Article:
Ruprecht, D (2014) Convergence of Parareal with spatial coarsening. PAMM, 14 (1). 1031 - 1034. ISSN 1617-7061

https://doi.org/10.1002/pamm.201410490

Convergence of Parareal with spatial coarsening

Daniel Ruprecht[1]*

[1] Institute of Computational Science, Università della Svizzera italiana, CH-6904 Lugano, Switzerland

The effect is investigated of using a reduced spatial resolution in the coarse propagator of the time-parallel Parareal method for a finite difference discretization of the linear advection-diffusion equation. It is found that convergence can critically depend on the order of the interpolation used to transfer the coarse propagator solution to the fine mesh in the correction step. The effect also strongly depends on the employed spatial and temporal resolution.

1 Introduction

For the solution of time-dependent partial differential equations on parallel computers, spatial parallelization by means of mesh or domain decomposition is a standard and very successful approach. To advance the solution in time, time marching schemes like Runge-Kutta or multi-step methods are typically used that compute step after step in a serial fashion. Therefore, in a certain sense, the temporal direction constitutes a serial bottleneck: If one increases the spatial resolution for a given problem, the increase in computational cost *per time-step* can be compensated by using more processors for the spatial parallelization. However, the temporal resolution will have to be increased as well in order to maintain either accuracy or stability and therefore also *more* time-steps will be necessary. The resulting increase in computational cost by having to compute a larger number of steps cannot be mitigated by parallelization in space alone.

Time-parallel methods are parallel methods for the solution of initial value problems that feature concurrency along the temporal axis. A widely studied example is Parareal, introduced in [1] and comprehensively analyzed mathematically in [2]. Another approach is the "parallel full approximation scheme in space and time" (PFASST), introduced in [3]. PFASST has been shown to be able to provide significant additional speedup beyond the saturation point of spatial parallelization in runs using $262{,}144$ cores [4].

The idea of both Parareal and PFASST is to introduce a two-level (or even multi-level) hierarchy of decreasing accuracy and cost and to "shift" the serial dependency in time to the coarsest and therefore cheapest level. Accuracy is achieved by performing a number of iterations and successively correcting the solution. Parareal does this by means of two classical integration methods called the "fine" and the "coarse" integrator. The coarse method will typically be of lower order and use a much larger time-step than the fine. The achievable speedup depends on the one hand on the runtime ratio of the coarse to the fine level and on the other hand on the number of iterations required for convergence.

For time-dependent PDEs, it is possible to also reduce the accuracy of the spatial discretization on the coarse level. For multi-level SDC, a method closely related to PFASST, a number of such approaches have been studied in [5]. For Parareal, a lower order finite difference discretization for the coarse method is used e.g. in [6]. The possibility to use a coarsened spatial mesh in Parareal is studied in [7] in the context of finite elements. Transfer of the solution from the coarse to the fine mesh is done by interpolation as well as by L^2 projection. While projection worked well, simple interpolation is found to cause stability problems.

Here, the approach of using a lower spatial resolution for the coarse method in Parareal is studied for a finite difference discretization of the advection-diffusion equation and the effect of the order of the interpolation is investigated. Furthermore, two different ways to track convergence are briefly compared, one based on monitoring the norm of corrections from one iteration to the next, the other on a residual defined by interpreting Parareal as a preconditioned fixed point iteration.

2 Parareal

Consider an initial value problem

$$y_t(t) = f(y(t), t), \quad y(0) = b \in \mathbb{R}^D, \quad 0 \leq t \leq T. \tag{1}$$

Now let $t_0 := 0 < t_1 < \ldots < t_N = T$ be a decomposition of $[0, T]$ into N so-called time-slices $[t_n, t_{n-1}]$, $n = 1, \ldots, N$ of equal length (for the sake of simplicity). Further, let δt and Δt be a fine and coarse time-step size such that each time-slice consists of an integer number of δt or Δt steps. Denote by $\mathcal{F}_{\delta t}$ an accurate integration method, e.g. a higher-order Runge-Kutta method, using a time-step δt. Then, integrating (1) over all time-slices serially corresponds to evaluating

$$y_{n+1} = \mathcal{F}_{\delta t}(y_n, t_{n+1}, t_n),\ n = 0, \ldots, N-1. \tag{2}$$

* Corresponding author. Email: daniel.ruprecht@usi.ch

step-by-step with $y_0 := b$. Parareal replaces (2) by an iteration

$$y_{n+1}^{k+1} = \mathcal{G}_{\Delta t}(y_n^{k+1}, t_{n+1}, t_n) + \mathcal{F}_{\delta t}(y_n^k, t_{n+1}, t_n) - \mathcal{G}_{\Delta t}(y_n^k, t_{n+1}, t_n), \tag{3}$$

where $\mathcal{G}_{\Delta t}$ is a coarser method of typically lower order and with time-step $\Delta t \gg \delta t$. Note that in (3), once the y_n^k are known, the evaluation of the fine method can be parallelized. When a coarsened spatial mesh is to be used for $\mathcal{G}_{\Delta t}$, the iteration becomes

$$\mathbf{y}_{n+1}^{k+1} = \mathbf{I}\mathcal{G}_{\Delta t}(\mathbf{R}y_n^{k+1}) + \mathcal{F}_{\delta t}(y_n^k) - \mathbf{I}\mathcal{G}_{\Delta t}(\mathbf{R}y_n^k) \tag{4}$$

see [7], where $\mathbf{R}$ and $\mathbf{I}$ denote interpolation and restriction operators between the two spatial meshes. In the example below, aligned meshes are used where the coarse mesh consists simply of every second point of the fine one. Lagrangian interpolation of orders $p = 1, 3, 5, 7$ is used for $\mathbf{I}$ while $\mathbf{R}$ is always simple injection. However, using $\mathbf{R} := c\mathbf{I}^{\mathrm{T}}$ with normalization c such that row sums are unity did not much change the results.

2.1 Convergence criteria

The quality of the solution provided by Parareal can be measured by the defect at the end of each time-slice between the parallel solution and the solution provided by running $\mathcal{F}_{\delta t}$ in serial

$$d_n^k := y_n^k - y_n,\ n = 1, \ldots, N. \tag{5}$$

Once $\left\|d_n^k\right\|$ for all n is smaller than the (estimated) discretization error of $\mathcal{F}_{\delta t}$, both solutions will typically be of comparable accuracy. All values reported here use the maximum norm $\|\cdot\|_\infty$. An easy way to track convergence of Parareal without needing the serial reference solution is to compute the norm of the corrections

$$c_n^k := y_n^k - y_n^{k-1},\ n = 1, \ldots, N \tag{6}$$

and iterate until the maximum $\left\|c_n^k\right\|$ is below a predefined threshold. As $c_n^k \to 0$, the $\mathcal{G}$-terms in (3) will cancel out and we get $y_{n+1}^{k+1} = \mathcal{F}(y_n^k)$ for $n = 1, \ldots, N$, that is the fine serial solution. As shown below, the corrections can give a good estimate of the defect.

A different estimate can be derived by considering Parareal as a fixed point iteration, see e.g. [8] or [9]. For a linear problem, the propagators $\mathcal{F}$ and $\mathcal{G}$ can be written as matrices $G, F \in \mathbb{R}^{D\times D}$. Then, integrating with the fine or coarse method over all time-slices can be written as size $(N+1)D \times (N+1)D$ matrices

$$\mathbf{M}_f = \begin{bmatrix} I & \cdots & & \\ -F & I & & \\ & \ddots & \ddots & \\ & & -F & I \end{bmatrix}, \quad \mathbf{M}_g = \begin{bmatrix} I & \cdots & & \\ -G & I & & \\ & \ddots & \ddots & \\ & & -G & I \end{bmatrix}. \tag{7}$$

Therefore, serially computing the fine solution is equivalent to a block-wise solution of $\mathbf{M}_{\mathrm{f}}\mathbf{y} = \mathbf{b}$ where $\mathbf{y} = (y_0, \ldots, y_N)^{\mathrm{T}}$ and $\mathbf{b} = (b, 0, \ldots, 0)^{\mathrm{T}}$. The Parareal iteration (3) can now be written as the preconditioned fixed point iteration

$$\mathbf{M}_g\mathbf{y}^{k+1} = (\mathbf{M}_g - \mathbf{M}_f)\,\mathbf{y}^k + \mathbf{b}, \tag{8}$$

iteratively computing the solution of $\mathbf{M}_{\mathrm{f}}\mathbf{y} = \mathbf{b}$. Inverting $\mathbf{M}_g$ by block-wise elimination corresponds to a serial run of the coarse propagator. The proper residual for Parareal is therefore

$$\mathbf{r}^k := \mathbf{b} - \mathbf{M}_f\mathbf{y}^k \quad \text{or component wise} \quad r_0^k = 0,\ r_n^k := Fy_{n-1}^k - y_n^k,\ n = 1, \ldots, N. \tag{9}$$

It is shown below that $\left\|\mathbf{r}^k\right\|_\infty$ can give a slightly more accurate estimate of the norm of the defect than $\left\|\mathbf{c}^k\right\|_\infty$. The drawback of (9) is that in order to compute r_n^k, the fine propagator has to be run first to provide Fy_n^k, while the correction $c_n^k := y_n^k - y_n^{k-1}$ is available directly after y_n^k has been computed.

3 Numerical results

The benchmark problem here is the one-dimensional linear advection-diffusion equation

$$u_t(x,t) + au_x(x,t) = \nu u_{xx}(x,t), \quad u(x,0) = u_0(x) = \sin(2\pi x) \tag{10}$$

on $[0, 1]$ with periodic boundary conditions. Both gradient and Laplacian are discretized using second-order centered finite differences. A second-order implicit trapezoidal rule is used for $\mathcal{F}_{\delta t}$ and a first order implicit Euler for $\mathcal{G}_{\Delta t}$. Parameters for

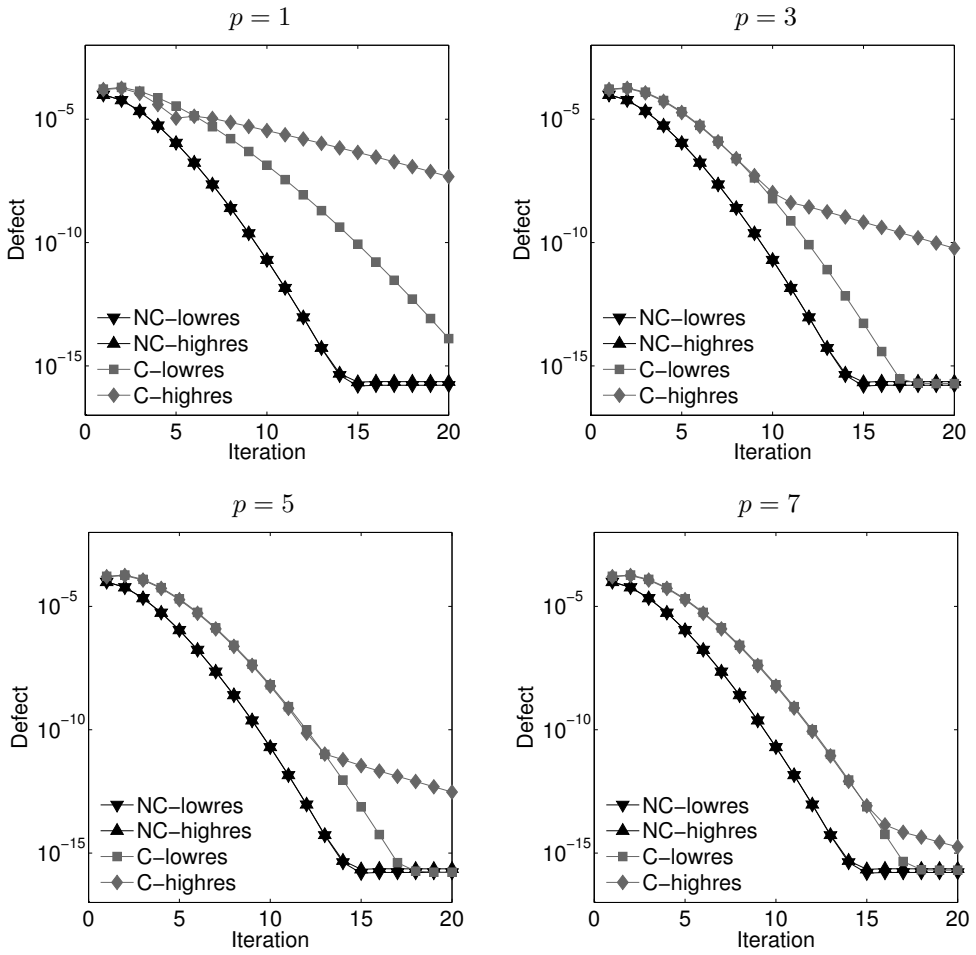

Fig. 1 Maximum norm of defect $\left\|\mathbf{d}^k\right\|$ versus k. NC indicates that no spatial coarsening is used, C indicates spatial coarsening. In addition, *lowres* means a spatial resolution of $N_x = 40$ on the fine level while *highres* means $N_x = 80$. The order of the interpolation operator $\mathbf{I}$ is indicated by p. While for Parareal without coarsening (NC) convergence is identical for *lowres* and *highres* Parareal with spatial coarsening (C) shows significant differences in convergence behavior depending on resolution and interpolation order.

advection and diffusion are $a = \nu = 0.25$ and the problem is integrated until $T = 1$. The interval $[0, 1]$ is decomposed into $P = 40$ time-slices of length 0.025. The coarse time-step is $\Delta t = 0.025$ (that is, one step per time-slice), the fine time-step $\delta t = 0.005$. Two setups are used, a low resolution run using $N_x = 40$ nodes in space (labeled *lowres*) and a high resolution run with $N_x = 80$ (labeled *highres*). The resulting CFL numbers on the fine level for diffusion and advection are

$$C_{\text{diff}}^{\text{lowres}} = 2,\ C_{\text{adv}}^{\text{lowres}} = 0.05 \quad \text{and} \quad C_{\text{diff}}^{\text{highres}} = 8,\ C_{\text{adv}}^{\text{highres}} = 0.1. \tag{11}$$

Each setup is run once with no spatial coarsening (NC) for reference and then with spatial coarsening (C) for interpolation of order $p = 1, 3, 5, 7$. Figure 1 shows the maximum norm of the defect, that is $\left\|\mathbf{d}^k\right\|_\infty$ versus the iteration number k for the different runs. A number of observations can be made:

1. Without coarsening, convergence of Parareal does not seem to be affected by the spatial resolution. The lines for NC-*lowres* and NC-*highres* essentially coincide.

2. The runs using spatial coarsening (C-*lowres* and C-*highres*) always converge slower than the NC runs, since the coarse solver $\mathcal{G}$ is less accurate.

3. For the $N_x = 40$ runs with coarsening (C-*lowres*), linear interpolation ($p = 1$) reduces convergence speed somewhat compared to higher order interpolation ($p = 3, 5, 7$). However, for $p \geq 3$ there is essentially no difference.

4. The high-resolution runs with coarsening (C-*highres*) show a strong dependence on the order of interpolation: Depending on p, there is a defect level after which convergence becomes significantly slower. The higher p is chosen, the lower this level becomes. For $p = 7$, convergence is again essentially the same for C-*highres* and C-*lowres*, up to very small values of $\left\|\mathbf{d}^k\right\| \sim 10^{-13}$. Because the effect is much more pronounced for the *highres* case and can be counteracted to some extend by using a smaller δt (see comment below), it seems to be linked to the CFL number.

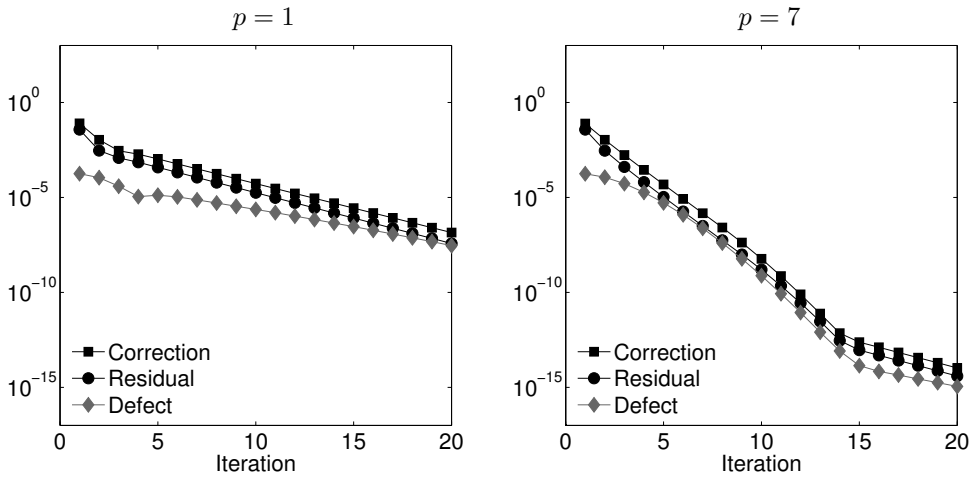

Fig. 2 Maximum norm of defect $\left\|\mathbf{d}^k\right\|_\infty$, see (5), correction $\left\|\mathbf{c}^k\right\|_\infty$, see (6), and residual $\left\|\mathbf{r}^k\right\|_\infty$, see (9), versus k for interpolation operators $\mathbf{I}$ of order $p = 1$ (left) and $p = 7$ (right) for the C-*highres* run. Note that the norm of the residual gives a slightly better estimate for the defect than the norm of the corrections and how the high order interpolation leads to a massive improvement in convergence.

These results cover only a very small range of possible parameters and thus cannot offer a systematic study of how spatial and temporal resolution affect the impact of spatial coarsening. Further tests not documented here suggest that e.g. a smaller δt reduces the impact of the interpolation order. For $\delta t = 0.001$, an interpolation order of $p = 3$ suffices to eliminate the difference in convergence between C-*lowres* and C-*highres*. On the other hand, the order of the spatial discretization seems to have no big effect: If one uses fourth-order centered differences, the results are very similar to what is shown in Figure 1. Although preliminary, the results strongly suggest that applying spatial coarsening in Parareal might not be straightforward and that higher-order interpolation might be required to achieve good convergence, at least for some setups. It is interesting to note that the necessity for higher-order interpolation for MLSDC was also recognized in [5].

Figure 2 shows again the maximum norm of the defect for the C-*highres* run versus the iteration number for $p = 1$ (left) and $p = 7$ (right). In addition, it also shows the norm of the residual $\left\|\mathbf{r}^k\right\|_\infty$ as defined in (9) and the iterative correction $\left\|\mathbf{c}^k\right\|_\infty$. Comparing the left and right figure shows again the drastic improvement in convergence by using a high-order interpolation. Furthermore, in both cases, residual as well as correction provide a good estimate of the actual defect, with the residual giving slightly more accurate values.

4 Conclusion

The paper investigates how using a reduced spatial resolution in the coarse method in Parareal affects convergence. It is found that the order of the employed interpolation can have a significant influence. The effect however strongly depends on the used resolutions in both space and time. An in-depth investigation including a detailed explanation is left for future work.

Acknowledgements This work was supported by the Swiss National Science Foundation (SNSF) under the lead agency agreement through the project "ExaSolvers" within the Priority Programme 1648 “Software for Exascale Computing” (SPPEXA) of the Deutsche Forschungsgemeinschaft (DFG).

References

[1] J. L. Lions, Y. Maday, and G. Turinici, Comptes Rendus de l’Académie des Sciences - Series I - Mathematics **332**, 661–668 (2001).
[2] M. J. Gander and S. Vandewalle, SIAM Journal on Scientific Computing **29**(2), 556–578 (2007).
[3] M. Emmett and M. L. Minion, Communications in Applied Mathematics and Computational Science **7**, 105–132 (2012).
[4] R. Speck, D. Ruprecht, R. Krause, M. Emmett, M. Minion, M. Winkel, and P. Gibbon, A massively space-time parallel N-body solver, in: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, , SC ’12 (IEEE Computer Society Press, Los Alamitos, CA, USA, 2012), pp. 92:1–92:11.
[5] R. Speck, D. Ruprecht, M. Emmett, M. Minion, M. Bolten, and R. Krause, A multi-level spectral deferred correction method, arXiv:1307.1312 [math.NA], 2013.
[6] D. Ruprecht and R. Krause, Computers & Fluids **59**(0), 72 – 83 (2012).
[7] P. F. Fischer, F. Hecht, and Y. Maday, A parareal in time semi-implicit approximation of the Navier-Stokes equations, in: Domain Decomposition Methods in Science and Engineering, edited by R. Kornhuber and et al., Lecture Notes in Computational Science and Engineering Vol. 40 (Springer, Berlin, 2005), pp. 433–440.
[8] Y. Maday and G. Turinici, Comptes Rendus Mathématique **335**(4), 387 – 392 (2002).
[9] P. Amodio and L. Brugnano, Applied Numerical Mathematics **59**(3-4), 424 – 435 (2009).