Space-time parallel solvers based on multigrid-reduction-in-time and operator splitting methods

Laura Portero

Public University of Navarre, Spain
INAMAT2

Joint work with: A. Arrarás, F.J. Gaspar and I. Jimenez

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Universidad Zaragoza
1542

Innovative & Efficient Strategies for Stiff Differential Equations

Contents

Contents

Introduction

- Evolutionary problems with fine resolution in space and time arise in many applications and are computationally demanding.
- As increasing clock speeds becomes harder, modern architectures add more cores, making parallel algorithms more important.
- First, spatial parallel computing methods were developed. Time discretization was sequential.
- For evolutionary PDEs, spatial parallelism alone is often insufficient and space-time parallelism is needed to fully exploit computing resources.
- In this talk, spatial parallelism will be achieved using operator splitting time integrators.
- Time parallelism will be obtained using Parareal and Multigrid-Reduction-in-Time (MGRIT) algorithms.

Contents

The continuous problem

- Let us consider a linear parabolic initial-boundary value problem

$$\begin{aligned} u_t &= L\,u + f, && \text{in } \Omega \times (0, T], \\ u &= g, && \text{on } \Gamma \times (0, T], \\ u &= u_0, && \text{in } \bar{\Omega} \times \{0\}, \end{aligned}$$

  where L is an elliptic operator and $\Omega \subset \mathbb{R}^d$ is a bounded open domain, $d = 2, 3$, with boundary $\Gamma = \partial\Omega$.

- For instance, we may consider a reaction-diffusion problem

$$Lu = \nabla \cdot (K \nabla u) - cu,$$

  where $K(\boldsymbol{x}) \in \mathbb{R}^{d \times d}$ is an SPD full tensor, and the functions $c(\boldsymbol{x}) \geq 0$, $f(\boldsymbol{x}, t)$, $g(\boldsymbol{x}, t)$ and $u_0(\boldsymbol{x})$ are sufficiently smooth.

Operator splitting and method of lines

- We consider a partition of the elliptic operator into M terms:

$$L = L_1 + L_2 + \ldots + L_M.$$

- Operators L_i are simpler than L and permit spatial parallelization.
- Following the method of lines, we obtain a partitioned stiff IVP

$$\begin{aligned} U'(t) &= (A_1 + \ldots + A_M)\, U(t) + F(t), \qquad t \in (0, T], \\ U(0) &= U_0, \end{aligned}$$

by using a suitable spatial discretization method (e.g., finite differences, finite volumes or finite elements).

- U_0 is an appropriate restriction or projection of u_0, $U(t) \approx u(\boldsymbol{x}, t)$ and $A_i \approx L_i(\boldsymbol{x})$. Define $A = A_1 + \ldots + A_M$. F comprises the approximation of f and the contribution of the boundary conditions.
- Although not clearly specified, h denotes the spatial grid size.

Splitting time integrators

- Consider a uniform partition of the time interval $[0, T]$ into N_t subintervals of length $\delta t = T/N_t$. The time grid is given by

$$\{t_0, t_1, \ldots, t_{N_t}\}, \text{ with } t_i = i\,\delta t.$$

- Let us denote by $U_n \approx U(t_n)$ and $F_n \approx F(t_n)$.
- The fractional implicit Euler (FIE) scheme is an L-stable and first-order convergent method[1] defined as:

$$\begin{cases} (I - \delta t\, A_1) U_{n,1} = U_n + \delta t\, F_{n+1}, \\ (I - \delta t\, A_j) U_{n,j} = U_{n,j-1}, \qquad j = 2, 3, \ldots, M, \\ U_{n+1} = U_{n,M}, \end{cases}$$

whose stability function is given by $R_{FIE}(z_1, \ldots, z_M) = \prod_{j=1}^{M} \frac{1}{(1-z_j)}$.

[1] Hundsdorfer et al. Springer (2003)

Splitting time integrators

- The Douglas-Rachford (DR) scheme is an A-stable (not L-stable) and first-order convergent method defined as:

$$\begin{cases} (I - \delta t\, A_1) U_{n,1} = (I + \delta t\, A)\, U_n + \delta t\, F_n, \\ (I - \delta t\, A_j) U_{n,j} = U_{n,j-1}, & j = 2, 3, \ldots, M, \\ U_{n+1} = U_{n,M}, \end{cases}$$

  The stability function is given by

$$R_{DR}(z_1, \ldots, z_M) = 1 + \frac{\sum_{j=1}^{M} z_j}{\prod_{j=1}^{M} (1 - z_j)}.$$

- The splitting error with respect to Implicit Euler (IE) is first-order for FIE and second-order for DR.

Partitioning techniques and parallelization capabilities

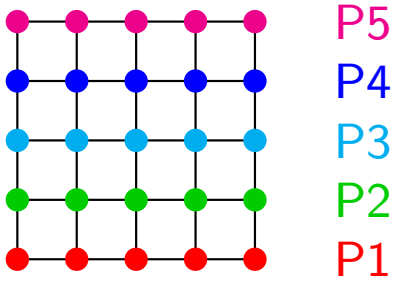
- There are different ways to split the elliptic operator in M terms, as:
  - Dimensional splitting: related to the spatial variables (requires Cartesian grids and no mixed derivatives).
  - Domain decomposition splitting: related to a decomposition of the spatial domain.
- At each internal stage, the resulting linear systems are spatially uncoupled and can be solved in parallel.
- Unlike other conventional parallel-in-space strategies, no iterative procedure is required to converge.

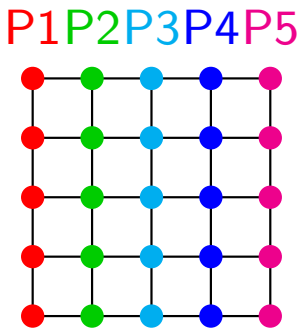
Dimensional splitting technique

- Assume that $\Omega \subset \mathbb{R}^d$ admits a Cartesian grid and K is a diagonal $d \times d$ matrix with diagonal elements $k_1, \ldots, k_d$.
- The dimensional splitting in $M = d$ terms of operator $L = L_1 + \ldots + L_M$ is given by the terms:

$$L_j u = \frac{\partial}{\partial x_j}(k_j \frac{\partial u}{\partial x_j}) - \frac{1}{M} cu, \quad \text{for } j = 1, \ldots, M.$$

- The resulting linear systems are essentially 1D and uncoupled.





Domain decomposition splitting technique

- Let $\Omega = \bigcup_{k=1}^{M} \Omega_k$ be an overlapping decomposition[2], where

$$\Omega_k = \bigcup_{l=1}^{q} \Omega_{kl}, \qquad \Omega_{ki} \cap \Omega_{kj} = \emptyset \text{ for } i \neq j.$$

- Let $\{\rho_k(\boldsymbol{x})\}_{k=1}^{M}$ be a smooth partition of unity such that

$$\begin{cases} \text{supp}\,(\rho_k(\boldsymbol{x})) \subset \overline{\Omega}_k, & \text{for } k = 1, \ldots, M, \\ 0 \leq \rho_k(\boldsymbol{x}) \leq 1, & \text{for } k = 1, \ldots, M, \\ \rho_1(\boldsymbol{x}) + \rho_2(\boldsymbol{x}) + \ldots + \rho_M(\boldsymbol{x}) = 1. \end{cases}$$

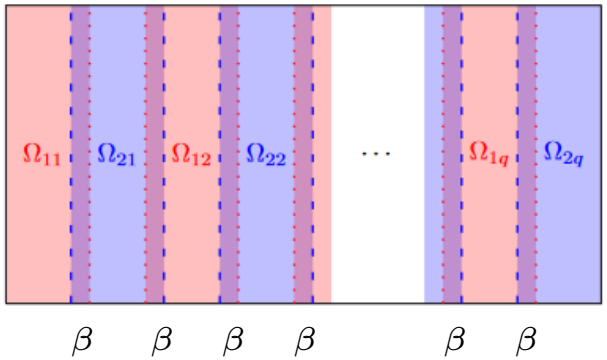
- The domain decomposition splitting in M terms of operator $L = L_1 + \ldots + L_M$ is given by the terms:

$$L_j u = \nabla \cdot (\rho_j K \nabla u) - \rho_j c u, \quad \text{for } j = 1, \ldots, M.$$

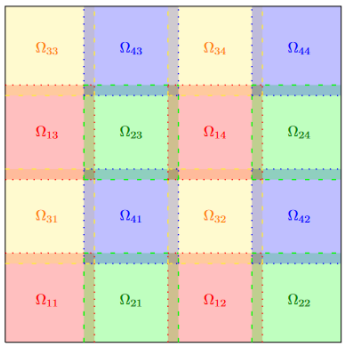
[2] Mathew et al. SIAM J. Sci. Comput. (1998)

Domain decomposition splitting technique

- Example with $M = 2$ overlapping subdomains and q general.



- Example with $M = 4$ overlapping subdomains and $q = 4$.



Connection to linear systems

- For the linear IVP

$$U'(t) = (A_1 + \ldots + A_M)\, U(t) + F(t), \qquad t \in (0, T],$$
$$U(0) = U_0,$$

a general one-step time integrator can be written as

$$U_0 = U(0),$$
$$U_i = \Phi_{\delta t}\, U_{i-1} + b_i, \;\; i = 1, \ldots, N_t,$$

which is equivalent to the linear system of equations

$$C\,\mathbf{U} \equiv \begin{bmatrix} I & & & \\ -\Phi_{\delta t} & I & & \\ & \ddots & \ddots & \\ & & -\Phi_{\delta t} & I \end{bmatrix} \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_{N_t} \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{N_t} \end{bmatrix} = \mathbf{b}, \text{ with } b_0 = U(0).$$

Connection to linear systems

- Traditional time marching corresponds to a block forward solve of this system, which is completely sequential

$$\begin{bmatrix} I & & & \\ -\Phi_{\delta t} & I & & \\ & \ddots & \ddots & \\ & & -\Phi_{\delta t} & I \end{bmatrix} \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_{N_t} \end{bmatrix} = \mathbf{b}.$$

- Examples of $\Phi_{\delta t}$:
  - $\Phi_{\delta t}^{IE} = (I - \delta t\, A)^{-1}$.
  - $\Phi_{\delta t}^{FIE} = (I - \delta t\, A_M)^{-1} (I - \delta t\, A_{M-1})^{-1} \ldots (I - \delta t\, A_1)^{-1}$.

Contents

Time-parallel time integration methods

- In this work we consider the Multigrid-reduction in time (MGRIT) method to further parallelize in time.
- The MGRIT algorithm is a multilevel iterative method for solving time-stepping problems.
- MGRIT achieves parallelism combining the original time-stepping problem with a coarser approximate representation.
- Parareal method can be interpreted as a two-level MGRIT scheme.
- Both MGRIT and parareal are non-intrusive iterative methods that converge in a finite number of iterations.
- They can also be applied to nonlinear problems.

Contents

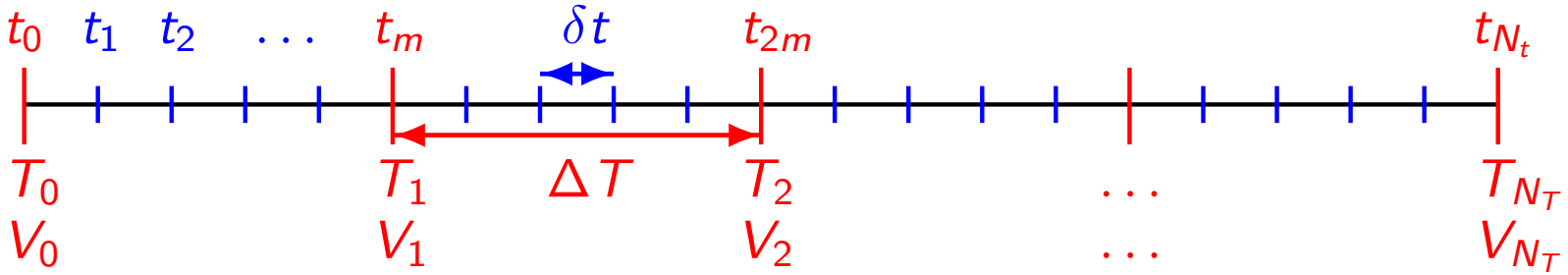
The parareal method

- The parareal method was first introduced by Lions, Maday and Turinici (2001) as a *parallel-in-time* iterative algorithm involving two propagators:
  - $\mathcal{F}_{\delta t}$, a fine propagator (expensive, accurate, parallel) with a small stepsize δt;
  - $\mathcal{G}_{\Delta T}$, a coarse propagator (inexpensive, inaccurate, sequential) with a large stepsize $\Delta T = m\,\delta t$ ($m \equiv$ coarsening factor).
- The parareal method can be interpreted[3] as a predictor-corrector method, a multiple shooting method, a fixed-point iteration method or a two-level multigrid method in time.

[3] Gander et al. SIAM J. Sci. Comput (2007)

The parareal method: time discretization

- Given $m > 1$ a coarsening factor, the idea of the parareal method is:
  - Consider a fine grid $\{t_0, t_1, \ldots, t_{N_t}\}$, with step-size $\delta t := T/N_t$ and a fine solution $U_n \approx U(t_n)$ for $n = 0, 1, \ldots, N_t$.
  - Consider a coarse grid $\{T_0, T_1, \ldots, T_{N_T}\}$, with step-size $\Delta T = m\,\delta t$, where $N_T = N_t/m$, and a coarse solution $V_n \approx U(T_n)$ for $n = 0, 1, \ldots, N_T$.



The parareal method: algorithm

- Initialization:

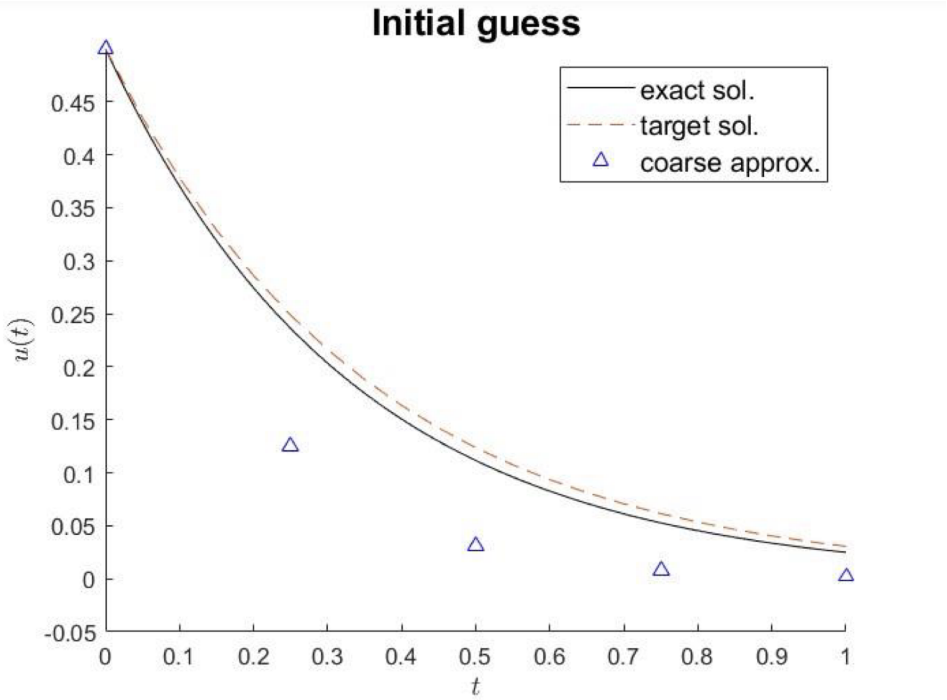
$$\begin{cases} V_0^{(0)} = V_0, \\ V_{n+1}^{(0)} = \mathcal{G}_{\Delta T}(T_n, T_{n+1}, V_n^{(0)}), \quad \text{for } n = 0, 1, \ldots, N_T - 1. \end{cases}$$

- For $k = 0, 1, \ldots,$ until convergence, do:

$$\begin{cases} V_0^{(k+1)} = V_0, \\ V_{n+1}^{(k+1)} = \mathcal{G}_{\Delta T}(T_n, T_{n+1}, V_n^{(k+1)}) + \mathcal{F}_{\delta t}^m(T_n, T_{n+1}, V_n^{(k)}) - \\ \qquad\qquad \mathcal{G}_{\Delta T}(T_n, T_{n+1}, V_n^{(k)}), \quad \text{for } n = 0, 1, \ldots, N_T - 1. \end{cases}$$

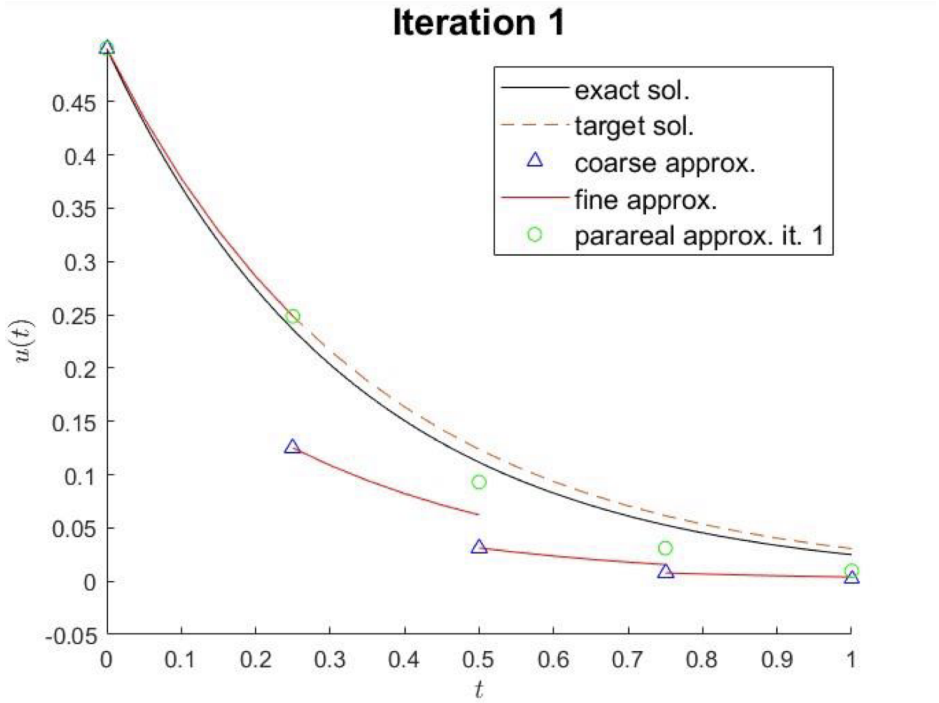
The parareal method: two graphical examples

- IVP: $y'(t) = -3\,y(t),\ t \in (0,1],\ y(0) = 1/2$
- Coarse/fine prop.: Forward/Backward Euler $\Delta T = \dfrac{T}{4}$ and $\delta t = \dfrac{T}{20}$.



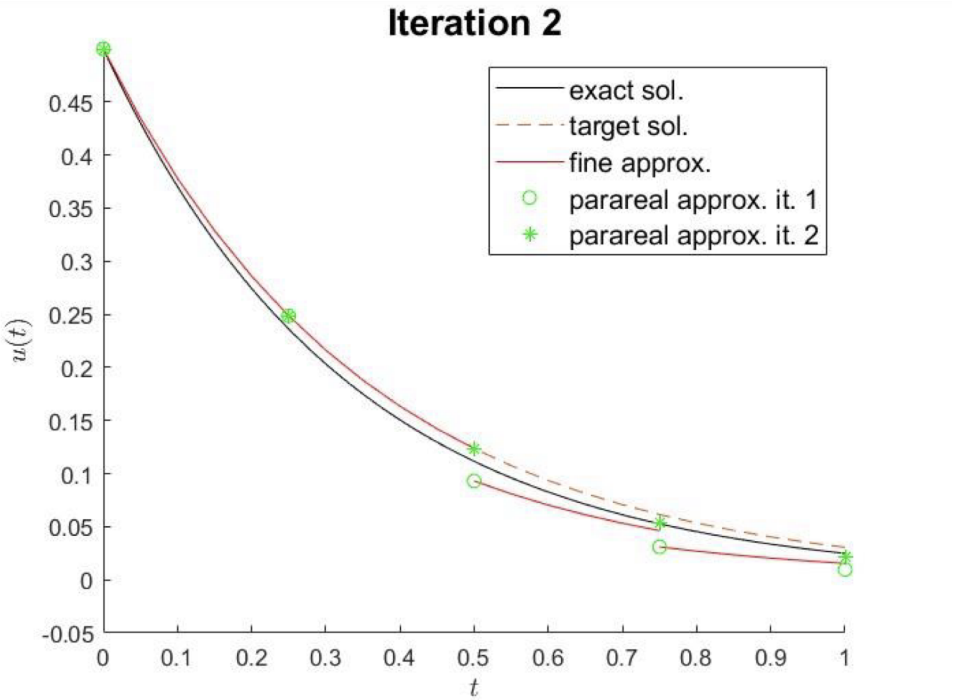
The parareal method: two graphical examples

- IVP: $y'(t) = -3\,y(t),\ t \in (0,1],\ y(0) = 1/2$
- Coarse/fine prop.: Forward/Backward Euler $\Delta T = \dfrac{T}{4}$ and $\delta t = \dfrac{T}{20}$.



The parareal method: two graphical examples

- IVP: $y'(t) = -3\,y(t),\ t \in (0,1],\ y(0) = 1/2$
- Coarse/fine prop.: Forward/Backward Euler $\Delta T = \dfrac{T}{4}$ and $\delta t = \dfrac{T}{20}$.

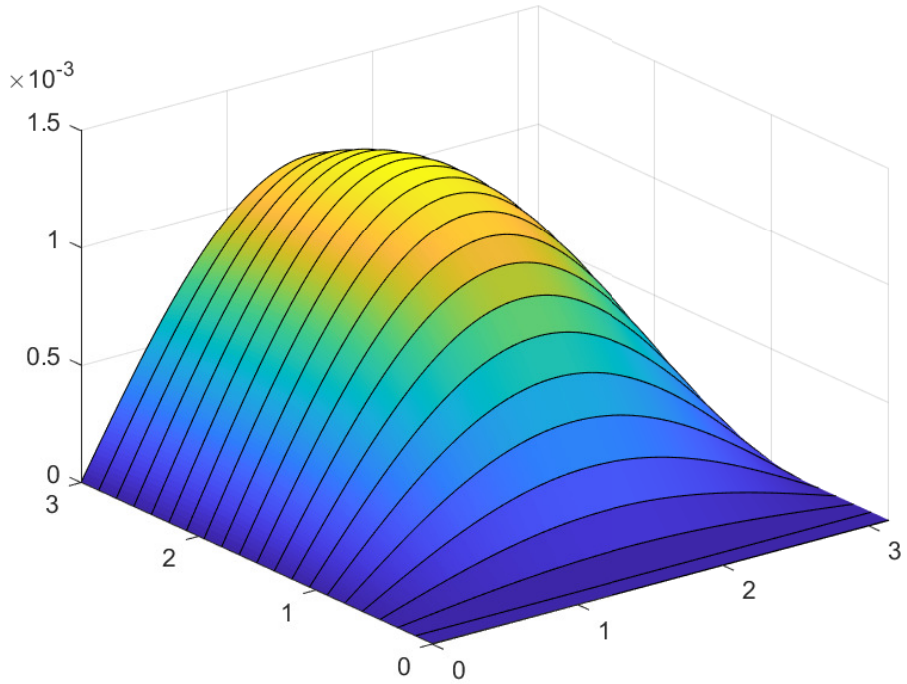


The parareal method: two graphical examples

- 1D Heat equation on $\Omega \times (0, T] \equiv (0, \pi) \times (0, 3]$, initial condition $u(0, x) = \sin x$, homogeneous Dirichlet b.c.
- Second-order finite differences with $h = \frac{\pi}{200}$.
- Coarse propagator: Implicit Euler with $\Delta T = \frac{3}{20}$.
- Fine propagator: SDIRK(2,2) ($\gamma = \frac{2-\sqrt{2}}{2}$) with $\delta t = \frac{3}{200}$.
- The error of the fine solution is $\approx 4 \cdot 10^{-6}$.
- After 6 parareal iterations, the error w.r.t. the fine solution at the coarse points is $\approx 5 \cdot 10^{-11}$.

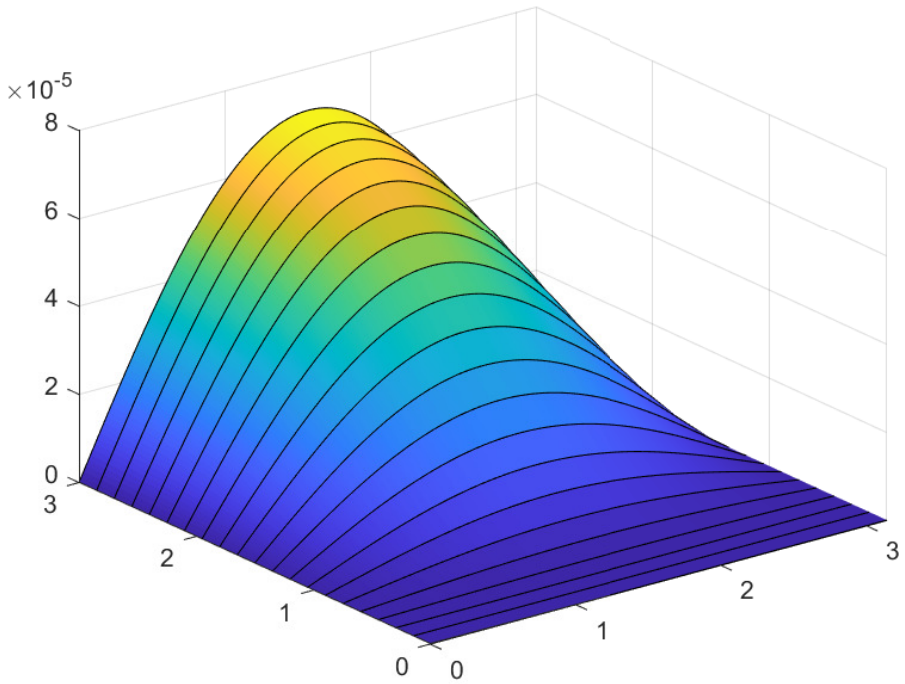
The parareal method: two graphical examples

Error at iteration 1:



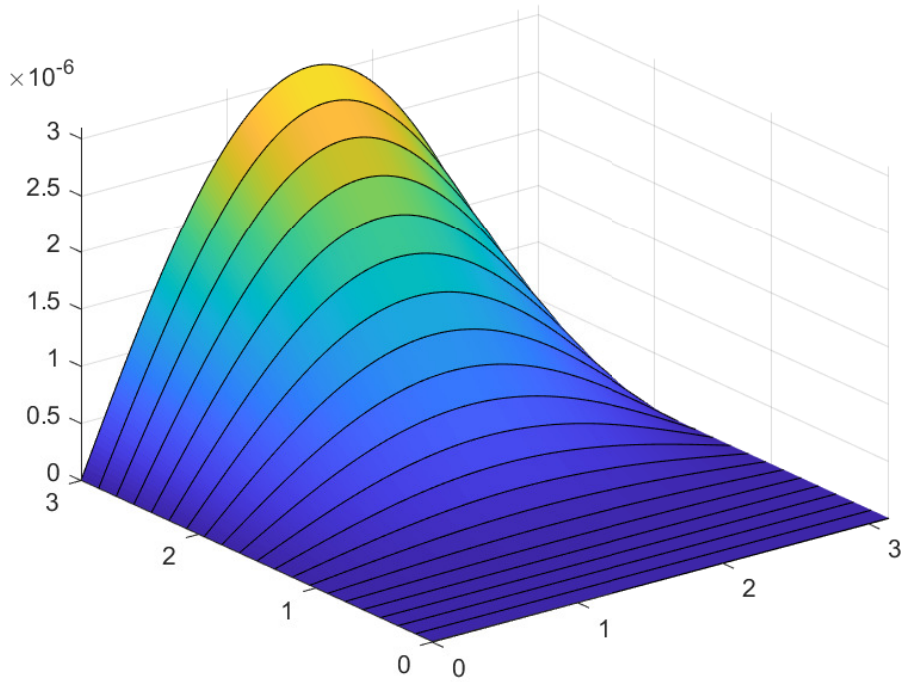
The parareal method: two graphical examples

Error at iteration 2:



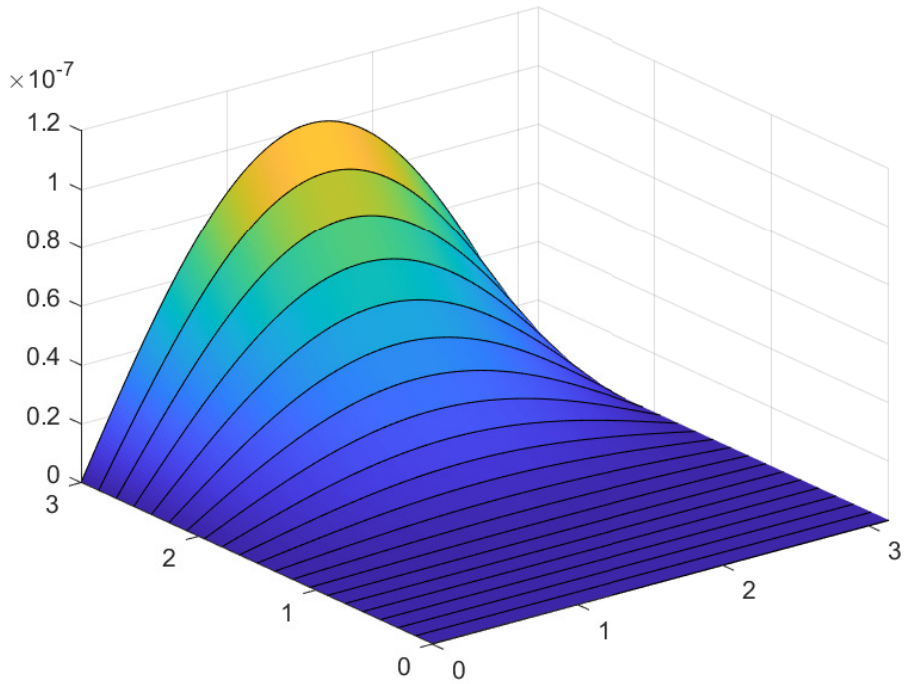
The parareal method: two graphical examples

Error at iteration 3:



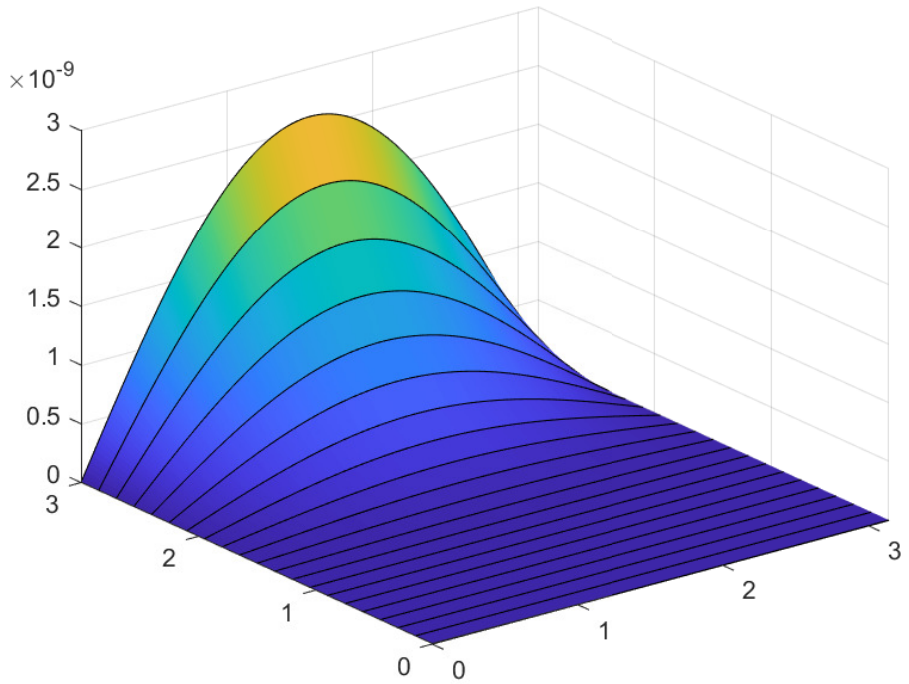
The parareal method: two graphical examples

Error at iteration 4:



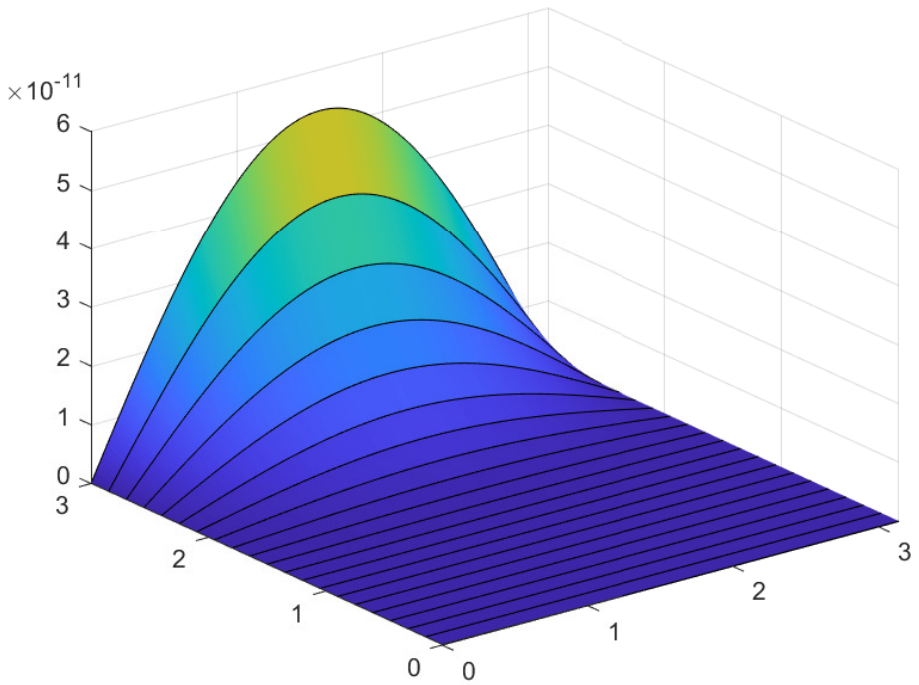
The parareal method: two graphical examples

Error at iteration 5:



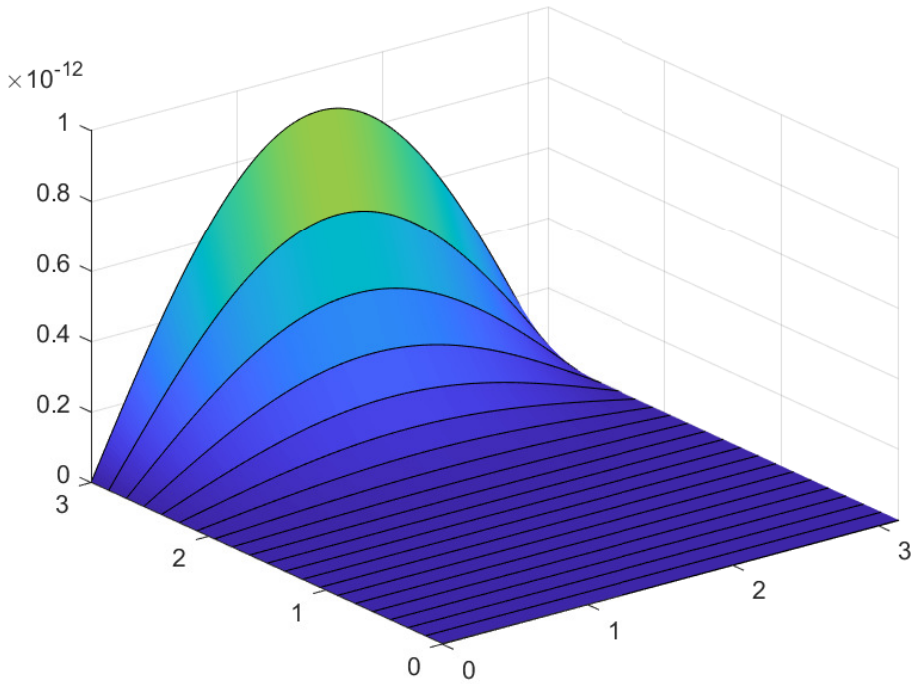
The parareal method: two graphical examples

Error at iteration 6:



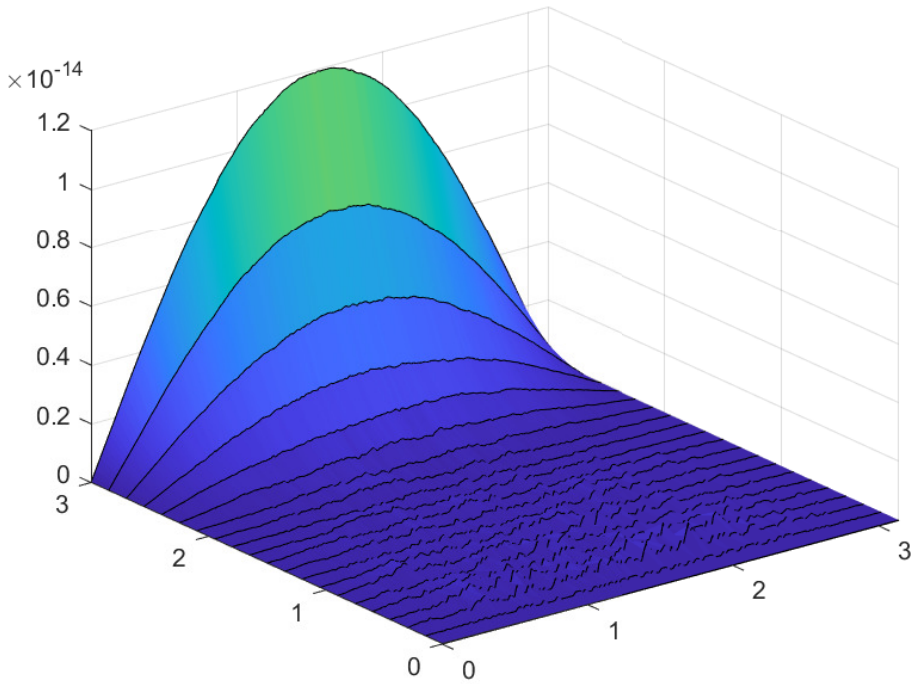
The parareal method: two graphical examples

Error at iteration 7:



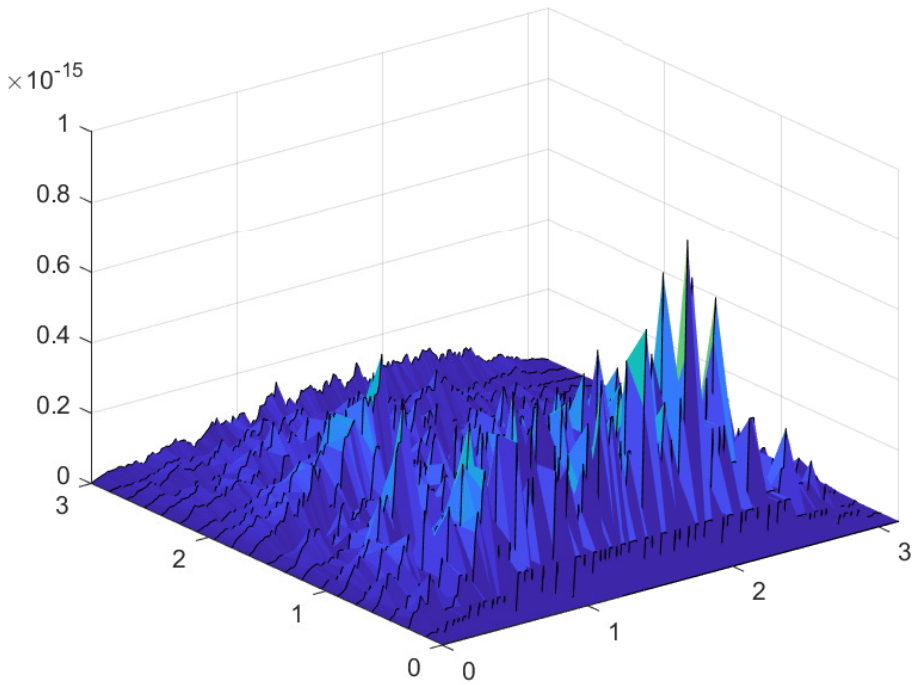
The parareal method: two graphical examples

Error at iteration 8:



The parareal method: two graphical examples

Error at iteration 9:



The parareal method: some properties

- If there are N_T large time subintervals, $V_n^{(k)}$ converges in exactly N_T iterations to the serial solution obtained with the fine propagator $\mathcal{F}$.
- Thus, the method is useful if we can stop the algorithm after $k << N_T$ iterations, obtaining a good approximation.
- We are free to choose the ode-solver we prefer for $\mathcal{F}$ and G. The chosen methods should be convergent and stable for the specified stepsizes δt and ΔT. In fact, L-stability is a key property for the coarse propagator[4].

[4] Friedhoff et al. Numer. Linear Algebra Appl. (2021)

Our proposal

- We propose to use splitting time integrators as the fine and coarse propagators, $\mathcal{F}$ and $\mathcal{G}$, in the parareal method.
- The idea is to design methods that allow for parallelization in both space and time.
- Unlike classical parareal, the new schemes allow both propagators to run in parallel.
- Unlike related existing methods (e.g., parareal Schwarz waveform relaxation methods), these schemes do not require any iteration to adjust the boundary conditions of the subdomains.
- Space-time parallel solvers considered:
  - Parareal FIE-FIE method.
  - Parareal FIE-DR method.

Convergence for the scalar case

- Let us consider the partitioned Dahlquist test problem for M splitting terms

$$\begin{cases} y'(t) = (\lambda_1 + \cdots + \lambda_M) y(t), & t \in (0, T], \\ y(0) = y_0, \end{cases}$$

with $\lambda_i \in \mathbb{C}$, for $i \in \{1, \ldots, M\}$.

- The convergence factor of the parareal algorithm is defined as

$$\mathcal{K}(z_1, \ldots, z_M, m) = \frac{|R_{\mathcal{F}}(z_1/m, \ldots, z_M/m)^m - R_{\mathcal{G}}(z_1, \ldots, z_M)|}{1 - |R_{\mathcal{G}}(z_1, \ldots, z_M)|},$$

where $R_{\mathcal{G}}$ and $R_{\mathcal{F}}$ are the stability functions of the coarse and fine propagators and $m = \Delta T / \delta t$.

Convergence for the scalar case

Theorem

The following bounds can be shown for the two parareal methods:

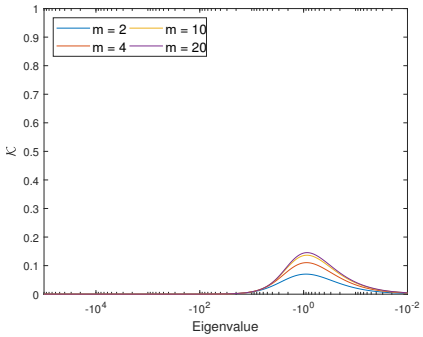
$$\mathcal{K}_{FIE\text{-}FIE}(z_1, \ldots, z_M, m) \leq 1/3,$$
$$\mathcal{K}_{FIE\text{-}DR}(z_1, \ldots, z_M, m) < 1,$$

for every $M \geq 2$, $(z_1, \ldots, z_M) \in (-\infty, 0)^M$ *and* $m > 1$.

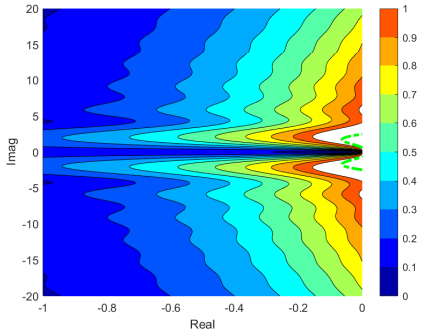
Proof.

The theorem is shown for $M = 2$ by using standard analysis tools for functions depending on two variables. The extension to the general case $M > 2$ is proven following an induction procedure. □

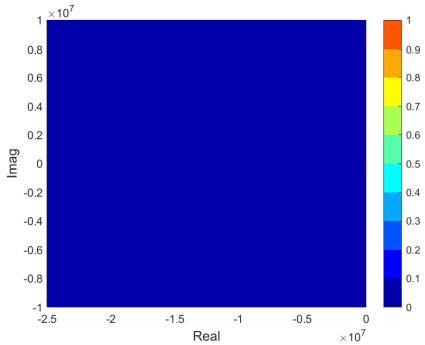
Convergence factor for Parareal FIE-FIE (M=2)



$\mathcal{K}_{\text{FIE-FIE}}(z, z, m)$ over $\mathbb{R}^-$

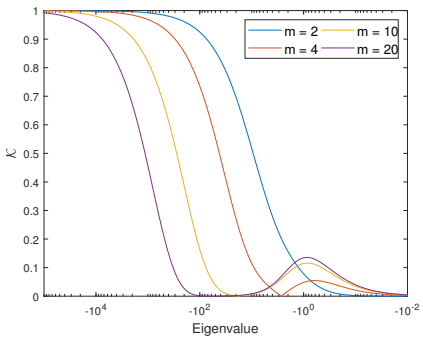


$\mathcal{K}_{\text{FIE-FIE}}(z, z, 1000)$ in $(-1, 0) \times (-20, 20)$

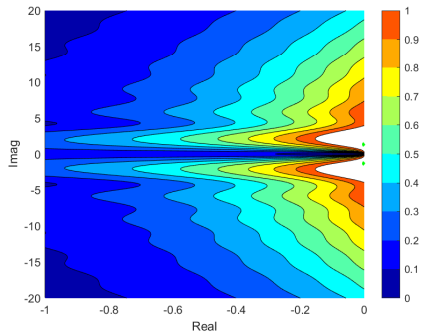


$\mathcal{K}_{\text{FIE-FIE}}(z, z, 1000)$ in $(-2.5 \cdot 10^7, 0) \times (-10^7, 10^7)$

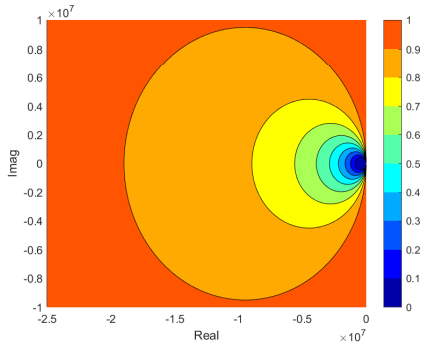
Convergence factor for Parareal FIE-DR (M=2)



$\mathcal{K}_{\text{FIE-DR}}(z, z, m)$ over $\mathbb{R}^-$



$\mathcal{K}_{\text{FIE-DR}}(z, z, 1000)$ in $(-1, 0) \times (-20, 20)$



$\mathcal{K}_{\text{FIE-DR}}(z, z, 1000)$ in $(-2.5 \cdot 10^7, 0) \times (-10^7, 10^7)$

Convergence for the matrix case

Theorem

The previous results can be generalized for matrix differential problems of the form

$$\begin{cases} y'(t) = (A_1 + A_2 + \ldots + A_M), y(t), & t \in (0, T], \\ y(0) = y_0, \end{cases}$$

with $A_i \in \mathbb{R}^{m \times m}$, *for* $i \in \{1, \ldots, M\}$ *if the set* $\{A_1, A_2, \ldots, A_M\}$ *is simultaneously diagonalizable (i.e. there exists an invertible matrix* $V \in \mathbb{R}^{m \times m}$ *such that* $V^{-1} A_i V$ *is a diagonal matrix for* $i = 1, 2, \ldots, M$*).*

Numerical experiments

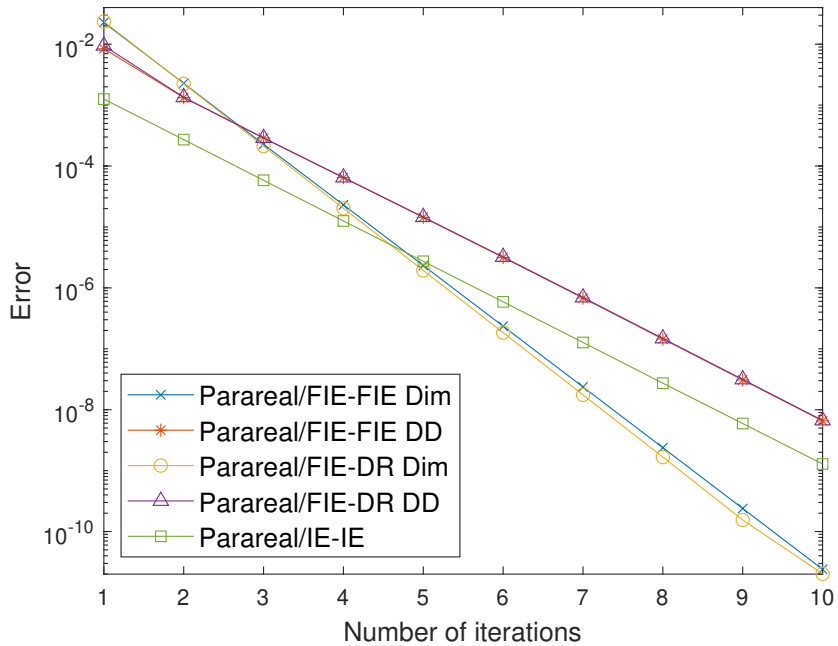
- Let us consider the 2D heat equation on $\Omega \times [0, T] = [0, 1]^3$, with homogeneous initial and Dirichlet boundary conditions, and source term f such that the exact solution is

$$u(x, y, t) = \sin(2\pi t)\, \sin(2\pi x)\, \sin(2\pi y).$$

- We consider a finite difference spatial discretization on a Cartesian mesh with size $h = 10^{-3}$.
- We consider $\Delta T = 5 \cdot 10^{-2}$.
- We consider splittings in $M = 2$ terms. For the DD splitting, we consider $q = 2$ and overlapping size between subdomains is $1/8$.

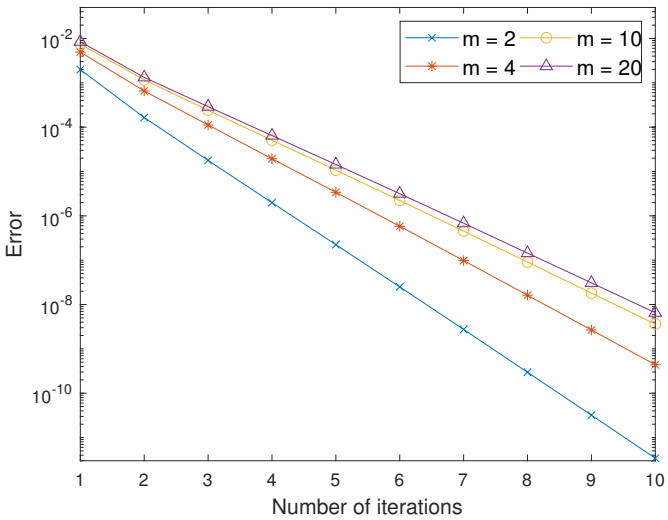
Numerical experiments

- Error with respect to the fine approximation
  $(m = 20,\ \Delta T = 5 \cdot 10^{-2},\ \delta t = 2.5 \cdot 10^{-3})$

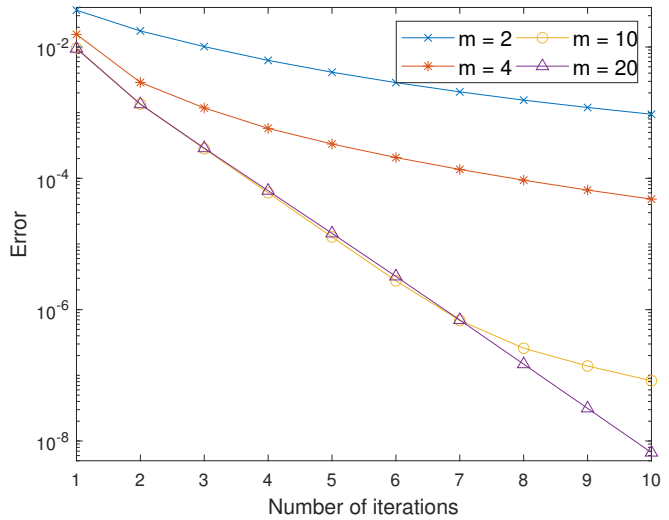


Numerical experiments

- Influence of $m = \Delta T/\delta t$ on the convergence for DD splitting



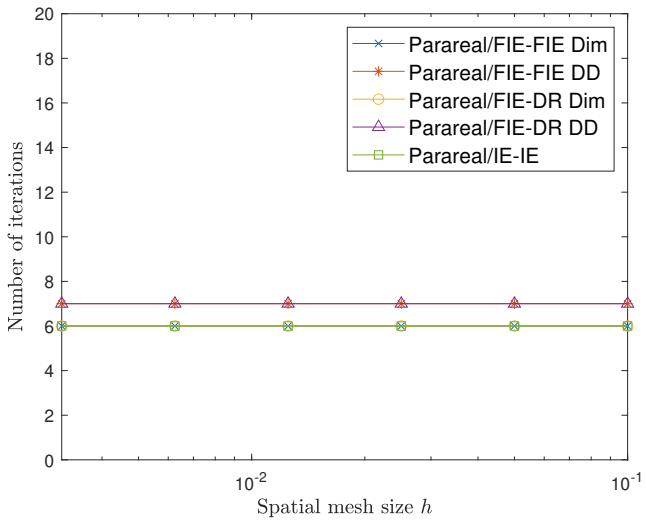
Parareal/FIE-FIE DD



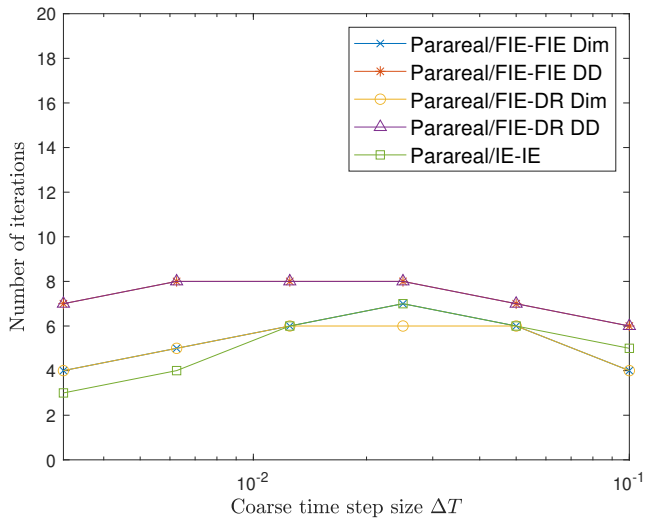
Parareal/FIE-DR DD

Numerical experiments

- Robustness with respect to h and ΔT: (Tol$= 10^{-6}$).



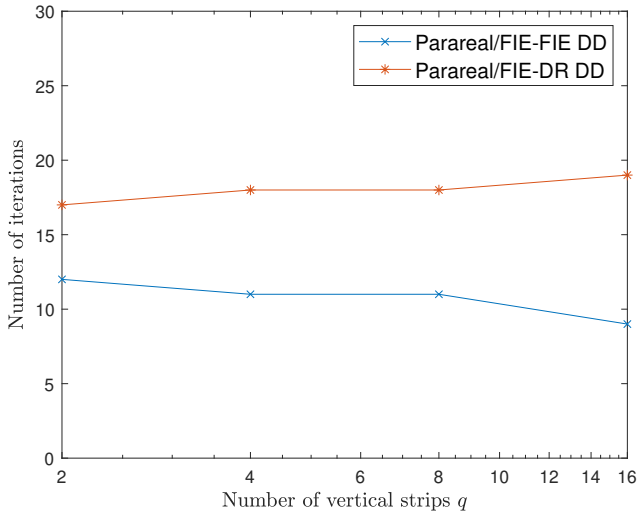
$\Delta T = 5 \cdot 10^{-2}$



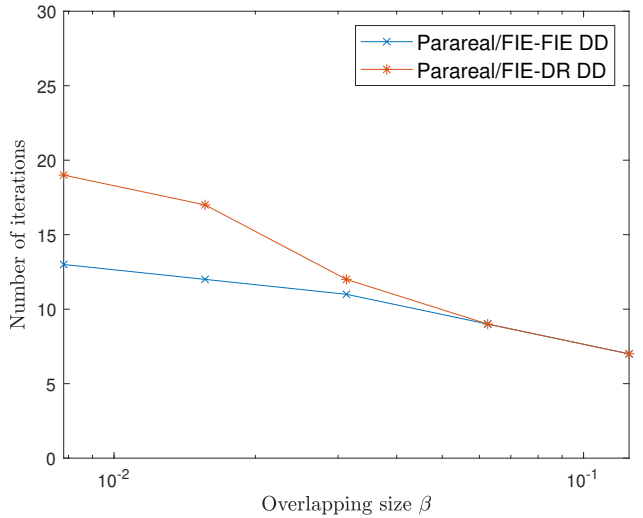
$h = 10^{-2}$

Numerical experiments

- How do the sizes of q and β affect the number of iterations?



$\beta = 1/64$

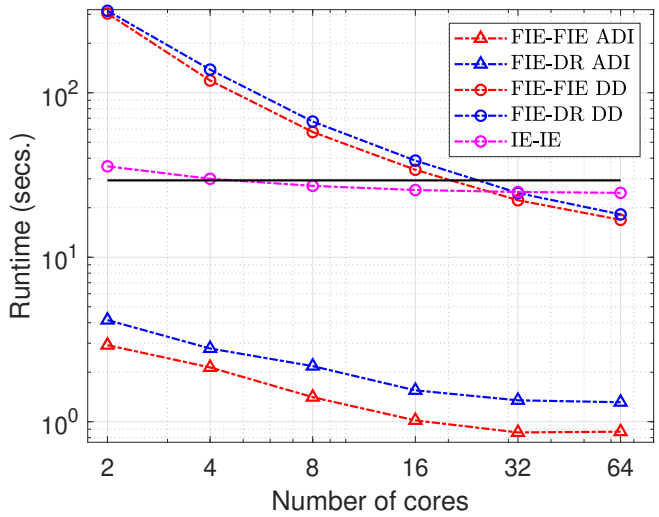


$q = 2$

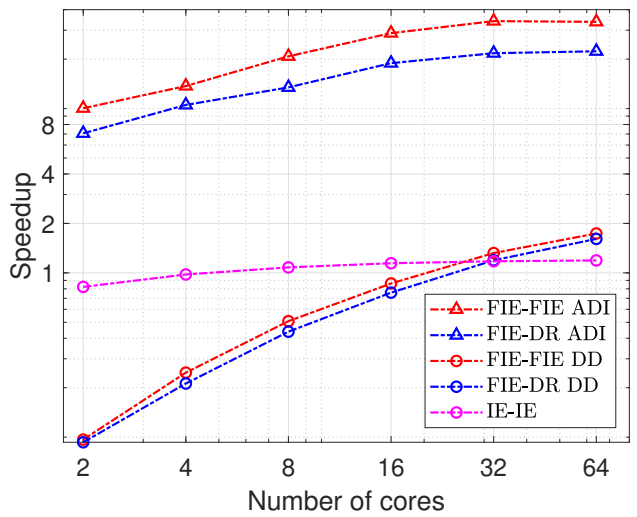
Numerical experiments

- Scalability test: $h = 5 \cdot 10^{-3}, \Delta T = 1/128, \delta t = 1/1280$. Coarse propagator: $q = 6$, $\beta = 1/12$; Fine propagator: $q = 4$, $\beta = 1/8$. Speedup: $S(n_p) = T_s / T_p(n_p)$.

CPU time



Speedup



Interpretations of parareal

1. Let the restriction operator be the $(N_T + 1) \times (N_t + 1)$ block-diagonal matrix $R = \text{diag}(R_I, R_I, \ldots, R_I, I)$, with

$$R_I = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}.$$

2. Let the prolongation operator be a block-diagonal matrix with $(N_t + 1) \times (N_T + 1)$ blocks, such that $P = \text{diag}(P_\Phi, \ldots, P_\Phi, I)$, with

$$P_\Phi = \begin{bmatrix} I \\ \Phi \\ \cdots \\ \Phi^{m-1} \end{bmatrix}.$$

Parareal as a fixed-point iteration method

1. Recall the connection of time-stepping methods with linear systems

$$C\,\mathbf{U} \equiv \begin{bmatrix} I & & & \\ -\Phi_{\delta t} & I & & \\ & \ddots & \ddots & \\ & & -\Phi_{\delta t} & I \end{bmatrix} \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_{N_t} \end{bmatrix} = \mathbf{b}.$$

2. Considering only the coarse time points, the previous system is equivalent to

$$C_F\,\mathbf{V} = \begin{bmatrix} I & & & \\ -\Phi_{\delta t}^{m} & I & & \\ & \ddots & \ddots & \\ & & -\Phi_{\delta t}^{m} & I \end{bmatrix} \begin{bmatrix} V_0 \\ V_1 \\ \vdots \\ V_{N_T} \end{bmatrix} = \mathbf{b}_F.$$

Parareal as a fixed-point iteration method

1. Parareal consists of solving the system

$$C_F\,\mathbf{V} = \begin{bmatrix} I & & & \\ -\Phi_{\delta t}^m & I & & \\ & \ddots & \ddots & \\ & & -\Phi_{\delta t}^m & I \end{bmatrix} \begin{bmatrix} V_0 \\ V_1 \\ \vdots \\ V_{N_T} \end{bmatrix} = \mathbf{b}_F$$

   with a fixed-point method based on a matrix splitting.

2. In particular, we consider $C_F = C_G - (C_G - C_F)$, where

$$C_G = \begin{bmatrix} I & & & \\ -\Phi_{\Delta T} & I & & \\ & \ddots & \ddots & \\ & & -\Phi_{\Delta T} & I \end{bmatrix}.$$

Parareal as a fixed-point iteration method

1. Defining $\mathbf{V} = R\mathbf{U}$, compute $V^{(0)}$ as the solution to

$$C_G\,\mathbf{V}^{(0)} = \mathbf{b}_G$$

For $k = 0, 1, 2, \ldots$ compute

$$C_G\,\mathbf{V}^{(k+1)} = (C_G - C_F)\,\mathbf{V}^{(k)} + \mathbf{b}_G$$

or, equivalently[5],

$$\mathbf{V}^{(k+1)} = (I - C_G^{-1}\,C_F)\mathbf{V}^{(k)} + C_G^{-1}\,\mathbf{b}_G.$$

2. $\mathbf{U}^{(k+1)} = P\,\mathbf{V}^{(k+1)}$.

[5] Ruprecht, Comput. Vis. Sci. (2018)

Parareal as a two-level MGRIT method

1. Taking into account that $\mathbf{V}^{(k)} = R\,\mathbf{U}^{(k)}$ and $C_F = R\,C\,P$ (Schur complement), the iteration can be written in terms of MGRIT method as:

$$\mathbf{U}^{(k+1)} = P\,\mathbf{V}^{(k+1)} = (I - P\,C_G^{-1}\,R\,C)\,P\,R\,\mathbf{U}^{(k)} + P\,C_G^{-1}\mathbf{b}_G.$$

2. Defining the error $\mathbf{e}^{(k+1)} = \mathbf{U}^{(k+1)} - \mathbf{U}$, it can be shown that

$$\mathbf{e}^{(k+1)} = E\,\mathbf{e}^{(k)} = \ldots = E^{k+1}\,\mathbf{e}^{(0)},$$

where

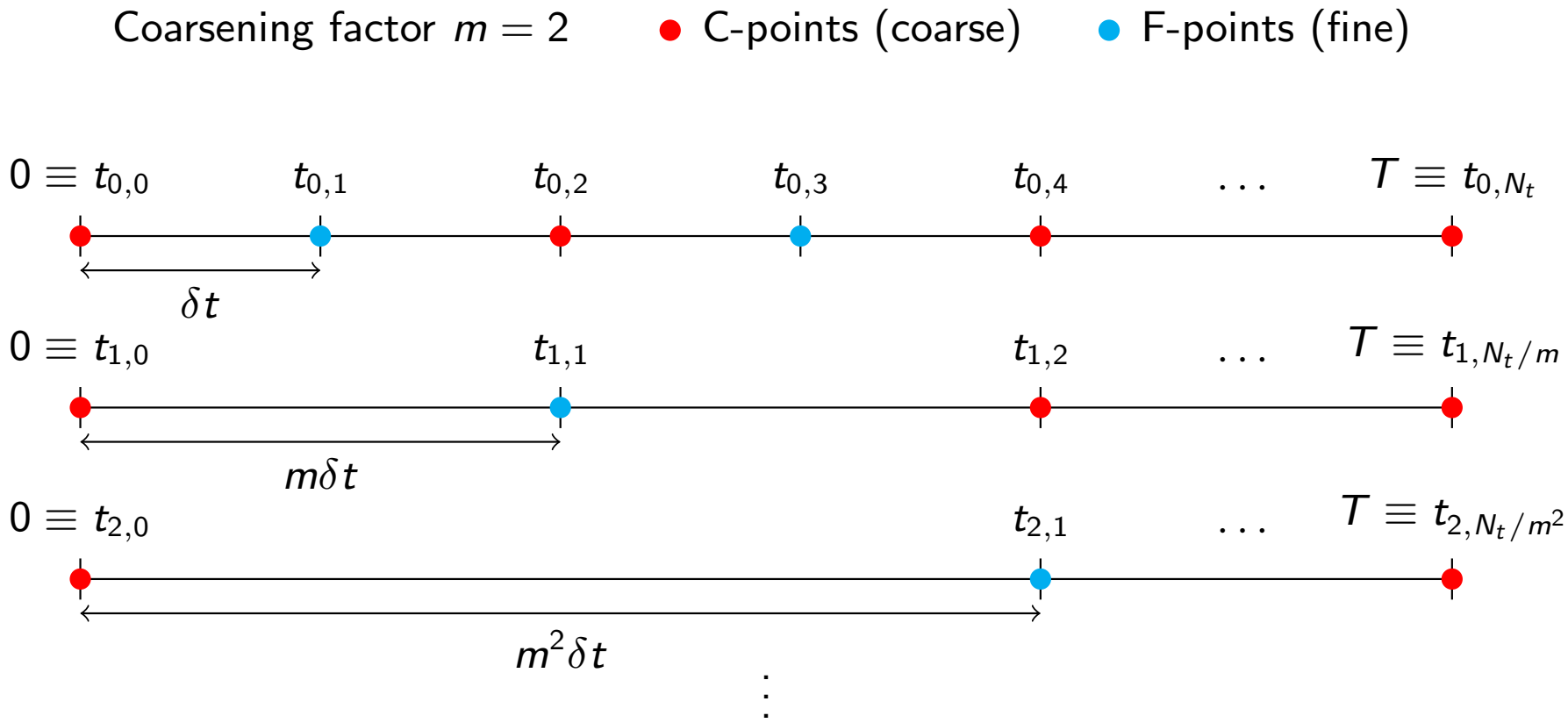
$$E = (I - P\,C_G^{-1}\,R\,C)\,P\,R,$$

is a nilpotent matrix.

3. The method can be extended with additional levels to reduce the bottleneck from sequential coarse-grid solves.

Contents

Partitions of the time interval



Time integration process

- For each level, the time integration process

$$C_\ell \mathbf{U}_\ell = \begin{bmatrix} I & & & \\ -\Phi_\ell & I & & \\ & \ddots & \ddots & \\ & & -\Phi_\ell & I \end{bmatrix} \begin{bmatrix} U_{\ell,0} \\ U_{\ell,1} \\ \vdots \\ U_{\ell,N_t/m^\ell} \end{bmatrix} = \mathbf{b}_\ell$$

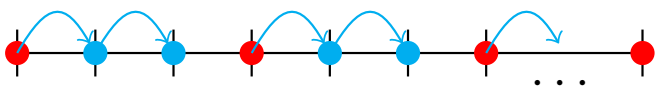
  is considered, $\ell \in \{0, 1, \ldots, \ell_{\max}\}$.

- Φ_ℓ is a matrix defined by a suitable integrator, e.g., the FIE scheme gives rise to

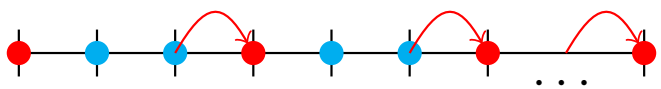
$$\Phi_\ell^{\mathrm{FIE}} = (I + m^\ell \delta t A_d)^{-1}(I + m^\ell \delta t A_{d-1})^{-1} \cdots (I + m^\ell \delta t A_1)^{-1}.$$

Relaxations

- The MGRIT method[6] is based on the multigrid-reduction methods. Thus, its components are similar.
- Relaxation: we consider F-relaxation and FCF-relaxation (combination of F-relaxation and C-relaxation). Both can be done in parallel.



F-relax.

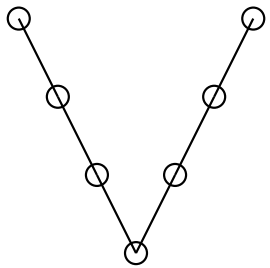


C-relax.

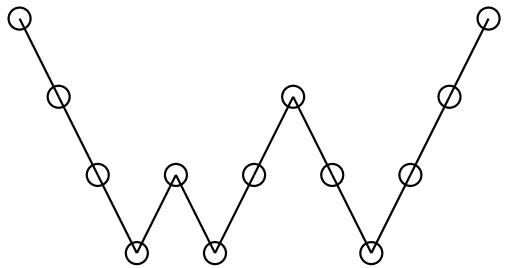
[6] Falgout et al., SIAM J. Sci. Comput. (2014)

Other components

- Restriction operator: injection (denoted by R).
- Prolongation operator: ideal interpolation (denoted by P).
- We mainly focus on V-cycles and F-cycles[7].



V-cycle



F-cycle

[7] Trottenberg et al. Elsevier Academic Press (2001)

The algorithm

1: **if** $\ell = \ell_{\max}$ **then**
2: Solve the problem on the coarsest level $\ell_{\max}$.
3: **else**
4: Relax on $C_\ell \mathbf{U}_\ell = \mathbf{b}_\ell$ using F(CF)$^\nu$-relaxation.
5: Compute and restrict the residual using injection, i.e., $\mathbf{b}_{\ell+1} \leftarrow R(\mathbf{b}_\ell - C_\ell \mathbf{U}_\ell)$.
6: Solve on the next level by the MGRIT algorithm at level $\ell + 1$
7: Correct and prolongate by ideal interpolation (F-relaxation),
$\mathbf{U}_\ell \leftarrow \mathbf{U}_\ell + P\mathbf{U}_{\ell+1}$.
8: **end if**
9: **return** $\mathbf{U}_\ell$

Iteration matrix

- Let the restriction operator be the $(N_t/m^{\ell+1}+1)\times(N_t/m^{\ell}+1)$ block-diagonal matrix $R_\ell = \text{diag}(R_{\ell,I}, R_{\ell,I}, \ldots, R_{\ell,I}, I)$, with

$$R_{\ell,I} = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}.$$

- Let the prolongation operator be a block-diagonal matrix with $(N_t/m^{\ell}+1)\times(N_t/m^{\ell+1}+1)$ blocks, such that $P_\ell = \text{diag}(P_{\ell,\Phi}, \ldots, P_{\ell,\Phi}, I)$, with

$$P_{\ell,\Phi} = \begin{bmatrix} I \\ \Phi_\ell \\ \vdots \\ \Phi_\ell^{m-1} \end{bmatrix}.$$

Iteration matrix

- Then, the iteration matrix is recursively defined by

$$M_{\ell_{\max}} = 0,$$
$$M_\ell = (I_\ell - P_\ell(I_{\ell+1} - M_{\ell+1}^\gamma)C_{\ell+1}^{-1}R_\ell C_\ell)P_\ell(I_\ell - R_\ell C_\ell P_\ell)^\nu R_\ell,$$

  for $\ell \in \{0, 1, \ldots, \ell_{\max} - 1\}$, where $\nu = 0$ for F-relax. and $\nu = 1$ for FCF-relax., whereas $\gamma = 1$ for V-cycles and $\gamma = 2$ for W-cycles.

- We only consider the two-level case ($\ell_{\max} = 1$) for the local Fourier analysis of the MGRIT-FIE-Dim method.

Space-time LFA

- We consider the d-dimensional heat equation with no source term and $K = I$.
- Let us define an infinite mesh in space with size h:

$$G_h = \{h\mathbf{j} \ : \ \mathbf{j} \in \mathbb{Z}^d\}.$$

- For the two-level MGRIT, two infinite meshes are considered in time,

$$G_{\delta t} = \{k\delta t \ : \ k \in \mathbb{N} \cup \{0\}\} \text{ and } G_{m\delta t} = \{km\delta t \ : \ k \in \mathbb{N} \cup \{0\}\}.$$

- The space-time infinite uniform fine and coarse grids are defined as $\Omega_{h,\delta t} = G_h \otimes G_{\delta t}$ and $\Omega_{h,m\delta t} = G_h \otimes G_{m\delta t}$, respectively.

Space-time LFA

- The space-time Fourier modes[8] are given by

$$\psi(\boldsymbol{\theta}, \omega, \mathbf{x}, t) = e^{\iota \boldsymbol{\theta} \cdot \mathbf{x}/h} e^{\iota \omega t/\delta t},$$

for $\boldsymbol{\theta} \in (-\pi, \pi]^d$ and $\omega \in \left(-\frac{\pi}{m}, 2\pi - \frac{\pi}{m}\right]$.

- We consider the space with coarsening in time

$$\mathcal{E}^{\boldsymbol{\theta},\omega_0} = \text{span}\{e^{\iota \boldsymbol{\theta} \cdot \mathbf{x}/h}\} \otimes \text{span}_{1 \le r \le m-1}\{e^{\iota \omega_0 t_{Lm+r}/\delta t} \mathbf{e}_r\},$$

for $\theta \in (-\pi, \pi]^d$ and $\omega_0 \in \left(-\frac{\pi}{m}, \frac{\pi}{m}\right]$, where $\mathbf{e}_r$ denotes a vector whose components are equal to 1 at t_{Lm+r} ($L \in \mathbb{N}$), and 0 otherwise.

[8] De Sterck et al., Numer. Linear Algebra Appl. (2020)

Space-time LFA

- If L_h is an infinite-grid Toeplitz operator, the space-time Fourier symbol $\hat{L}_h(\boldsymbol{\theta}, \omega_0)$ is given by

$$(L_h e(\boldsymbol{\theta}, \omega_0))_{\mathbf{j},k} = \hat{L}_h(\boldsymbol{\theta}, \omega_0) e_{\mathbf{j},k}(\boldsymbol{\theta}, \omega_0),$$

  where $e_{\mathbf{j},k} \in \mathcal{E}^{\boldsymbol{\theta},\omega_0}$.

- Accordingly, the space-time Fourier symbol of M_0 is defined as

$$\hat{M}_0 = (I - \hat{P}_0 \hat{C}_1^{-1} \hat{R}_0 \hat{C}_0) \hat{P}_0 (1 - \hat{R}_0 \hat{C}_0 \hat{P}_0)^{\nu} \hat{R}_0.$$

- Thus, we need to compute the symbols of C_0, C_1, R_0 and P_0 that depend on the spatial symbols of the time integration method

$$\tilde{\Phi}^{FIE}_{\ell,\boldsymbol{\theta}} = \prod_{j=1}^{d} \frac{1}{1 + 2\lambda\, m^{\ell}(1 - \cos\theta_j)}, \text{ for } \ell = 0, 1, \text{ where } \lambda = \delta t / h^2,$$

  if we consider ADI splitting and second-order finite differences.

Space-time LFA

- The asymptotic convergence factor is given by

$$\rho_{\text{LFA}}(M_0) = \sup\left\{\rho(\hat{M}_0(\boldsymbol{\theta},\omega)) \ : \ \boldsymbol{\theta} \in (-\pi,\pi]^d, \omega_0 \in \left(-\frac{\pi}{m},\frac{\pi}{m}\right]\right\}.$$

- We also consider the convergence bound given by Dobrev[9]

$$\rho_{\text{Dob}} = \max_{\mathbf{z}} \frac{|R(\mathbf{z}) - R(\mathbf{z}/m)^m|}{1-|R(\mathbf{z})|}|R(\mathbf{z}/m)^m|^\nu,$$

where $\mathbf{z} \in (-\infty, 0)^d$, and R is the stability function of the integrator.

[9] Dobrev et al. SIAM J. Sci. Comput. (2017).

Comparison of convergence factors for MGRIT-FIE method

- Consider the 3D heat equation with $T = 1$, $h = 2^{-7}$, $\delta t = 2^{-10}$, $u_0 \equiv 0$, and random initial guess (Tol: Machine precission).

			Experimental	ρ_{LFA}	ρ_{Dob}
F-relax.	$m = 2$	FIE	1.1074e-01	1.2482e-01	1.2500e-01
		IE	1.3182e-01	1.2482e-01	1.2500e-01
	$m = 4$	FIE	1.7589e-01	2.0094e-01	2.0386e-01
		IE	2.0033e-01	2.0257e-01	2.0386e-01
	$m = 32$	FIE	1.3154e-01	2.8256e-01	2.8557e-01
		IE	2.0952e-01	2.8256e-01	2.8557e-01
FCF-relax.	$m = 2$	FIE	4.2854e-02	5.1876e-02	5.2734e-02
		IE	5.9979e-02	5.2364e-02	5.2734e-02
	$m = 4$	FIE	5.9766e-02	7.7918e-02	8.1327e-02
		IE	8.0560e-02	7.8937e-02	8.1327e-02
	$m = 32$	FIE	2.7767e-02	1.0116e-01	1.0766e-01
		IE	7.6979e-02	1.0116e-01	1.0766e-01

Numerical experiment

- We consider the 3D heat equation with homogeneous Dirichlet boundary conditions, such that the exact solution is

$$u(x, y, z, t) = \sin(2\pi t)\sin(2\pi x)\sin(2\pi y)\sin(2\pi z),$$

  where $(x, y, z) \in \Omega = (0, 1)^3$, $t \in [0, T]$.
- Second-order central finite differences are chosen for the spatial discretization.
- The time integrator is chosen to be FIE with dimensional splitting (into 3 terms) or IE.
- We denote $M = 1/h$ and $N = 1/\delta t$ as the discretization parameters.

Robustness considering F-relaxation

Tol for the residual=10^{-9}

			$M \times N$	$2^5 \times 2^6$	$2^6 \times 2^8$	$2^7 \times 2^{10}$
F-relax	two-level	$m = 2$	FIE	8	8	7
			IE	11	11	8
		$m = 4$	FIE	7	10	8
			IE	11	15	11
	V-cycle	$m = 2$	FIE	8	11	13
			IE	14	23	30
		$m = 4$	FIE	7	10	11
			IE	11	17	22
	F-cycle	$m = 2$	FIE	8	8	6
			IE	11	10	8
		$m = 4$	FIE	7	10	8
			IE	11	15	10

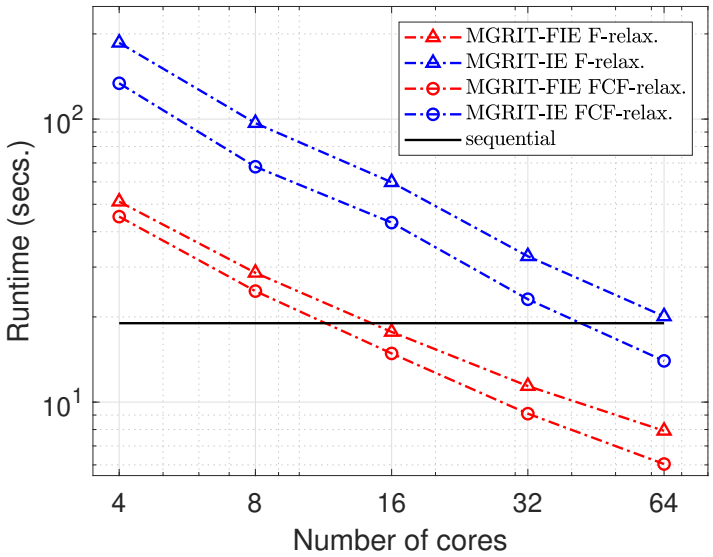
Robustness considering FCF-relaxation

			$M \times N$	$2^5 \times 2^6$	$2^6 \times 2^8$	$2^7 \times 2^{10}$
FCF-relax	two-level	$m = 2$	FIE	4	7	6
			IE	6	8	8
		$m = 4$	FIE	3	6	7
			IE	4	8	10
	V-cycle	$m = 2$	FIE	4	7	8
			IE	6	9	11
		$m = 4$	FIE	3	6	8
			IE	4	8	10
	F-cycle	$m = 2$	FIE	4	7	6
			IE	6	8	8
		$m = 4$	FIE	3	6	7
			IE	4	8	10

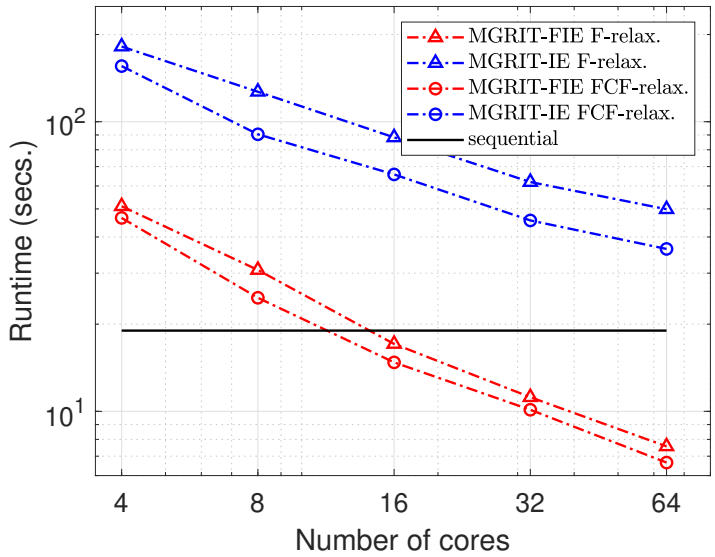
Scalability

- Simple scalability test for $m = 16$, $M = 2^6$ and $N = 2^{10}$. Up to 8 processors are devoted to parallelization in space.

Two-level MGRIT



V-cycle MGRIT



This seems nice but...

- We should take into account the splitting error generated by the described partitioning.
- The error of the target solution may increase significantly and finer time steps may be required.
- If the splitting error is too large, another simple alternative is to consider the integration on the finest level by the implicit Euler, and the rest of the levels by the FIE.
- The resulting method looses space parallelization on the finest grid.
- However, the obtained approximation could be far more accurate than the one obtained by the MGRIT and FIE.

Robustness

		$2^6 \times 2^8$	FIE	IE	FIE+IE
F-relax	two-level	$m = 2$	8	11	8
		$m = 4$	10	15	7
	V-cycle	$m = 2$	11	23	9
		$m = 4$	10	17	8
	F-cycle	$m = 2$	8	10	8
		$m = 4$	10	15	7
FCF-relax	two-level	$m = 2$	7	8	7
		$m = 4$	6	8	5
	V-cycle	$m = 2$	7	9	6
		$m = 4$	6	8	5
	F-cycle	$m = 2$	7	8	7
		$m = 4$	6	8	5

Contents

Conclusions and work in progress

- We designed and analyzed space-time parallel methods that consider first-order splitting time integrators.
- Compared to classical parareal/MGRIT methods, the presented schemes also use parallel computations in the coarsest level.
- Regarding CPU times, the new schemes can outperform the classical Parareal and MGRIT method considering the IE scheme.
- The number of iterations for reaching certain tolerance can decrease (especially for V-cycles and F-relax.).
- In progress:
  - Extend the scalability analyses considering more cores.
  - Analyze the MGRIT-FIE method with DD splitting (more flexible).
  - Consider higher-order splitting time integrators.
  - Test the solvers in complex problems (e.g., porous media).

Some bibliography

- A. Arrarás, F.J. Gaspar, I. Jimenez, L.P. *Space-time parallel iterative solvers for the integration of parabolic problems*, arXiv:2502.08370 [math.NA].
- T. Buvoli, M. Minion (2021). *IMEX Runge-Kutta Parareal for non-diffusive equations*. In B. Ong et al. (eds.), Parallel-in-Time Integration Methods. Springer Proc. Math. Stat., vol. 356, 95–127.
- H. De Sterck, S. Friedhoff, A. J. M. Howse, S. P. MacLachlan, *Convergence analysis for parallel-in-time solution of hyperbolic systems*, Numer. Linear Algebra Appl., 27 (2020), e2271.
- V. A. Dobrev, T. Kolev, N. A. Petersson, J. B. Schroder, *Two-level convergence theory for multigrid reduction in time (MGRIT)*, SIAM J. Sci. Comput., 39 (2017), pp. 5501–5527.
- R. D. Falgout, S. Friedhoff, T. V. Kolev, S. P. MacLachlan, J. B. Schroder, *Parallel time integration with multigrid*, SIAM J. Sci. Comput., 36 (2014), pp. C635–C661.
- S. Friedhoff, B. S. Southworth, *On "optimal" h-independent convergence of Parareal and multigrid-reduction-in-time using Runge-Kutta time integration*, Numer. Linear Algebra Appl., 28 (2021).
- M. Gander, S. Vandewalle (2007). *Analysis of the parareal time-parallel time-integration method*. SIAM J. Sci. Comput., 29(2): 556–578.
- W. Hundsdorfer, J. Verwer (2003). *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations.* In Springer Ser. Comput. Math., vol. 33. Springer-Verlag, Berlin/Heidelberg.
- J.L. Lions, Yvon Maday, G. Turinici (2001). *Résolution d'EDP par un schéma en temps "pararéel"*. C. R. Acad. Sci. (1) - Math., 332(7): 661–668.
- T.P. Mathew, P.L. Polyakov, G. Russo, J. Wang (1998). *Domain decomposition operator splittings for the solution of parabolic equations*. SIAM J. Sci. Comput., 19(3): 912–932.
- D. Ruprecht (2018). *Wave propagation characteristics of Parareal*. Comput. Vis. Sci., 19: 1–17.
- U. Trottenberg, C. W. Oosterlee, A. Schüller, *Multigrid*, Elsevier Academic Press, San Diego, 2001.

Space-time parallel solvers based on multigrid-reduction-in-time and operator splitting methods

Laura Portero

Public University of Navarre, Spain
INAMAT2

Joint work with: A. Arrarás, F.J. Gaspar and I. Jimenez

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Universidad Zaragoza
1542

Innovative & Efficient Strategies for Stiff Differential Equations