Asymptotic convergence of the parallel full approximation scheme in space and time for linear problems

Matthias Bolten[1], Dieter Moser[2], Robert Speck[2*]

[1]*Department of Mathematics, Universität Kassel, Germany.*
[2] *Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH, Germany.*

SUMMARY

For time-dependent partial differential equations, parallel-in-time integration using the "parallel full approximation scheme in space and time" (PFASST) is a promising way to accelerate existing space-parallel approaches beyond their scaling limits. Inspired by the classical Parareal method and multigrid ideas, PFASST allows to integrate multiple time-steps simultaneously using a space-time hierarchy of spectral deferred correction sweeps. While many use cases and benchmarks exist, a solid and reliable mathematical foundation is still missing. Very recently, however, PFASST for linear problems has been identified as a multigrid method. in this paper, we will use this multigrid formulation and in particular PFASST's iteration matrix to show that in the non-stiff as well as in the stiff limit PFASST indeed is a convergent iterative method. We will provide upper bounds for the spectral radius of the iteration matrix and investigate how PFASST performs for increasing numbers of parallel time-steps. Finally, we will demonstrate that the results obtained here indeed relate to actual PFASST runs. Copyright © 0000 John Wiley & Sons, Ltd.

1. INTRODUCTION

With the advent of supercomputing architectures featuring millions of processing units, classical parallelization techniques used to accelerate the solution of discretized partial differential equations face new challenges. For fixed-size problems, communication starts to dominate eventually, when only small portions of data are left for computation on each unit. This "trap of strong scaling" leads to severe and inevitable upper limits for speedup obtainable with parallelization in space, leaving large parts of extreme scale supercomputers unexploited. If weak scaling is the target, this may not be an issue, but for time-dependent problems stability considerations often lead to an increase in the number of time-steps as the problem is refined in space. This is not mitigated by spatial parallelization alone, yielding the "trap of weak scaling". Thus, the challenges arising from the extreme levels of parallelism required by today's and future high-performance computing systems mandates the development of new numerical methods that feature a maximum degree of concurrency.

For time-dependent problems, in particular for time-dependent partial differential equations, approaches for the parallelization along the temporal dimension have become increasingly popular over the last years. In his seminal work in 2015 Gander lists over 25 approaches to parallelize the

---

*Correspondence to: E-mail: r.speck@fz-juelich.de

seemingly serial process of time integration [1]. In particular, the invention of the Parareal method in 2001 [2] alone sparked a multitude of new developments in this area. This "parallelization across the step" approach allows to integrate many time-steps simultaneously. This can work on top of already existing parallelization strategies in space. The idea is to derive a coarser, less expensive time-integration scheme for the problem at hand and use this so-called coarse propagator to quickly and serially propagate information forward in time. The original integrator, in this context often called the fine propagator, is then used in parallel-in-time using the initial values the coarse scheme provided. This cycle of fine and coarse, parallel and serial time-stepping is repeated and upon convergence, Parareal is as accurate as the fine propagator run in serial. This way, the costs of the expensive fine scheme are distributed, while the serial-in-time part is kept small using a cheap propagator. This predictor-corrector approach, being easy to implement and easy to apply, has been analyzed extensively. It has been identified as a multiple shooting method or as an FAS multigrid scheme [3] and convergence has been proven under various conditions, see e.g. [3, 4, 5, 6, 7].

Yet, a key drawback of Parareal is the severe upper bound on parallel efficiency. If K iterations are required for convergence, the efficiency is bounded by $1/K$. Perfect linear speedup cannot be expected due to the serial coarse propagator, but efficiencies of a few percent are also not desirable. Therefore, many researchers started to enhance the Parareal idea with the goal of loosening this harsh bound on parallel efficiency. One idea is to replace the fine and the coarse propagators by iterative solvers and coupling their "inner" iteration with the "outer" Parareal iteration. A first step in this direction was done in [8], where spectral deferred correction methods (SDC, see [9]) were used within Parareal. This led to the "parallel full approximation scheme in space and time" (PFASST), which augmented this approach by ideas from non-linear multigrid methods [10, 11]. In these original papers from 2012 and 2014, the PFASST algorithm was introduced, its implementation was discussed and it was applied to first problems. In the following years, PFASST has been applied to more and more problems and coupled to different space-parallel solvers, ranging from a Barnes-Hut tree code to geometric multigrid, see [12, 13, 14, 15]. Together with spatial parallelization, PFASST was demonstrated to run and scale on up to $458{,}752$ cores of an IBM Blue Gene/Q installation. Yet, while applications, implementation and improvements are discussed frequently, a solid and reliable convergence theory is still missing. While for Parareal many results exist and provide a profound basis for a deep understanding of this algorithm, this is by far not the case for PFASST. Very recently, however, PFASST for linear problems was identified as a multigrid method in [16, 17] and the definition of its iteration matrix yielded a new understanding of the algorithm's components and their mechanics. This understanding allows to analyze the method using the established Local Fourier Analysis (LFA) technique. LFA has been introduced to study smoothers in [18], later it was extended to study the whole multigrid algorithm [19] and since then it has become a standard tool for the analysis of multigrid. For a detailed introduction see [20, 21]. In the context of space-time multigrid the results obtained using plain LFA are less meaningful because of the non-normality due to the discretization of the time domain. To overcome this limitation the semi-algebraic mode analysis (SAMA) has been introduced in [22]. A kindred idea has been used to analyze PFASST in [16, 17]. Although this careful block Fourier mode analysis already revealed many interesting features and also limitations, a rigorous proof of convergence has not been provided so far.

In this paper, we will use the multigrid formulation of PFASST for linear problems and in particular the iteration matrix to show that in the non-stiff as well as in the stiff limit PFASST indeed is a convergent iterative method. We will provide upper bounds for the spectral radius of the iteration matrix and show that under certain assumptions, PFASST also satisfies the approximation property of standard multigrid theory. In contrast, the smoothing property does not hold, but we will state a modified smoother which allows to satisfy also this property. We will further investigate how PFASST performs for increasing numbers of parallel time-steps. Finally, we will demonstrate that the results obtained here indeed relate to actual PFASST runs. We start with a brief summary of the results found in [16], describing PFASST as a multigrid method.

2. A MULTIGRID PERSPECTIVE ON PFASST

We focus on linear, autonomous systems of ordinary differential equations (ODEs) with

$$u_t(t) = \mathbf{S}u(t) \text{ for } t \in [0, T],$$
$$u(0) = u_0 \tag{1}$$

with $u(t) \in \mathbb{C}^N$, $T > 0$, initial value $u_0 \in \mathbb{C}^N$ and "spatial" matrix $\mathbf{S} \in \mathbb{C}^{N \times N}$, stemming from e.g. a spatial discretization of a partial differential equation (PDE). Examples include the heat or the advection equation, but also the wave equation and other types of linear PDEs and ODEs.

2.1. The collocation problem and SDC

We decompose the time interval into L subintervals $[t_l, t_{l+1}]$, $l = 0, ..., L-1$ and rewrite the ODE for such a time-step in Picard formulation as

$$u(t) = u_l + \int_{t_l}^t \mathbf{S}u(s)ds, \ t \in [t_l, t_{l+1}],$$

where u_l is the initial condition for this time-step, e.g. coming from a time-stepping scheme. Introducing M quadrature nodes $\tau_1, ..., \tau_M$ with $t_l \leq \tau_1 < ... < \tau_M = t_{l+1}$, we can approximate the integrals from t_l to these nodes τ_m using spectral quadrature like Gauß-Radau or Gauß-Lobatto quadrature, such that

$$u_m = u_l + \Delta t \sum_{j=1}^M q_{m,j}\mathbf{S}u_j \approx u_l + \int_{t_l}^{\tau_m} \mathbf{S}u(s)ds, \text{ for } m = 1, ..., M,$$

where $u_m \approx u(\tau_m)$, $\Delta t = t_{l+1} - t_l$ and $q_{m,j}$ represent the quadrature weights for the interval $[t_l, \tau_m]$ with

$$q_{m,j} := \int_{t_l}^{\tau_m} \ell_j(s)ds, \quad m, j = 1, \ldots, M,$$

where ℓ_j are the Lagrange polynomials to the points τ_m. Note that for the quadrature rule on each subinterval $[\tau_m, \tau_{m+1}]$, $m = 1, ..., M-1$ all collocation nodes are taken into account, even if they do not belong to the subinterval under consideration. Combining this into one set of linear equations yields

$$U = U_l + \Delta t \left(\mathbf{Q} \otimes \mathbf{S}\right) U \quad \text{or} \quad \left(\mathbf{I}_{MN} - \Delta t \mathbf{Q} \otimes \mathbf{S}\right) U = U_l \tag{2}$$

for vectors $U = (u_1, ..., u_M)^T$, $U_l = (u_l, ..., u_l)^T \in \mathbb{C}^{MN}$ and quadrature matrix $\mathbf{Q} = (q_{m,j}) \in \mathbb{R}^{M \times M}$. This is the so-called "collocation problem" and it is equivalent to a fully implicit Runge-Kutta method. Before we proceed with describing the solution strategy for this problem, we slightly change the notation: Instead of working with the term $\Delta t \mathbf{Q} \otimes \mathbf{S}$, we introduce the "CFL number" μ (sometimes called the "discrete dispersion relation number") to absorb the time-step size Δt, problem-specific parameters like diffusion coefficients as well as the spatial mesh size Δx, if applicable. We write

$$\Delta t \mathbf{S} = \mu \mathbf{A},$$

where the matrix $\mathbf{A}$ is the normalized description of the spatial problem or system of ODEs. For example, for the heat and the advection equation, the parameter μ is defined by

$$\mu_{\text{diff}} = \nu \frac{\Delta t}{\Delta x^2}, \quad \mu_{\text{adv}} = c \frac{\Delta t}{\Delta x}$$

with diffusion coefficient ν and advection speed c. Then, Equation (2) reads

$$\left(\mathbf{I}_{MN} - \mu \mathbf{Q} \otimes \mathbf{A}\right) U = U_l \tag{3}$$

and we will use this form for the remainder of this paper.

This system of equations is dense and a direct solution is not advisable, in particular if the right-hand side of the ODE is non-linear. While the standard way of solving this is a simplified Newton approach [23], the more recent development of spectral deferred correction methods (SDC, see [9]) provides an interesting and very flexible alternative. In order to present this approach, we follow the idea of preconditioned Picard iteration as found e.g. in [24, 25, 26]. The key idea here is to provide a flexible preconditioner based on a simpler quadrature rule for the integrals. More precisely, the iteration k is given by

$$U^{k+1} = U^k + \left(\mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1} \left(U_l - \left(\mathbf{I}_{MN} - \mu \mathbf{Q} \otimes \mathbf{A}\right) U^k\right), \text{ for } k = 0, ...K,$$

with $K \in \mathbb{N}$ and where the matrix $\mathbf{Q}_\Delta \in \mathbb{R}^{M \times M}$ gathers the weights of this simpler quadrature rule. Examples are the implicit right-hand side rule or the explicit left-hand side rule, both yielding lower triangular matrices, which make the inversion of the preconditioner straightforward using simple forward substitution. More recently, Weiser [25] defined $\mathbf{Q}_\Delta = \mathbf{U}^{\mathbf{T}}$ for $\mathbf{Q}^T = \mathbf{LU}$ and showed superior convergence properties of SDC for stiff problems. This approach has become the de-facto standard for SDC preconditioning and is colloquially known as St. Martin's or LU trick. Now, for each time-step, SDC can be used to generate an approximate solution of the collocation problem (3). As soon as SDC has converged (e.g. the residual of the collocation problem is smaller than a prescribed threshold), the solution at τ_M is used as initial condition for the next time-step. In order to parallelize this, the "parallel full approximation scheme in space and time" (PFASST, see [10]) makes use of a space-time hierarchy of SDC iterations (called "sweeps" in this context), using the coarsest level to propagate information quickly forward in time. This way, multiple time-steps can be integrated simultaneously, where on each local time interval SDC sweeps are used to approximate the collocation problem. We follow [16, 17] to describe the PFASST algorithm more formally.

2.2. The composite collocation problem and PFASST

The problem PFASST is trying to solve is the so called "composite collocation problem" for $L \in \mathbb{N}$ time-steps with

$$\begin{pmatrix} \mathbf{I}_{MN} - \mu \mathbf{Q} \otimes \mathbf{A} & & & \\ -\mathbf{H} & \mathbf{I}_{MN} - \mu \mathbf{Q} \otimes \mathbf{A} & & \\ & \ddots & \ddots & \\ & & -\mathbf{H} & \mathbf{I}_{MN} - \mu \mathbf{Q} \otimes \mathbf{A} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_L \end{pmatrix} = \begin{pmatrix} U_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The system matrix consists of L collocation problems on the block diagonal and a transfer matrix $\mathbf{H} = \mathbf{N} \otimes \mathbf{I}_N \in \mathbb{R}^{MN \times MN}$ on the lower diagonal, which takes the last value of each time-step and makes it available as initial condition for the next one. With the nodes we choose here, $\mathbf{N}$ is simply given by

$$\mathbf{N} = \begin{pmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{M \times M}.$$

For collocation nodes with $\tau_M < t_{l+1}$, i.e. with the last node τ_M not being identical with the subinterval boundary point t_{l+1}, the matrix $\mathbf{N}$ would contain an extrapolation rule for the solution

value at t_{l+1} in each line. Note that instead of extrapolation the collocation formulation could be used as well. More compactly and more conveniently, the composite collocation problem can be written as

$$\mathbf{C}\vec{U} = \vec{U}_0$$

with space-time-collocation vectors $\vec{U} = (U_1, ..., U_L)^T, \vec{U}_0 = (U_0, 0, ..., 0)^T \in \mathbb{R}^{LMN}$ and system matrix $\mathbf{C} = \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H} \in \mathbb{C}^{LMN \times LMN}$, where the matrix $\mathbf{E} \in \mathbb{R}^{L \times L}$ simply has ones on the first lower subdiagonal and zeros elsewhere.

The key idea of PFASST is to solve the composite collocation problem using a multigrid scheme. If the right-hand side of the ODE (1) is non-linear, a non-linear FAS multigrid is used. Although our analysis is focused on linear problems, we use FAS terminology to formulate PFASST to remain consistent with the literature. Also, we limit ourselves to a two-level scheme in order to keep the notation as simple as possible. Three components are needed to describe the multigrid scheme used to solve the composite collocation problem: (1) a smoother on the fine level, (2) a solver on the coarse level and (3) level transfer operators. In order to obtain parallelism, the smoother we choose is an approximative block Jacobi smoother, where the entries on the lower subdiagonal are omitted. The idea is to use SDC within each time-step (this is why it is an "approximative" Jacobi iteration), but omit the transfer matrices $\mathbf{H}$ on the lower diagonal. In detail, the smoother is defined by the preconditioner

$$\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A} & & & \\ & \mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A} & & \\ & & \ddots & \\ & & & \mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A} \end{pmatrix}$$

or, more compactly, $\hat{\mathbf{P}} = \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}$. Inversion of this matrix can be done on all L time-steps simultaneously, leading to L decoupled SDC sweeps. Note that typically this is done only once or twice on the fine level. In contrast, the solver on the coarse level is given by an approximative block Gauß-Seidel preconditioner. Here, SDC is used for each time-step, but the transfer matrix $\mathbf{H}$ is included. This yields for the preconditioner

$$\tilde{\mathbf{P}} = \begin{pmatrix} \mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A} & & & \\ -\mathbf{H} & \mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A} & & \\ & \ddots & \ddots & \\ & & -\mathbf{H} & \mathbf{I}_{MN} - \mu \mathbf{Q}_\Delta \otimes \mathbf{A} \end{pmatrix}$$

or $\tilde{\mathbf{P}} = \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H}$. Inversion of this preconditioner is inherently serial, but the goal is to keep this serial part as small as possible by applying it on the coarse level only, just as the Parareal method does. Thus, we need coarsening strategies in place to reduce the costs on the coarser levels [27]. To this end, we introduce block-wise restriction and interpolation $\mathbf{T}_F^C$ and $\mathbf{T}_C^F$, which coarsen the problem in space and reduce the number of quadrature nodes but do not coarsen in time, i.e., the number of time steps is not reduced. We note that the latter is also possible in this formal notation, but so far no PFASST implementation is working with this. Also, the theory presented here makes indeed use of the fact that coarsening across time-steps is not applied. A first discussion on this topic can be found in [17]. Let $\tilde{N} \in \mathbb{N}$ be the number of degrees of freedom on the coarse level and $\tilde{M} \in \mathbb{N}$ the number of collocation nodes on the coarse level. Restriction and interpolation operators are then given by

$$\mathbf{T}_F^C = \mathbf{I}_L \otimes \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{F,A}^C \in \mathbb{R}^{L\tilde{M}\tilde{N} \times LMN},$$
$$\mathbf{T}_C^F = \mathbf{I}_L \otimes \mathbf{T}_{C,Q}^F \otimes \mathbf{T}_{C,A}^F \in \mathbb{R}^{LMN \times L\tilde{M}\tilde{N}},$$

where the matrices $\mathbf{T}_{F,Q}^C$ and $\mathbf{T}_{C,Q}^F$ represent restriction and interpolation on the quadrature nodes, while $\mathbf{T}_{F,A}^C$ and $\mathbf{T}_{C,A}^F$ operate on spatial degrees-of-freedom. Within PFASST, these operators are typically standard Lagrangian-based restriction and interpolation. We use the tilde symbol to denote matrices on the coarse level, so that the approximative block Gauß-Seidel preconditioner is actually given by

$$\tilde{\mathbf{P}} = \mathbf{I}_{L\tilde{M}\tilde{N}} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}} \in \mathbb{R}^{L\tilde{M}\tilde{N} \times L\tilde{M}\tilde{N}},$$

where $\tilde{\mathbf{H}} = \tilde{\mathbf{N}} \otimes \mathbf{I}_{\tilde{N}} \in \mathbb{R}^{\tilde{M}\tilde{N} \times \tilde{M}\tilde{N}}$. Note that this preconditioner is typically applied only once or twice on the coarse level, too. In addition, the composite collocation problem has to be modified on the coarse level. This is done by the τ-correction of the FAS scheme and we refer to [16, 17] for details on this, since the actual formulation does not matter here. We now have all ingredients for one iteration of the two-level version of PFASST using post-smoothing:

1. restriction to the coarse level including the formulation of the τ-correction,
2. serial approximative block Gauß-Seidel iteration on the modified composite collocation problem on the coarse level,
3. coarse-grid correction of the fine-level values,
4. smoothing of the composite collocation problem on the fine level using parallel approximative block Jacobi iteration.

Thus, one iteration of PFASST can simply be written as

$$\vec{U}^{k+1/2} = \vec{U}^k + \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \left(\vec{U}_0 - \mathbf{C} \vec{U}^k \right),$$
$$\vec{U}^{k+1} = \vec{U}^{k+1/2} + \hat{\mathbf{P}}^{-1} \left(\vec{U}_0 - \mathbf{C} \vec{U}^{k+1/2} \right).$$

Beside this rather convenient and compact form, this formulation has the great advantage of providing the iteration matrix of PFASST, which paves the way to a comprehensive analysis.

2.3. Overview and notation

In the following, we summarize the results described above and state the iteration matrix of PFASST.

Theorem 1
Let $\mathbf{T}_F^C$ and $\mathbf{T}_C^F$ be block-wise defined transfer operators, which treat the subintervals $[t_l, t_{l+1}]$, $l = 0, ..., L - 1$ independently from each other (i.e. which do not coarsen or refine across subinterval boundaries). For a CFL number $\mu > 0$ we define the composite collocation problem as

$$\mathbf{C} = \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H},$$

with collocation matrix $\mathbf{Q}$, spatial matrix $\mathbf{A}$ (for more details see (3)) and

$$\mathbf{H} = \mathbf{N} \otimes \mathbf{I}_N \in \mathbb{R}^{NM} \quad \text{with} \quad \mathbf{N} = \begin{pmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{M \times M}$$

and $\mathbf{E} \in \mathbb{R}^{L \times L}$ being a matrix which has ones on the first subdiagonal and zeros elsewhere. We further define by $\hat{\mathbf{P}}$ the approximative block Jacobi preconditioner on the fine level and by $\tilde{\mathbf{P}}$ the approximative block Gauss-Seidel preconditioner on the coarse level (the tilde symbol always indicates the coarse level operators), i.e.

$$\tilde{\mathbf{P}} = \mathbf{I}_{L\tilde{M}\tilde{N}} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}},$$
$$\hat{\mathbf{P}} = \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}$$

where $\mathbf{Q}_\Delta$ corresponds to a simple quadrature rule. For given $\mathbf{H}$ and $\tilde{\mathbf{H}}$ we require that the restriction operator $\mathbf{T}_F^C$ satisfies $(\mathbf{E} \otimes \tilde{\mathbf{H}})\mathbf{T}_F^C = \mathbf{T}_F^C(\mathbf{E} \otimes \mathbf{H})$. Then, the PFASST iteration matrix is given by the product of the smoother's and the coarse-grid correction's iteration matrix with

$$\mathbf{T}_{\text{PFASST}} = \mathbf{T}_S \mathbf{T}_{\text{CGC}} = \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}^{-1}\mathbf{C}\right)\left(\mathbf{I}_{LMN} - \mathbf{T}_C^F\tilde{\mathbf{P}}^{-1}\mathbf{T}_F^C\mathbf{C}\right).$$

We note that we use the same CFL number μ for both coarse and fine level and absorb constant factors between the actual CFL numbers of the coarse and the fine problem into the operators $\mathbf{A}$ and $\tilde{\mathbf{A}}$.

Proof
This is taken from [16, 17] and a much more detailed derivation and discussion can be found there. □

Note that we assume here and in the following that the inverses of $\hat{\mathbf{P}}$ and $\tilde{\mathbf{P}}$ both exist. This corresponds to the fact that time-stepping via $\mathbf{Q}_\Delta$ is possible on each subinterval.

In what follows, we are interested in the behavior of PFASST's iteration matrix for the non-stiff as well as for the stiff limit. More precisely, we will look at the case where the CFL number μ either goes to zero or to infinity. While the first case represents the analysis for smaller and smaller time-steps, the second one covers scenarios where e.g. the mesh- or element-size Δx goes to zero. Alternatively, problem parameters like the diffusion coefficient or the advection speed could become very small or very large, while Δt and Δx are fixed.

3. THE NON-STIFF LIMIT

This section is split into three parts. We look at the iteration matrices of the smoother and the coarse-grid correction separately and then analyze the full iteration matrix of PFASST. While for the smoother we introduce the main idea behind the asymptotic convergence analysis, the analysis of the coarse-grid correction is dominated by the restriction and interpolation operators. For PFASST's full iteration matrix we then combine both results in a straightforward manner.

3.1. The smoother

We first consider the iteration matrix $\mathbf{T}_S$ of the smoother with

$$\begin{aligned}
\mathbf{T}_S &= \mathbf{I}_{LMN} - \hat{\mathbf{P}}^{-1}\mathbf{C} \\
&= \mathbf{I}_{LMN} - \left(\mathbf{I}_{LMN} - \mu\mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}\left(\mathbf{I}_{LMN} - \mu\mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H}\right).
\end{aligned}$$

We write $\mathbf{T}_S = \mathbf{T}_S(\mu)$, so that

$$\mathbf{T}_S(0) = \mathbf{E} \otimes \mathbf{H}.$$

Therefore, we have

$$\begin{aligned}
\mathbf{T}_S(\mu) - \mathbf{T}_S(0) &= \hat{\mathbf{P}}^{-1}\left(\hat{\mathbf{P}} - \mathbf{C} - \hat{\mathbf{P}}\left(\mathbf{E} \otimes \mathbf{H}\right)\right) \\
&= \mu\hat{\mathbf{P}}^{-1}\left(\mathbf{I}_L \otimes \left(\mathbf{Q} - \mathbf{Q}_\Delta\right) \otimes \mathbf{A} + \mathbf{E} \otimes \mathbf{Q}_\Delta\mathbf{N} \otimes \mathbf{A}\right).
\end{aligned} \tag{4}$$

Moreover, if μ is smaller than a given value $\mu_S^* > 0$, the norm of $\hat{\mathbf{P}}^{-1}$ can be bounded by

$$\left\|\hat{\mathbf{P}}^{-1}\right\| = \left\|\hat{\mathbf{P}}^{-1}(\mu)\right\| = \left\|\left(\mathbf{I}_{LMN} - \mu\mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}\right\| \le c_1(\mu_{S,0}^*) = c_1, \tag{5}$$

for a constant $c_1(\mu_{S,0}^*)$ independent of μ, since the function $\mu \mapsto \left\|\hat{\mathbf{P}}^{-1}(\mu)\right\|$ is continuous on the closed interval $[0, \mu_{S,0}^*]$. The norm $\|.\|$ can be any induced matrix norm unless stated otherwise.

Together with (4) we obtain

$$\big\| \mathbf{T}_\mathrm{S}(\mu) - \mathbf{T}_\mathrm{S}(0) \big\| \leq c_2 \mu$$

since the last factor of (4) does not depend on μ.

Therefore, the matrix $\mathbf{T}_\mathrm{S}(\mu)$ converges to $\mathbf{T}_\mathrm{S}(0) = \mathbf{E} \otimes \mathbf{H}$ linearly as $\mu \to 0$, so that we can write

$$\mathbf{T}_\mathrm{S}(\mu) = \mathbf{T}_\mathrm{S}(0) + \mathcal{O}(\mu) = \mathbf{E} \otimes \mathbf{H} + \mathcal{O}(\mu). \tag{6}$$

This leads us to the following lemma:

Lemma 1
The smoother of PFASST converges for linear problems and Gauß-Radau nodes, if the CFL number μ is small enough. More precisely, for L time-steps the spectral radius of the smoother is bounded by

$$\rho \left(\mathbf{T}_\mathrm{S}(\mu) \right) \leq c \mu^{\frac{1}{L}} \tag{7}$$

for a constant $c > 0$ independent of μ, if $\mu < \mu_{\mathrm{S},0}^*$ for a fixed value $\mu_{\mathrm{S},0}^* > 0$.

Proof
The matrix $\mathbf{T}_\mathrm{S}(\mu)$ can be seen as a perturbation of $\mathbf{T}_\mathrm{S}(0)$, where the perturbation matrix $\mathbf{D}(\mu)$ is of the order of $\mathcal{O}(\mu)$. The eigenvalues of the unperturbed matrix $\mathbf{T}_\mathrm{S}(0)$ are all zero (since it only has entries strictly below the diagonal). With

$$\mathbf{T}_\mathrm{S}(0) = \mathbf{E} \otimes \mathbf{H} = \mathbf{E} \otimes \mathbf{N} \otimes \mathbf{I},$$

its Jordan canonical form also consists for three parts. Obviously, the canonical form of $\mathbf{E}$ consists of a single Jordan block of size L for the eigenvalue 0, while $\mathbf{I}$ has N Jordan blocks of size 1, with N being the number of degrees-of-freedom in space. For M quadrature nodes, the canonical form of $\mathbf{N}$ consists of $M - 1$ blocks of size 1 for the eigenvalue 0 and one block of size 1 for the eigenvalue 1. Therefore, the canonical form of $\mathbf{T}_\mathrm{S}(0)$ consists of MN blocks of size L for the eigenvalue 0. Since the perturbation matrix $\mathbf{D}(\mu)$ does not have a particular structure other than its linear dependence on μ, the difference between the eigenvalues of the perturbed matrix $\mathbf{T}_\mathrm{S}(\mu)$ and the unperturbed matrix $\mathbf{T}_\mathrm{S}(0)$ is of the order of $\mathcal{O}(\mu^{\frac{1}{\alpha}})$, where $\alpha = L$ is the size of the largest Jordan block, see [28], p. 77ff. Thus,

$$\rho \left(\mathbf{T}_\mathrm{S}(\mu) \right) \leq c \mu^{\frac{1}{L}} \tag{8}$$

and especially $\rho \left(\mathbf{T}_\mathrm{S}(\mu) \right) < 1$ for μ small enough. $\qquad\square$

Figure 1 shows for Dahlquist's test problem $u_t = \lambda u$, $u(0) = 1$, that this estimate is severely over-pessimistic, if L is small, but becomes rather accurate, if L becomes larger. This does not change significantly when choosing an imaginary value for λ of the test equation, as Figure 1b shows. Note, however, that for these large numbers of time-steps L the numerical computation of the spectral radius may be prone to rounding errors. We refer to the discussion in [17] for more details.

Yet, while this lemma gives a proof of the convergence of the smoother, it cannot be used to estimate the speed of convergence. The standard way of providing such an estimate would be to bound the norm of the iteration matrix $\mathbf{T}_\mathrm{S}(\mu)$ by something smaller than 1. However, even in the limit $\mu \to 0$ the norm of $\mathbf{T}_\mathrm{S}(\mu) = \mathbf{T}_\mathrm{S}(0)$ is still larger than or equal to 1 in all feasible matrix norms. Alternatively, we can look at the kth power of the iteration matrix, which corresponds to k applications of the smoother.

Remark 1
For all $k \in \mathbb{N}$ we have

$$\mathbf{T}_\mathrm{S}(\mu)^k = \left(\mathbf{T}_\mathrm{S}(0) + \mathcal{O}(\mu) \right)^k = \mathbf{T}_\mathrm{S}(0)^k + \mathcal{O}(\mu)$$

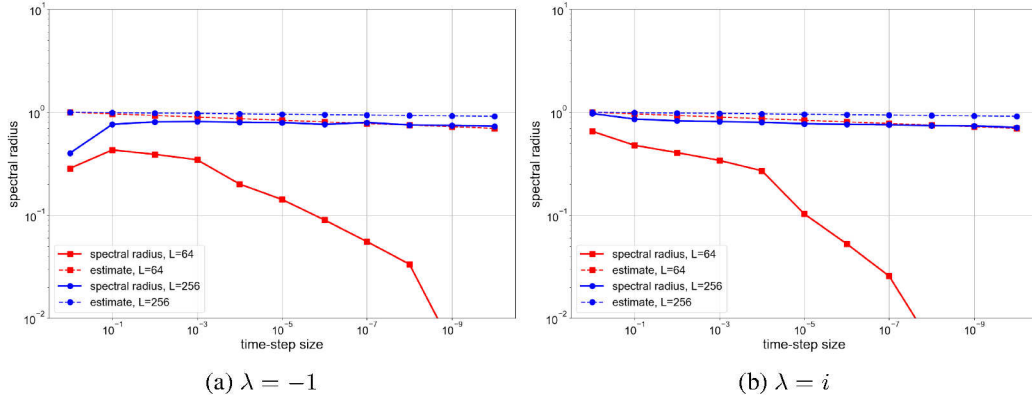(a) $\lambda = -1$

(b) $\lambda = i$

Figure 1. Convergence of the spectral radius of the smoother vs. the estimation (7) for Dahlquist's test problem $u_t = \lambda u$, $u(0) = 1$. Left: negative real value $\lambda = -1$, right: purely imaginary value $\lambda = i$. 3 Gauss-Radau nodes are taken and the LU trick is used for $\mathbf{Q}_\Delta$, see [25].

for some perturbations of order $\mathcal{O}(\mu)$. The matrix $\mathbf{T}_S(0) = \mathbf{E} \otimes \mathbf{H}$ is nilpotent with $\mathbf{T}_S(0)^L = \mathbf{0}$, because $\mathbf{E}^L = 0$. Thus, for $k \geq L$,

$$\left\| \mathbf{T}_S(\mu)^k \right\| \leq \left\| \mathbf{T}_S(0)^k \right\| + \mathcal{O}(\mu) = \mathcal{O}(\mu).$$

This shows that for enough iterations of the smoother, the error is reduced in the order of μ. However, for $k < L$, this is not true, since there is still a term larger than or equal to 1 coming from $\mathbf{E} \otimes \mathbf{H}$. Of course, doing as many smoothing steps as we have time-steps is not feasible, so that this result is rather pathological.

3.2. The coarse-grid correction

We now consider the iteration matrix $\mathbf{T}_{\mathrm{CGC}}$ of the coarse-grid correction with

$$\mathbf{T}_{\mathrm{CGC}} = \mathbf{I}_{LMN} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \mathbf{C} = \mathbf{I}_{LMN} - \mathbf{T}_C^F \left(\mathbf{I}_{L\tilde{M}\tilde{N}} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}} \right)^{-1} \cdot$$
$$\mathbf{T}_F^C \left(\mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H} \right)$$

and write $\mathbf{T}_{\mathrm{CGC}} = \mathbf{T}_{\mathrm{CGC}}(\mu)$. Then,

$$\begin{aligned} \mathbf{T}_{\mathrm{CGC}}(0) &= \mathbf{I}_{LMN} - \mathbf{T}_C^F \left(\mathbf{I}_{L\tilde{M}\tilde{N}} - \mathbf{E} \otimes \tilde{\mathbf{H}} \right)^{-1} \mathbf{T}_F^C \left(\mathbf{I}_{LMN} - \mathbf{E} \otimes \mathbf{H} \right) \\ &= \mathbf{I}_{LMN} - \mathbf{T}_C^F \mathbf{T}_F^C \left(\mathbf{I}_{LMN} - \mathbf{E} \otimes \mathbf{H} \right)^{-1} \left(\mathbf{I}_{LMN} - \mathbf{E} \otimes \mathbf{H} \right) \\ &= \mathbf{I}_{LMN} - \mathbf{T}_C^F \mathbf{T}_F^C \end{aligned}$$

according to Theorem 1. Therefore, we get

$$\mathbf{T}_{\mathrm{CGC}}(\mu) - \mathbf{T}_{\mathrm{CGC}}(0) = \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \left(\tilde{\mathbf{P}} \mathbf{T}_F^C - \mathbf{T}_F^C \mathbf{C} \right).$$

Then, with $\mathbf{T}_F^C = \mathbf{I}_L \otimes \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{F,A}^C$ we have

$$\begin{aligned} \tilde{\mathbf{P}} \mathbf{T}_F^C &= \left(\mathbf{I}_{L\tilde{M}\tilde{N}} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}} \right) \mathbf{T}_F^C \\ &= \mathbf{T}_F^C \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \mathbf{T}_{F,Q}^C \otimes \tilde{\mathbf{A}} \mathbf{T}_{F,A}^C - \mathbf{T}_F^C \mathbf{E} \otimes \mathbf{H} \end{aligned}$$

and therefore

$$\mathbf{T}_{\mathrm{CGC}}(\mu) - \mathbf{T}_{\mathrm{CGC}}(0) = \mu \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \left(\mathbf{I}_L \otimes \left(\mathbf{T}_{F,Q}^C \mathbf{Q} \otimes \mathbf{T}_{F,A}^C \mathbf{A} - \tilde{\mathbf{Q}}_\Delta \mathbf{T}_{F,Q}^C \otimes \tilde{\mathbf{A}} \mathbf{T}_{F,A}^C \right) \right). \quad (9)$$

As before, the norm of the inverse of the coarse-level preconditioner can be bounded by

$$\left\|\tilde{\mathbf{P}}^{-1}\right\| = \left\|\tilde{\mathbf{P}}^{-1}(\mu)\right\| = \left\|\left(\mathbf{I}_{L\tilde{M}\tilde{N}} - \mu\mathbf{I}_L \otimes \tilde{\mathbf{Q}}_{\boldsymbol{\Delta}} \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}}\right)^{-1}\right\| \leq c_3(\mu^*_{\text{CGC},0}) = c_3 \quad (10)$$

if μ is smaller than a given value $\mu^*_{\text{CGC},0} > 0$, where the constant $c_3(\mu^*_{\text{CGC},0})$ is again independent of μ. Together with (9) this leads to

$$\left\|\mathbf{T}_{\text{CGC}}(\mu) - \mathbf{T}_{\text{CGC}}(0)\right\| \leq c_4\mu$$

as for the smoother. This allows us to write the iteration matrix of the coarse-grid correction as

$$\mathbf{T}_{\text{CGC}}(\mu) = \mathbf{T}_{\text{CGC}}(0) + \mathcal{O}(\mu) = \mathbf{I}_{LMN} - \mathbf{T}_C^F\mathbf{T}_F^C + \mathcal{O}(\mu). \quad (11)$$

While the eigenvalues of $\mathbf{T}_{\text{CGC}}(\mu)$ again converge to the eigenvalues of $\mathbf{T}_{\text{CGC}}(0)$, the eigenvalues of the latter are not zero anymore. For a partial differential equation in one dimension half of the eigenvalues of $\mathbf{T}_C^F\mathbf{T}_F^C$ are zero for standard Lagrangian interpolation and restriction, because the dimension of the coarse space is only of half size. Therefore, the limit matrix $\mathbf{T}_{\text{CGC}}(0)$ has a spectral radius of at least 1.

3.3. PFASST

We now couple both results to analyze the full iteration matrix $\mathbf{T}_{\text{PFASST}} = \mathbf{T}(\mu)$ of PFASST. Using (6) and (11), we obtain

$$\begin{aligned}
\mathbf{T}(\mu) = \mathbf{T}_{\text{S}}(\mu)\mathbf{T}_{\text{CGC}}(\mu) &= \left(\mathbf{E} \otimes \mathbf{H} + \mathcal{O}(\mu)\right)\left(\mathbf{I}_{LMN} - \mathbf{T}_C^F\mathbf{T}_F^C + \mathcal{O}(\mu)\right) \\
&= \left(\mathbf{E} \otimes \mathbf{H}\right)\left(\mathbf{I}_{LMN} - \mathbf{T}_C^F\mathbf{T}_F^C\right) + \mathcal{O}(\mu) \\
&= \mathbf{E} \otimes \left(\mathbf{H}\left(\mathbf{I}_{MN} - \mathbf{T}_{C,Q}^F\mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{C,A}^F\mathbf{T}_{F,A}^C\right)\right) + \mathcal{O}(\mu) = \mathbf{T}(0) + \mathcal{O}(\mu)
\end{aligned}$$

Again, the eigenvalues of $\mathbf{T}(0)$ are all zero, because the eigenvalues of $\mathbf{E}$ are all zero. We can therefore extend Lemma 1 to cover the full iteration matrix of PFASST.

Theorem 2
PFASST converges for linear problems and Gauß-Radau nodes, if the CFL number μ is small enough. More precisely, for L time-steps the spectral radius of the iteration matrix is bounded by

$$\rho\left(\mathbf{T}_{\text{PFASST}}(\mu)\right) \leq c\mu^{\frac{1}{L}}$$

for a constant $c > 0$ independent of μ, if $\mu < \mu^*_0$ for a fixed value $\mu^*_0 > 0$.

Proof
The estimate for the spectral radius can be shown in an analogous way as the one for the smoother in Lemma 1. μ^* is given by $\mu^*_0 = \min\{\mu^*_{\text{S},0}, \mu^*_{\text{CGC},0}\}$, where $\mu^*_{\text{S},0}$ comes from (5) and $\mu^*_{\text{CGC},0}$ from (10). $\qquad\square$

We note that this result indeed proves convergence for PFASST in the non-stiff limit, but it does not provide an indication of the speed of convergence. It rather ensures convergence for the asymptotic case of $\mu \to 0$. To get an estimate of the speed of convergence of PFASST, a bound for the norm of the iteration matrix would be necessary. Yet, we could make the same observation as in Remark 1, but this is again not a particularly useful result.

4. THE STIFF LIMIT

Working with larger and larger CFL numbers μ requires an additional trick and poses limitations on the spatial problem. Again, we will analyze the iteration matrices of the smoother, the coarse-grid correction and the full PFASST iteration separately. In the last section we will then discuss the norms of the iteration matrices, which in contrast to the non-stiff limit case leads to interesting new insights.

4.1. The smoother

In order to calculate the stiff limit $\mathbf{T}_S(\infty)$ of the iteration matrix of the smoother $\mathbf{T}_S = \mathbf{T}_S(\mu)$ we write

$$\begin{aligned}
\mathbf{T}_S &= \mathbf{I}_{LMN} - \hat{\mathbf{P}}^{-1}\mathbf{C} \\
&= \mathbf{I}_{LMN} - \left(\mathbf{I}_{LMN} - \mu\mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}\left(\mathbf{I}_{LMN} - \mu\mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H}\right)
\end{aligned}$$

This can be written as

$$\mathbf{T}_S = \mathbf{I}_{LMN} - \left(\frac{1}{\mu}\mathbf{I}_{LMN} - \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}\left(\frac{1}{\mu}\mathbf{I}_{LMN} - \frac{1}{\mu}\mathbf{E} \otimes \mathbf{H} - \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A}\right)$$

and for $\mu \to \infty$ we would expect the limit matrix $\mathbf{T}_S(\infty)$ to be defined as

$$\mathbf{T}_S(\infty) = \mathbf{I}_L \otimes \left(\mathbf{I}_M - \mathbf{Q}_\Delta^{-1}\mathbf{Q}\right) \otimes \mathbf{I}_N.$$

We need to abbreviate this a bit to see the essential details in the following calculation. We define:

$$\begin{aligned}
\mathbf{M}_1 &= \mu\mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}, \\
\mathbf{M}_2 &= \mu\mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A}, \\
\mathbf{M}_3 &= \mathbf{I}_{LMN} - \mathbf{E} \otimes \mathbf{H},
\end{aligned}$$

so that

$$\begin{aligned}
\mathbf{T}_S(\mu) &= \mathbf{I} - \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\left(\mathbf{M}_3 - \mathbf{M}_2\right), \\
\mathbf{T}_S(\infty) &= \mathbf{I} - \left(\mathbf{M}_1\right)^{-1}\mathbf{M}_2,
\end{aligned}$$

omitting the dimension index at the identity matrix. Then we have

$$\begin{aligned}
\mathbf{T}_S(\mu) - \mathbf{T}_S(\infty) &= \mathbf{I} - \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\left(\mathbf{M}_3 - \mathbf{M}_2\right) - \mathbf{I} + \mathbf{M}_1^{-1}\mathbf{M}_2 \\
&= \mathbf{M}_1^{-1}\mathbf{M}_2 + \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_2 - \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_3 \\
&= \left(\mathbf{M}_1^{-1} + \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\right)\mathbf{M}_2 - \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_3 \\
&= \left(\mathbf{M}_1^{-1}\left(\mathbf{I} - \mathbf{M}_1\right) + \mathbf{I}\right)\left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_2 - \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_3 \\
&= \mathbf{M}_1^{-1}\left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_2 - \left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{M}_3.
\end{aligned}$$

In the first term, the matrix $\mathbf{M}_1^{-1}$ has a factor $\frac{1}{\mu}$ which is removed by the factor μ of the matrix $\mathbf{M}_2$. In the second term, the matrix $\mathbf{M}_3$ does not depend on μ, so that in both cases only $\left(\mathbf{I} - \mathbf{M}_1\right)^{-1}$ adds a dependence on μ. More precisely, following the same argument as for non-stiff part,

$$\begin{aligned}
\left\|\left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\right\| &= \left\|\left(\mathbf{I}_{LMN} - \mu\mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}\right\| \\
&= \frac{1}{\mu}\left\|\left(\frac{1}{\mu}\mathbf{I}_{LMN} - \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}\right\| \leq \frac{1}{\mu}c_4(\mu_{S,\infty}^*) = \frac{1}{\mu}c_4
\end{aligned}$$

for all $\mu \geq \mu_{S,\infty}^* > 0$, as long as $\left(\mathbf{Q}_\Delta \otimes \mathbf{A}\right)^{-1}$ exists. Then we obtain

$$\left\|\mathbf{T}_S(\mu) - \mathbf{T}_S(\infty)\right\| \leq c_5\frac{1}{\mu}, \tag{12}$$

so that the iteration matrix $\mathbf{T}_S(\mu)$ converges linearly to $\mathbf{T}_S(\infty)$ as $\mu \to \infty$ and we can write

$$\mathbf{T}_S(\mu) = \mathbf{T}_S(\infty) + \mathcal{O}\left(\frac{1}{\mu}\right) = \mathbf{I}_L \otimes \left(\mathbf{I}_M - \mathbf{Q}_\Delta^{-1}\mathbf{Q}\right) \otimes \mathbf{I}_N + \mathcal{O}\left(\frac{1}{\mu}\right)$$

for μ large enough. This result looks indeed very similar to the one obtained for the non-stiff limit, but now the eigenvalues of the limit matrix $\mathbf{T}_S(\infty)$ are not zero anymore, at least not for arbitrary choices of the preconditioner $\mathbf{Q}_\Delta$. Yet, if we choose to define this preconditioner using the LU trick [25], i.e. $\mathbf{Q}_\Delta = \mathbf{U}^T$ for $\mathbf{Q}^T = \mathbf{LU}$, we can prove the following analog of Lemma 1, provided that we couple the number of smoothing steps to the number of collocation nodes.

Lemma 2
The smoother of PFASST converges for linear problems and Gauß-Radau nodes with preconditioning using LU, if the CFL number μ is large enough and at least M iterations are performed. We further require that $\mathbf{A}$ is invertible. More precisely, the spectral radius of the iteration matrix $\mathbf{T}_S(\mu, k)$ is then bounded by

$$\rho\left(\mathbf{T}_S(\mu, k)\right) \leq c\frac{1}{\mu}$$

for a constant $c > 0$ independent of μ but depending on k, if $k \geq M$ and $\mu > \mu_{S,\infty}^*$ for a fixed value $\mu_{S,\infty}^* > 0$.

Proof
For $k \in \mathbb{N}$ iterations of the smoother we have

$$\mathbf{T}_S(\mu, k) = \left(\mathbf{T}_S(\infty) + \mathcal{O}\left(\frac{1}{\mu}\right)\right)^k = \left(\mathbf{T}_S(\infty)\right)^k + \mathcal{O}\left(\frac{1}{\mu}\right)$$

under the conditions of the lemma. With $\mathbf{Q}_\Delta = \mathbf{U}^T$ and $\mathbf{Q}^T = \mathbf{LU}$ it is $\mathbf{I}_M - \mathbf{Q}_\Delta^{-1}\mathbf{Q} = \mathbf{I}_M - \mathbf{L}^T$, so that $\mathbf{T}_S(\infty)$ is strictly upper diagonal and therefore nilpotent with $\left(\mathbf{T}_S(\infty)\right)^M = \mathbf{0}$. Therefore, after at least M iterations the stiff limit matrix of the smoother is actually $\mathbf{0}$ and so are its eigenvalues. Instead of using the general perturbation result as in the non-stiff limit, we now can apply Elsner's theorem [29], stating that for a perturbation $\mathbf{D} \in \mathbb{C}^{N \times N}$ of a matrix $\mathbf{T} \in \mathbb{C}^{N \times N}$, i.e. for $\hat{\mathbf{T}} = \mathbf{T} + \mathbf{D}$ it is

$$\max_{i=1,\ldots,N} \min_{j=1,\ldots,N} |\lambda_i(\hat{\mathbf{T}}) - \lambda_j(\mathbf{T})| \leq \left(\|\hat{\mathbf{T}}\|_2 + \|\mathbf{T}\|_2\right)^{1-1/N} \|\mathbf{D}\|_2^{1/N}, \tag{13}$$

where $\lambda_i(\hat{\mathbf{T}})$ corresponds to the ith eigenvalue of $\hat{\mathbf{T}}$ and $\lambda_j(\mathbf{T})$ to the jth eigenvalue of $\mathbf{T}$. In our case, $\mathbf{T} = \left(\mathbf{T}_S(\infty)\right)^k = \mathbf{0}$ for $k \geq M$, so that the left-hand side of this inequality is just the spectral radius of $\hat{\mathbf{T}} = \mathbf{T}_S(\mu, k)$. The norm of the perturbation matrix $\mathbf{D}$ is bounded via (12), so that for L steps, M nodes, N degrees-of-freedom, $k \geq M$ iterations we have

$$\rho\left(\mathbf{T}_S(\mu, k)\right) \leq \left(\left\|\mathbf{T}_S(\mu, k)\right\|_2 + \left\|\mathbf{T}_S(\infty, k)\right\|_2\right)^{1-\frac{1}{LMN}} \|\mathbf{D}\|_2^{\frac{1}{LMN}}$$

$$\leq \left(2\|\mathbf{0}\|_2 + \|\mathbf{D}\|_2\right)^{1-\frac{1}{LMN}} \|\mathbf{D}\|_2^{\frac{1}{LMN}} = \|\mathbf{D}\|_2$$

$$= \left\|\mathbf{T}_S(\mu, k) - \mathbf{T}_S(\infty, k)\right\| \leq c\frac{1}{\mu}.$$

by using $\|\mathbf{T}_S(\mu, k)\| \leq \|\mathbf{T}_S(\infty, k)\| + \|\mathbf{D}\|$ for the second inequality. $\square$

This result shows that the spectral radius goes to zero with the same speed as μ goes to infinity, independently of the number of time-step, the number of collocation nodes or the spatial resolution. Yet, the condition of requiring at least M smoothing steps is rather severe since standard simulations with PFASST typically perform only one or two steps here. We show in Figure 2 the spectral radii of the iteration matrix of the smoother for Dahlquist's test problem $u_t = \lambda u$, $u(0) = 1$ with $\lambda = -1$,

(a) $M = 3$ and $\lambda = -1$

(b) $M = 7$ and $\lambda = -1$

(c) $M = 3$ and $\lambda = i$
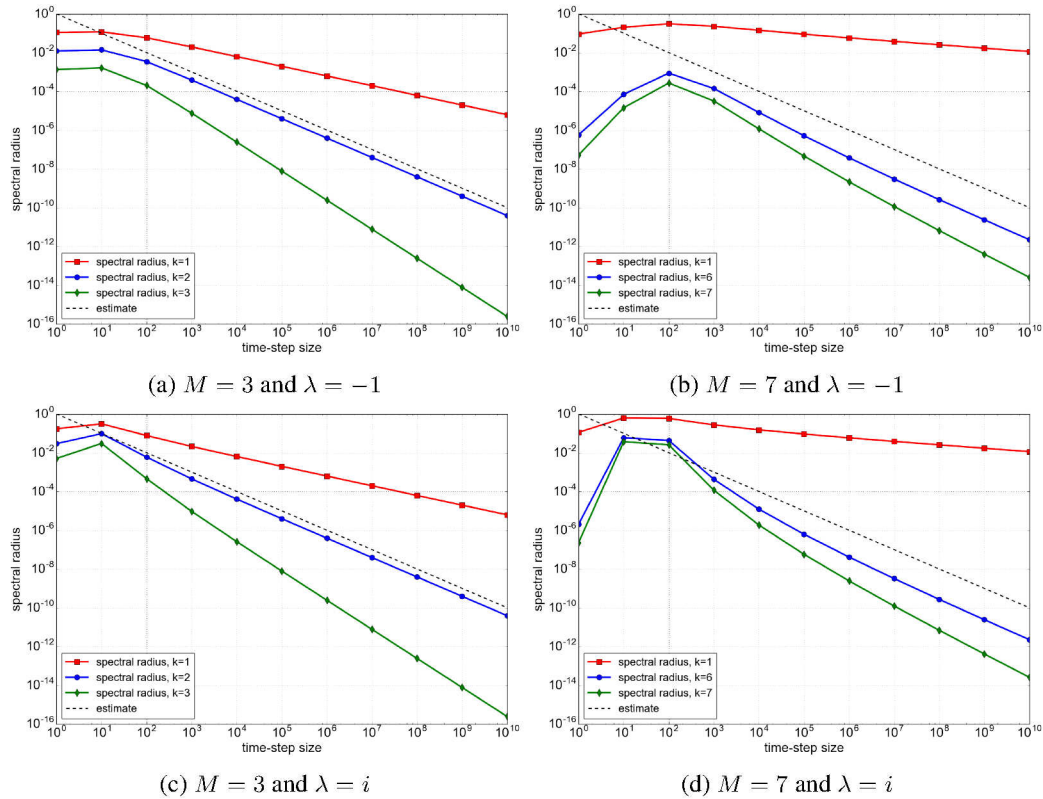
(d) $M = 7$ and $\lambda = i$

Figure 2. Spectral radius of the iteration matrix of the smoother with LU decomposition for $\lambda = -1$ (upper) and $\lambda = i$ (lower), $L = 1$. Left: $M = 3$, right: $M = 7$.

$\lambda = i$ (upper and lower figures) and larger and larger time-step sizes Δt (with $\mu = \lambda \Delta t$), using $M = 3$ and $M = 7$ Gauß-Radau nodes (left and right figures). There are a few interesting things to note here: First, for the asymptotics it does not matter which $\lambda \in \mathbb{C}$ of the two we choose, the plots look the same. Only for small values of Δt the spectral radii differ significantly, leading to a severely worse convergence behavior for complex eigenvalues. Second, the result does not depend on the number of time-steps L, so choosing $L = 1$ is reasonable here (and the plots do not change for $L > 1$, which is a key difference to the results for $\mu \to 0$). Third, the estimate of the spectral radius is rather pessimistic, again. More precisely, already for $k = M - 1$ the linear slope of the estimation is met (and, actually, surpassed), while for $k > M - 1$ the convergence is actually faster. This is not reflected in the result above, but it shows that this estimate is only a rather loose bound. However, we note that in Section 6 we see a different outcome, more in line with the theoretical estimate. Note that in Figure 2, the constant c is not included so that the estimate depicted there is not necessarily an upper bound, but reflects the asymptotics.

For a more complete overview of the smoother, Figure 3 shows the spectral radius of its iteration matrix for fixed L, M and Dahlquist's test for different values $\mu = \Delta t \lambda$, choosing the LU trick (Fig. 3a) and the classical implicit Euler method or right-hand side rule (Fig. 3b) for $\mathbf{Q}_\triangle$. Following Theorem 2, we choose $k = M = 3$ smoothing steps. Note that the scale of the colors is logarithmic, too, and all values for the spectral radius are considerably below 1. The largest for the LU trick is at about 0.05 and for the implicit Euler at about 0.17. We can nicely see how the smoother converges for $\mu \to 0$ as well as for $\mu \to \infty$ in both cases, although the LU case is much more pronounced. However, we also see that there is a banded area, where the spectral radius is significantly higher than in the regions of large or small μ. This is most likely due to the properties of the underlying preconditioner and can be observed on the real axis for standard SDC as well, see [25]. Oddly, this band does not describe a circle but rather a box, hitting the axes with its maximum at around -10
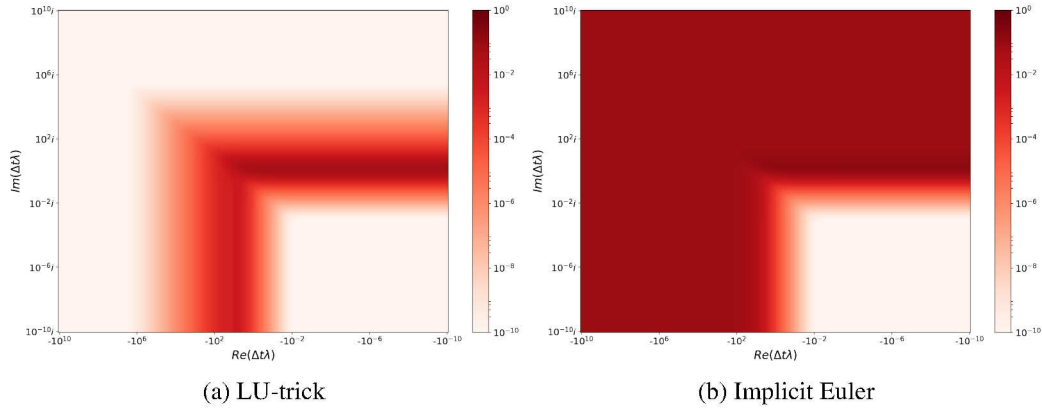
(a) LU-trick (b) Implicit Euler

Figure 3. Spectral radius of the iteration matrix of the smoother for different complex values of $\Delta t \lambda$. Fixed values of $M = 3$ and $L = 4$, $k = M = 3$ smoothing steps. Left: LU-trick for $\mathbf{Q}_\Delta$, right: implicit Euler for $\mathbf{Q}_\Delta$.

and $10i$. It seems unlikely that a closed description of this area, i.e. filling the gap between $\mu \to 0$ and $\mu \to \infty$ is easily possible.

4.2. The coarse-grid correction

We now focus on the iteration matrix $\mathbf{T}_{\mathrm{CGC}} = \mathbf{T}_{\mathrm{CGC}}(\mu)$ of the coarse-grid correction with

$$\mathbf{T}_{\mathrm{CGC}} = \mathbf{I}_{LMN} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \mathbf{C} = \mathbf{I}_{LMN} - \mathbf{T}_C^F \left(\mathbf{I}_{L\tilde{M}\tilde{N}} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}} \right)^{-1} \cdot$$
$$\mathbf{T}_F^C \left(\mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \mathbf{E} \otimes \mathbf{H} \right).$$

Using abbreviations

$$\tilde{\mathbf{M}}_1 = \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}},$$
$$\mathbf{M}_2 = \mu \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A},$$
$$\mathbf{M}_3 = \mathbf{I}_{LMN} - \mathbf{E} \otimes \mathbf{H},$$
$$\tilde{\mathbf{M}}_3 = \mathbf{I}_{L\tilde{M}\tilde{N}} - \mathbf{E} \otimes \tilde{\mathbf{H}}$$

we have

$$\mathbf{T}_{\mathrm{CGC}}(\mu) = \mathbf{I} - \mathbf{T}_C^F \left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1 \right)^{-1} \mathbf{T}_F^C \left(\mathbf{M}_3 - \mathbf{M}_2 \right).$$

If $\left(\tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} \right)^{-1}$ exists, we define

$$\mathbf{T}_{\mathrm{CGC}}(\infty) = \mathbf{I}_{LMN} - \mathbf{T}_C^F \left(\mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} \right)^{-1} \mathbf{T}_F^C \left(\mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} \right)$$

and write it more briefly as

$$\mathbf{T}_{\mathrm{CGC}}(\infty) = \mathbf{I} - \mathbf{T}_C^F \tilde{\mathbf{M}}_1^{-1} \mathbf{T}_F^C \mathbf{M}_2.$$

With the same ideas and rearrangements we have used for the smoother in the previous section, we obtain after somewhat lengthy algebra

$$\mathbf{T}_{\mathrm{CGC}}(\mu) - \mathbf{T}_{\mathrm{CGC}}(\infty) = \mathbf{T}_C^F \tilde{\mathbf{M}}_1^{-1} \tilde{\mathbf{M}}_3 \left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1 \right)^{-1} \mathbf{T}_F^C \mathbf{M}_2 - \mathbf{T}_C^F \left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1 \right)^{-1} \mathbf{T}_F^C \mathbf{M}_3.$$

In the first term, $\tilde{\mathbf{M}}_1^{-1}$ adds a factor $\frac{1}{\mu}$, which is removed by μ from $\mathbf{M}_2$. $\mathbf{M}_3$ does not have a dependence on μ, so for both terms only $\left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1\right)^{-1}$ is relevant. We have

$$
\begin{aligned}
\left\| \left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1\right)^{-1} \right\| &= \left\| \left(\mathbf{I}_{L\tilde{M}\tilde{N}} - \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \mathbf{E} \otimes \tilde{\mathbf{H}}\right)^{-1} \right\| \\
&= \frac{1}{\mu} \left\| \left(\frac{1}{\mu}\mathbf{I}_{L\tilde{M}\tilde{N}} - \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}} - \frac{1}{\mu}\mathbf{E} \otimes \tilde{\mathbf{H}}\right)^{-1} \right\| \quad (14) \\
&\leq \frac{1}{\mu} c_6(\mu^*_{\mathrm{CGC},\infty}) = \frac{1}{\mu} c_6
\end{aligned}
$$

for all $\mu \geq \mu^*_{\mathrm{CGC},\infty} > 0$ with the same argument as before. Therefore, $\mathbf{T}_{\mathrm{CGC}}(\mu)$ converges to $\mathbf{T}_{\mathrm{CGC}}(\infty)$ linearly as $\mu \to \infty$ and we can write

$$
\begin{aligned}
\mathbf{T}_{\mathrm{CGC}}(\mu) &= \mathbf{T}_{\mathrm{CGC}}(\infty) + \mathcal{O}\left(\frac{1}{\mu}\right) \\
&= \mathbf{I}_{LMN} - \mathbf{T}_C^F \left(\mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}}\right)^{-1} \mathbf{T}_F^C \left(\mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A}\right) + \mathcal{O}\left(\frac{1}{\mu}\right),
\end{aligned}
$$

for μ large enough.

4.3. PFASST

We can now combine both parts to obtain an analog of Theorem 2.

Theorem 3
PFASST converges for linear problems and Gauß-Radau nodes with preconditioning using LU, if the CFL number μ is large enough and at least M smoothing steps are performed. We further require that $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are invertible. More precisely, the spectral radius of the iteration matrix $\mathbf{T}_{\mathrm{PFASST}}(\mu, k)$ is then bounded by

$$
\rho\left(\mathbf{T}_{\mathrm{PFASST}}(\mu, k)\right) \leq c\frac{1}{\mu}
$$

for a constant $c > 0$ independent of μ but depending on k, if $k \geq M$ and $\mu > \mu^*_\infty$ for a fixed value $\mu^*_\infty > 0$.

Proof
For $k \in \mathbb{N}$ iterations of the smoother we have with Lemma 2 and the derivations from the previous section

$$
\mathbf{T}_{\mathrm{PFASST}}(\mu, k) = \left(\mathbf{T}_S(\mu)\right)^k \mathbf{T}_{\mathrm{CGC}}(\mu) = \left(\mathbf{T}_S(\infty)\right)^k \mathbf{T}_{\mathrm{CGC}}(\infty) + \mathcal{O}\left(\frac{1}{\mu}\right) = \mathcal{O}\left(\frac{1}{\mu}\right).
$$

The estimate of the spectral radius uses again Elsner's theorem from [29], see the proof of Lemma 2. □

Remark 2
Since the inverses of $\mathbf{Q}_\Delta$, $\tilde{\mathbf{Q}}_\Delta$, $\mathbf{A}$ and $\tilde{\mathbf{A}}$ appear in the limit matrices, their existence is essential and the assumptions of this lemma are a common theme in this section. We need to point out, however, that assuming the existence of these inverses is rather restrictive, since problems like the 1D heat equation on periodic boundaries are already excluded. However, this assumption can be weakened by considering a pseudoinverse that acts on the orthogonal complement of the null space of $\mathbf{A}$ or $\tilde{\mathbf{A}}$, only. In the case of e.g. periodic boundary conditions this results in considering the whole space except for vectors representing a constant function, see also the discussion in Section 6.

4.4. Smoothing and approximation property

For the stiff limit case, bounding the norms of the iteration matrices actually provides insight into the relationship between PFASST and standard multigrid theory. Following [30] we now analyze the smoothing and the approximation property of PFASST. Note that we now interpret $\mu \to \infty$ as $\Delta x \to 0$ so that in the following results the appearance of μ in the denominator and numerator is counterintuitive. We start with the approximation property, which is straightforward to show.

Lemma 3

The coarse-grid correction of PFASST for linear problems satisfies the approximation property if the CFL number μ is large enough and if $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are invertible, i.e. it holds

$$\left\| \mathbf{C}^{-1} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \right\| \leq c \frac{1}{\mu}$$

for μ large enough, with a constant $c > 0$ independent of μ.

Proof

As in Section 4.2 we make use of the abbreviations

$$\tilde{\mathbf{M}}_1 = \mu \mathbf{I}_L \otimes \tilde{\mathbf{Q}}_\Delta \otimes \tilde{\mathbf{A}},$$
$$\mathbf{M}_2 = \mu \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A},$$
$$\mathbf{M}_3 = \mathbf{I}_{LMN} - \mathbf{E} \otimes \mathbf{H},$$
$$\tilde{\mathbf{M}}_3 = \mathbf{I}_{L\tilde{M}\tilde{N}} - \mathbf{E} \otimes \tilde{\mathbf{H}}$$

and write

$$\mathbf{C}^{-1} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C = \left(\mathbf{M}_3 - \mathbf{M}_2 \right)^{-1} - \mathbf{T}_C^F \left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1 \right)^{-1} \mathbf{T}_F^C.$$

With (14) the second term can be bounded by

$$\left\| \mathbf{T}_C^F \left(\tilde{\mathbf{M}}_3 - \tilde{\mathbf{M}}_1 \right)^{-1} \mathbf{T}_F^C \right\| \leq c_6 \frac{1}{\mu}$$

if μ is large enough. In the very same way we can also bound the first term, i.e.

$$\left\| \left(\mathbf{M}_3 - \mathbf{M}_2 \right)^{-1} \mathbf{T}_F^C \right\| \leq c_7 \frac{1}{\mu},$$

for μ large enough, so that

$$\left\| \mathbf{C}^{-1} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \mathbf{T}_F^C \right\| \leq c_6 \frac{1}{\mu} + c_7 \frac{1}{\mu} \leq c \frac{1}{\mu}.$$

$\square$

A natural question to ask is whether the smoother satisfies the smoothing property, which would make PFASST an actual multigrid algorithm in the classical sense. However, this is not the case, as we can also observe numerically [16]. Still, we can bound the norm of the iteration matrix.

Lemma 4

For the CFL number μ large enough and $k \in \mathbb{N}$ iterations, we have

$$\left\| \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}^{-1} \mathbf{C} \right)^k \right\| \leq \mu \left(c \left\| \mathbf{I}_L \otimes \left(\mathbf{I}_M - \mathbf{Q}_\Delta^{-1} \mathbf{Q} \right)^k \otimes \mathbf{I}_N \right\| + \mathcal{O} \left(\frac{1}{\mu} \right) \right),$$

if $\mathbf{A}^{-1}$ exists and the LU trick is used for $\mathbf{Q}_\Delta$.

Proof
Showing this is rather straightforward using the derivation of Lemma 2 and by realizing that

$$\mathbf{C} = \mu \left(\frac{1}{\mu} \mathbf{I}_{LMN} - \mathbf{I}_L \otimes \mathbf{Q} \otimes \mathbf{A} - \frac{1}{\mu} \mathbf{E} \otimes \mathbf{H} \right)$$

and therefore $\|\mathbf{C}\| \leq c_8 \mu$ for μ large enough. $\qquad\square$

At first glance this does look like the standard smoothing property for multigrid methods after all, where we would expect

$$\left\| \mathbf{C} \left(\mathbf{T}_{\mathrm{S}} \left(\mu \right) \right)^k \right\| = \left\| \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}^{-1} \mathbf{C} \right)^k \right\| \leq \mu g(k)$$

with $g(k) \to 0$ for $k \to \infty$ independent of μ. In our case, however, $g(k) = \mathcal{O}\left(\frac{1}{\mu} \right)$ if k is larger than the number of quadrature nodes M, see Lemma 2. Even worse, this implies that

$$\left\| \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}^{-1} \mathbf{C} \right)^k \right\| \leq c$$

for some constant $c > 0$, if $k \geq M$, so that in this case the norm of the smoother does not even converge to zero, if μ goes to infinity. It is not even guaranteed that this constant c is below 1.

However, we can couple Lemmas 3 and 4 to bound the norm of the full iteration matrix of PFASST.

Theorem 4
If Lemmas 3 and 4 hold true, then the norm of the iteration matrix $\mathbf{T}_{\mathrm{PFASST}} \left(\mu, k \right)$ of PFASST with k pre-smoothing steps can be bounded by

$$\left\| \mathbf{T}_{\mathrm{PFASST}} \left(\mu, k \right) \right\| \leq c \left\| \mathbf{I}_L \otimes \left(\mathbf{I}_M - \mathbf{Q}_\Delta^{-1} \mathbf{Q} \right)^k \otimes \mathbf{I}_N \right\| + \mathcal{O}\left(\frac{1}{\mu} \right)$$

so that for $k \geq M$ the norm of the iteration matrix goes to zero as $\mu \to \infty$.

Proof
The proof is straightforward, but we note that we used pre-smoothing here instead of post-smoothing as in Theorem 1, which in the norm does not matter. $\qquad\square$

We see that this gives nearly μ-independent convergence of PFASST as for classical multigrid methods. The last term in the bound becomes smaller and smaller when μ becomes larger and larger so that at least asymptotically convergence speed is increased for $\mu \to \infty$.

Thus, the only fundamental difference between PFASST and classical linear multigrid methods lies in the smoothing property. For the standard approximative block Jacobi method, this does not seem to hold. Yet, there is another approach which helps us here, namely Reusken's Lemma [31, 32]. To apply this, we define the modified approximative block Jacobi preconditioner by

$$\hat{\mathbf{P}}_\omega = \mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \omega \mathbf{Q}_\Delta \otimes \mathbf{A}$$

and the corresponding iteration matrix $\mathbf{T}_\omega$ by

$$\mathbf{T}_\omega = \mathbf{I}_{LMN} - \hat{\mathbf{P}}_\omega^{-1} \mathbf{C}$$

for some parameter $\omega > 0$.

Lemma 5
We assume that the number of quadrature nodes M is small enough, $\mathbf{Q}_\Delta$ is given by the LU trick, $\mathbf{A}$ is invertible and μ is large enough. Then the iteration matrix $\mathbf{T}_2$ satisfies the smoothing property. More precisely, we have

$$\left\| \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}_2^{-1} \mathbf{C} \right)^k \right\| \leq c \sqrt{\frac{8}{k\pi}} \mu$$

for k smoothing steps, μ large enough and $M \leq M^*$, where M^* depends on the matrix norm.

Proof
The basis for this proof as well as for the choice of the parameter ω is Reusken's Lemma, stating that for some invertible matrix $\mathbf{P}$ and an iteration matrix $\mathbf{T} = \frac{1}{2}(\mathbf{I} + \mathbf{B})$ with $\mathbf{B} = \mathbf{I} - \mathbf{P}^{-1}\mathbf{C}$ it is

$$\left\| \mathbf{C}\mathbf{T}^k \right\| \leq \sqrt{\frac{8}{k\pi}} \left\| \mathbf{P} \right\|$$

if $\|\mathbf{B}\| \leq 1$. In our case we have

$$\mathbf{B} = \mathbf{B}_\omega = \mathbf{I} - 2\hat{\mathbf{P}}_\omega^{-1}\mathbf{C} \quad \text{and} \quad \mathbf{P} = \mathbf{P}_\omega = \frac{1}{2}\hat{\mathbf{P}}_\omega$$

and we write

$$\mathbf{B}_\omega = \mathbf{B}_\omega(\mu) = \mathbf{B}_\omega(\infty) + \mathcal{O}\left(\frac{1}{\mu}\right) = \mathbf{I}_L \otimes \left(\mathbf{I}_M - \frac{2}{\omega} \mathbf{Q}_\Delta^{-1} \mathbf{Q} \right) \otimes \mathbf{I}_N + \mathcal{O}\left(\frac{1}{\mu}\right),$$

with $\mathbf{B}_\omega(\infty)$ analogously derived as $\mathbf{T}_S(\infty)$ in Section 4.1. Now, bounding the norm of $\mathbf{B}_\omega(\infty)$ is by far not straightforward and we fall back on numerical calculations to find scenarios where $\|\mathbf{B}_\omega(\infty)\| < 1$. In particular, if we choose $\omega = 2$, $\|.\|$ as the infinity-norm and $M \leq 5$, then $\|\mathbf{B}_\omega(\infty)\| \approx 0.8676$. In the 2-norm, $M = 6, 7$ is also allowed and choosing $\omega > 2$ extends this range further. Anyway, we are able to bound $\|\mathbf{B}_2(\infty)\|$ and thus $\|\mathbf{B}_2\|$ by one, provided we have chosen the parameters carefully. Then, Reusken's Lemma is applicable and we have

$$\left\| \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}_2^{-1}\mathbf{C} \right)^k \right\| \leq \sqrt{\frac{8}{k\pi}} \left\| \mathbf{P}_2 \right\|.$$

Now, the norm of $\mathbf{P}_2$ can be bounded by

$$\left\| \mathbf{P}_2 \right\| = \left\| \frac{1}{2} \left(\mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes 2\mathbf{Q}_\Delta \otimes \mathbf{A} \right) \right\| \leq \mu \left\| \frac{1}{2\mu} \mathbf{I}_{LMN} - \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A} \right\| \leq c\mu$$

if $\mu \geq \mu^*$ for some μ^*, which concludes the proof. □

Although the assumptions in Lemma 5 are more restrictive that those of Lemma 4, we now have an algorithm satifying both smoothing and approximation property, i.e. a multigrid algorithm with μ-independent convergence.

Theorem 5
If Lemmas 3 and 5 hold true, then the norm of the iteration matrix $\mathbf{T}_{\text{PFASST}}(\mu, k)$ of PFASST with k modified approximative block Jacobi pre-smoothing steps can be bounded by

$$\left\| \mathbf{T}_{\text{PFASST}}(\mu, k) \right\| \leq g(k)$$

with $g(k) = ck^{-\frac{1}{2}} \to 0$ as $k \to \infty$, independently of μ.

Proof
This immediately follows from Lemmas 3 and 5 by writing

$$
\begin{aligned}
\left\| \mathbf{T}_{\text{PFASST}}\left(\mu, k\right) \right\| &= \left\| \left(\mathbf{C}^{-1} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \right) \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}_2^{-1} \mathbf{C} \right) \right\| \\
&\leq \left\| \left(\mathbf{C}^{-1} - \mathbf{T}_C^F \tilde{\mathbf{P}}^{-1} \mathbf{T}_F^C \right) \right\| \left\| \mathbf{C} \left(\mathbf{I}_{LMN} - \hat{\mathbf{P}}_2^{-1} \mathbf{C} \right) \right\| \\
&\leq c \frac{1}{\mu} \cdot \sqrt{\frac{8}{k\pi}} \mu = c \cdot \sqrt{\frac{8}{k\pi}}
\end{aligned}
$$

Note again that we used pre- instead of post-smoothing here to stay consistent with the standard multigrid literature. □

While the theoretical estimate of this theorem is much more convenient than the one of Theorem 4, practical implementations do not share this preference. In all tests we have done so far, using the damping factor of 2 (or any other factor other than 1) yields much worse convergence rates for PFASST, see also Section 6. We also note that the same result could have been obtained by using classical damping, i.e. by using

$$
\omega \hat{\mathbf{P}} = \omega \left(\mathbf{I}_{LMN} - \mu \mathbf{I}_L \otimes \mathbf{Q}_\Delta \otimes \mathbf{A} \right),
$$

because the limit matrix $\mathbf{B}_\omega(\infty)$ is the same for both approaches. Yet, the convergence results are even worse for this choice.

As a conclusion, the question of whether or not to use damping for the smoother in PFASST has two answers: yes, if PFASST should be a real multigrid solver and no, if PFASST should be a fast (multigrid-like) solver.

5. INCREASING THE NUMBER OF TIME-STEPS

This last scenario, where the number of time-steps is going to infinity, is actually twofold: (1) the time interval is fixed and (2) the time interval increases with the number of time-steps L. In the first case, a fixed interval $[0, T]$ of length T is divided into more and more time-steps, so that this is a special case of $\mu \to 0$: here, the time-step size Δt is going to 0 as the number of time-steps goes to infinity. The second case, in contrast, keeps μ constant, since neither Δt nor any other parameter is adapted. Solely the number of time-steps and therefore the length of the time interval under consideration is increasing. Yet, for both cases the dimensions of all matrices change with L and we make use of their periodic stencils in the sense of [33] to find bounds for their spectral radii. In particular, Lemma A.2 of [33] states that the spectral radius of an infinite block Toeplitz matrix $\mathbf{A}$ is equal to the essential supremum of the matrix-valued generating symbol of $\mathbf{A}$, providing the limit that is needed in the following.

5.1. Fixed time interval

For the first scenario, we again make use of the perturbation results in Lemma 1 and Theorem 2. Following Eq. (6), we write

$$
\lim_{L \to \infty} \mathbf{T}_{\text{S}}(\mu) = \lim_{L \to \infty} \mathbf{T}_{\text{S}}(0) + \lim_{L \to \infty} \mu \mathbf{D}
$$

for some perturbation matrix $\mathbf{D} \in \mathbb{R}^{LMN \times LMN}$, which is bounded for μ small enough (i.e. for L large enough), see the discussion leading to (6). Now, this matrix $\mathbf{D}$ is bounded for all L, so that because $\mu \to 0$ as $L \to \infty$, we have

$$
\lim_{L \to \infty} \mathbf{T}_{\text{S}}(\mu) = \lim_{L \to \infty} \mathbf{T}_{\text{S}}(0)
$$

The matrix $\mathbf{T}_S(0) = \mathbf{E} \otimes \mathbf{H}$ is a block Toeplitz matrix with periodic stencil and symbol

$$\widehat{\mathbf{T}_S(0)}(x) = e^{-ix}\mathbf{H}$$

in the sense of [33].

Then, we have

$$\rho\left(\lim_{L \to \infty} \mathbf{T}_S(0)\right) = \sup_{x \in [-\pi, \pi]} \rho\left(e^{-ix}\mathbf{H}\right) = 1,$$

since the eigenvalues of $e^{-ix}\mathbf{H}$ are $(M-1)N$-times 0 and M-times e^{-ix}, see Theorem 1. Thus, the spectral radius of $\mathbf{T}_S(\mu)$ converges to 1 for $L \to \infty$. Therefore, in this limit the smoother does not converge, or, more precisely, the smoother will converge slower and slower the larger L becomes. This is because for finite matrices, the spectral radius of the symbol serves as upper limit, so that the spectral radius of the iteration matrix converges to 1 from below. Also, this confirms the heuristic extension of Lemma 1 to infinite-sized operators, where the upper limit of the spectral radius goes to 1 for $L \to \infty$, too.

For the iteration matrix $\mathbf{T}_{\mathrm{CGC}}$ of the coarse-grid correction, the limit matrix $\mathbf{T}_{\mathrm{CGC}}(0)$ is already block-diagonal, see Section 3.2. We have also seen that the spectral radius of this matrix is at least 1 and due to the block-diagonal structure, this does not change for $L \to \infty$. Thus, not surprisingly, also the coarse-grid correction does not converge for $L \to \infty$.

Now, it seems obvious that PFASST itself will not converge, since both components alone fail to do so. The next theorem shows that this is indeed the case, but the proof is slightly more involved.

Theorem 6
For a fixed time interval with L time steps and CFL number $\mu = \mu(L)$, the iteration matrix of PFASST satisfies

$$\rho\left(\lim_{L \to \infty} \mathbf{T}(\mu)\right) \geq 1.$$

Proof
The full iteration matrix $\mathbf{T}(\mu)$ of PFASST can be written as

$$\mathbf{T}(\mu) = \mathbf{T}(0) + \mathcal{O}(\mu) = \mathbf{E} \otimes \left(\mathbf{H}\left(\mathbf{I}_{MN} - \mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C\right)\right) + \mathcal{O}(\mu)$$

see the discussion leading to Theorem 2. As before, this yields

$$\rho\left(\lim_{L \to \infty} \mathbf{T}(\mu)\right) = \rho\left(\lim_{L \to \infty} \mathbf{T}(0)\right) = \sup_{x \in [-\pi, \pi]} \rho\left(\widehat{\mathbf{T}(0)}(x)\right),$$

following [33].

Now, the symbol of the limit matrix $\mathbf{T}(0)$ is given by

$$\widehat{\mathbf{T}(0)}(x) = e^{-ix}\mathbf{H}\left(\mathbf{I}_{MN} - \mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C\right),$$

which makes the computation of the eigenvalues slightly more intricate. Using Theorem 1, we write

$$\mathbf{H}\left(\mathbf{I}_{MN} - \mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C\right) = \mathbf{N} \otimes \mathbf{I}_N - \mathbf{N}\mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C$$

and note that

$$\mathbf{N}\mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C = \begin{pmatrix} t_{M,1} & ... & t_{M,M} \\ \vdots & & \vdots \\ t_{M,1} & ... & t_{M,M,} \end{pmatrix},$$

where $t_{i,j}$ are the entries of the matrix $\mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C$. Thus, we have

$$\mathbf{N} \otimes \mathbf{I}_N - \mathbf{N}\mathbf{T}_{C,Q}^F \mathbf{T}_{F,Q}^C \otimes \mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C = \begin{pmatrix} -t_{M,1}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C & ... & \mathbf{I}_N - t_{M,M}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C \\ \vdots & & \vdots \\ -t_{M,1}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C & ... & \mathbf{I}_N - t_{M,M}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C, \end{pmatrix}$$

and the eigenvalues of this $MN \times MN$ matrix are all zero except for N eigenvalues λ_n given by the eigenvalues of

$$\mathbf{K} = -\sum_{m=1}^{M-1} t_{M,m}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C + \mathbf{I}_N - t_{M,M}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C$$

$$= \mathbf{I}_N - \sum_{m=1}^{M} t_{M,m}\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C = \mathbf{I}_N - c\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C$$

for a constant c representing the sum over all $t_{M,m}$. This holds since for a rank-1 matrix $\mathbf{B} = uv^T$ all eigenvalues are 0 except for the eigenvalue $v^T u$. In Section 3.2 we already discussed that for standard Lagrangian interpolation and restriction half of the eigenvalues of multiplications like $\mathbf{T}_{C,A}^F \mathbf{T}_{F,A}^C$ are zero. Thus, half of the eigenvalues of $\mathbf{K}$ are one, so that half of the eigenvalues λ_n are one. Therefore, the spectral radius of the symbol of the limit matrix can simply be bounded by

$$\rho\left(\widehat{\mathbf{T}(0)}(x)\right) \geq \left|e^{-ix} \cdot 1\right| = 1$$

for all $x \in [-\pi, \pi]$, so that

$$\rho\left(\lim_{L \to \infty} \mathbf{T}(\mu)\right) = \sup_{x \in [-\pi,\pi]} \rho\left(\widehat{\mathbf{T}(0)}(x)\right) \geq 1,$$

which ends the proof. $\qquad\square$

Therefore, PFASST itself does not converge in the limit of $L \to \infty$, if the time interval is fixed. Also, for finite numbers of time-steps, the spectral radius of the iteration matrix is bounded by 1, so that the spectral radius converges to 1 from below, making PFASST slower and slower for increasing numbers of time-steps.

5.2. Extending time interval

In this scenario, the parameter μ does not change for $L \to \infty$. Thus, applying the perturbation argument we frequently used in this work is not possible and fully algebraic bounds for the spectral radii of the iteration matrices do not seem possible. However, we can make use of the results found in [16], where a block-wise Fourier mode analysis is applied to reduce the computational effort required to compute eigenvalues and spectral radii numerically.

More precisely, there exist a transformation matrix $\mathcal{F}$, consisting of a permutation matrix for the Kronecker product as well as a Fourier matrix decomposing the spatial problem, such that for the smoother we have

$$\mathbf{T}_{\mathrm{S}} = \mathcal{F}\mathrm{diag}\left(\mathbf{B}_1, ..., \mathbf{B}_N\right)\mathcal{F}^{-1}$$

for blocks

$$\mathbf{B}_n = \mathbf{I}_{LM} - \left(\mathbf{I}_{LM} - \mu\lambda_n\mathbf{I}_L \otimes \mathbf{Q}_\Delta\right)^{-1}\left(\mathbf{I}_{LM} - \mu\lambda_n\mathbf{I}_L \otimes \mathbf{Q} - \mathbf{E} \otimes \mathbf{N}\right) \in \mathbb{R}^{LM \times LM}.$$

This corresponds to the iteration matrix of the smoother for a single eigenvalue $\lambda_n(\mathbf{A}) = \lambda_n$ of the spatial matrix $\mathbf{A}$. These blocks are block Toeplitz matrices themselves and their symbol is given by

$$\widehat{\mathbf{B}_n}(x) = \mathbf{I}_M - \left(\mathbf{I}_M - \mu\lambda_n\mathbf{Q}_\Delta\right)^{-1}\left(\mathbf{I}_M - \mu\lambda_n\mathbf{Q}\right) + e^{-ix}\mathbf{N}$$
$$= \left(\mathbf{I}_M - \mu\lambda_n\mathbf{Q}_\Delta\right)^{-1}\mu\lambda_n\left(\mathbf{Q} - \mathbf{Q}_\Delta\right) + e^{-ix}\mathbf{N}.$$

Then, for $L \to \infty$ we again use [33] and obtain

$$\rho\left(\lim_{L\to\infty}\mathbf{T}_{\mathrm{S}}\right) = \max_{n=1,\ldots,N}\rho\left(\lim_{L\to\infty}\mathbf{B}_n\right) = \max_{n=1,\ldots,N}\sup_{x\in[-\pi,\pi]}\rho\left(\widehat{\mathbf{B}_n}(x)\right)$$

However, although being only of size $M \times M$, it is unknown how to compute the spectral radius of these symbols for fixed μ, so that numerical computation is required to find these values for a given problem.

Even worse, for the coarse-grid correction the blocks are at least of size $2LM \times 2LM$ due to mode mixing in space (and time, if coarsening in the nodes is applied) and they are not given by periodic stencils as for the smoother. Thus, the results in [33] cannot be applied. Clearly, the same is true for the full iteration matrix of PFASST and so neither the perturbation argument nor the analysis of the symbols provide conclusive bounds or limits for the spectral radius if μ is fixed. We refer to [16, 17] for detailed numerical studies of these iteration matrices and their action on error vectors.

6. SUMMARY AND DISCUSSION

While the parallel full approximation scheme in space and time has been used successfully with advanced space-parallel solvers on advanced HPC machines around the world, a solid mathematical analysis as well as a proof of convergence was still missing. The algorithm in its original form is rather complex and not even straightforward to write down, posing a severe obstacle for any attempt to even formulate a conclusive theory. Yet, with the formal equivalence to multigrid methods as shown in [16], a mathematical framework now indeed exists which allows to use a broad range of established methods for the analysis of PFASST, at least for linear problems. While in [16] a detailed, semi-algebraic Fourier mode analysis revealed many interesting features and also limitations of PFASST, a rigorous convergence proof has not been given so far. In the present paper, we used the iteration matrices of PFASST, its smoother and the coarse-grid correction to establish an asymptotic convergence theory for PFASST. In three sections, we analyzed the convergence of PFASST for the non-stiff and the stiff limit as well as its behavior for increasing numbers of time-steps. For small enough CFL numbers (or dispersion relations) μ, we proved an upper limit for the spectral radius of PFASST's iteration matrix, which goes to zeros for smaller and smaller μ (see Theorem 2). In turn, if μ becomes large, we showed in Theorem 3 that the spectral radius of the iteration matrix is also bounded, but only when the parameters of the smoother are chosen appropriately. In this stiff limit, PFASST also satisfies the standard approximation property of linear multigrid methods, see Lemma 3, and despite the missing smoothing property, a weakened form of μ-independent convergence was proven in Theorem 4. However, in order to satisfy the smoothing property and to make PFASST a pure multigrid method, the smoother needs a damping parameter. Then, using Reusken's Lemma, the smoothing property can be shown and fully μ-independent convergence is achieved, see Theorem 5. For all these results, we used a perturbation argument for the iteration matrices which allowed us to extract their limit matrices and to show convergence towards those. Finally, we investigated PFASST for increasing numbers L of time-steps and showed that PFASST does not converge in the limit case. Even worse, convergence is expected to degenerate for L larger and larger. However, this applies only to a fixed time interval, i.e. to the case where $\mu \to 0$ as $L \to \infty$. For an extending time interval, no analytic bound has been established.

The results presented here contain the first convergence proofs for PFASST for both non-stiff as well as stiff limit cases. The lemmas and theorems of this paper are of theoretical nature and, frankly, quite technical. Thus, the obvious question to ask is whether the results of this paper relate to actual computations with PFASST. In order to provide first answers to this question, we used two standard test cases:

Test A 1D heat equation with $\nu > 0$

$$u_t = \nu \Delta u \quad \text{on } [0,1] \times [0,T],$$
$$u(0,t) = 0, \ u(1,t) = 0,$$
$$u(x,0) = \text{ random}$$

Test B 1D advection equation with $c > 0$

$$u_t = c \nabla u \quad \text{on } [0,1] \times [0,T],$$
$$u(0,t) = u(1,t)$$
$$u(x,0) = \sin(64\pi x)$$

For both cases and all runs we used $M = 3$ Gauß-Radau nodes with the LU trick, $N = 127$ for Test A and $N = 128$ for Test B. The time interval was fixed to $[0,T] = [0,1]$. For both problems we used centered finite differences for the differential operators, yielding real negative eigenvalues for Test A and imaginary eigenvalues for Test B. With Dirichlet boundary conditions in the diffusive case, the spatial matrix $\mathbf{A}$ is invertible and we can allow all modes to be present using a random initial guess. In the advective case, however, $\mathbf{A}$ is not invertible, since two eigenvalues are zero. Thus, we did not use random initial data but a rather oscillatory initial sine wave. For PFASST, we removed the initial prediction phase on the coarse level [10] and set all variables to zero at the beginning of each run, except for $u(x,0)$ (this corresponds to no spreading). Coarsening was done in space only and we terminated the iteration when an absolute residual tolerance of 10^{-8} was met. All results were generated using the pySDC framework and the codes can be found online, see [34].

The first thing to analyze is whether PFASST indeed converges for the non-stiff as well as for the stiff limit. To test this, we chose $\Delta t = 0.25$ fixed and varied ν or c such that the CFL number μ varied between 10^{-4} and 10^{11}. The results for Test A ("diffusion") and Test B ("advection") are shown in Figure 4a for $k = M = 3$ smoothing steps and in Figure 4b for $k = 2$ smoothing steps. We observe how for both problems the numbers of iterations go down for μ small and μ large, at least if $k = M$. As predicted by Theorem 3 and Lemma 4, we need as many smoothing steps as there are quadrature nodes in order to see the number of iterations of PFASST to go down in the stiff limit while convergence in the non-stiff limit is not affected. We also observe how the iteration counts increase for medium sizes of μ and peak at about $\mu \approx 10$, very much in line with the observations made in Figure 3.

The second experiment concerns the behavior of PFASST for increasing numbers of time-steps. For Figure 5 we fixed $\nu = 0.1$ and $c = 0.1$ as well as all other parameters and used 3 smoothing steps. Then, we increased the number of time-steps for PFASST from 1 to 4096 and counted the number of iterations. We did this for the standard, undamped smoother ("LU") as well as for the damped smoother ("LU2") proposed in Lemma 5, in Figure 5a for Test A and in Figure 5b for Test B. This experiment shows three things: first, we see that indeed PFASST's convergence degenerates when more and more time-steps are considered, even if the time interval is fixed. Second, the advective case performs much worse than the diffusive case, which is to be expected from a generic time-parallel method. Third, PFASST with damped smoothing has much worse convergence rates than the unmodified PFASST algorithm and despite being an actual multigrid solver, iteration counts increase for $L \to \infty$. Although the increase in iteration counts is not as severe as for unmodified PFASST, there is no scenario where iteration counts are lower, making this a rather poor parallel-in-time integrator.
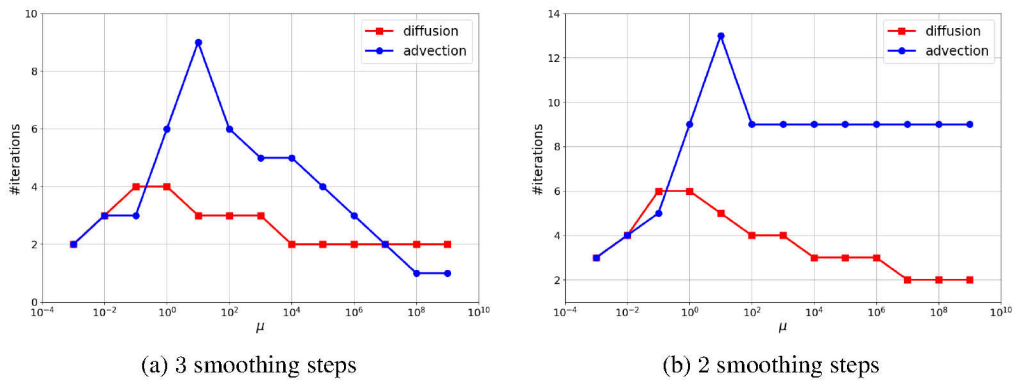
(a) 3 smoothing steps

(b) 2 smoothing steps

Figure 4. Iteration counts for the diffusion (Test A) and advection (Test B) problem using varying parameters ν and c while keeping the other parameters fixed. Left: $k = 3 = M$ smoothing steps, right: $k = 2 < M$ smoothing steps.
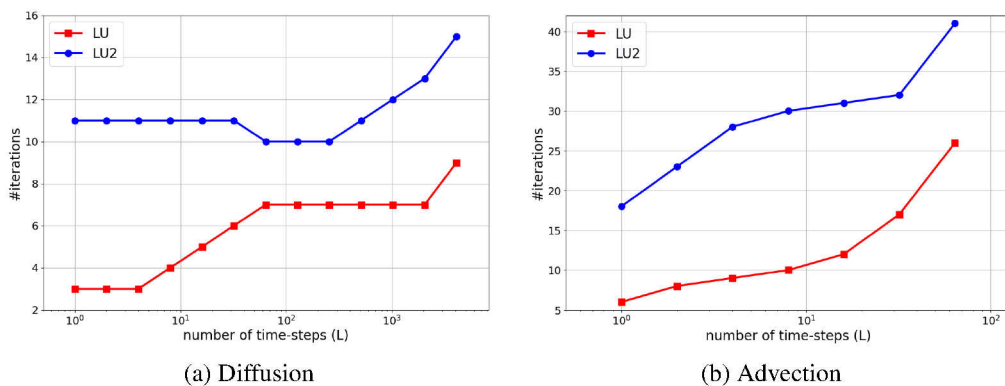

(a) Diffusion

(b) Advection

Figure 5. Mean number of iterations over all time-steps for the diffusion (left) and advection (right) problem on a fixed time interval with increasing numbers of time-steps. PFASST with standard smoothing ("LU") and damped smoothing ("LU2") is shown.

In this paper, we presented the first convergence proofs for the parallel full approximation scheme in space and time, covering asymptotic cases for linear problems. We have seen that the results obtained here are well reflected in actual PFASST runs and can help to understand convergence behavior in realistic scenarios. While this establishes a rather broad convergence theory, one key property of any parallel-in-time integration method cannot be investigated with this directly: parallel performance. In the original papers on PFASST, first theoretical efficiency limits and expected speedups were already derived, but only for fixed numbers of iterations [8, 10]. Yet, the formulation of PFASST as a multigrid method now allows us not only to prove convergence but also to estimate precisely the expected parallel performance for linear problems. We will report on this in a subsequent paper.

REFERENCES

1. Gander MJ. *50 Years of Time Parallel Time Integration*. Springer International Publishing: Cham, 2015; 69–113, doi:10.1007/978-3-319-23321-5_3. URL http://dx.doi.org/10.1007/978-3-319-23321-5_3.
2. Lions JL, Maday Y, Turinici G. A "parareal" in time discretization of PDE's. *Comptes Rendus de l'Acadmie des Sciences - Series I - Mathematics* 2001; **332**:661–668. URL http://dx.doi.org/10.1016/S0764-4442(00)01793-6.
3. Gander MJ, Vandewalle S. Analysis of the Parareal Time-Parallel Time-Integration Method. *SIAM Journal on Scientific Computing* 2007; **29**(2):556–578. URL http://dx.doi.org/10.1137/05064607X.
4. Gander MJ, Vandewalle S. On the Superlinear and Linear Convergence of the Parareal Algorithm. *Domain Decomposition Methods in Science and Engineering, Lecture Notes in Computational Science and Engineering,*

vol. 55, Widlund OB, Keyes DE (eds.). Springer Berlin Heidelberg, 2007; 291–298. URL http://dx.doi.org/10.1007/978-3-540-34469-8_34.

5. Gander MJ. Analysis of the Parareal Algorithm Applied to Hyperbolic Problems using Characteristics. *Bol. Soc. Esp. Mat. Apl.* 2008; **42**:21–35.

6. Gander MJ, Hairer E. Nonlinear Convergence Analysis for the Parareal Algorithm. *Domain Decomposition Methods in Science and Engineering*, *Lecture Notes in Computational Science and Engineering*, vol. 60, Langer U, Widlund O, Keyes D (eds.), Springer, 2008; 45–56. URL http://dx.doi.org/10.1007/978-3-540-75199-1_4.

7. Staff GA, Rnquist EM. Stability of the parareal algorithm. *Domain Decomposition Methods in Science and Engineering*, *Lecture Notes in Computational Science and Engineering*, vol. 40, Kornhuber R, et al (eds.), Springer: Berlin, 2005; 449–456. URL http://dx.doi.org/10.1007/3-540-26825-1_46.

8. Minion ML. A Hybrid Parareal Spectral Deferred Corrections Method. *Communications in Applied Mathematics and Computational Science* 2010; **5**(2):265–301. URL http://dx.doi.org/10.2140/camcos.2010.5.265.

9. Dutt A, Greengard L, Rokhlin V. Spectral deferred correction methods for ordinary differential equations. *BIT Numerical Mathematics* 2000; **40**(2):241–266. URL http://dx.doi.org/10.1023/A:1022338906936.

10. Emmett M, Minion ML. Toward an Efficient Parallel in Time Method for Partial Differential Equations. *Communications in Applied Mathematics and Computational Science* 2012; **7**:105–132. URL http://dx.doi.org/10.2140/camcos.2012.7.105.

11. Emmett M, Minion ML. Efficient implementation of a multi-level parallel in time algorithm. *Domain Decomposition Methods in Science and Engineering XXI*, *Lecture Notes in Computational Science and Engineering*, vol. 98, Springer International Publishing, 2014; 359–366. URL http://dx.doi.org/10.1007/978-3-319-05789-7_33.

12. Speck R, Ruprecht D, Krause R, Emmett M, Minion ML, Winkel M, Gibbon P. A massively space-time parallel N-body solver. *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '12, IEEE Computer Society Press: Los Alamitos, CA, USA, 2012; 92:1–92:11. URL http://dx.doi.org/10.1109/SC.2012.6.

13. Speck R, Ruprecht D, Emmett M, Bolten M, Krause R. A space-time parallel solver for the three-dimensional heat equation. *Parallel Computing: Accelerating Computational Science and Engineering (CSE)*, *Advances in Parallel Computing*, vol. 25, IOS Press, 2014; 263–272. URL http://dx.doi.org/10.3233/978-1-61499-381-0-263.

14. Speck R, Ruprecht D, Krause R, Emmett M, Minion ML, Winkel M, Gibbon P. Integrating an N-body problem with SDC and PFASST. *Domain Decomposition Methods in Science and Engineering XXI*, *Lecture Notes in Computational Science and Engineering*, vol. 98, Springer International Publishing, 2014; 637–645. URL http://dx.doi.org/10.1007/978-3-319-05789-7_61.

15. Minion ML, Speck R, Bolten M, Emmett M, Ruprecht D. Interweaving PFASST and parallel multigrid. *SIAM Journal on Scientific Computing* 2015; **37**:S244 – S263. URL http://dx.doi.org/10.1137/14097536X.

16. Bolten M, Moser D, Speck R. A multigrid perspective on the parallel full approximation scheme in space and time. *Numerical Linear Algebra with Applications* 2017; **24**(6):e2110, doi:10.1002/nla.2110. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.2110.

17. Moser D. A multigrid perspective on the parallel full approximation scheme in space and time. PhD Thesis, Bergische Universität Wuppertal, Germany 2017.

18. Brandt A. Multi-level adaptive solutions to boundary-value problems. *Math. Comp.* 1977; **31**(138):333–390.

19. Stüben K, Trottenberg U. Multigrid methods: fundamental algorithms, model problem analysis and applications. *Multigrid methods (Cologne, 1981)*, *Lecture Notes in Math.*, vol. 960. Springer, Berlin-New York, 1982; 1–176.

20. Trottenberg U, Oosterlee CW, Schller A. *Multigrid*. Academic Press, 2001.

21. Wienands R, Joppich W. *Practical Fourier Analysis For Multigrid Methods*. Chapman Hall/CRC Press, 2005.

22. Friedhoff S, MacLachlan S. A generalized predictive analysis tool for multigrid methods. *Numer. Linear Algebra Appl.* 2015; **22**(4):618–647, doi:10.1002/nla.1977. URL http://dx.doi.org/10.1002/nla.1977.

23. Hairer E, Wanner G. *Solving ordinary differential equations. II. Stiff and differential-algebraic problems*. Springer series in computational mathematics, Springer: Heidelberg, New York, 2010. URL http://opac.inria.fr/record=b1130632.

24. Huang J, Jia J, Minion M. Accelerating the convergence of spectral deferred correction methods. *Journal of Computational Physics* 2006; **214**(2):633 – 656. URL http://dx.doi.org/10.1016/j.jcp.2005.10.004.

25. Weiser M. Faster SDC convergence on non-equidistant grids by DIRK sweeps. *BIT Numerical Mathematics* 2014; **55**(4):1219–1241, doi:10.1007/s10543-014-0540-y. URL http://dx.doi.org/10.1007/s10543-014-0540-y.

26. Ruprecht D, Speck R. Spectral deferred corrections with fast-wave slow-wave splitting. *SIAM Journal on Scientific Computing* 2016; **38**(4):A2535–A2557, doi:10.1137/16M1060078. URL http://dx.doi.org/10.1137/16M1060078.

27. Speck R, Ruprecht D, Emmett M, Minion M, Bolten M, Krause R. A multi-level spectral deferred correction method. *BIT Numerical Mathematics* 2015; **55**(3):843–867, doi:10.1007/s10543-014-0517-x. URL http://dx.doi.org/10.1007/s10543-014-0517-x.

28. Wilkinson JH ((ed.)). *The Algebraic Eigenvalue Problem*. Oxford University Press, Inc.: New York, NY, USA, 1988.

29. Stewart GW, Sun Jg. *Matrix perturbation theory*. Academic Press: Boston, 1990.

30. Hackbusch W. *Multi-grid methods and applications*. Springer series in computational mathematics, Springer-Verlag, 1985. URL https://books.google.de/books?id=Vf8_AQAAIAAJ.

31. Reusken AA. A new lemma in multigrid convergence theory. *Rana: reports on applied and numerical analysis; vol. 9107*, Technische Universiteit Eindhoven 1991.

32. Braess D. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 2001. URL https://books.google.de/books?id=DXin1jNR4ioC.
33. Bolten M, Rittich H. Fourier analysis of periodic stencils in multigrid methods. *submitted* 2017; .
34. Speck R. Parallel-in-time/pysdc: Asymptotic convergence Mar 2017, doi:10.5281/zenodo.420034. URL https://doi.org/10.5281/zenodo.420034.