Convergence Bounds for Parareal with Spatial Coarsening

Aušra Pogoželskytė and Martin J. Gander

1 Introduction

In the times of exascale computing, very efficient algorithms for space parallelism exist and communication between processors has become a bottleneck; parallelism in the time dimension is necessary. The Parareal algorithm [10] allows us to do this in a non-intrusive fashion. However, it is limited by the sequential nature of its coarse operator. A solution would be to consider a multilevel version of Parareal such as Multigrid Reduction in Time (MGRIT) [3], but it also suffers from scalability issues due to coarsening only in the time dimension. There are some algorithms that are scalable such as the Space-Time Multigrid algorithm of Gander and Neumüller [8], but they are more intrusive.

Parareal with spatial coarsening offers a solution that would be both scalable and non-intrusive. It has been first introduced in [4] in order to preserve the stability of the scheme while solving the Navier-Stokes equations, but it has only been numerically studied in [11].

Many bounds for Parareal exist [5, 7, 9], but they do not provide useful information in the case where spatial coarsening is used, or have assumptions that cannot be satisfied such as simultaneous diagonalizability in [2, 12]. We present here two convergence bounds for Parareal with spatial coarsening, and illustrate our results using the heat equation.

Aušra Pogoželskytė, Martin J. Gander
University of Geneva, rue du Conseil-Général 7-9, Geneva, e-mail: ausra.pogozelskyte@unige.ch, martin.gander@unige.ch

2 Parareal with spatial coarsening

Let $A \in \mathbb{R}^{n_x \times n_x}$ be a matrix coming from a uniform n_x point spatial discretization of a partial differential equation and define $f\colon \mathbb{R} \to \mathbb{R}^{n_x}$ to be a source term. Consider the initial value problem,

$$y'(t) = A\,y(t) + f(t),\ t \in (0,T], \quad y(0) = y_0\ . \tag{1}$$

Note that the assumptions on the linearity of the problem and the time independence of A are only necessary to facilitate the analysis. The problem (1) can be then discretized in time on a uniform $n_t + 1$ point mesh using a one-step method. This results in the time-stepping iteration,

$$y_{n+1} = \Phi\, y_n + f_n, \quad n = 0, \ldots, n_t\ . \tag{2}$$

Now, coarsen the spatial and time mesh by a factor of m_x and m_t such that the fine mesh can be seen as a refinement of the coarse mesh. Define u and g to be the restrictions of y and f to the coarse grid. Proceeding in the same way, a time-stepping iteration can be defined on the coarse mesh,

$$u_{n+1} = \Psi\, u_n + g_n, \quad n = 0, \ldots, n_t/m_t\ , \tag{3}$$

where Ψ is an operator depending on the coarse time-step and the spatial discretization matrix of the coarse problem. Define also the operators $F := \Phi^{m_t}$ and $G := \Psi$.

Example: Discretization of the heat equation

Consider the one-dimensional heat equation,

$$\begin{cases} \partial_t y(x,t) = c^2 \partial_{xx} y(x,t) + f(x,t)\ , & (x,t) \in (0,L) \times (0,T], \\ y(x,0) = y_0(x)\ , & x \in (0,L), \\ y(0,t) = y(L,t) = 0\ , & t \in (0,T]\ . \end{cases} \tag{4}$$

Denote $\{x_i\}_{i=0}^{n_x}$ and $\{X_j\}_{j=0}^{N_x}$ to be fine and coarse discretizations of $[0,L]$ such that $x_i = i\Delta x$ and $X_j = j m_x \Delta x$ where $\Delta x = L/n_x$ and $n_x = m_x N_x$. This leads to a system of the form (1) where A and A_Δ are discrete Laplacians on the fine and coarse grids,

$$A := \frac{c^2}{\Delta x^2} \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix}, \quad A_\Delta := \frac{c^2}{(m_x \Delta x)^2} \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix}.$$

Using a Runge-Kutta method whose stability function is given by R, one gets the operators $F = R(\Delta t A)^{m_t}$ and $G = R(m_t \Delta t A_\Delta)$.

The time stepping (2) can be parallelized in the time dimension using the Parareal algorithm with spatial coarsening. It was first introduced by Fischer, Hecht and Maday in [4] in order to preserve the stability of their numerical scheme when solving the Navier-Stokes equations using Parareal as well as reducing the serial overhead of the coarse operator. Define $I \in \mathbb{R}^{n_x \times N_x}$ and $R \in \mathbb{R}^{N_x \times n_x}$ to be interpolation and restriction operators. The algorithm is initialized as

$$Y_0^0 = y_0 \,, \quad Y_n^0 = (I\,G\,R) Y_{n-1}^0 \,, \qquad n = 1, \ldots, N_t \tag{5}$$

and its kth iteration is given by

$$Y_0^k = y_0 \,, \quad Y_n^k = F\,Y_{n-1}^{k-1} - (I\,G\,R) Y_{n-1}^{k-1} + (I\,G\,R) Y_{n-1}^k \,, \quad n = 1, \ldots, N_t \tag{6}$$

Remark 1 The iteration (6) is the same as traditional Parareal, except that the coarse operator $\widetilde{G} := IGR$ is used. This means, in particular, that properties such as finite time convergence [6, Theorem 5] are conserved. The finite time convergence property states that the algorithm will converge in the worst case at iteration $k = N_t$.

3 Analysis of parareal with spatial coarsening

To get insight on the convergence of the algorithm, we will investigate how it behaves for each spatial mode of the error. Therefore, the spatial dimension will be transformed in a Fourier basis, leaving the time dimension as is. For simplicity of the analysis, we will assume that the coarsening factor in space is $m_x = 2$.

3.1 Derivation of the error propagation operator

The first step of the analysis is to derive the error propagation operator. Let Y_n be the fine solution of (1) and $e_n^k = Y_n - Y_n^k$ denote the error of the Parareal algorithm. Then, the error verifies the recurrence relation

$$\begin{aligned} e_{n+1}^{k+1} &= F\,Y_n - F\,Y_n^k + (I\,G\,R) Y_n^k - (I\,G\,R) Y_n^{k+1} + (I\,G\,R) Y_n - (I\,G\,R) Y_n \\ &= (F - I\,G\,R)\, e_n^k + (I\,G\,R)\, e_n^{k+1} \,. \end{aligned} \tag{7}$$

Let $\boldsymbol{e}^k := [e_0^k, \ldots, e_{n_t}^k]^\top$ be the vector of the errors at iteration k for all time steps. Define Γ_{-i} to be a matrix with ones on its ith subdiagonal and zeros elsewhere. Then, the relation (7) can be written as the linear system

$$(I_t \otimes I_x - \Gamma_{-1} \otimes (I\,G\,R))\, \boldsymbol{e}^{k+1} = (\Gamma_{-1} \otimes (F - I\,G\,R))\, \boldsymbol{e}^k \,, \tag{8}$$

which can be solved to get the iteration $\boldsymbol{e}^{k+1} = E\,\boldsymbol{e}^{k}$. The matrix E is called the *error propagation operator* and describes the evolution of the error between two iterations for all time (and space) steps.

Lemma 1 *The error propagation operator E is given by*

$$E = \left(\sum_{i=0}^{n_t-1} \Gamma_{-i-1} \otimes (I\,G\,R)^i\right)(I_t \otimes (F - I\,G\,R))\ .$$

Proof The result is obtained by direct computation, see [1, Lemma 2.1] for similar computations. □

3.2 Analysis of the error propagation operator in Fourier space

Let $U \in \mathbb{R}^{n_x \times n_x}$ be a basis that diagonalizes the fine operator F. The columns of the matrix U will be referred to as *modes* and can be further split into *low* and *high* frequency modes $\check{u}_\kappa$ and $\hat{u}_\kappa$. Low modes are defined to be the modes that can be well represented on the considered mesh, and high modes those who can not.

When considering coarsening in space, the issue of *aliasing* arises, see Figure 1. That means that there are pairs of low and high frequency modes $\{\check{u}_\kappa, \hat{u}_\kappa\}$ that are mapped to the same coarse mode u_κ on the coarse grid. Those pairs are said to be *harmonic* of each other. We can define the *space of harmonics* as the space spanned by pairs of such modes (see [13] for a more detailed discussion).

Assume that the transfer operators R and I keep the spaces of harmonics invariant[1]. Then, there exist symbols $\check{\iota}_\kappa$, $\hat{\iota}_\kappa$, $\check{r}_\kappa$ and $\hat{r}_\kappa$ such that

$$R\check{u}_\kappa = \check{r}_\kappa u_\kappa\ , \quad R\hat{u}_\kappa = \hat{r}_\kappa u_\kappa\ , \quad Iu_\kappa = \check{\iota}\check{u}_\kappa + \hat{\iota}\hat{u}_\kappa\ .$$

Likewise, under our hypothesis on the coarse grid, the coarse modes u_κ will diagonalize the operator G,

$$Gu_\kappa = \mu_\kappa u_\kappa, \quad F\check{u}_\kappa = \check{\lambda}_\kappa \check{u}_\kappa, \quad F\hat{u}_\kappa = \hat{\lambda}_\kappa \hat{u}_\kappa\ .$$

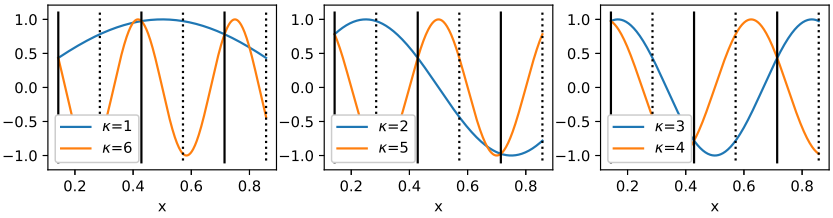

Fig. 1 Example of aliasing for the heat equation with $n_x = 7$. Points on the fine and coarse grid intersect the dotted and continuous lines. One can observe pairs of modes: one low (blue) and one high (orange), that get mapped to a single coarse mode on the coarse grid.

[1] This is true for common Multigrid transfer operators, see [13].

Example: Modes and Fourier symbols for the heat equation

For the heat equation (4), due to Dirichlet boundary conditions, the fine operator is diagonalized by a basis of sines, thus the columns of U are given by

$$u_\kappa(x) = \sin\left(\frac{\kappa\pi x}{L}\right) , \quad x \in \{x_j\}_{j=0}^{n_x}, \quad \kappa = 1, \dots, n_x .$$

The corresponding low (coarse) and high modes are for $\kappa = 1, \dots, N_x$

$$\check{u}_\kappa(x) = \sin\left(\frac{\kappa\pi x}{L}\right) \equiv u_\kappa(x) , \quad \hat{u}_\kappa(x) = \sin\left(\frac{(n_x-\kappa)\pi x}{L}\right) \equiv u_{n_x-\kappa}(x) .$$

Consider linear interpolation and full-weighting restriction, defined as

$$(R_{FW}u)_j = \frac{1}{4}(u_{2j-1} + 2u_{2j} + u_{2j+1}), \quad j = 0, \dots, n/2-1 ,$$
$$(I_{lin}v)_j = \begin{cases} v_{j/2} & \text{if } j \text{ is even,} \\ \frac{1}{2}(v_{(j-1)/2} + v_{(j+1)/2}) & \text{if } j \text{ is odd.} \end{cases}$$

Then the associated Fourier symbols are given by

$$\check{\iota}_\kappa = \check{r}_\kappa = \frac{1}{2}(1 + \cos(\pi x/L)) , \quad \hat{\iota}_\kappa = \hat{r}_\kappa = \frac{1}{2}(1 - \cos(\pi x/L)) .$$

Lemma 2 *The error propagation operator keeps the space of harmonics invariant.*

Proof By Lemma 1, the error propagation operator depends on the fine and coarse propagators F and G as well as on the transfer operators R and I, and powers thereof. Indeed,

$$IGR\check{u}_\kappa = IG\check{r}_\kappa u_\kappa = I\mu_\kappa\check{r}_\kappa u_\kappa = \check{\iota}_\kappa\mu_\kappa\check{r}_\kappa\check{u}_\kappa + \hat{\iota}_\kappa\mu_\kappa\check{r}_\kappa\hat{u}_\kappa,$$
$$IGR\hat{u}_\kappa = IG\hat{r}_\kappa u_\kappa = I\mu_\kappa\hat{r}_\kappa u_\kappa = \check{\iota}_\kappa\mu_\kappa\hat{r}_\kappa\check{u}_\kappa + \hat{\iota}_\kappa\mu_\kappa\hat{r}_\kappa\hat{u}_\kappa.$$

This can be summarized as

$$IGR[\check{u}_\kappa, \hat{u}_\kappa] = [\check{u}_\kappa, \hat{u}_\kappa]\mu_\kappa \begin{pmatrix} \check{\iota}_\kappa\check{r}_\kappa & \check{\iota}_\kappa\hat{r}_\kappa \\ \hat{\iota}_\kappa\check{r}_\kappa & \hat{\iota}_\kappa\hat{r}_\kappa \end{pmatrix} =: [\check{u}_\kappa, \hat{u}_\kappa]\mu_\kappa\Pi_\kappa . \tag{9}$$

Similarly, we can write an analog expression for the fine operator F,

$$F[\check{u}_\kappa, \hat{u}_\kappa] = [\check{u}_\kappa, \hat{u}_\kappa] \begin{pmatrix} \check{\lambda}_\kappa & 0 \\ 0 & \hat{\lambda}_\kappa \end{pmatrix} =: [\check{u}_\kappa, \hat{u}_\kappa]\Lambda_\kappa . \tag{10}$$

Thus, as the space of harmonics is invariant for the operators IGR and F, it is also invariant for the error propagation operator. □

Lemma 2 says that the error propagator E can be transformed in the Fourier space to an operator that is diagonal by blocks, where each block is given by

$$\widetilde{E}_\kappa = \left(\sum_{i=0}^{n_t-1} \Gamma_{-i-1} \otimes (\mu_\kappa \Pi_\kappa)^i \right) (I_t \otimes (\Lambda_\kappa - \mu_\kappa \Pi_\kappa)) \ . \tag{11}$$

Denote by $\tilde{e}_n^k(\kappa)$ the error at time-step n and iteration k of Parareal in Fourier space for the wavenumber κ and its corresponding high frequency mode. Using the relationship (11), we can get a relationship on the pairs of harmonics that can be bounded in the 2-norm as

$$\|\tilde{e}_{n+1}^{k+1}(\kappa)\|_2 = \|\Lambda_\kappa - \mu_\kappa \Pi_\kappa\|_2 \|\tilde{e}_n^k(\kappa)\|_2 + \|\mu_\kappa \Pi_\kappa\|_2 \, \|\tilde{e}_n^{k+1}(\kappa)\|_2 \ . \tag{12}$$

To shorten the notation, we will omit κ in $\tilde{e}_n^k(\kappa)$ and refer to it as $\tilde{e}_n^k$, and all the norms in what follows will be 2-norms, i.e. $\|\cdot\| := \|\cdot\|_2$.

Theorem 1 (Linear bound)

Let Λ_κ and $\mu_\kappa \Pi_\kappa$ be defined as in (9) *and* (10)*. Then, for a given wavenumber κ, the error of Parareal with spatial coarsening in Fourier space is bounded as*

$$\max_{1 \le n \le N_t} \|\tilde{e}_n^{k+1}\| \le \frac{1 - \|\mu_\kappa \Pi_\kappa\|^{N_t}}{1 - \|\mu_\kappa \Pi_\kappa\|} \|\Lambda_\kappa - \mu_\kappa \Pi_\kappa\| \max_{1 \le n \le N_t} \|\tilde{e}_n^k\| \ . \tag{13}$$

Proof Let $\alpha = \|\Lambda_\kappa - \mu_\kappa \Pi_\kappa\|_2$ and $\beta = \|\mu_\kappa \Pi_\kappa\|_2$. Using iteration (12) and iterating on the term $\|\tilde{e}_n^{k+1}\|$ yields

$$\|\tilde{e}_{n+1}^{k+1}\| \le \alpha \|\tilde{e}_n^k\| + \beta\big(\alpha \|\tilde{e}_{n-1}^k\| + \beta \|\tilde{e}_{n-1}^{k+1}\|\big) \le \ldots \le \sum_{i=0}^{n-1} \beta^i \alpha \|\tilde{e}_{n-i}^k\| + \beta^n \|\tilde{e}_0^{k+1}\| \ .$$

As the error at the initial time step is $e_0^{k+1} = 0$, we have $\tilde{e}_0^{k+1} = 0$. Taking the maximum over n and using the closed form of geometric series concludes the proof.

Remark 2 One can notice that (13) is very similar to the bound found in [2] where no spatial coarsening is considered. The only addition is this matrix Π_κ which accounts for the effect of transfer operators on the Fourier modes.

Remark 3 Note that in Theorem 1 the inequality step is due to the submultiplicative property of the matrix norm and the fact that

$$\|(\mu_\kappa \Pi_\kappa)^i\| \le \|\mu_\kappa \Pi_\kappa\|^i \ . \tag{14}$$

However, the inequality (14) is an equality, as Π_κ is a rank one matrix. Indeed, there exists a matrix Π_0 such that $\Pi^k = \mathrm{tr}(\Pi)^k \Pi_0$. In our case, $\Pi_0 := \frac{1}{\mathrm{tr}(\Pi)} \Pi$.

Lemma 3 (Theorem 2.10 in [7])

Let $\alpha, \beta \in \mathbb{R}$, then the recurrence relation $e_{n+1}^{k+1} \leq \alpha e_n^k + \beta e_n^{k+1}$ implies

$$e_{n+1}^k \leq \frac{\alpha^k}{(k-1)!} \sum_{i=0}^{n-k} \prod_{l=1}^{k-1} (i+l) \beta^i \max_n e_n^0 \,. \tag{15}$$

Theorem 2 (Superlinear bound)

Let Λ_κ and $\mu_\kappa \Pi_\kappa$ be defined as in (9) *and in* (10)*. Then, for a given wavenumber κ, the error of Parareal with spatial coarsening in Fourier space is bounded by*

$$\|\tilde{e}_{n+1}^k\| \leq \frac{\|\Lambda_\kappa - \mu_\kappa \Pi_\kappa\|^k}{(k-1)!} \sum_{i=0}^{N_t-k} \prod_{l=1}^{k-1} (i+l) \, \|\mu_\kappa \Pi_\kappa\|^i \max_n \|\tilde{e}_n^0\| \,.$$

Proof From (12) and setting $\alpha = \|\Lambda_\kappa - \mu_\kappa \Pi_\kappa\|$ and $\beta = \|\mu_\kappa \Pi_\kappa\|$, we obtain the relationship in Lemma 3, which concludes the proof. □

Remark 4 As we have a bound for the error in Fourier space, we can deduce a bound for the error in the real space e_n^k. Indeed, by the discrete equivalent of the Perseval theorem, we get

$$\|e_n^k\|_2^2 = \frac{1}{n_x} \sum_{\kappa=1}^{n_x} |\tilde{e}_n^k|^2 = \frac{1}{m_x N_x} \sum_{\kappa=1}^{N_x} (|\check{e}_n^k|^2 + |\hat{e}_n^k|^2) = \frac{1}{m_x N_x} \sum_{\kappa=1}^{N_x} \|\tilde{e}_n^k\|_2^2 \,,$$

where the second equality is obtained by splitting the error modes into high and low frequencies, and the third by grouping them into pairs. In particular,

$$\|e_n^k\|^2 \leq \frac{1}{m_x} \max_\kappa \|\tilde{e}_n^k\|^2 \,.$$

We show in Figure 2 a numerical illustration of the linear and superlinear bounds from Theorem 1 and Theorem 2. We see that our convergence bounds for Parareal with

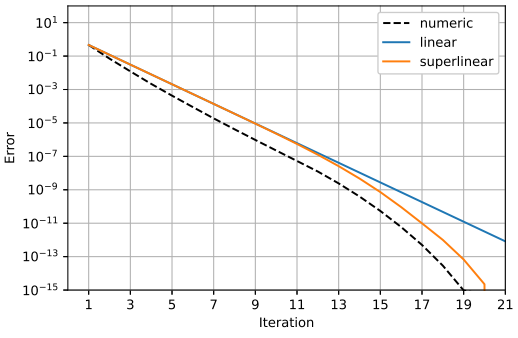

Fig. 2 Numerical error for Parareal with spatial coarsening, and linear (Theorem 1) and superlinear (Theorem 2) convergence bounds. The numerical errors are measured in the 2-norm in space and ∞-norm in time when solving the heat equation with $n_x = n_t = 200$, $m_x = 2$, $m_t = 10$ on $[0, 10^{-1}]$.

spatial coarsening are indeed capturing the convergence behavior of the algorithm. Our results are thus an important step for fully understanding Parareal with spatial coarsening. In particular, they emphasize the crucial role of transfer operators in the convergence of the algorithm, through the matrix Π.

Acknowledgements This project has received funding from the Federal Ministry of Education and Research and the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955701, Time-X. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, Switzerland.

References

1. Angel, J., Götschel, S., and Ruprecht, D. Impact of spatial coarsening on Parareal convergence (2021). URL http://arxiv.org/abs/2111.10228.
2. Dobrev, V. A., Kolev, T., Petersson, N. A., and Schroder, J. B. Two-Level Convergence Theory for Multigrid Reduction in Time (MGRIT). *SIAM J. Sci. Comput.* **39**(5), S501–S527 (2017).
3. Falgout, R. D., Friedhoff, S., Kolev, T. V., MacLachlan, S. P., and Schroder, J. B. Parallel Time Integration with Multigrid. *SIAM J. Sci. Comput.* **36**(6), C635–C661 (2014).
4. Fischer, P. F., Hecht, F., and Maday, Y. A Parareal in Time semi-implicit Approximation of the Navier-Stokes Equations. In: *Domain Decomposition Methods in Science and Engineering*, vol. 40, 433–440. Springer, Berlin (2005).
5. Gander, M. J. and Hairer, E. Nonlinear Convergence Analysis for the Parareal Algorithm. In: *Domain Decomposition Methods in Science and Engineering XVII*, vol. 60, 45–56. Springer, Berlin (2008).
6. Gander, M. J., Kwok, F., and Zhang, H. Multigrid interpretations of the parareal algorithm leading to an overlapping variant and MGRIT. *Comput. Visual Sci.* **19**(3), 59–74 (2018).
7. Gander, M. J., Lunet, T., Ruprecht, D., and Speck, R. A unified analysis framework for iterative parallel-in-time algorithms (2022). URL http://arxiv.org/abs/2203.16069.
8. Gander, M. J. and Neumüller, M. Analysis of a New Space-Time Parallel Multigrid Algorithm for Parabolic Problems. *SIAM J. Sci. Comput.* **38**(4), A2173–A2208 (2016).
9. Gander, M. J. and Vandewalle, S. Analysis of the Parareal Time-Parallel Time-Integration Method. *SIAM J. Sci. Comput.* **29**(2), 556–578 (2007).
10. Lions, J.-L., Maday, Y., and Turinici, G. Résolution d'EDP par un schéma en temps "pararéel". *C. R. Acad. Sci. Paris Sér. I Math.* **332**(7), 661–668 (2001).
11. Ruprecht, D. Convergence of Parareal with spatial coarsening. *PAMM* **14**(1), 1031–1034 (2014).
12. Southworth, B. S. Necessary Conditions and Tight Two-level Convergence Bounds for Parareal and Multigrid Reduction in Time. *SIAM J. Matrix Anal. Appl.* **40**(2), 564–608 (2019).
13. Trottenberg, U., Oosterlee, C. W., and Schüller, A. *Multigrid*. Academic Press, Inc., San Diego, CA (2001).