Improved Paradiag
via low-rank updates and interpolation

Daniel Kressner

Institute of Mathematics / EPFL
daniel.kressner@epfl.ch
anchp.epfl.ch

Joint work with
Stefano Massei   Junli Zhu

PinT 2022
11th Conference on Parallel-in-Time Integration

CIRM

EPFL

Space-time discretization

Model problem on $\Omega \subset \mathbb{R}^{1,2,3}$:

$$\begin{aligned} \frac{\partial}{\partial t} u(t,x) + \Delta u(t,x) &= f(t,x) & x \in \Omega, t \in (0,T] \\ u(x,t) &= 0, & x \in \partial\Omega, t \in (0,T] \\ u(x,0) &= u_0(x) & x \in \Omega. \end{aligned}$$

Space-time discretization

Finite difference/finite element discretization (in space) with n degrees of freedom $\leadsto$

$$\begin{aligned} M\dot{\mathbf{u}}(t) + A\mathbf{u}(t) &= \mathbf{f}(t) \qquad\qquad t \in (0, T], \\ \mathbf{u}(0) &= \mathbf{x}_0. \end{aligned}$$

with

- $n \times n$ symm pos def mass matrix M
- $n \times n$ symm pos def (ill-conditioned) stiffness matrix A

Space-time discretization

Implicit Euler with time step size $\Delta t = T/n_t \rightsquigarrow$

$$(\Delta t^{-1} M + A)\mathbf{x}_k = \mathbf{f}_k + \Delta t^{-1} M \mathbf{x}_{k-1}, \quad k = 1, \ldots, n_t,$$

Sequence of n_t sparse linear systems.

Space-time discretization

Stacking these equations yields an $nn_t \times nn_t$ linear system of the form

$$\begin{bmatrix} \Delta t^{-1}M + A & & & \\ -\Delta t^{-1}M & \Delta t^{-1}M + A & & \\ & \ddots & \ddots & \\ & & -\Delta t^{-1}M & \Delta t^{-1}M + A \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{n_t} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 + \Delta t^{-1}M\mathbf{x}_0 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_{n_t} \end{bmatrix}$$

To create (additional) opportunities for:

- **Parallelism:**
  Parareal [Lions/Maday/Turinici'2007], [Gander/Vandewalle'2007], ...
  Paradiag [Maday/Rønquist'2008], [Gander et al.'2016–2020], [McDonald et al.'2018], [Liu'2021], [Wu/Zhu'2021], [Liu/Wu'2021], ...
  In particular: Gander/Liu/Wu/Yue/Zhou. Paradiag: Parallel-in-time algorithms based on the diagonalization technique. ArXiv'2020.
  Space-time Petrov-Galerkin [Andreev'2013], [Falgout et al.'2014], [Steinbach'2015–], [Dörfler et al.'2019], *many more*
- **Low-rank approximation:**
  [Andreev/Tobler'2015], [Dolgov/Khoromskij'2015], [Stoll/Breiten'2015], [Breiten/Simoncini/Stoll'2016], [Benner et al.'2016], [Palita'2021], ...

Paradiag

Matrix equations

$$\begin{bmatrix} \Delta t^{-1}M + A & & & \\ -\Delta t^{-1}M & \Delta t^{-1}M + A & & \\ & \ddots & \ddots & \\ & & -\Delta t^{-1}M & \Delta t^{-1}M + A \end{bmatrix} = I \otimes A + B \otimes M$$

with

$$B = \frac{1}{\Delta t} \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

$\otimes$ denotes the Kronecker product with the property

$$(I \otimes A + B \otimes M)\mathrm{vec}(X) = \mathrm{vec}(AX + MXB).$$

$\rightsquigarrow$ BIG linear system is equivalent to linear matrix equation

$$AX + MXB^T = F$$

with unknown $X = \left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_t}\right] \in \mathbb{R}^{n \times n_t}$.

Matrix equations

Assuming w.l.o.g. $M = I \rightsquigarrow$

$$AX + XB^T = F$$

Let us diagonalize $B = T\Lambda T^{-1}$ with $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{n_t})$:

$$A\widetilde{X} + \widetilde{X}\Lambda = \widetilde{F}.$$

This is equivalent to solving (in parallel!) n_d linear systems

$$\begin{aligned}
(A + \lambda_1 I)\widetilde{\mathbf{x}}_1 &= \widetilde{\mathbf{f}}_1 \\
(A + \lambda_2 I)\widetilde{\mathbf{x}}_2 &= \widetilde{\mathbf{f}}_2 \\
&\vdots \\
(A + \lambda_{n_d} I)\widetilde{\mathbf{x}}_{n_d} &= \widetilde{\mathbf{f}}_{n_d}
\end{aligned}$$

Turned a sequence of n_d linear systems into n_d decoupled linear systems.

Wait a second..

Let us diagonalize $B = T\Lambda T^{-1}$...

$$B = \frac{1}{\Delta t}\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

This is one big Jordan block!

Wait a second..

Let us diagonalize $B = T\Lambda T^{-1}$...

$$B = \frac{1}{\Delta t}\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

This is one big Jordan block!

Ways to fix this:

1. Choose non-uniform time steps [Maday/Rønquist'2008]
2. Perturb B into a diagonalizable matrix [Bertaccini'2000], [Gander et al.'2016–2020], [McDonald/Pestana/Wathen'2018]
3. Our approaches (stay tuned)

Approaches 1 and 2 may still face ill-conditioned T. Ours do not.

Perturbing B

$$B_{\alpha} = \frac{1}{\Delta t} \begin{bmatrix} 1 & & & -\alpha \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

B_α is α-circulant and becomes circulant after diagonal scaling:

$$D_\alpha B_\alpha D_\alpha^{-1} = \text{circulant matrix}$$

with

$$D_\alpha = \text{diag}(1, \alpha^{\frac{1}{n_t}}, \ldots, \alpha^{\frac{n_t - 1}{n_t}}).$$

Note that $\kappa(D_\alpha) \approx 1/\alpha$ for small α.

Paradiag

1. Compute eigenvalues $\lambda_1^{(\alpha)}, \dots, \lambda_{n_t}^{(\alpha)}$ of B_α by FFT.
2. Compute $\widetilde{F} \leftarrow F D_\alpha \overline{\Omega}$ by scaling and FFT.
3. Solve in parallel

$$
\begin{aligned}
(A + \lambda_1^{(\alpha)} I)\widetilde{\mathbf{x}}_1 &= \widetilde{\mathbf{f}}_1 \\
&\vdots \\
(A + \lambda_{n_d}^{(\alpha)} I)\widetilde{\mathbf{x}}_{n_d} &= \widetilde{\mathbf{f}}_{n_d}
\end{aligned}
$$

4. Compute $X \leftarrow \widetilde{X} \Omega D_\alpha^{-1}$ by FFT and column scaling.

Procedure needs to be applied with $\alpha \not\approx 0$ as a preconditioner P_α:

- Preconditioned GMRES [Bertaccini'2000, McDonald/Pestana/Wathen'2018, Gander et al.'2020].
- Preconditioned stationary iteration [Liu and Wu'2020].

Even with optimal choice of α usually $\#\text{it} > 2$ iterations needed
$\rightsquigarrow$ overhead factor $\#\text{it}$ compared to sequential!

Low-Rank Update Approach

Low-rank relations

Idea: Getting away with $\alpha = 1$ essentially at the cost of $\#\text{it} = 1$ iteration using

$$\Delta t(B_1 - B) = \begin{bmatrix} 1 & & & -1 \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & & & 0 \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}$$

Low-rank relations

Idea: Getting away with $\alpha = 1$ essentially at the cost of $\#\text{it} = 1$ iteration using

$$\Delta t(B_1 - B) = \begin{bmatrix} 0 & & & -1 \\ 0 & 0 & & \\ & \ddots & \ddots & \\ & & 0 & 0 \end{bmatrix}$$

Low-rank relations

Idea: Getting away with $\alpha = 1$ essentially at the cost of $\#\text{it} = 1$ iteration using

$$\Delta t(B_1 - B) = \text{rank one.}$$

Analogous statements hold more generally for implicit Runge-Kutta and multistep (BDF) methods.

Low-rank relations

Idea: Getting away with $\alpha = 1$ essentially at the cost of $\#\text{it} = 1$ iteration using

$$\Delta t(B_1 - B) = \text{rank one.}$$

In terms of matrix equations:

$$AX_1 + X_1 B_1 = F \quad (1)$$
$$AX + XB = F \quad (2)$$

Subtracting (2)–(1) yields matrix equation for correction $\delta X := X - X_1$:

$$A\,\delta X + \delta X B = \text{rank one.}$$

Summary of our approach:
Compute B_1 ($\#\text{it} = 1$ iteration of Paradiag) and subtract δX.

Key advantage: Rhs of correction eqn has rank one *independent* of F.

Basis of Divide-and-Conquer for matrix equations [K./Massei/Robol'2019].

Correction equation

$$A\,\delta X + \delta X\,B = \text{rank one.}$$

Recall that

- A is $n \times n$ symmetric pos definite and large+sparse;
- B is an $n_t \times n_t$ big Jordan block.

Necessary condition for most large-scale solvers:

δX should admit good low-rank approximation

Connection to Zolotarev approximation problems.

SIAM REVIEW
Vol. 61, No. 2, pp. 319–344

Bounds on the Singular Values of Matrices with Displacement Structure*

Bernhard Beckermann[†]
Alex Townsend[‡]

Correction equation

$$A\,\delta X + \delta X\,B = \text{rank one}.$$

Recall that

- A is $n \times n$ symmetric pos definite and large+sparse;
- B is an $n_t \times n_t$ big Jordan block.

Necessary condition for most large-scale solvers:

δX should admit good low-rank approximation

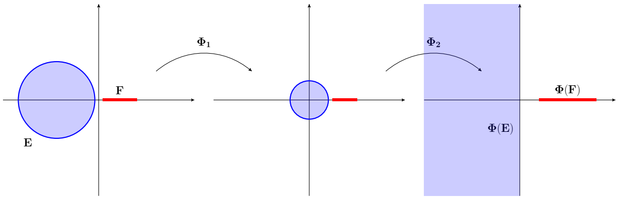
Connection to Zolotarev numbers [Beckermann/Townsend'2019]:

$$\sigma_{j+1}(\delta X) \lesssim Z_j(E,F), \qquad Z_j(E,F) := \min_{r(z)\in\mathcal{R}_{j,j}} \frac{\max_{z\in E}|r(z)|}{\min_{z\in F}|r(z)|},$$

- E = numerical range of $-B$ = disc with center $-1/\triangle$ and radius $\approx -1/\triangle$
- F = numerical range of A = real positive interval
- Optimal rational function r can be used to approximate solution, e.g., via ADI / rational Krylov for solving correction equation.

Solution of Zolotarev problem

Zolotarev numbers are invariant under Moebius maps:



$$Z_j(E,F) = Z_j(\mathrm{i}\mathbb{R}, [\tilde{a}, \tilde{b}]) \stackrel{?^1}{\leq} \sqrt{Z_j([-\tilde{b}, -\tilde{a}], [\tilde{a}, \tilde{b}])} \leq 2\eta^{-j}$$

with

$$\eta = \exp\left(\frac{\pi^2}{2\log(4\tilde{b}/\tilde{a})}\right) > 1.$$

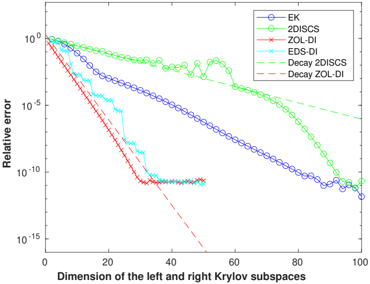
Related work in [Le Bailly/Thiran'2000], [Druskin/Knizhnerman/Zaslavsky'2009], [Beckermann/Townsend'2019].

[1] Bernd Beckermann: This is an equality.

Back to model problem

Correction equation is $A\,\delta X + \delta X B = \mathrm{rank}\ 1$ for 1D heat equation and finite differences:

$$A = \Delta t(n+1)^2 \mathrm{tridiag}(-1, 2, -1), \qquad B = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$



Convergence of rational Krylov subspace method with poles $0/\infty$ (EK), poles from Zolotarev on two discs (2DISCS), poles from Zolotarev on disc-interval (ZOL-DI), nested poles (EDS).

Evaluation-Interpolation Approach

Interpolation on the unit circle

Recall that

$$AX(\alpha) + X(\alpha)B_\alpha^T = F, \quad B_\alpha = \frac{1}{\Delta t}\begin{bmatrix} 1 & & & -\alpha \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

Want to know $X(0)$ but can only evaluate $X(\alpha)$ for $\alpha \neq 0$.

Idea:
Obtain $X(0)$ from interpolating matrix-valued function $X(\alpha)$ for d *different* α on circle of radius $\rho > 0$.

Evaluation-interpolation algorithm

1. **for** $j = 0, \dots, d-1$
2. $\quad \zeta_j = \exp(2\pi \mathbf{i}/d)$
3. $\quad$ Compute $X_j \leftarrow X(\zeta_j \rho)$ (Paradiag)
4. **endfor**
5. **return** $\widetilde{X} \leftarrow (X_0 + \dots X_{d-1})/d$

Theorem

Let $X(z)$ be analytic on $\{|z| < R\}$. Then for $\rho < R$, we have

$$\|X - \widetilde{X}\| \lesssim \left(\frac{\rho}{R}\right)^d .$$

Model problem

Relative residual for different values of d, ρ and $n = n_t$:

n	ρ	$d = 1$	$d = 2$	$d = 3$	$d = 4$
1000	$1e+00$	$1.59e-01$	$6.47e-06$	$3.51e-10$	$1.61e-12$
	$1e-02$	$1.59e-03$	$6.48e-10$	$1.82e-11$	$1.59e-11$
	$1e-04$	$1.59e-05$	$1.17e-09$	$9.69e-10$	$8.31e-10$
	$1e-06$	$1.93e-07$	$7.84e-08$	$6.23e-08$	$5.53e-08$
	$1e-08$	$7.74e-06$	$5.44e-06$	$4.45e-06$	$3.83e-06$
	$1e-10$	$5.78e-04$	$4.11e-04$	$3.27e-04$	$2.84e-04$
	$1e-12$	$4.50e-02$	$3.16e-02$	$2.63e-02$	$2.19e-02$
2000	$1e+00$	$1.70e-01$	$6.70e-06$	$3.55e-10$	$7.06e-12$
	$1e-02$	$1.70e-03$	$6.75e-10$	$6.66e-11$	$5.73e-11$
	$1e-04$	$1.70e-05$	$3.79e-09$	$3.07e-09$	$2.60e-09$
	$1e-06$	$3.62e-07$	$2.25e-07$	$1.86e-07$	$1.58e-07$
	$1e-08$	$2.15e-05$	$1.52e-05$	$1.25e-05$	$1.08e-05$
	$1e-10$	$1.58e-03$	$1.12e-03$	$9.11e-04$	$7.96e-04$
	$1e-12$	$1.25e-01$	$9.00e-02$	$7.09e-02$	$6.38e-02$
4000	$1e+00$	$2.14e-01$	$9.97e-06$	$5.23e-10$	$3.96e-11$
	$1e-02$	$2.14e-03$	$1.04e-09$	$2.43e-10$	$2.12e-10$
	$1e-04$	$2.14e-05$	$1.46e-08$	$1.18e-08$	$1.03e-08$
	$1e-06$	$1.34e-06$	$9.35e-07$	$7.65e-07$	$6.63e-07$
	$1e-08$	$9.56e-05$	$6.73e-05$	$5.51e-05$	$4.78e-05$
	$1e-10$	$7.32e-03$	$5.15e-03$	$4.21e-03$	$3.66e-03$
	$1e-12$	$5.98e-01$	$4.23e-01$	$3.39e-01$	$2.98e-01$

Numerical Comparison

Numerical comparison

Consider 2D convection diffusion problem [McDonald et al.'2018]:

$$\begin{cases} u_t - \epsilon\Delta u + \mathbf{w}\cdot\nabla u = 0, & \Omega\times(0,1),\ \Omega = (-1,1)^2, \\ u(x,y,0) = \chi_{\{x=1\}}, & \\ u(x,y,t) = \chi_{\{x=1\}}, & \partial\Omega\times(0,1), \end{cases}$$

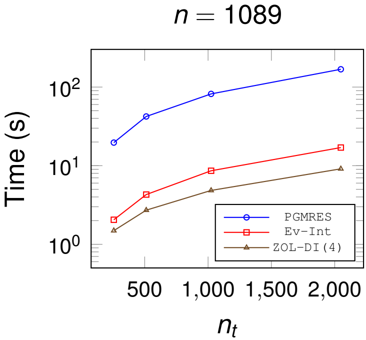
with $\epsilon = 0.005$, indicator function $\chi_{\{x=1\}}$, and $\mathbf{w} = (2y(1-x^2), -2x(1-y^2))$.

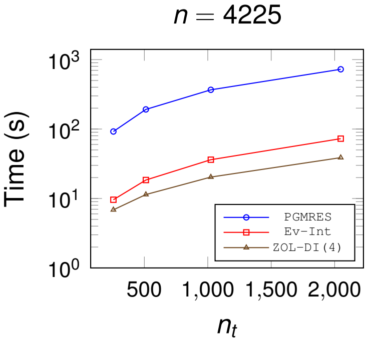
Spatial discretization with Q1 finite elements and time discretization with implicit Euler.

Resulting matrix equation:

$$AX + MXB_1^T = F.$$

Numerical comparison





- PGMRES = Existing method from [McDonald et al.'2018]
- ZOL-DI(4) = Rational Krylov with 4 poles from Zolotarev for $j = 4$
- Ev-Int = Evaluation-interpolation approach with $d = 2$ and $\rho = 5 \cdot 10^{-4}$

Numerical comparison

Similar results for:

- 2D fractional space diffusion with implicit Euler in time
- 2D wave equation with implicit leap-frog scheme
- Implicit Runge-Kutta methods (matrix equations obtained via embedding to arrive at)
- Fractional time diffusion (low-rank update does not apply, Ev-Int performs well)

See

- Kressner, Massei, Zhu. *Improved parallel-in-time integration via low-rank updates and interpolation*. arXiv 2022.

Conclusions

- Two novel approaches to avoid need for iterative refinement in existing ParaDiag algorithms.
- Evaluation-interpolation approach is highly parallelizable.
- Low-rank update approach is cheaper but correction equation harder to parallelize.
- Extensions of these approaches possible for: implicit Runge-Kutta, implicit leap-frog for wave equations, fractional-in-time equations
- Possible extension to non-linear problems via (quasi-)Newton iteration.

References

- Gander et. al. *ParaDiag: Parallel-in-Time Algorithms Based on the Diagonalization Technique*. arXiv 2021.
- Wu. *Toward parallel coarse grid correction for the parareal algorithm*. SISC 2018.
- Maday, Rønquist. *Parallelization in time through tensor-product space-time solvers*. Comptes Rendus Mathematique 2008.
- Bertaccini. *The spectrum of circulant-like preconditioners for some general linear mul- tistep formulas for linear boundary value problems*. SINUM 2002.
- McDonald, Pestana, Wathen. *Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations*. SISC 2018.
- Palitta. *Matrix equation techniques for certain evolutionary partial differential equations*. J. Scientific Computing, 2021.
- Kressner, Massei, Zhu. *Improved parallel-in-time integration via low-rank updates and interpolation*. arXiv 2022.