once system for the Runge–Kutta discretization in time can be written in matrix form as follows:

$$\underbrace{\begin{bmatrix} \Phi & \Psi_1 \\ \Psi_2 & \Theta \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} \mathbf{v} \\ \mathbf{k} \end{bmatrix} = \mathbf{b}, \tag{1}$$

where the blocks of the matrix $\mathcal{A}$ are given by

$$\begin{aligned} \Phi &= \begin{bmatrix} \mathbf{M} & & & \\ -\mathbf{M} & \ddots & & \\ & \ddots & \ddots & \\ & & -\mathbf{M} & \mathbf{M} \end{bmatrix}, \qquad \Psi_1 = -\begin{bmatrix} 0 & & & \\ \tau \mathbf{b}_{\mathrm{RK}}^\top \otimes \mathbf{M} & & & \\ & & \ddots & \\ & & & \tau \mathbf{b}_{\mathrm{RK}}^\top \otimes \mathbf{M} \end{bmatrix}, \\ \Psi_2 &= \begin{bmatrix} \mathbf{e} \otimes \mathbf{D}_1 & & & \\ \mathbf{e} \otimes \mathbf{D}_2 & & & \\ & \ddots & & \\ & & \mathbf{e} \otimes \mathbf{D}_1 & 0 \\ & & \mathbf{e} \otimes \mathbf{D}_2 & 0 \end{bmatrix}, \quad \Theta = I_{n_t} \otimes \begin{bmatrix} I_s \otimes \mathbf{M} + \tau A_{\mathrm{RK}} \otimes \mathbf{D}_1 \\ \tau A_{\mathrm{RK}} \otimes \mathbf{D}_2 \end{bmatrix}. \end{aligned} \tag{2}$$

Note that Θ is block-diagonal.

3 An SVD preconditioner for the all-at-once Stokes formulation

Given a domain $\Omega \subset \mathbb{R}^d$, with $d = 2, 3$, and a final time $t_f > 0$, we consider the Stokes equations defined as follows:

$$\begin{cases} \frac{\partial \vec{v}}{\partial t} - \nabla^2 \vec{v} + \nabla p = \vec{f}(\mathbf{x}, t) & \text{in } \Omega \times (0, t_f), \\ -\nabla \cdot \vec{v} = 0 & \text{in } \Omega \times (0, t_f), \\ \vec{v}(\mathbf{x}, t) = \vec{g}(\mathbf{x}, t) & \text{on } \partial\Omega \times (0, t_f), \\ \vec{v}(\mathbf{x}, 0) = \vec{v}_0(\mathbf{x}) & \text{in } \Omega, \end{cases}$$

where the functions $\vec{f}$ and $\vec{g}$ are known. In addition, the initial condition $\vec{v}_0(\mathbf{x})$ is also given.

In what follows, we will denote with $\mathbf{K}_v$ and $\mathbf{M}_v$ (resp., $\mathbf{K}_p$ and $\mathbf{M}_p$) the *(vector)-stiffness* and *(vector)-mass* (resp., *stiffness* and *mass*) matrices, respectively, and with B (B^T) the discrete *negative* divergence (gradient) operator.
The all-at-once system is given by

$$\begin{bmatrix} \Phi & \Psi_1 \\ \Psi_2 & \Theta_S \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{k} \end{bmatrix} = \mathbf{b}, \tag{3}$$

where Φ and Ψ_1 are defined as in (2), with $\mathbf{M} = \begin{bmatrix} \mathbf{M}_v & 0 \\ 0 & \mathbf{M}_p \end{bmatrix}$, and Ψ_2 and Θ_S are defined, respectively, as

$$\Psi_2 = \begin{bmatrix} \widehat{\Psi}_2 & & \\ & \ddots & \\ & & \widehat{\Psi}_2 \;\; 0 \end{bmatrix}, \quad \widehat{\Psi}_2 = \begin{bmatrix} \mathbf{e} \otimes \mathbf{K}_v & \mathbf{e} \otimes B^\top \\ \mathbf{e} \otimes B & 0 \end{bmatrix},$$

and

$$\Theta_S = I_{n_t} \otimes \underbrace{\begin{bmatrix} I_s \otimes \mathbf{M}_v + \tau A_{\mathrm{RK}} \otimes \mathbf{K}_v & \tau A_{\mathrm{RK}} \otimes B^\top \\ \tau A_{\mathrm{RK}} \otimes B & 0 \end{bmatrix}}_{\widehat{\Theta_S}}.$$

We consider as a preconditioner for the system (3) the following matrix:

$$\mathcal{P} = \begin{bmatrix} S & \Psi_1 \\ 0 & \Theta_S \end{bmatrix}, \tag{4}$$

where $S = \Phi - \Psi_1 \Theta_S^{-1} \Psi_2$ is the Schur complement, with Φ, Ψ_1, Ψ_2, and Θ_S defined above. Specifically, we have

$$S = \begin{bmatrix} \mathbf{M} & & & \\ -\mathbf{M} + X & \ddots & & \\ & \ddots & \ddots & \\ & & -\mathbf{M} + X & \mathbf{M} \end{bmatrix},$$

where

$$X = \tau \begin{bmatrix} b_1 \mathbf{M} & \dots & b_s \mathbf{M} \end{bmatrix} \widehat{\Theta}_S^{-1} \begin{bmatrix} \mathbf{e} \otimes \mathbf{K}_v & \mathbf{e} \otimes B^\top \\ \mathbf{e} \otimes B & 0 \end{bmatrix},$$

Applying preconditioner $\mathcal{P}$ rests on efficiently approximating matrix $\widehat{\Theta}_S$:

$$\widehat{\Theta}_S = \begin{bmatrix} I_s \otimes \mathbf{M}_v + \tau A_{\mathrm{RK}} \otimes \mathbf{K}_v & \tau A_{\mathrm{RK}} \otimes B^\top \\ \tau A_{\mathrm{RK}} \otimes B & 0 \end{bmatrix} = \begin{bmatrix} \widehat{\Theta}_{11} & \tau A_{\mathrm{RK}} \otimes B^\top \\ \tau A_{\mathrm{RK}} \otimes B & 0 \end{bmatrix}$$

We approximate this matrix by the optimal block triangular preconditioner

$$\mathcal{P}_{\mathrm{RK}} = \begin{bmatrix} I_s \otimes \mathbf{M}_v + \tau A_{\mathrm{RK}} \otimes \mathbf{K}_v & 0 \\ \tau A_{\mathrm{RK}} \otimes B & S_{\mathrm{RK}} \end{bmatrix} \equiv \begin{bmatrix} \widehat{\Theta}_{11} & 0 \\ \tau A_{\mathrm{RK}} \otimes B & S_{\mathrm{RK}} \end{bmatrix}$$

where

$$S_{\mathrm{RK}} = -\tau^2 (A_{\mathrm{RK}} \otimes B)\widehat{\Theta}_{11}^{-1}(A_{\mathrm{RK}} \otimes B^\top).$$

To apply the preconditioner $\mathcal{P}_{\mathrm{RK}}$ we need to develop efficient preconditioners for its diagonal blocks. To approximately invert S_{RK} we use properties of the Kronecker product, and write

$$\begin{aligned} S_{\mathrm{RK}} &= -\tau^2 (A_{\mathrm{RK}} \otimes I_{n_p})(I_s \otimes B)\widehat{\Theta}_{11}^{-1}(I_s \otimes B^\top)(A_{\mathrm{RK}} \otimes I_{n_p}) \\ &\approx -\tau^2 (A_{\mathrm{RK}} \otimes I_{n_p})(I_s \otimes \mathbf{K}_p)(I_s \otimes \mathbf{M}_p + \tau A_{\mathrm{RK}} \otimes \mathbf{K}_p)^{-1}(I_s \otimes \mathbf{M}_p)(A_{\mathrm{RK}} \otimes I_{n_p}). \end{aligned}$$

employing the block-commutator argument derived in [4], based on the fact that $BB^T \approx \mathbf{K}_p\mathbf{M}_p$.

3.1 SVD-based preconditioner for $\widehat{\Theta}_{11}$

To derive a preconditioner for matrix $\widehat{\Theta}_{11}$, we consider an SVD decomposition of matrix $A_{\mathrm{RK}} = U\Sigma V^\top$, where U and V are unitary matrix whose columns are the left and right singular vectors of A_{RK}, respectively, and Σ is a diagonal matrix with entries the singular values of A_{RK} (we will drop the subscript $_v$ in the sequel):

$$\widehat{\Theta}_{11} = I_s \otimes \mathbf{M} + \tau A_{\mathrm{RK}} \otimes \mathbf{K} = (U \otimes I_{n_x})[(U^\top V) \otimes \mathbf{M} + \tau\Sigma \otimes \mathbf{K}](V^\top \otimes I_{n_x}).$$

Since the eigenvalues of $U^\top V$ have all modulus 1, we approximate $U^\top V$ with the identity matrix and define a preconditioner for $\widehat{\Theta}_{11}$ as

$$\mathcal{P}_{\mathrm{SVD}} = (U \otimes I_{n_x})(I_s \otimes \mathbf{M} + \tau\Sigma \otimes \mathbf{K})(V^\top \otimes I_{n_x}). \tag{5}$$

The following Theorem gives the optimality of the proposed preconditioner, under reasonable assumptions.

Theorem 3.1 *Let be A_{RK} be the matrix representing the coefficients of a Runge–Kutta method. Let $A_{\mathrm{RK}} = U\Sigma V^\top$ be an SVD of the matrix A_{RK}. Suppose that the real part of the Rayleigh quotient $\frac{\mathbf{x}^*(U^\top V)\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$ is positive, for any $\mathbf{x} \in \mathbb{C}^s \setminus \{\mathbf{0}\}$. Then, the eigenvalues of the matrix $\mathcal{P}_{\mathrm{SVD}}^{-1}(I_s \otimes \mathbf{M} + \tau A_{\mathrm{RK}} \otimes \mathbf{K})$ lie all in the right-half of the unit circle centered at the origin of the complex plane.*

See [3] for the proof of this Theorem.

We report in Figure 1 the eigenvalue distribution of the matrices $\mathcal{P}_{\mathrm{SVD}}^{-1}(I_s \otimes \mathbf{M} + \tau A_{\mathrm{RK}} \otimes \mathbf{K})$ and $U^\top V$ employing $\mathbf{Q}_2$ elements, for 5-stages and 9-stages Radau IIA, with $\tau = 0.2$ and level of refinement $\mathtt{l} = 4$. Here, $\mathtt{l}$ represents a spatial uniform grid of mesh size $h = 2^{-\mathtt{l}}$, in each dimension. Further, in green, we plot the unit circle centered at the origin of the complex plane.

It is possible to show that the eigenvalues of the preconditioned $\widehat{\Theta}_{11}$ are bounded away from zero. The eigenvalues of $\mathcal{P}_{\mathrm{SVD}}^{-1}\widehat{\Theta}_{11}$ solve the following generalized eigenvalue problem:

$$(U^\top V + \tau\bar{\lambda}\Sigma)\mathbf{x} = \lambda(I_s + \tau\bar{\lambda}\Sigma)\mathbf{x},$$

where $\bar{\lambda}$ is an eigenvalue of $\mathbf{M}^{-\frac{1}{2}}\mathbf{K}\mathbf{M}^{-\frac{1}{2}}$. Denoted with $\mu_1, \ldots, \mu_s$ the eigenvalues of $U^\top V$, this can be seen as a perturbation of

$$\left(\Lambda + \tau\bar{\lambda}\Sigma\right)\mathbf{x} = \lambda\left(I_s + \tau\bar{\lambda}\Sigma\right)\mathbf{x}, \qquad \Lambda = \mathrm{diag}(\mu_1, \ldots, \mu_s).$$

For every j, setting $\mu_j \equiv a + \mathbf{i}b$, $c = \tau\bar{\lambda}\sigma_j$, with $a \geq \mu_{\min} > 0$, we have

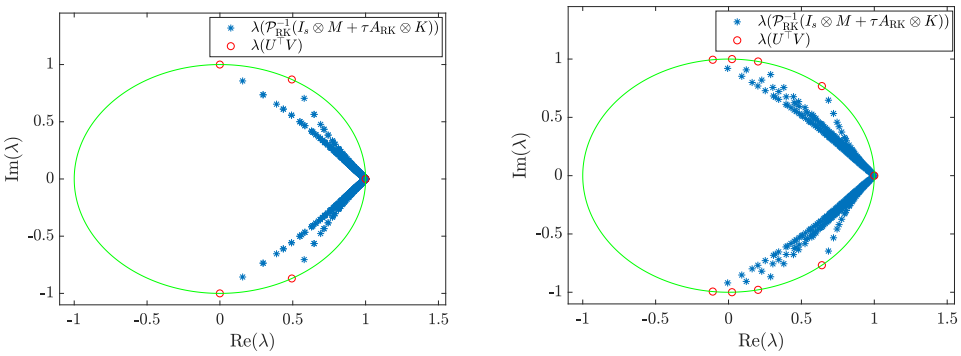
$$\lambda = \frac{a + \mathbf{i}b + c}{1 + c} = \frac{a + c}{1 + c} + \mathbf{i}\frac{b}{1 + c}, \qquad |\lambda|^2 = \frac{(a + c)^2 + b^2}{(1 + c)^2}.$$

Whence

$$|\lambda|^2 = \frac{1 + 2ac + c^2}{(1 + c)^2} = 1 + 2(a - 1)\frac{c}{(1 + c)^2} \equiv \varphi(c) \geq \varphi(1) = \frac{1 + a}{2} \geq \frac{1 + \mu_{\min}}{2} \geq \frac{1}{2},$$

which shows that the eigenvalues are outside the circle of center 0 and radius $\frac{\sqrt{2}}{2}$.

Figure 1: Eigenvalue distribution of $\mathcal{P}_{\text{SVD}}^{-1}(I_s \otimes \mathbf{M} + \tau A_{\text{RK}} \otimes \mathbf{K})$ and of $U^\top V$, for 5-stages and 9-stages Radau IIA, and $\texttt{l} = 4$. In green, the unit circle centered at the origin of the complex plane.



3.2 Alternative preconditioners for $\widehat{\Theta}_{11}$

In a recent work [6] a preconditioner for the $\widehat{\Theta}_{11}$ matrix has been proposed, based on the LU factorization of A_{RK}^{-1}, with $\text{diag}(U) = I_s$. The L factor is then spectrally decomposed into

$$L = G\Gamma G^{-1},$$

and the preconditioner is defined as

$$\mathcal{P}_{\text{Munch}}^{-1} = (G \otimes I_n)\left(\Gamma \otimes \mathbf{M} + \tau I_s \otimes \mathbf{K}\right)^{-1}\left(G^{-1} \otimes I_n\right).$$

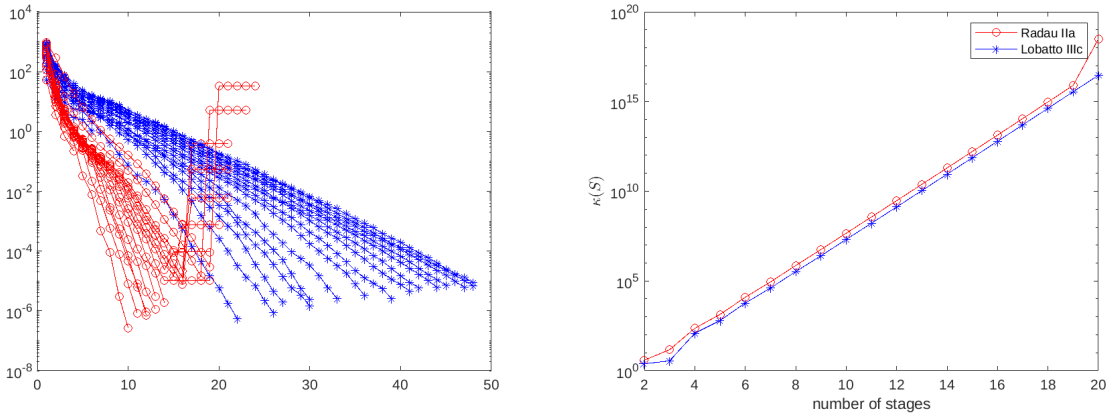
We will present some comparisons between our SVD-based preconditioner and the one proposed by Munch and coauthors. From our experiments, we could observed that the eigenvalue distribution of the preconditioned matrix with Munch's preconditioner is more favorable than that produced by our SVD-based preconditioner. However, we found that the main drawback of Munch's approach resides in the ill-conditioning of the eigenvectors matrices G which grows exponentially with the number of stages. In Figure 2 we report the number of iterations to solve a linear system with $\widehat{\Theta}_{11}$ for different numbers of stages with both approaches and Lobatto IIC (left), and the condition number of G for different stage numbers for both Lobatto IIC and Radau IIIA (right). Increasing the number of stages, GMRES with $\mathcal{P}_{\text{Munch}}$ can not reach the prescribed tolerance due to ill-conditioning of the eigenvector matrix G.

4 Numerical Results

We will present numerical results of both sequential and parallel implementations of our parallel-in-time solver for the Stokes problem. Comparisons with the well-known `ParaDiag` [2] all-at-once solver will be presented showing the superiority of the proposed approach for very fine spatial discretizations.

The MPI-based Fortran 90 implementation of the previously described solver also employs the `XBraid` [7] multigrid-in-time package to solve for the Schur complement S (the $(1,1)$ block in (4)). With our approach, we were able to solve problems with up to 16 thousand timesteps and more than 5 billion unknowns on the M100 supercomputer located at CINECA (Bologna, Italy). The obtained strong scalability is very satisfactory, with a parallel efficiency always larger than 25% up to 4096 computing cores.

Figure 2: Left: Number of GMRES iteration to reach the tolerance of 10^{-6} in the solution of a system with $\widehat{\Theta}_{11}$ with $\mathcal{P}_{\text{SVD}}$ (blue) and with $\mathcal{P}_{Munch}$ (red). Right: $\kappa(G)$ vs number of stages.



Acknowledgments

We acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and technical support. AM and LB gratefully acknowledge the financial support of the INdAM-GNCS Project CUP_E53C22001930001. JWP gratefully acknowledges financial support from the Engineering and Physical Sciences Research Council (EPSRC) UK grant EP /S027785/1. The work of AM was carried out within the PNRR research activities of the consortium iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS_00000043). This manuscript reflects only the Authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

[1] R. D. Falgout, S. Friedhoff, T. V. Kolev, S. P. MacLachlan, and J. B. Schröder, *Parallel time integration with multigrid*, SIAM J. Sci. Comput., 36, pp. C635–C661, 2014.

[2] M. J. Gander, J. Liu , S.-L. Wu, X. Yue, and T. Zhou. *ParaDiag: parallel-in-time algorithms based on the diagonalization technique*. arXiv preprint arXiv:2005.09158 (2021).

[3] S. Leveque, L. Bergamaschi, Á. Martínez, and J. W. Pearson. *Parallel-in-time solver for the all-at-once Runge–Kutta discretization*, 2024, Submitted.

[4] S. Leveque and J. W. Pearson, *Parameter-robust preconditioning for Oseen iteration applied to stationary and instationary Navier–Stokes control*, SIAM J. Sci. Comput., 44, pp. B694–B722, 2022.

[5] E. McDonald, J. Pestana, and A. Wathen, *Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations*, SIAM J. Sci. Comput., 40, pp. A1012–A1033, 2018.

[6] P. Munch, E. Dravins, M. Kronbichler, M. Neytcheva, *Stage-parallel fully implicit Runge-Kutta implementations with optimal multilevel preconditioners at the scaling limit*, Numer. Lin. Alg. Appl., e2532, 2023.

[7] *XBraid: Parallel multigrid in time.* `http://llnl.gov/casc/xbraid`.