

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/176760>

How to cite:

The repository item page linked to above, will contain details on accessing citation guidance from the publisher.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>.



Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Error bound analysis of the stochastic parareal algorithm

Kamran Pentland¹ Massimiliano Tamborrino² T. J. Sullivan^{1,3}

March 13, 2023

Abstract. Stochastic parareal (SParareal) is a probabilistic variant of the popular parallel-in-time algorithm known as parareal. Similarly to parareal, it combines fine- and coarse-grained solutions to an ordinary differential equation (ODE) using a predictor-corrector (PC) scheme. The key difference is that carefully chosen random perturbations are added to the PC to try to accelerate the location of a stochastic solution to the ODE. In this paper, we derive superlinear and linear mean-square error bounds for SParareal applied to nonlinear systems of ODEs using different types of perturbations. We illustrate these bounds numerically on a linear system of ODEs and a scalar nonlinear ODE, showing a good match between theory and numerics.

Keywords. Parallel-in-time • stochastic parareal • error bounds • ordinary differential equations

2020 Mathematics Subject Classification. 65L70 • 65Y05 • 65C99

1. Introduction

In recent years, there has been an increasing need to develop faster numerical integration methods for a broad range of time-dependent ordinary, partial, and stochastic differential equations (ODEs/PDEs/SDEs) that form the building blocks of mathematical models of real-world systems. Advances in high performance computing have fuelled the development of new parallel algorithms for solving such initial value problems (IVPs), with one particular area of increasing focus being parallel-in-time (PinT) methods¹. Given the causal nature of time, PinT algorithms provide a way to, either iteratively or directly, simulate the solution states of an IVP across the entire time interval of integration concurrently. Since the seminal work of Nievergelt (1964), a variety of different PinT methods have been developed over the last 60 years—refer to Gander (2015) and Ong and Schroder (2020) for reviews.

Of particular interest here is the increasingly popular multiple shooting-type/multigrid PinT algorithm known as parareal (Lions et al., 2001). Dividing the time interval into N ‘slices’, typically of fixed size ΔT , parareal iteratively combines solutions calculated by two serial numerical integrators (one coarse- and one fine-grained) on each slice using a predictor-corrector

¹ Department of Mathematics, University of Warwick, Coventry, CV4 7AL, United Kingdom
(kamran.pentland@warwick.ac.uk, t.j.sullivan@warwick.ac.uk)

² Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom
(massimiliano.tamborrino@warwick.ac.uk)

³ Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom

¹A collection of PinT research papers and software can found at <https://parallel-in-time.org/>.

(PC) scheme. Using N processors, one for each time slice, parareal runs the fine (expensive) computations in parallel, locating a solution to the IVP in a fixed number of iterations $k \leq N$, after which one says the algorithm has ‘converged’. Parallel speedup from parareal is bounded above by N/k and so, ideally, the algorithm finds a solution in $k \ll N$ iterations. Since it was proposed, parareal has been shown to provide speedup for a range of different IVPs in areas ranging from molecular dynamics (Engblom, 2009; Legoll et al., 2020, 2022) to plasma physics (Grigori et al., 2021; Samaddar et al., 2010, 2019)—refer to (Ong and Schroder, 2020, Sec. 2) for an overview. Variants have also been developed to tackle issues with Hamiltonian systems (Bal and Wu, 2008; Dai et al., 2013), task scheduling (Elwasif et al., 2011), and IVP stiffness (Maday and Mula, 2020). Others, aimed at reducing the number iterations in parareal, have also emerged. Approaches include learning the correction term in the PC using Gaussian process emulators (Pentland et al., 2022b) and building the coarse solver using a Krylov subspace of the set of PC solutions (Gander and Petcu, 2008).

The focus of this paper is the *stochastic parareal* algorithm (henceforth SParareal), developed by Pentland et al. (2022a), in which random perturbations are added to the PC to increase the probability of locating the solution to the IVP in fewer iterations. In each time slice, M samples are taken from probability distributions constructed using different combinations of coarse and fine solution information (known as “sampling rules”) and propagated forward in time in parallel. From this set of perturbed solutions, the ones which generate the most continuous (fine) trajectory in solution space, starting from the exact known initial condition, are selected. Numerical simulations showed that as the number of samples increased, the probability of locating the exact solution increased and so convergence would occur in fewer iterations and therefore faster wallclock time. Numerical simulations also showed that, upon multiple simulations of SParareal, one could obtain a distribution of solutions to the IVP which were numerically accurate with respect to the serially located fine solution. Conrad et al. (2017) and Lie et al. (2019, 2022) developed similar (non-time-parallel) numerical schemes in which they perturb the solution states generated by serial numerical integrators to incorporate a measure of uncertainty quantification in the solution of ODEs.

1.1. Contribution and outline

Deriving rigorous error bounds for SParareal (and other PinT methods in general) is important in demonstrating that numerical solutions obtained in parallel are meaningful, accurate, and that they can be compared to one another (Gander et al., 2022). The first convergence results for parareal assumed that k was fixed, showing that the error of the scheme approached zero as the time slice size $\Delta T \rightarrow 0$ (Bal, 2005; Bal and Maday, 2002; Lions et al., 2001). Here, we are interested in studying how numerical errors behave when ΔT is fixed and the number of iterations k grows. Gander and Vandewalle (2007) investigated this, deriving both linear and superlinear error bounds for the scalar linear ODE problem on unbounded and bounded time intervals, respectively. Following this, Gander and Hairer (2008) used the generating function method to derive a superlinear bound for nonlinear systems of ODEs on bounded intervals. Whereas work has been done to derive error bounds for parareal applied to certain SDEs (Bal, 2006; Engblom, 2009), where the IVP itself contains randomness, those results cannot be applied here, as it is the SParareal scheme itself that contains randomness, not the solvers or the IVP.

In this paper, we extend the qualitative discussion of numerical convergence in Pentland et al. (2022a) by making use of convergence analysis from both the parareal literature (see above) and the sampling-based ODE solver presented in Lie et al. (2019). We derive explicit mean-square error bounds for SParareal applied to nonlinear systems of ODEs (over a finite time interval), using both *state-independent* and *state-dependent* perturbations. The rest of the

paper is organised as follows. In [Section 2](#), we set up the IVP and provide an overview of the (classic) parareal algorithm. We then detail the SParareal scheme in [Section 3](#), outlining how it works and defining the (state-dependent) sampling rules. In [Section 4](#), we begin by outlining the assumptions on the fine and coarse integrators required to derive the error bounds. We first consider the *state-independent* setting, where the random perturbations do not depend on the current time step or iteration and are assumed to have bounded absolute moments. In this setting we derive our main result ([Theorem 4.6](#)), a superlinear bound on the mean-square error depending on the time step and iteration. Using this result, we can maximise the error over time (fixing the iteration), to derive a linear error bound ([Corollary 4.7](#)). In the *state-dependent* setting, we allow the perturbations to depend on the current state of the system, i.e. the coarse and fine solution information at the current time step and iteration, to analyse the convergence of the sampling rules proposed in the original work. We derive linear bounds ([Corollaries 4.13](#) and [4.14](#)) in this case. Following this, we verify these bounds by applying them numerically to a linear system of ODEs and a nonlinear scalar ODE in [Section 5](#). We conclude with some brief remarks on the theoretical and numerical results and discuss possible directions for future study in [Section 6](#). In the Appendix, we provide some technical results used to derive the aforementioned error bounds.

1.2. Notation

Let \mathbb{E} denote the expected value of a random variable. Variables $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ will be d -dimensional real-valued vectors unless otherwise stated. Denote the component-wise absolute value of a vector as $|\mathbf{u}| = (|u_1|, \dots, |u_d|)^\top$ and the Hadamard (component-wise) product as $\mathbf{u} \circ \mathbf{v} = (u_1 v_1, \dots, u_d v_d)^\top$. Also note that \mathbf{u}^2 will correspond to component-wise squaring, i.e. $\mathbf{u}^2 = \mathbf{u} \circ \mathbf{u}$. We let $\|\mathbf{u}\|$ denote the infinity (or uniform) norm, i.e. $\|\mathbf{u}\| = \max_{i=1, \dots, d} |u_i|$. The d -dimensional vector of ones and the identity matrix will be written as $\mathbf{1}$ and I_d , respectively. Non-negative constants will be denoted throughout by C_1, C_2, \dots .

2. The parareal scheme

In this section, we describe the IVP under consideration and define the classic parareal scheme.

2.1. Problem setup

Consider the following system of $d \in \mathbb{N}$ autonomous ODEs

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(\mathbf{u}(t)) \quad \text{over } t \in [t_0, T], \quad \text{with } \mathbf{u}(t_0) = \mathbf{u}_0, \quad (2.1)$$

where $\mathbf{f}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a nonlinear vector field, $\mathbf{u}: [t_0, T] \rightarrow \mathbb{R}^d$ is the time-dependent solution, and $\mathbf{u}_0 \in \mathbb{R}^d$ is the initial value at time t_0 . We assume \mathbf{f} is sufficiently smooth such that [\(2.1\)](#) has a unique solution for all initial conditions of interest. We seek numerical solutions $\mathbf{U}_n \approx \mathbf{u}(t_n)$ to the IVP [\(2.1\)](#) on a pre-defined mesh $\mathbf{t} = (t_0, \dots, t_N)$, where $t_n = t_0 + n\Delta T$ for fixed $\Delta T = (T - t_0)/N$. N is the number of processors required for parareal to compute a solution in parallel, i.e. one processor is assigned to each time slice $[t_n, t_{n+1}]$, $n = 0, \dots, N - 1$. Note that everything that follows extends to the nonautonomous case but is not discussed here to simplify explanation and notation.

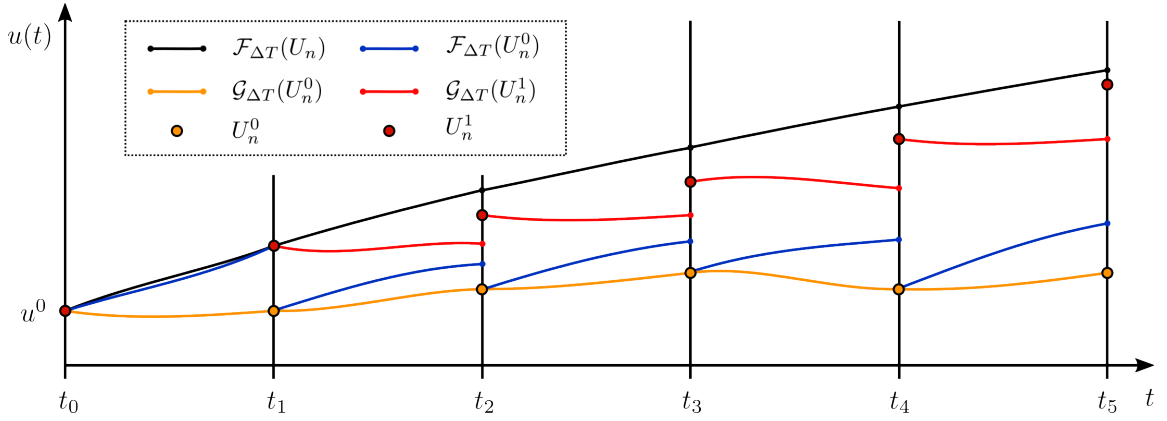


Figure 2.1. First iteration of parareal that numerically approximates the exact solution (black line), obtained via repeated serial applications of the fine solver, i.e. (2.2) in the scalar case. The coarse initial guess found using $\mathcal{G}_{\Delta T}$ (yellow lines and dots) is followed by the parallel runs of $\mathcal{F}_{\Delta T}$ from these guesses (blue lines). The coarse predictions from $\mathcal{G}_{\Delta T}$ (red lines) are then used in the prediction-correction update (2.5) (red dots). Figure adapted from Pentland et al. (2022a) (Fig. 2).

2.2. The scheme

The parareal algorithm uses N processors and two deterministic numerical flow maps (solvers) to locate solutions \mathbf{U}_n^k , at iteration k and time t_n , to system (2.1). These maps take an initial state \mathbf{U}_n^k at time t_n and propagate it, over a time slice of size ΔT , to a terminal state \mathbf{U}_{n+1}^k at time t_{n+1} . The *fine* solver, denoted as the flow map $\mathcal{F}_{\Delta T}: \mathbb{R}^d \rightarrow \mathbb{R}^d$, returns a terminal state with high numerical accuracy at very high computational cost. The cost is high enough that trying to use $\mathcal{F}_{\Delta T}$ to solve (2.1) sequentially, i.e. calculate

$$\mathbf{U}_{n+1} = \mathcal{F}_{\Delta T}(\mathbf{U}_n) \quad \text{for } n = 0, \dots, N-1, \quad (2.2)$$

is computationally infeasible. The *coarse* solver, denoted similarly by $\mathcal{G}_{\Delta T}: \mathbb{R}^d \rightarrow \mathbb{R}^d$, returns a terminal state with a much lower numerical accuracy, at a smaller computational cost than $\mathcal{F}_{\Delta T}$. Therefore, $\mathcal{G}_{\Delta T}$ is fast and allowed to run serially at any time, whilst $\mathcal{F}_{\Delta T}$ is expensive and must only be run in parallel.

Definition 2.1 (Parareal). For the two numerical flow maps $\mathcal{F}_{\Delta T}$ and $\mathcal{G}_{\Delta T}$ described above, the parareal scheme is given by

$$\mathbf{U}_0^0 = \mathbf{u}_0, \quad (2.3)$$

$$\mathbf{U}_{n+1}^0 = \mathcal{G}_{\Delta T}(\mathbf{U}_n^0), \quad 0 \leq n \leq N-1, \quad (2.4)$$

$$\mathbf{U}_{n+1}^{k+1} = \mathcal{G}_{\Delta T}(\mathbf{U}_n^{k+1}) + \mathcal{F}_{\Delta T}(\mathbf{U}_n^k) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^k), \quad 0 \leq k \leq n \leq N-1. \quad (2.5)$$

Parareal begins by propagating the exact initial condition (2.3) across $[t_0, t_N]$ serially using $\mathcal{G}_{\Delta T}$ (2.4). Each of the solution states found in (2.4) are then propagated in parallel using $\mathcal{F}_{\Delta T}$. Equation (2.5), often referred to as the PC, is then run serially to update the solution states at each t_n . This process of propagating using $\mathcal{F}_{\Delta T}$ and updating using the PC can be repeated iteratively, stopping after k iterations, once a pre-specified stopping criterion is met. The stopping criteria can take many different forms, one popular choice being that the time slices up to t_I are considered converged if

$$\|\mathbf{U}_n^k - \mathbf{U}_n^{k-1}\| < \varepsilon \quad \forall n \leq I, \quad (2.6)$$

for some user-specified tolerance $\varepsilon > 0$. Once $I = N$, parareal is said to have converged in k iterations. An illustration of the first iteration of parareal is given in [Figure 2.1](#).

Remark 2.2. It is often assumed, and will be assumed in this paper, that $\mathcal{F}_{\Delta T}$, if it were computationally feasible to apply serially over $[t_0, T]$, returns the ‘exact’ solution to (2.1), i.e. it returns $\mathbf{U}_{n+1} = \mathbf{u}(t_{n+1}) = \mathcal{F}_{\Delta T}(\mathbf{u}(t_n))$, $n = 0, \dots, N - 1$. Therefore, after k iterations, parareal returns the exact solution $\mathbf{U}_n^k = \mathbf{U}_n$ for $k \geq n$, meaning that parareal returns the exact solution over $[t_0, T]$ in at most N iterations—albeit achieving no parallel speedup in this case. This is an important feature of parareal which is also preserved by SParareal.

3. The stochastic parareal scheme

In this section, we outline the SParareal scheme, how it works, and how we have slightly modified the scheme, compared to the original formulation in [Pentland et al. \(2022a\)](#), to carry out the error bound analysis in [Section 4](#). Following this, we describe the sampling rules tested in the original work.

3.1. The scheme

The intuition behind SParareal is to perturb the solution states \mathbf{U}_n^k in the classic parareal scheme, i.e. the PC (2.5), with some additive noise to reduce the number of iterations k taken until the stopping tolerance (2.6) is met. Recall that a reduction in k by even a single iteration can correspond to a large increase in numerical speedup—by approximately the runtime of a single run of $\mathcal{F}_{\Delta T}$.

First, let us formally define the scheme.

Definition 3.1 (SParareal). For the two numerical flow maps, $\mathcal{F}_{\Delta T}$ and $\mathcal{G}_{\Delta T}$, described in [Section 2](#), the stochastic parareal scheme is given by

$$\mathbf{U}_0^0 = \mathbf{u}_0, \tag{3.1}$$

$$\mathbf{U}_{n+1}^0 = \mathcal{G}_{\Delta T}(\mathbf{U}_n^0), \quad 0 \leq n \leq N - 1, \tag{3.2}$$

$$\mathbf{U}_{n+1}^1 = \mathcal{G}_{\Delta T}(\mathbf{U}_n^1) + \mathcal{F}_{\Delta T}(\mathbf{U}_n^0) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^0), \quad 0 \leq n \leq N - 1, \tag{3.3}$$

$$\mathbf{U}_{n+1}^{k+1} = \mathcal{G}_{\Delta T}(\mathbf{U}_n^{k+1}) + \mathcal{F}_{\Delta T}(\mathbf{U}_n^k) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^k) + \boldsymbol{\xi}_n^k(\mathbf{U}_n^k), \quad 1 \leq k \leq n \leq N - 1, \tag{3.4}$$

where $\boldsymbol{\xi}_n^k(\mathbf{U}_n^k)$ are (possibly state-dependent) random variables. Note that $\boldsymbol{\xi}_n^k(\mathbf{U}_n^k) \equiv \mathbf{0}$ when $n = k$.

The first three stages of the scheme (3.1)-(3.3) are identical to the ‘zeroth’ and first iteration of parareal. The exact initial condition (3.1) is propagated forward in time using the coarse solver (3.2), then there is a first pass of the PC (3.3). Following this, the stochastic iterations begin (3.4), whereby random perturbations, i.e. a single draw from the random variable $\boldsymbol{\xi}_n^k(\mathbf{U}_n^k)$, are added to the PC solution. Notice that no random perturbation is added when $n = k$ to ensure that SParareal returns the exact solution up to time t_k after k iterations, just as parareal does, recall [Remark 2.2](#). The reason the random perturbations are only included from iteration $k \geq 1$ onward is because $\boldsymbol{\xi}_n^k(\mathbf{U}_n^k)$ may depend on solution information from iteration $k - 1$. In [Section 3.2](#), we will discuss the construction of these random variables in more detail. Note that the parareal scheme ([Definition 2.1](#)) can be recovered by setting $\boldsymbol{\xi}_n^k(\mathbf{U}_n^k) \equiv \mathbf{0}$ for $n \geq k$ in (3.4).

The scheme in [Definition 3.1](#) looks slightly different to the one presented in [Pentland et al. \(2022a\)](#), where random perturbations were instead incorporated via random variables, denoted

by α_n^k , in the correction term, i.e. equation (3.4) was instead written as

$$\mathbf{U}_{n+1}^{k+1} = \mathcal{G}_{\Delta T}(\mathbf{U}_n^{k+1}) + \mathcal{F}_{\Delta T}(\alpha_n^k) - \mathcal{G}_{\Delta T}(\alpha_n^k), \quad 1 \leq k \leq n \leq N-1. \quad (3.5)$$

The different state-dependent forms that α_n^k can take are defined through the ‘‘sampling rules’’ in Section 3.2. Also note that $\alpha_n^k = \mathbf{U}_n^k$ when $n = k$, which is equivalent to the condition that $\xi_n^k(\mathbf{U}_n^k) \equiv \mathbf{0}$ when $n = k$ in Definition 3.1. This version 3.5 of the scheme was designed so that $M \geq 1$ samples could be drawn from each of the random variables α_n^k to increase the probability of locating the exact solution state \mathbf{U}_n in fewer iterations. From the sets of samples generated at each t_n , all having been propagated in parallel using $\mathcal{F}_{\Delta T}$ and $\mathcal{G}_{\Delta T}$ (recall (3.5)), those generating the most continuous $\mathcal{F}_{\Delta T}$ trajectory across $[t_0, t_N]$ were then chosen as the ‘‘best’’ perturbations $\hat{\alpha}_n^k$. Numerical experiments illustrated that increasing M led to further and further reductions in k , albeit at the cost of requiring more processors, specifically $\mathcal{O}(MN)$ in SPareal vs $\mathcal{O}(N)$ in parareal.

To enable us to carry out the convergence analysis, we move the random perturbations in (3.5) outside the correction term, and express them using $\xi_n^k(\mathbf{U}_n^k)$ in (3.4). This new scheme is equivalent to the old scheme in the case where one sample ($M = 1$) is drawn at each time t_n . To obtain the appropriate values for $\xi_n^k(\mathbf{U}_n^k)$, we equate (3.4) and (3.5) to find

$$\xi_n^k(\mathbf{U}_n^k) = (\mathcal{F}_{\Delta T}(\alpha_n^k) - \mathcal{G}_{\Delta T}(\alpha_n^k)) - (\mathcal{F}_{\Delta T}(\mathbf{U}_n^k) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^k)). \quad (3.6)$$

Using the scheme in Definition 3.1, we can derive error bounds for state-independent perturbations, i.e. $\xi_n^k(\mathbf{U}_n^k) = \xi_n^k$, that have bounded absolute moments. Using this result, we can then derive corresponding bounds for the state-dependent sampling rules summarised in Table 3.1 using relation (3.6). All of these bounds are derived assuming one sample ($M = 1$) is drawn from each $\xi_n^k(\mathbf{U}_n^k)$. The M sample case is much more complex and out of the scope of the present work.

3.2. Sampling Rules

The sampling rules presented by Pentland et al. (2022a) describe the probability distributions that α_n^k follow in the SPareal algorithm, see Table 3.1. These distributions were designed to vary with both iteration k and time step n , so that as the solution states \mathbf{U}_n^k get closer to \mathbf{U}_n , their variances would decrease. The benefit of this property will be highlighted in Section 5. The different rules were constructed to assess the performance of SPareal when the perturbations had different distribution families, marginal means, or correlations.

Table 3.1. Sampling rules that the random variables α_n^k follow. The quantities $\mathbf{z}_n^k \sim \mathcal{N}(\mathbf{0}, I_d)$ and $\mathbf{w}_n^k \sim \mathcal{U}([0, 1]^d)$ are d -dimensional Gaussian and uniform random vectors, respectively, whilst $\sigma_n^k = |\mathcal{G}_{\Delta T}(\mathbf{U}_{n-1}^k) - \mathcal{G}_{\Delta T}(\mathbf{U}_{n-1}^{k-1})|$.

Sampling rule	α_n^k
1	$\mathcal{F}_{\Delta T}(\mathbf{U}_{n-1}^{k-1}) + (\sigma_n^k \circ \mathbf{z}_n^k)$
2	$\mathbf{U}_n^k + (\sigma_n^k \circ \mathbf{z}_n^k)$
3	$\mathcal{F}_{\Delta T}(\mathbf{U}_{n-1}^{k-1}) + (\sqrt{3}\sigma_n^k \circ (2\mathbf{w}_n^k - \mathbf{1}))$
4	$\mathbf{U}_n^k + (\sqrt{3}\sigma_n^k \circ (2\mathbf{w}_n^k - \mathbf{1}))$

Sampling rules 1 and 2 correspond to multivariate Gaussian perturbations with marginal means $\mathcal{F}_{\Delta T}(\mathbf{U}_{n-1}^{k-1})$ and \mathbf{U}_n^k , respectively, and marginal standard deviations $\sigma_n^k = |\mathcal{G}_{\Delta T}(\mathbf{U}_{n-1}^k) -$

$\mathcal{G}_{\Delta T}(\mathbf{U}_{n-1}^{k-1})$. The variable $\mathbf{z}_n^k \sim \mathcal{N}(\mathbf{0}, I_d)$ is a standard d -dimensional Gaussian random vector. Sampling rules 3 and 4 correspond to perturbations following a multivariate uniform distribution with the same marginal means and standard deviations as rules 1 and 2, respectively. The variable $\mathbf{w}_n^k \sim \mathcal{U}([0, 1]^d)$ is a standard d -dimensional uniformly distributed random vector with independent components. Note that [Pentland et al. \(2022a\)](#) considered both correlated and uncorrelated random variables $\boldsymbol{\alpha}_n^k$ in their experiments, whereas here we carry out our analysis only for the uncorrelated case for simplicity.

4. Error bound analysis

In this section, we analyse the mean-square error

$$e_n^k := \mathbb{E}[\|\mathbf{u}(t_n) - \mathbf{U}_n^k\|^2] \quad (4.1)$$

between the exact solutions $\mathbf{u}(t_n)$ and the stochastic numerical solutions \mathbf{U}_n^k located by the SParareal scheme. We also define the maximal mean-square error (over time) at iteration k to be

$$\hat{e}^k := \max_{1 \leq n \leq N} \{e_n^k\}. \quad (4.2)$$

Specifically, we analyse the mean-square error e_n^k for the nonlinear autonomous system of ODEs in (2.1), first deriving superlinear ([Theorem 4.6](#)) and linear ([Corollary 4.7](#)) bounds using state-independent perturbations in SParareal. Then, using these results, we derive linear bounds for the state-dependent sampling rules 2 and 4 ([Corollary 4.13](#)) and 1 and 3 ([Corollary 4.14](#)). In the following, we introduce some assumptions on the flow maps ([Gander and Hairer, 2008](#)) and perturbations ([Lie et al., 2019](#)) required to derive the error bounds.

Assumption 4.1 (Exact flow map). The flow map $\mathcal{F}_{\Delta T}$ solves (2.1) exactly such that

$$\mathbf{u}(t_{n+1}) = \mathcal{F}_{\Delta T}(\mathbf{u}(t_n)). \quad (4.3)$$

This assumption is made for simplicity, since SParareal is essentially trying to locate the solution that would be obtained by running the fine solver serially, i.e. (2.2), in parallel. If instead we were to consider $\mathcal{F}_{\Delta T}$ to be a numerical method with some (very small) numerical error with respect to the exact solution, then the accuracy of $\mathcal{F}_{\Delta T}$ would provide a lower bound on the accuracy of the SParareal scheme as a whole.

Assumption 4.2 (Coarse flow map). The flow map $\mathcal{G}_{\Delta T}$ is a one-step numerical method with uniform local truncation error $\mathcal{O}(\Delta T^{p+1})$, for $p \geq 1$, such that

$$\mathcal{F}_{\Delta T}(\mathbf{u}) - \mathcal{G}_{\Delta T}(\mathbf{u}) = c_1(\mathbf{u})\Delta T^{p+1} + c_2(\mathbf{u})\Delta T^{p+2} + \dots, \quad (4.4)$$

for $\mathbf{u} \in \mathbb{R}^d$ and continuously differentiable functions $c_i(\mathbf{u})$. Taking the difference of (4.4) evaluated at states $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, then applying norms and the triangle inequality, we can write

$$\|(\mathcal{F}_{\Delta T}(\mathbf{u}) - \mathcal{G}_{\Delta T}(\mathbf{u})) - (\mathcal{F}_{\Delta T}(\mathbf{v}) - \mathcal{G}_{\Delta T}(\mathbf{v}))\| \leq C_1 \Delta T^{p+1} \|\mathbf{u} - \mathbf{v}\|, \quad (4.5)$$

where $C_1 > 0$ is the Lipschitz constant for c_1 and we absorb terms $\mathcal{O}(\Delta T^{p+2})$ into C_1 .

Assumption 4.3 (Lipschitz coarse flow). The flow map $\mathcal{G}_{\Delta T}$ satisfies the Lipschitz condition

$$\|\mathcal{G}_{\Delta T}(\mathbf{u}) - \mathcal{G}_{\Delta T}(\mathbf{v})\| \leq L_G \|\mathbf{u} - \mathbf{v}\|, \quad (4.6)$$

for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ and Lipschitz constant $L_G > 0$.

Note that these assumptions do not restrict the choice of $\mathcal{G}_{\Delta T}$, as they are met when choosing any Runge-Kutta or Taylor method (Hairer et al., 1993).

In addition to assumptions on the flow maps, we require an assumption on the absolute moments of the (state-independent) random variables, which will be needed to prove [Theorem 4.6](#).

Assumption 4.4 (Bounded absolute moments of ξ_n^k). For $q \geq 0$, $\tilde{r} \in \mathbb{N} \cup \{\infty\}$, and $C_2 > 0$ independent of n , k , and ΔT , the r -th absolute moments of $\xi_n^k(\mathbf{U}_n^k)$ satisfy

$$\mathbb{E}[\|\xi_n^k\|^r] \leq (C_2 \Delta T^{q+\frac{1}{2}})^r, \quad 1 \leq r \leq \tilde{r}. \quad (4.7)$$

This assumption enables flexibility in defining the state-independent perturbations, in the sense that it does not require that the random variables be independent, identically distributed, or centred (Lie et al., 2019). It also means that each ξ_n^k could follow a different probability distribution, with the only requirement being that they share a common maximal bound on their absolute moments with respect to the norm. Note that we assume $\Delta T < 1$ without loss of generality here, so that for increasing q , the perturbations get smaller and smaller. For $\Delta T > 1$, we can simply take q to be negative.

These assumptions will enable us to derive error bounds in the state-independent and state-dependent cases (using sampling rules 2 and 4). The sampling rule 1 and 3 case requires the following as well.

Assumption 4.5 (Lipschitz exact flow). The flow map $\mathcal{F}_{\Delta T}$ satisfies the Lipschitz condition

$$\|\mathcal{F}_{\Delta T}(\mathbf{u}) - \mathcal{F}_{\Delta T}(\mathbf{v})\| \leq L_{\mathcal{F}} \|\mathbf{u} - \mathbf{v}\|, \quad (4.8)$$

for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ and constant $L_{\mathcal{F}} > 0$.

4.1. State-independent perturbations

In this section, we derive error bounds for SPareal when using the state-independent perturbations $\xi_n^k(\mathbf{U}_n^k) = \xi_n^k$.

Theorem 4.6 (Superlinear error bound for state-independent perturbations). *Suppose the SPareal scheme (3.1)-(3.4) with $\xi_n^k(\mathbf{U}_n^k) = \xi_n^k$ satisfies [Assumptions 4.1](#), [4.2](#), [4.3](#) and [4.4](#). Then, the mean-square error (4.1) of the solution to the nonlinear ODE system (2.1) at iteration k and time t_n satisfies*

$$e_n^k \leq DA^{k-1} \sum_{\ell=0}^{n-k} \binom{\ell+k-1}{\ell} B^\ell + \Lambda \sum_{j=0}^{k-2} \sum_{\ell=0}^{n-(j+1)} \binom{\ell+j}{\ell} A^j B^\ell,$$

for $2 \leq k < n \leq N$ and constants $A = C_1^2 \Delta T^{2p+2} (2 + \Delta T^{-1})$, $B = L_{\mathcal{G}}^2 (1 + 2\Delta T)$, $\Lambda = C_2^2 \Delta T^{2q+1} (2 + \Delta T^{-1})$, and $D = A\hat{e}^0$.

Proof. Using (3.4), that $\mathcal{F}_{\Delta T}$ is the exact solver (4.3), and adding and subtracting $\mathcal{G}_{\Delta T}(\mathbf{u}(t_n))$, we see that

$$\begin{aligned} e_{n+1}^{k+1} &= \mathbb{E}[\|\mathcal{F}_{\Delta T}(\mathbf{u}(t_n)) - (\mathcal{G}_{\Delta T}(\mathbf{U}_n^{k+1}) + \mathcal{F}_{\Delta T}(\mathbf{U}_n^k) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^k) + \xi_n^k(\mathbf{U}_n^k)) \pm \mathcal{G}_{\Delta T}(\mathbf{u}(t_n))\|^2] \\ &= \mathbb{E}[\|W_1 + W_2 + W_3\|^2], \end{aligned}$$

where W_1 , W_2 , and W_3 are given by

$$\begin{aligned} W_1 &= \mathcal{F}_{\Delta T}(\mathbf{u}(t_n)) - \mathcal{G}_{\Delta T}(\mathbf{u}(t_n)) - (\mathcal{F}_{\Delta T}(\mathbf{U}_n^k) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^k)), \\ W_2 &= \mathcal{G}_{\Delta T}(\mathbf{u}(t_n)) - \mathcal{G}_{\Delta T}(\mathbf{U}_n^{k+1}), \\ W_3 &= -\xi_n^k. \end{aligned}$$

Then, using the triangle inequality and (A.1) for the cross terms (Engblom, 2009, Sec. 4.2), we obtain

$$e_{n+1}^{k+1} \leq (1 + \delta_1^{-1} + \delta_2^{-1})\mathbb{E}[\|W_1\|^2] + (1 + \delta_1 + \delta_3^{-1})\mathbb{E}[\|W_2\|^2] + (1 + \delta_2 + \delta_3)\mathbb{E}[\|W_3\|^2]. \quad (4.9)$$

Using (4.5), we can bound

$$\mathbb{E}[\|W_1\|^2] \leq C_1^2 \Delta T^{2(p+1)} e_n^k. \quad (4.10)$$

Applying the Lipschitz condition (4.6), we obtain

$$\mathbb{E}[\|W_2\|^2] \leq L_G^2 e_n^{k+1}. \quad (4.11)$$

Using (4.7) with $r = 2$, we obtain

$$\mathbb{E}[\|W_3\|^2] \leq C_2^2 \Delta T^{2q+1}. \quad (4.12)$$

Plugging (4.10)-(4.12) into (4.9) and choosing $\delta_1 = \Delta T$, $\delta_2 = 1$, and $\delta_3 = \Delta T^{-1}$, we obtain the double recursion

$$e_{n+1}^{k+1} \leq A e_n^k + B e_n^{k+1} + \Lambda, \quad e_{n+1}^1 \leq D + B e_n^1, \quad (4.13)$$

where $A = C_1^2 \Delta T^{2p+2}(2 + \Delta T^{-1})$, $B = L_G^2(1 + 2\Delta T)$, $\Lambda = C_2^2 \Delta T^{2q+1}(2 + \Delta T^{-1})$, and $D = A \hat{e}^0$. Solving (4.13) using the generating function method in Lemma B.1 (see Appendix), we obtain the result. \blacksquare

One can make alternative choices for δ_1 , δ_2 , and δ_3 , however, the choices given in the proof above seem to yield the smallest error bounds. If we were to maximise (4.13) over n , we obtain the following linear error bound in the case that $B < 1$, i.e. $L_G < (1 + 2\Delta T)^{-1/2}$.

Corollary 4.7 (Linear error bound for state-independent perturbations). *Suppose the SParareal scheme (3.1)-(3.4) with $\xi_n^k(\mathbf{U}_n^k) = \xi_n^k$ satisfies Assumptions 4.1, 4.2, 4.3 and 4.4. Then, the maximal mean-square error (4.2) of the solution to the nonlinear ODE system (2.1) at iteration k satisfies*

$$\hat{e}^k \leq \hat{e}^1 \left(\frac{A}{1-B} \right)^{k-1} + \frac{\Lambda}{1-B} \sum_{j=0}^{k-2} \left(\frac{A}{1-B} \right)^j, \quad \text{if } B < 1,$$

for $2 \leq k \leq N$ and constants $A = C_1^2 \Delta T^{2p+2}(2 + \Delta T^{-1})$, $B = L_G^2(1 + 2\Delta T)$, and $\Lambda = C_2^2 \Delta T^{2q+1}(2 + \Delta T^{-1})$.

Proof. Following the proof of Theorem 4.6, we maximise (4.13) over n to obtain

$$\hat{e}^{k+1} \leq \tilde{A} \hat{e}^k + \tilde{\Lambda}, \quad (4.14)$$

where

$$\tilde{A} = \frac{A}{1-B} \quad \text{and} \quad \tilde{\Lambda} = \frac{\Lambda}{1-B}.$$

Solving recursion (4.14) with initial condition \hat{e}^1 , we obtain the desired result. \blacksquare

Remark 4.8. The bounds in [Theorem 4.6](#) and [Corollary 4.7](#) hold for $2 \leq k < n \leq N$ due to the design of the SParareal scheme. We can recover the bound for iteration $k = 1$ (which is deterministic) by solving the second recursion in [\(4.13\)](#) with initial value $e_1^1 = 0$ such that

$$e_n^1 \leq \hat{e}^0 A \sum_{i=0}^{n-2} B^i, \quad 1 \leq n \leq N. \quad (4.15)$$

For the case when $k = n$, the numerical error is zero, as $\mathcal{F}_{\Delta T}$ will have propagated the exact initial condition at t_0 forward k times without any perturbations, just like parareal (recall [Remark 2.2](#)).

Remark 4.9. The bound in [Theorem 4.6](#) (similarly for [Corollary 4.7](#)) can be written as

$$e_n^k \leq C_{k,n} \max\{\Delta T^{(2p+1)k}, \Delta T^{2q}\}, \quad (4.16)$$

where $C_{k,n}$ is a function of n and k . Assuming $\Delta T < 1$ and that $C_{k,n}$ is non-increasing in k , the accuracy of SParareal should increase with each iteration proportional to the local truncation error of \mathcal{G} (i.e. the term $\Delta T^{(2p+1)k}$) up until the errors induced by the perturbations (i.e. ΔT^{2q}) become dominant. We illustrate this property numerically in [Section 5](#).

Remark 4.10. As $\Delta T \rightarrow 0$, both error bounds go to zero as expected, as can be seen clearly in [\(4.16\)](#). The intuition being that as $\Delta T \rightarrow 0$, the local truncation error of $\mathcal{G}_{\Delta T}$ goes to zero, i.e. it ‘becomes’ the exact flow map $\mathcal{F}_{\Delta T}$, see [\(4.5\)](#).

Remark 4.11. If we send $q \rightarrow \infty$ (assuming $\Delta T < 1$), the second moments of the random variables vanish and we recover the classic parareal scheme. This can be seen in both [Theorem 4.6](#) and [Corollary 4.7](#), where Λ vanishes as $q \rightarrow \infty$, leading to bounds similar to those for classic parareal. These bounds are not identical to those calculated by [Gander and Vandewalle \(2007\)](#), [Gander and Hairer \(2008\)](#), and [Gander et al. \(2022\)](#) because we are working with the mean-square error, not the (mean) absolute error.

Remark 4.12. If we additionally assume that the random variables ξ_n^k are centred, i.e. $\mathbb{E}[\xi_n^k] = \mathbf{0}$, and work in the 2-norm, i.e. $\|\mathbf{u}\|_2^2 = \langle \mathbf{u}, \mathbf{u} \rangle = \sum_{i=1}^d u_i^2$, in the proof of [Theorem 4.6](#), we can write [\(4.9\)](#) as

$$e_{n+1}^{k+1} \leq (1 + \delta_1^{-1})\mathbb{E}[\|W_1\|^2] + (1 + \delta_1)\mathbb{E}[\|W_2\|^2] + \mathbb{E}[\|W_3\|^2] + 2\mathbb{E}[\langle W_1, W_3 \rangle] + 2\mathbb{E}[\langle W_2, W_3 \rangle],$$

where the final two terms are equal to zero by independence of W_1 and W_2 with W_3 and using the fact that each ξ_n^k is centred. Continuing the proof, we obtain the same bounds for [Theorem 4.6](#) and [Corollary 4.7](#) with slightly altered constants $A = C_1^2 \Delta T^{2p+2}(1 + \Delta T^{-1})$, $B = L_{\mathcal{G}}^2(1 + \Delta T)$, $\Lambda = C_2^2 \Delta T^{2q+1}$, and $D = A\hat{e}^0$.

4.2. State-dependent perturbations (sampling rules)

We now use the previous results to derive the corresponding error bounds for the state-dependent sampling rules defined in [Table 3.1](#).

Corollary 4.13 (Linear error bound for state-dependent sampling rules 2 and 4). *Suppose the SParareal scheme [\(3.1\)](#)-[\(3.4\)](#) satisfies [Assumptions 4.1](#), [4.2](#) and [4.3](#), with $\xi_n^k(\mathbf{U}_n^k)$ defined using sampling rule 2 or 4. Then, the maximal mean-square error [\(4.2\)](#) of the solution to the nonlinear ODE system [\(2.1\)](#) at iteration k satisfies*

$$\hat{e}^k \leq \hat{e}^0 \left[\frac{A + \Lambda_1 + \sqrt{(A + \Lambda_1)^2 + 4\Lambda_2(1 - B)}}{2(1 - B)} \right]^k, \quad \text{if } B < 1,$$

for $2 \leq k \leq N$ and constants $A = C_1^2 \Delta T^{2p+2} (2 + \Delta T^{-1})$, $B = L_G^2 (1 + 2\Delta T)$, $\Lambda_1 = C_1^2 \Delta T^{2p+2} L_G^2 (1 + \Delta T^{-1})$, and $\Lambda_2 = C_1^2 \Delta T^{2p+2} L_G^2 (1 + \Delta T)$.

Proof for sampling rule 2. The proof follows in the same fashion as [Theorem 4.6](#). Instead of using the bound [\(4.12\)](#), we obtain, using [\(3.6\)](#) and applying [\(4.5\)](#),

$$\mathbb{E}[\|W_3\|^2] \leq C_1^2 \Delta T^{2(p+1)} \mathbb{E}[\|\alpha_n^k - \mathbf{U}_n^k\|^2]. \quad (4.17)$$

Substituting in α_n^k for sampling rule 2 ([Table 3.1](#)) we get

$$\begin{aligned} \mathbb{E}[\|W_3\|^2] &\leq C_1^2 \Delta T^{2(p+1)} \mathbb{E}[\|\sigma_n^k \circ \mathbf{z}_n^k\|^2] \\ &\leq C_1^2 \Delta T^{2(p+1)} \mathbb{E}[\|\sigma_n^k\|^2] \mathbb{E}[\|\mathbf{z}_n^k\|^2] \\ &\leq C_1^2 \Delta T^{2(p+1)} L_G^2 \mathbb{E}[\|\mathbf{U}_{n-1}^k - \mathbf{U}_{n-1}^{k-1}\|^2]. \end{aligned}$$

The second inequality follows by Cauchy-Schwarz and independence of σ_n^k and \mathbf{z}_n^k . The third follows by plugging in σ_n^k , applying [\(4.6\)](#), and noting that $\mathbb{E}[\|\mathbf{z}_n^k\|^2] = 1$. Next, we add and subtract $\mathbf{u}(t_{n-1})$ inside the expectation and then apply [\(A.1\)](#), with $\delta = \Delta T$, to get

$$\mathbb{E}[\|W_3\|^2] \leq C_1^2 \Delta T^{2(p+1)} L_G^2 ((1 + \Delta T^{-1})e_{n-1}^k + (1 + \Delta T)e_{n-1}^{k-1}). \quad (4.18)$$

Using the new bound for $\mathbb{E}[\|W_3\|^2]$ in [\(4.9\)](#), we obtain the double recurrence

$$e_{n+1}^{k+1} \leq A e_n^k + B e_n^{k+1} + \Lambda_1 e_{n-1}^k + \Lambda_2 e_{n-1}^{k-1}, \quad (4.19)$$

where $A = C_1^2 \Delta T^{2p+2} (2 + \Delta T^{-1})$, $B = L_G^2 (1 + 2\Delta T)$, $\Lambda_1 = C_1^2 \Delta T^{2p+2} L_G^2 (1 + \Delta T^{-1})$, and $\Lambda_2 = C_1^2 \Delta T^{2p+2} L_G^2 (1 + \Delta T)$. Maximising over n , we obtain

$$\hat{e}^{k+1} \leq \tilde{A} \hat{e}^k + \tilde{B} \hat{e}^{k-1}, \quad (4.20)$$

where

$$\tilde{A} = \frac{A + \Lambda_1}{1 - B} \quad \text{and} \quad \tilde{B} = \frac{\Lambda_2}{1 - B}.$$

Recursion [\(4.20\)](#) can be solved using [Lemma B.2](#) (see Appendix), resulting in the desired bound. \blacksquare

Proof for sampling rule 4. The proof follows in the same way as the proof for sampling rule 2, except that $\mathbb{E}[\|\sqrt{3}(2\mathbf{w}_n^k - \mathbf{1})\|^2] = 1$ is used in place of $\mathbb{E}[\|\mathbf{z}_n^k\|^2] = 1$. \blacksquare

Corollary 4.14 (Linear error bound for state-dependent sampling rules 1 and 3). *Suppose the SParareal scheme [\(3.1\)](#)-[\(3.4\)](#) satisfies [Assumptions 4.1](#), [4.2](#), [4.3](#) and [4.5](#), with $\xi_n^k(\mathbf{U}_n^k)$ defined using sampling rule 1 or 3. Then, the maximal mean-square error [\(4.2\)](#) of the solution to the nonlinear ODE system [\(2.1\)](#) at iteration k satisfies*

$$\hat{e}^k \leq \hat{e}^0 \left[\frac{A + \Lambda_1 + \Lambda_3 + \sqrt{(A + \Lambda_1 + \Lambda_3)^2 + 4\Lambda_2(1 - B)}}{2(1 - B)} \right]^k, \quad \text{if } B < 1,$$

for $2 \leq k \leq N$ and constants $A = C_1^2 \Delta T^{2p+2} (2 + \Delta T^{-1})$, $B = L_G^2 (1 + 2\Delta T)$, $\Lambda_1 = 2C_1^2 \Delta T^{2p+2} L_G^2 (1 + \Delta T^{-1})$, $\Lambda_2 = 2C_1^2 \Delta T^{2p+2} (L_G^2 (1 + \Delta T) + 2L_F^2)$, and $\Lambda_3 = 4C_1^2 \Delta T^{2p+2}$.

Proof for sampling rule 1. The proof follows in the same fashion as [Corollary 4.13](#). Substituting α_n^k for sampling rule 1 ([Table 3.1](#)) in [\(4.17\)](#), we get

$$\begin{aligned} \mathbb{E}[\|W_3\|^2] &\leq C_1^2 \Delta T^{2(p+1)} \mathbb{E}[\|\mathcal{F}_{\Delta T}(\mathbf{U}_{n-1}^{k-1}) - \mathbf{U}_n^k + \sigma_n^k \circ \mathbf{z}_n^k\|^2] \\ &\leq 2C_1^2 \Delta T^{2(p+1)} \left(\underbrace{\mathbb{E}[\|\mathcal{F}_{\Delta T}(\mathbf{U}_{n-1}^{k-1}) - \mathbf{U}_n^k\|^2]}_{\text{1st Term}} + \underbrace{\mathbb{E}[\|\sigma_n^k \circ \mathbf{z}_n^k\|^2]}_{\text{2nd Term}} \right). \end{aligned} \quad (4.21)$$

The second inequality follows by applying [\(A.1\)](#) with $\delta = 1$. To bound the first term in [\(4.21\)](#), we add and subtract $\mathcal{F}_{\Delta T}(\mathbf{u}(t_{n-1}))$ inside the expectation and apply [\(A.1\)](#) again with $\delta = 1$, obtaining

$$\begin{aligned} \text{1st Term} &\leq 2(\mathbb{E}[\|\mathcal{F}_{\Delta T}(\mathbf{U}_{n-1}^{k-1}) - \mathcal{F}_{\Delta T}(\mathbf{u}(t_{n-1}))\|^2] + \mathbb{E}[\|\mathcal{F}_{\Delta T}(\mathbf{u}(t_{n-1})) - \mathbf{U}_n^k\|^2]) \\ &\leq 2(L_{\mathcal{F}}^2 e_{n-1}^{k-1} + e_n^k). \end{aligned}$$

The second inequality follows by applying the Lipschitz condition [\(4.8\)](#) and recalling that $\mathcal{F}_{\Delta T}$ is the exact solver [\(4.3\)](#). The second term in [\(4.21\)](#) can be bounded as in [\(4.18\)](#) in [Corollary 4.13](#),

$$\text{2nd Term} \leq L_{\mathcal{G}}^2 ((1 + \Delta T^{-1})e_{n-1}^k + (1 + \Delta T)e_{n-1}^{k-1}).$$

Combining both terms in [\(4.21\)](#), we obtain

$$\mathbb{E}[\|W_3\|^2] \leq \Lambda_1 e_{n-1}^k + \Lambda_2 e_{n-1}^{k-1} + \Lambda_3 e_n^k,$$

where $\Lambda_1 = 2C_1^2 \Delta T^{2p+2} L_{\mathcal{G}}^2 (1 + \Delta T^{-1})$, $\Lambda_2 = 2C_1^2 \Delta T^{2p+2} (L_{\mathcal{G}}^2 (1 + \Delta T) + 2L_{\mathcal{F}}^2)$, and $\Lambda_3 = 4C_1^2 \Delta T^{2p+2}$. Using the new bound for $\mathbb{E}[\|W_3\|^2]$ in [\(4.9\)](#), we obtain the following recurrence

$$e_{n+1}^{k+1} \leq (A + \Lambda_3)e_n^k + B e_n^{k+1} + \Lambda_1 e_{n-1}^k + \Lambda_2 e_{n-1}^{k-1}, \quad (4.22)$$

where $A = C_1^2 \Delta T^{2p+2} (2 + \Delta T^{-1})$ and $B = L_{\mathcal{G}}^2 (1 + 2\Delta T)$. Maximising over n , we obtain

$$\hat{e}^{k+1} \leq \tilde{A} \hat{e}^k + \tilde{B} \hat{e}^{k-1}, \quad (4.23)$$

where

$$\tilde{A} = \frac{A + \Lambda_1 + \Lambda_3}{1 - B} \quad \text{and} \quad \tilde{B} = \frac{\Lambda_2}{1 - B}.$$

Recursion [\(4.23\)](#) can be solved using [Lemma B.2](#) (see Appendix), resulting in the desired bound. ■

Proof for sampling rule 3. The proof follows in a similar fashion to the proof for sampling rule 1, with $\mathbb{E}[\|\sqrt{3}(2\mathbf{w}_n^k - \mathbf{1})\|^2] = 1$ being used in place of $\mathbb{E}[\|\mathbf{z}_n^k\|^2] = 1$. ■

Remark 4.15. In [Section 5](#), we can observe the behaviour of e_n^k (not just \hat{e}^k) for each of the sampling rules by solving the recursions [\(4.19\)](#) and [\(4.22\)](#) numerically. We do this by replacing the inequality with an equality, i.e. upper bounding the error estimate.

5. Numerical Experiments

Here, we present some experiments to compare the theoretical bounds derived in [Section 4](#) with the errors generated by running SParareal numerically. MATLAB code to reproduce the results below can be found at <https://github.com/kpentland/StochasticPararealAnalysis>.

5.1. System of linear ODEs

In the following experiments, we solve the linear system of ODEs

$$\frac{d\mathbf{u}}{dt} = Q\mathbf{u} \quad \text{over } t \in [0, T], \quad \text{with } \mathbf{u}(0) = \mathbf{u}_0, \quad (5.1)$$

where $Q \in \mathbb{R}^{d \times d}$. This system has the exact solution $\mathbf{u}(t) = \mathbf{u}_0 e^{Qt}$, where $e^{Qt} = \sum_{i=0}^{\infty} (Qt)^i / i!$ is the matrix exponential.

First, we examine the superlinear and linear bounds derived in [Theorem 4.6](#) and [Corollary 4.7](#), respectively, by running SParareal numerically with state-independent Gaussian perturbations

$$\boldsymbol{\xi}_n^k \sim \mathcal{N}(\mathbf{0}, \Delta T^{2q+1} I_d), \quad q \geq 0. \quad (5.2)$$

We solve (5.1) with $d = 100$ and $T = 2$, discretising the time interval into $N = 20$ time slices so that $\Delta T = 0.25$. We construct the matrix of coefficients Q such that $B < 1$ and also select initial conditions $\mathbf{u}_0 \in [-5, 5]^d$. We use the exact solver $\mathcal{F}_{\Delta T}(\mathbf{u}) = \mathbf{u}e^{Q\Delta T}$ and the forward Euler method $\mathcal{G}_{\Delta T}(\mathbf{u}) = (I_d + Q\Delta T)\mathbf{u}$. In [Figure 5.1](#), we plot the maximal theoretical bounds \hat{e}^k and numerical errors of SParareal as a function of k for different values of q when $B < 1$. These results illustrate how the errors decrease as k increases (except when $q = 0$), up until the error induced by the perturbations become dominant—exactly the effect described in [Remark 4.9](#). For all considered values of q , the error for $k \geq 2$ has a hard lower bound of $\mathcal{O}(\Delta T^{2q+1})$, i.e. the error cannot go below the second moments of the perturbations (indicated by the dashed black line in each case). By altering Q and running the same experiment, we see similar effects in the $B \geq 1$ case, see [Figure 5.2](#).

It can be seen that, regardless of B , using the state-independent perturbations may not be optimal because of the lower bound forced upon the errors. If they are to be used, then they would need to be chosen such that the second moments are smaller than the accuracy of the solutions sought. This approach, however, may not yield accelerated convergence over the classic parareal scheme. To avoid this (and the lower bound on accuracy), the perturbations need to be state-dependent and therefore able to adapt, i.e. the second moments need to decrease with k and scale with n . In [Figure 5.3](#), we illustrate how the second moments of the perturbations used in the

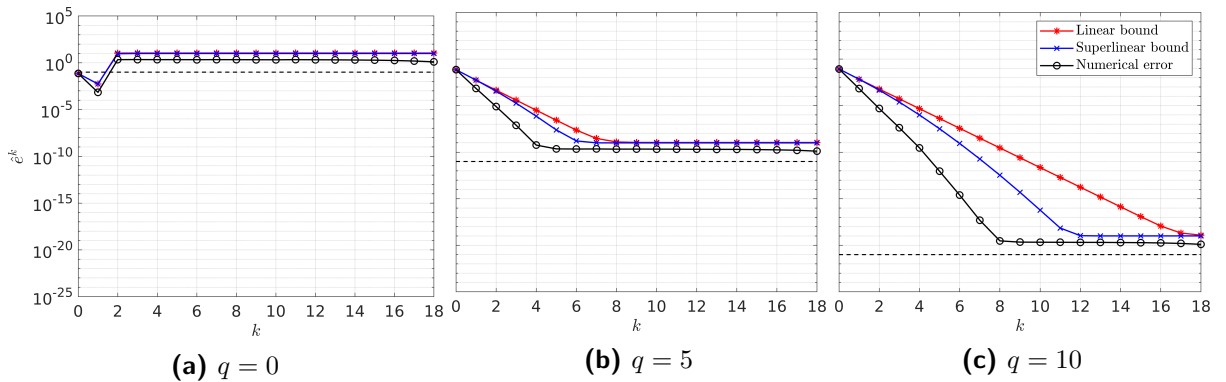


Figure 5.1. Theoretical bounds vs. numerical errors for SParareal applied to the linear system of ODEs (5.1) (with $B < 1$) using state-independent Gaussian perturbations (5.2). The superlinear bound ([Theorem 4.6](#)) is given in blue, the linear bound ([Corollary 4.7](#)) in red, the numerical error in black, and ΔT^{2q+1} in dashed black. Each plot corresponds to a different level of Gaussian noise: (a) $q = 0$, (b) $q = 5$, and (c) $q = 10$. Numerical errors were calculated by averaging over 500 independent realisations of SParareal.

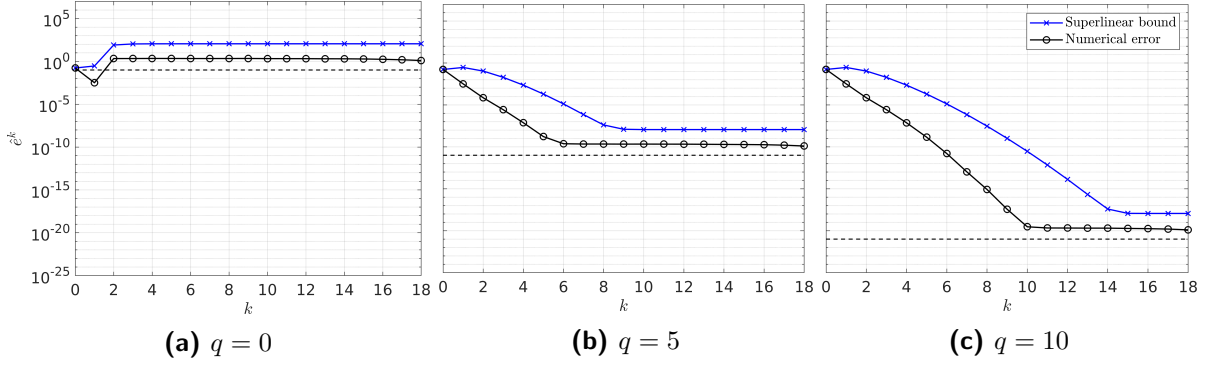


Figure 5.2. Theoretical bounds vs. numerical errors for SPareal applied to the linear system of ODEs (5.1) (with $B \geq 1$) using state-independent Gaussian perturbations (5.2). The superlinear bound (Theorem 4.6) is given in blue, the numerical error in black, and ΔT^{2q+1} in dashed black. Each plot corresponds to a different level of Gaussian noise: (a) $q = 0$, (b) $q = 5$, and (c) $q = 10$. Numerical errors were calculated by averaging over 500 independent realisations of SPareal.

state-dependent sampling rules decrease with k throughout the SPareal simulation, comparing these to the fixed second moments of the Gaussians (5.2) for each $q \in \{0, 5, 10\}$ (dashed lines). Using the sampling rules enables SPareal to sample from probability distributions that begin to “contract” around the exact solution states as the simulation progresses, i.e. as k increases. This results in high solution accuracy in very few iterations, as will be shown in Figure 5.4.

It should be noted that we could have also chosen a different distribution other than the Gaussian from which to sample each state-independent ξ_n^k , as long as Assumption 4.4 is satisfied. For example, choosing uniformly distributed perturbations $\xi_n^k \sim \mathcal{U}[-\sqrt{3}\Delta T^{q+\frac{1}{2}}, \sqrt{3}\Delta T^{q+\frac{1}{2}}]^d$ yielded almost identical results (not shown).

Next, we plot the linear bounds for perturbations defined by the sampling rules, i.e. Corollary 4.13 and Corollary 4.14, against the corresponding numerical errors in Figure 5.4 (for the

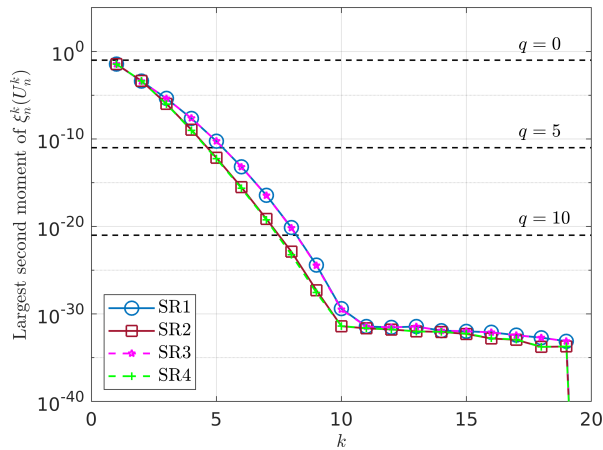


Figure 5.3. Largest second moments (over n) of $\xi_n^k(\mathbf{U}_n^k)$ for the sampling rules 1 to 4 (light blue, brown, purple, and green respectively) and the Gaussian perturbations (5.2) for $q \in \{0, 5, 10\}$ (dashed black), plotted against iteration number k . Second moments for the sampling rules were calculated by averaging over 500 independent realisations of SPareal.

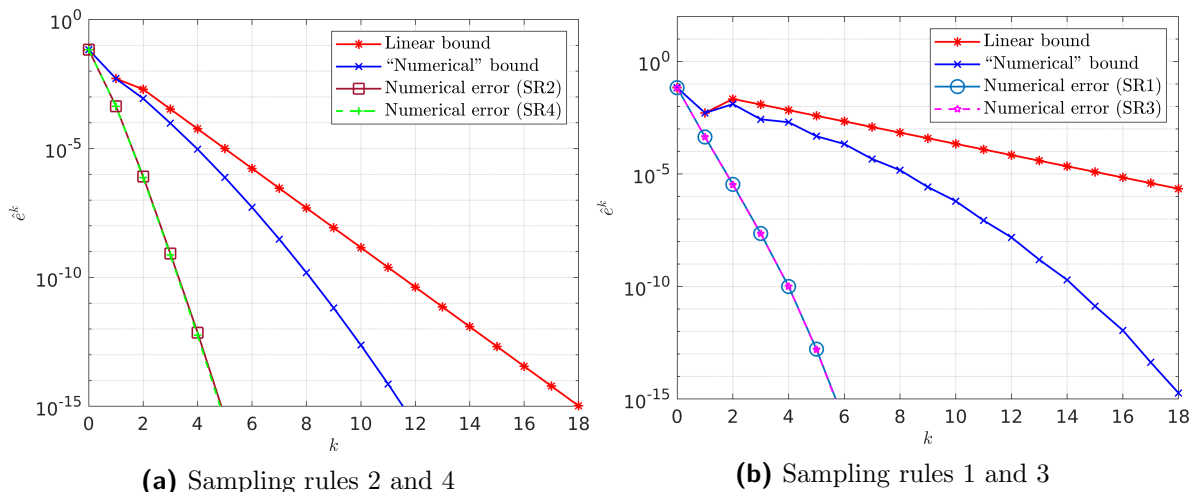


Figure 5.4. Theoretical bounds vs. numerical errors for SPareal applied to the linear system of ODEs (5.1) (with $B < 1$) using the state-dependent sampling rules. (a) The linear bound in Corollary 4.13 is shown in red, the numerically solved recursion (4.19) in blue, and the numerical errors for sampling rules 2 and 4 in brown and green, respectively. (b) The linear bound in Corollary 4.14 is shown in red, the numerically solved recursion (4.22) in blue, and the numerical errors for sampling rules 1 and 3 in light blue and purple, respectively. Numerical errors were calculated by averaging over 500 independent realisations of SPareal.

$B < 1$ problem). We observe that the linear bounds are not that tight due to the maximisation over n required to derive them. However, by solving recursions (4.19) and (4.22) numerically (recall Remark 4.15), we observe a tighter bound on the error. All that is required to calculate these “numerical” bounds are the errors at the ‘zeroth’ iteration (obtained from SPareal itself by just running \mathcal{G}), errors at the first iteration (recall (4.15)) and the constants C_1 and $L_{\mathcal{G}}$. Note

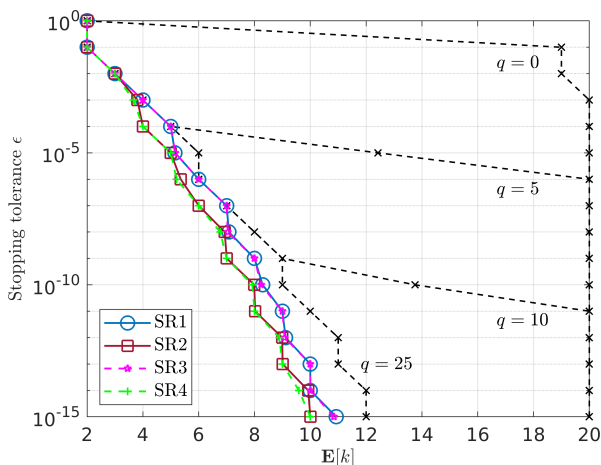


Figure 5.5. Expected number of iterations k taken to reach stopping tolerance ε (2.6) for SPareal applied to the linear system (5.1) (with $B < 1$). Results plotted using SPareal with each sampling rule (see legend) and the Gaussian perturbations (5.2) for $q \in \{0, 5, 10, 25\}$ (dashed black lines). $\mathbb{E}[k]$ calculated by averaging k over 500 independent realisations of SPareal.

that the numerical errors for sampling rules 2/4 and 1/3 overlap because the perturbations used in each scheme have almost identical second moments (recall [Figure 5.3](#)).

In [Figure 5.5](#), we compare the performance of the state-independent and -dependent perturbations by plotting the expected number of iterations $\mathbb{E}[k]$ taken to reach a pre-defined stopping tolerance ε , recall [\(2.6\)](#). We observe that, on average, the sampling rules reach tolerance in fewer iterations than the state-independent perturbations. The sampling rules also outperform classic parareal, which can be observed by comparing them to the state-independent perturbations for $q = 25$, for which perturbations are so small that SParareal is practically deterministic (i.e. parareal). Recall that reducing k by even a few iterations can significantly increase parallel speedup. The other advantage of using SParareal is that it returns a distribution of solution trajectories upon multiple realisations, instead of a single solution trajectory as parareal does, which can be interpreted as some form of uncertainty quantification over the solution to the IVP.

5.2. Scalar nonlinear ODE

In the following experiments, we solve the scalar nonlinear equation

$$\frac{du}{dt} = \sqrt{u^2 + 2} \quad \text{over } t \in [-1, 1], \quad \text{with } u(-1) = 5. \quad (5.3)$$

This equation has exact solution $u(t) = \sqrt{2} \sinh(t + 1 + \sinh^{-1}(5/2))$. We solve [\(5.3\)](#) using SParareal with $N = 20$ time slices (thus $\Delta T = 0.1$), exact solver $\mathcal{F}_{\Delta T}(u) = \sqrt{2} \sinh(\Delta T + \sinh^{-1}(u/\sqrt{2}))$, and forward Euler $\mathcal{G}_{\Delta T} = u + \Delta T \sqrt{u^2 + 2}$.

[Figure 5.6](#) illustrates a good match between the superlinear bound ([Theorem 4.6](#), $B \geq 1$) and the numerical errors when using SParareal and the Gaussian perturbations [\(5.2\)](#). One can see that at $k = 0$ the error is quite large, $\mathcal{O}(10^1)$, and so even when using the forward Euler method for $\mathcal{G}_{\Delta T}$, the SParareal error decreases rapidly (for sufficiently large q). Given the bounds in [Corollary 4.13](#) and [Corollary 4.14](#) only hold when $B < 1$, we again solve the respective recursions [\(4.19\)](#) and [\(4.22\)](#) numerically, obtaining a good match between theory and numerics when using the sampling rules (see [Figure 5.7](#)). [Figure 5.8](#) illustrates the performance of the state-independent perturbations and the sampling rules for varying stopping tolerance. As it did for the system of linear ODEs, using SParareal with state-dependent perturbations is more

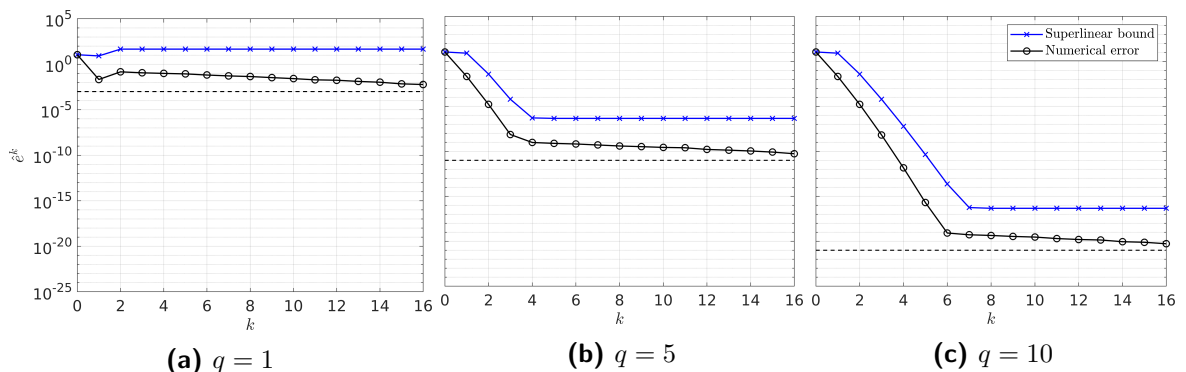


Figure 5.6. Theoretical bounds vs. numerical errors for SParareal applied to the nonlinear scalar ODE [\(5.3\)](#) (with $B \geq 1$) using state-independent Gaussian perturbations [\(5.2\)](#). The superlinear bound ([Theorem 4.6](#)) is given in blue, the numerical error in black, and ΔT^{2q+1} in dashed black. Each plot corresponds to a different level of Gaussian noise: (a) $q = 1$, (b) $q = 5$, and (c) $q = 10$. Numerical errors were calculated by averaging over 500 independent realisations of SParareal.

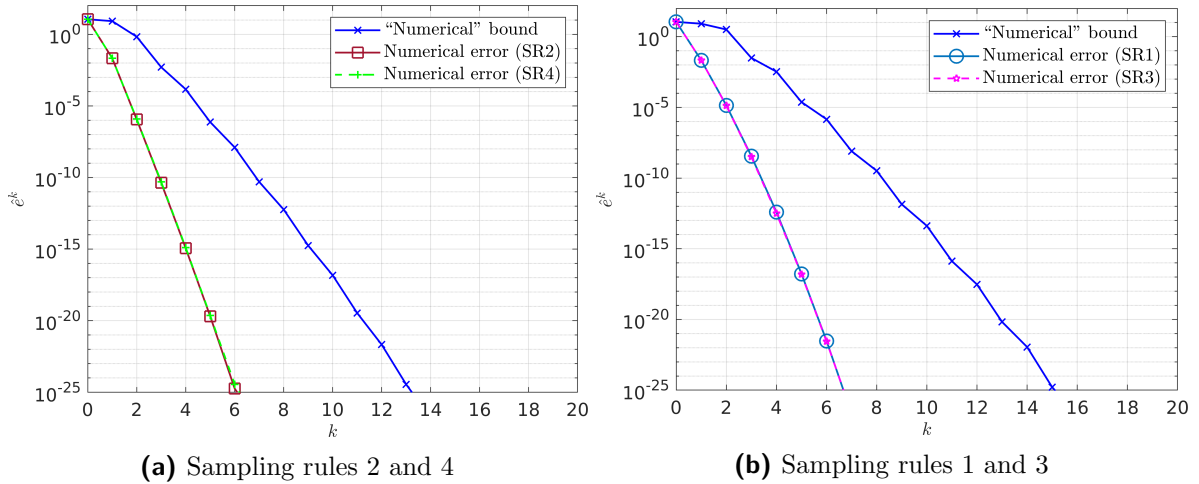


Figure 5.7. “Numerical” bounds vs. numerical errors for SPareal applied to the nonlinear scalar ODE (5.3) (with $B \geq 1$) using the state-dependent sampling rules. (a) The numerically solved recursion (4.19) is shown in blue and the numerical errors for sampling rules 2 and 4 in brown and green, respectively. (b) The numerically solved recursion (4.22) is shown in blue and the numerical errors for sampling rules 1 and 3 in light blue and purple, respectively. Numerical errors were calculated by averaging over 500 independent realisations of SPareal.

effective than with the state-independent perturbations, regardless of the value of chosen value of q for the Gaussian perturbations (recall that parareal can be recovered when choosing $q \geq 25$).

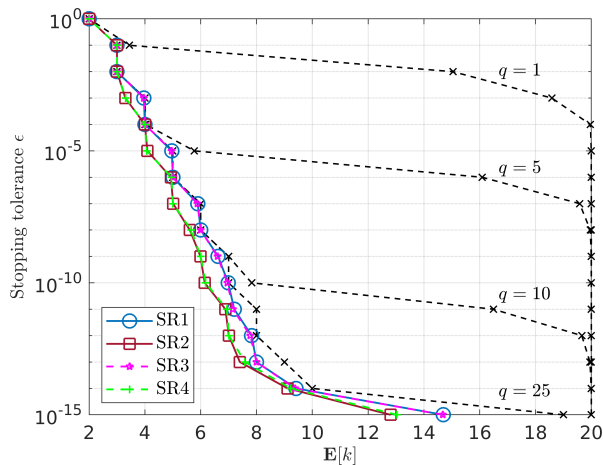


Figure 5.8. Expected number of iterations k taken to reach stopping tolerance ε (2.6) for SPareal applied to the nonlinear scalar ODE (5.3) (with $B \geq 1$). Results plotted using SPareal with each sampling rule (see legend) and the Gaussian perturbations (5.2) for $q \in \{0, 5, 10, 25\}$ (dashed black lines). $\mathbb{E}[k]$ calculated by averaging k over 500 independent realisations of SPareal.

6. Conclusions

The SParareal algorithm solves IVPs by perturbing solutions from the classic (deterministic) parareal scheme using a single sample drawn from a pre-specified probability distribution. This probabilistic time-parallel scheme generates stochastic solutions to the IVP. In this paper, we analyse the error of these stochastic numerical solutions by deriving mean-square error bounds for SParareal, equipped with different types of perturbations, applied to systems of nonlinear ODEs.

In [Section 4](#), we make assumptions about the fine ($\mathcal{F}_{\Delta T}$) and coarse ($\mathcal{G}_{\Delta T}$) numerical integrators used by SParareal, namely that $\mathcal{F}_{\Delta T}$ returns the exact solution to the ODE and that $\mathcal{G}_{\Delta T}$ has uniform local truncation error and satisfies a Lipschitz condition. Error bounds were then derived for two types of random perturbation, one in which the random variables do not depend on the current state of the system (state-independent) and one in which they do (state-dependent). In the state-independent case, where specific upper bounds were assumed on the second moments of the random variables, we derived both superlinear ([Theorem 4.6](#)) and linear ([Corollary 4.7](#)) bounds on the mean-square error. In the state-dependent case, where a number different perturbations were defined according to “sampling rules” (from the the original work), we derived linear bounds on the errors (see [Corollaries 4.13](#) and [4.14](#)).

In [Section 5](#), we illustrate these bounds, comparing them to the errors generated by running SParareal numerically. We demonstrate a good match between the theoretical bounds and numerical errors for systems of linear ODEs ([Section 5.1](#)) and a scalar nonlinear ODE ([Section 5.2](#)). Using the state-independent perturbations, we observed tight superlinear and linear bounds with respect to the numerical errors. However, because these perturbations do not adapt with iteration k and time step n , their practical usage faced limitations. They encoded a hard lower bound on solution accuracy (of the order of the size of the second moments, see [Remark 4.9](#)) and more iterations were typically required to reach stopping tolerance for larger perturbations. Instead, the sampling rules, shown to adapt with both k and n , did not suffer from these issues, as was previously discussed in [Pentland et al. \(2022a\)](#). The derivation of the linear bounds for the sampling rules did, however, require multiple applications of the Peter-Paul inequality, resulting in less tight bounds compared to those found in the state-independent case. Tighter bounds were observed by solving the double recursions ([4.19](#)) and ([4.22](#)) numerically. In addition, these linear bounds required the constant B to be less than one (restricting its use to problems where the Lipschitz constant for $\mathcal{G}_{\Delta T}$ is smaller than one) and that $\mathcal{F}_{\Delta T}$ be Lipschitz continuous for sampling rules 1 and 3 (an additional restriction). In the future, it would be interesting to see if these restrictions can be avoided or whether one can derive bounds without the Lipschitz assumptions.

As high performance computing technology advances, the demand for faster and more accurate time-parallel integration methods will increase. With SParareal, we have seen that introducing “local” perturbations into an existing time-parallel scheme can enable convergence in fewer iterations (using the sampling rules) and can, on average, result in higher accuracy solutions (refer to numerical experiments in [Pentland et al. \(2022a\)](#)). Following multiple realisations of SParareal, these solutions can form a measure of uncertainty, i.e. a “global” distribution, over the true solution, the accuracy of which can be estimated by the error bounds derived in this paper. Further work is required to investigate whether similar bounds can be derived for the originally presented SParareal scheme (where M samples can be taken from the probability distributions instead of just one) which is able to locate solutions with increasing accuracy and numerical speedup when increasing numbers of samples are taken. In addition, in most practical applications, the exact flow map $\mathcal{F}_{\Delta T}$ is unknown and so it would be advantageous to investigate what happens when one relaxes this assumption, taking $\mathcal{F}_{\Delta T}$ to be a numerical flow map.

Acknowledgements

The authors would like to thank Han Cheng Lie for helpful discussions. KP is funded by the Engineering and Physical Sciences Research Council through the MathSys II CDT (EP/S022244/1) as well as the Culham Centre for Fusion Energy. This work has partly been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014–2018 and 2019–2020 under grant agreement No. 633053. The views and opinions expressed herein do not necessarily reflect those of any of the above-named institutions or funding agencies.

References

- G. Bal. On the convergence and the stability of the parareal algorithm to solve partial differential equations. *Lecture Notes in Computational Science and Engineering*, 40:425–432, 2005. doi:10.1007/3-540-26825-1_43.
- G. Bal. Parallelization in time of (stochastic) ordinary differential equations, 2006. Pre-print: www.stat.uchicago.edu/guillaumbal.
- G. Bal and Y. Maday. A “parareal” time discretization for non-linear PDEs with application to the pricing of an american put. In *Recent Developments in Domain Decomposition Methods*, pages 189–202. Springer, Berlin, Heidelberg, 2002. doi:10.1007/978-3-642-56118-4_12.
- G. Bal and Q. Wu. Symplectic parareal. In *Lecture Notes in Computational Science and Engineering*, volume 60, pages 401–408. Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-75199-1_51.
- B. Carrel, M. J. Gander, and B. Vandereycken. Low-rank parareal: a low-rank parallel-in-time integrator, 2022. arXiv:2203.08455.
- P. R. Conrad, M. Girolami, S. Särkkä, A. Stuart, and K. Zygalakis. Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Statistics and Computing*, 27(4):1065–1082, 2017. doi:10.1007/s11222-016-9671-0.
- X. Dai, C. Le Bris, F. Legoll, and Y. Maday. Symmetric parareal algorithms for hamiltonian systems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(3):717–742, 2013. doi:10.1051/m2an/2012046.
- W. R. Elwasif, S. S. Foley, D. E. Bernholdt, L. A. Berry, D. Samaddar, D. E. Newman, and R. Sanchez. A dependency-driven formulation of parareal: Parallel-in-time solution of PDEs as a many-task application. In *MTAGS’11 - Proceedings of the 2011 ACM International Workshop on Many Task Computing on Grids and Supercomputers, Co-Located with SC’11*, pages 15–24, New York, NY, 2011. ACM Press. doi:10.1145/2132876.2132883.
- S. Engblom. Parallel in time simulation of multiscale stochastic chemical kinetics. *Multiscale Model. Simul.*, 8:46–68, 2009. doi:10.1137/080733723.
- M. Gander and M. Petcu. Analysis of a krylov subspace enhanced parareal algorithm for linear problems. *ESAIM: Proceedings*, 25:114–129, 2008. doi:10.1051/proc:082508.
- M. J. Gander. 50 Years of Time Parallel Time Integration. In *Multiple Shooting and Time Domain Decomposition Methods*, pages 69–113. Springer, 2015. doi:10.1007/978-3-319-23321-5_3.
- M. J. Gander and E. Hairer. Nonlinear convergence analysis for the parareal algorithm. In *Lecture Notes in Computational Science and Engineering*, volume 60, pages 45–56. Springer, 2008. doi:10.1007/978-3-540-75199-1_4.
- M. J. Gander and S. Vandewalle. Analysis of the parareal time-parallel time-integration method. *SIAM J. Sci. Comput.*, 29:556–578, 2007. doi:10.1137/05064607X.
- M. J. Gander, T. Lunet, D. Ruprecht, and R. Speck. A unified analysis framework for iterative parallel-in-time algorithms, 2022. arXiv:2203.16069.

- L. Grigori, S. A. Hirstoaga, V.-T. Nguyen, and J. Salomon. Reduced model-based parareal simulations of oscillatory singularly perturbed ordinary differential equations. *Journal of Computational Physics*, 436, 2021. ISSN 0021-9991. doi:10.1016/j.jcp.2021.110282.
- E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer Series in Computational Mathematics. Springer-Verlag, second edition, 1993. doi:10.1007/978-3-540-78862-1.
- F. Legoll, T. Lelièvre, K. Myerscough, and G. Samaey. Parareal computation of stochastic differential equations with time-scale separation: A numerical convergence study. *Comput. Vis. Sci.*, 23:1–18, 2020. doi:10.1007/s00791-020-00329-y.
- F. Legoll, T. Lelièvre, and U. Sharma. An adaptive parareal algorithm: Application to the simulation of molecular dynamics trajectories. *SIAM J. Sci. Comput.*, 44(1):B146–B176, 2022. ISSN 1064-8275. doi:10.1137/21M1412979.
- H. C. Lie, A. M. Stuart, and T. J. Sullivan. Strong convergence rates of probabilistic integrators for ordinary differential equations. *Statistics and Computing*, 29:1265–1283, 2019. doi:10.1007/s11222-019-09898-6.
- H. C. Lie, M. Stahn, and T. J. Sullivan. Randomised one-step time integration methods for deterministic operator differential equations. *Calcolo*, 59(1):13, 2022. doi:10.1007/s10092-022-00457-6.
- J. L. Lions, Y. Maday, and G. Turinici. Résolution d’EDP par un schéma en temps «pararéel». *Comptes Rendus Acad. Sci. Ser. I Math.*, 332(7):661–668, 2001. doi:10.1016/S0764-4442(00)01793-6.
- Y. Maday and O. Mula. An adaptive parareal algorithm. *J. Comput. Appl. Math.*, 377: 112915–112915, 2020. doi:10.1016/j.cam.2020.112915.
- J. Nievergelt. Parallel methods for integrating ordinary differential equations. *Communications of the ACM*, 7:731–733, 1964. doi:10.1145/355588.365137.
- B. W. Ong and J. B. Schroder. Applications of time parallelization. *Comput. Vis. Sci.*, 23, 2020. doi:10.1007/s00791-020-00331-4.
- K. Pentland, M. Tamborrino, D. Samaddar, and L. C. Appel. Stochastic parareal: An application of probabilistic methods to time-parallelization. *SIAM J. Sci. Comput.*, pages S82–S102, 2022a. ISSN 1064-8275. doi:10.1137/21M1414231.
- K. Pentland, M. Tamborrino, T. J. Sullivan, J. Buchanan, and L. C. Appel. GParareal: A time-parallel ODE solver using Gaussian process emulation, 2022b. arXiv:2201.13418.
- D. Samaddar, D. E. Newman, and R. Sánchez. Parallelization in time of numerical simulations of fully-developed plasma turbulence using the parareal algorithm. *J. Comput. Phys.*, 229: 6558–6573, 2010. doi:10.1016/j.jcp.2010.05.012.
- D. Samaddar, D. P. Coster, X. Bonnin, L. A. Berry, W. R. Elwasif, and D. B. Batchelor. Application of the parareal algorithm to simulations of ELMs in ITER plasma. *Comput. Phys. Commun.*, 235:246–257, 2019. doi:10.1016/j.cpc.2018.08.007.

A. Standard results

Here we state some results that we make repeated use of.

Lemma A.1 (Peter-Paul Inequality). *For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ and $\delta > 0$, we have that*

$$2\|\mathbf{u}\|\|\mathbf{v}\| \leq \delta\|\mathbf{u}\|^2 + \delta^{-1}\|\mathbf{v}\|^2. \quad (\text{A.1})$$

Theorem A.2 (Binomial Theorem). *For $|x| < 1$ and some $m \in \mathbb{N}$, we have that*

$$\frac{1}{(1-x)^m} = \sum_{i=0}^{\infty} \binom{i+m-1}{i} x^i. \quad (\text{A.2})$$

B. Generating Function Method

Here, we solve two recurrence relations using generating functions. These two variable (“double”) recurrences crop up often in convergence analysis involving parareal (and other PinT) algorithms and have been used in a number of settings—refer to [Gander and Hairer \(2008\)](#), [Carrel et al. \(2022\)](#), and [Gander et al. \(2022\)](#) for examples.

Lemma B.1. *Let e_n^k be a non-negative sequence and $A, B, D, \Lambda \in \mathbb{R}$ be non-negative constants. If e_n^k satisfies*

$$e_{n+1}^{k+1} \leq Ae_n^k + Be_n^{k+1} + \Lambda, \quad e_{n+1}^1 \leq D + Be_n^1, \quad (\text{B.1})$$

for $2 \leq k < n \leq N$ and $e_0^k = 0 \forall k \geq 0$, then

$$e_n^k \leq DA^{k-1} \sum_{\ell=0}^{n-k} \binom{\ell+k-1}{\ell} B^\ell + \Lambda \sum_{j=0}^{k-2} \sum_{\ell=0}^{n-(j+1)} \binom{\ell+j}{\ell} A^j B^\ell.$$

Proof. For $k \geq 1$, define the generating function for e_n^k as

$$g_k(x) = \sum_{n=1}^{\infty} e_n^k x^n. \quad (\text{B.2})$$

Multiply (B.1) by x^{n+1} and sum from n equals zero to infinity to obtain

$$\begin{aligned} \sum_{n=0}^{\infty} e_{n+1}^{k+1} x^{n+1} &\leq A \sum_{n=0}^{\infty} e_n^k x^{n+1} + B \sum_{n=0}^{\infty} e_n^{k+1} x^{n+1} + \Lambda \sum_{n=0}^{\infty} x^{n+1}, \\ \sum_{n=0}^{\infty} e_{n+1}^1 x^{n+1} &\leq D \sum_{n=0}^{\infty} x^{n+1} + B \sum_{n=0}^{\infty} e_n^1 x^{n+1}. \end{aligned}$$

Using (B.2), binomial theorem (A.2), and recalling that $e_0^k = 0 \forall k \geq 0$, we can write these expressions as

$$g_{k+1}(x) \leq \frac{Ax}{1-Bx} g_k(x) + \frac{\Lambda x}{(1-x)(1-Bx)}, \quad g_1(x) \leq \frac{Dx}{(1-x)(1-Bx)},$$

for $|x| < 1$, which can be solved iteratively to give

$$g_k(x) \leq \left(\frac{Ax}{1-Bx} \right)^{k-1} \frac{Dx}{(1-x)(1-Bx)} + \frac{\Lambda x}{(1-x)(1-Bx)} \sum_{j=0}^{k-2} \left(\frac{Ax}{1-Bx} \right)^j.$$

Re-arranging terms, this can be written as

$$g_k(x) \leq DA^{k-1}x^k \left(\frac{1}{1-Bx}\right)^k \frac{1}{(1-x)} + \Lambda \sum_{j=0}^{k-2} A^j x^{j+1} \left(\frac{1}{1-Bx}\right)^{j+1} \frac{1}{1-x}.$$

The first term can be expressed as

$$\begin{aligned} DA^{k-1}x^k \left(\frac{1}{1-Bx}\right)^k \frac{1}{(1-x)} &= DA^{k-1}x^k \left(\sum_{i=0}^{\infty} \binom{i+k-1}{i} (Bx)^i\right) \left(\sum_{i=0}^{\infty} x^i\right) \\ &= DA^{k-1}x^k \sum_{m=0}^{\infty} \left(\sum_{\ell=0}^m \binom{\ell+k-1}{\ell} B^\ell\right) x^m \\ &= DA^{k-1} \sum_{n=k}^{\infty} \left(\sum_{\ell=0}^{n-k} \binom{\ell+k-1}{\ell} B^\ell\right) x^n. \end{aligned}$$

The first line follows by applying (A.2) twice, the second line using the Cauchy product, and the third by setting $n = m + k$. The second term can be expressed as

$$\begin{aligned} \Lambda \sum_{j=0}^{k-2} A^j x^{j+1} \left(\frac{1}{1-Bx}\right)^{j+1} \frac{1}{1-x} &= \Lambda \sum_{j=0}^{k-2} A^j x^{j+1} \left(\sum_{i=0}^{\infty} \binom{i+j}{i} (Bx)^i\right) \left(\sum_{i=0}^{\infty} x^i\right) \\ &= \Lambda \sum_{j=0}^{k-2} A^j x^{j+1} \sum_{m=0}^{\infty} \left(\sum_{\ell=0}^m \binom{\ell+j}{\ell} B^\ell\right) x^m \\ &= \Lambda \sum_{n=j+1}^{\infty} \left(\sum_{j=0}^{k-2} \sum_{\ell=0}^{n-(j+1)} \binom{\ell+j}{\ell} A^j B^\ell\right) x^n. \end{aligned}$$

These steps follows as they did for the first term, except that we now set $n = m + j + 1$ instead of $n = m + k$ in the last step. Combining these expressions we get

$$g_k(x) = \sum_{n=1}^{\infty} e_n^k x^n \leq DA^{k-1} \sum_{n=k}^{\infty} \left(\sum_{\ell=0}^{n-k} \binom{\ell+k-1}{\ell} B^\ell\right) x^n + \Lambda \sum_{n=j+1}^{\infty} \left(\sum_{j=0}^{k-2} \sum_{\ell=0}^{n-(j+1)} \binom{\ell+j}{\ell} A^j B^\ell\right) x^n.$$

Equating the coefficients in x^n on both sides of the inequality we obtain the bound. ■

The initial condition for the recursion (B.1) can be written differently depending on the available information, i.e. one could instead use $e_{n+1}^1 \leq D := \hat{e}^1$, slightly altering the final bound obtained.

Lemma B.2. *Let \hat{e}^k be a non-negative sequence and $\tilde{A}, \tilde{B} \in \mathbb{R}$ be non-negative constants. If \hat{e}^k satisfies*

$$\hat{e}^{k+1} \leq \tilde{A}\hat{e}^k + \tilde{B}\hat{e}^{k-1}, \tag{B.3}$$

with initial conditions \hat{e}^0 and \hat{e}^1 , then

$$\hat{e}^k \leq \hat{e}^0 \left[\frac{\tilde{A} + \sqrt{\tilde{A}^2 + 4\tilde{B}}}{2} \right]^k.$$

Proof. Define the following generating function for \hat{e}^k :

$$g(x) = \sum_{k=0}^{\infty} \hat{e}^k x^k.$$

Multiply (B.3) by x^{k+1} and sum from k equals one to infinity to obtain

$$\sum_{k=1}^{\infty} \hat{e}^{k+1} x^{k+1} \leq \tilde{A} \sum_{k=1}^{\infty} \hat{e}^k x^{k+1} + \tilde{B} \sum_{k=1}^{\infty} \hat{e}^{k-1} x^{k+1}.$$

Shifting indices, rearranging, and using the initial conditions we get

$$g(x) = \sum_{k=0}^{\infty} \hat{e}^k x^k \leq \frac{\hat{e}^0(1 - \tilde{A}x) + \hat{e}^1 x}{1 - \tilde{A}x - \tilde{B}x^2}.$$

Expanding the right hand side in powers of x^k , the coefficients give us

$$\hat{e}^k \leq \frac{1}{2\sqrt{\tilde{A}^2 + 4\tilde{B}}} \left[(\tilde{A}\hat{e}^0 + \hat{e}^0\sqrt{\tilde{A}^2 + 4\tilde{B}} - 2\hat{e}^1)\lambda_1^k + (-\tilde{A}\hat{e}^0 + \hat{e}^0\sqrt{\tilde{A}^2 + 4\tilde{B}} + 2\hat{e}^1)\lambda_2^k \right],$$

where

$$\lambda_{1,2} = \frac{\tilde{A} \pm \sqrt{\tilde{A}^2 + 4\tilde{B}}}{2}.$$

Without loss of generality, we use that $\lambda_1 \geq \lambda_2$ to simplify the bound and obtain

$$\hat{e}^k \leq \hat{e}^0 \lambda_1^k,$$

which yields the desired result. ■