

Error estimation and adaptivity for differential equations with multiple scales in time

L. Lautsch* T. Richter†

April 28, 2021

We consider systems of ordinary differential equations with multiple scales in time. In general, we are interested in the long time horizon of a slow variable that is coupled to solution components that act on a fast scale. Although the fast scale variables are essential for the dynamics of the coupled problem, they are often of no interest in themselves. Recently we have proposed a temporal multiscale approach that fits into the framework of the heterogeneous multiscale method and that allows for efficient simulations with significant speedups. Fast and slow scales are decoupled by introducing local averages and by replacing fast scale contributions by localized periodic-in-time problems. Here, we generalize this multiscale approach to a larger class of problems but in particular, we derive an a posteriori error estimator based on the dual weighted residual method that allows for a splitting of the error into averaging error, error on the slow scale and error on the fast scale. We demonstrate the accuracy of the error estimator and also its use for adaptive control of a numerical multiscale scheme.

1 Introduction

We are interested in the efficient approximation of dynamical systems with multiple scales in time. Such problems appear in various applications such as material damage mechanics [19], astrophysics [5] or cardiovascular settings [13, 18]. Although multiscale problems are extensively studied in literature, see e.g. [6, 20], most works focus on problems where the multiscale character is in space but not in time. The *heterogeneous multiscale method (HMM)* [9, 8, 1, 10] is a very general framework and easily applied to temporal multiscale dynamics. Here, fast and slow problems are decoupled by means

*Otto-von-Guericke Universität Magdeburg, Germany leopold.lautsch@ovgu.de

†Otto-von-Guericke Universität Magdeburg, Germany thomas.richter@ovgu.de

of an averaging that gives an effective equation for the slow dynamics. The feedback between both scales is realized by localized fine scale problems that have to be approximated in every time step of the slow problem.

In [12, 18] we have developed such a multiscale approach with applications to medical flow problems, where the slow scale describes the growth of a stenosis and where the fast problem is the oscillatory dynamics coming from heart driven blood flow. For decoupling the scales local periodic-in-time solutions describing the fast dynamics are introduced and have to be solved once in each time step of the slow problem. An a priori error estimate for this multiscale scheme has been shown for a simple problem based on the Stokes equation. Numerically, significant acceleration and a reduction of the computational time by a factor of up to 10 000 is observed as compared to fully resolved simulations.

Algorithmically such multiscale schemes are complex, as they are based on multiple discretization schemes for slow and fast scales and as they require careful control of the transmission operator that carries information between these scales. Time step sizes must be chosen for the slow and the fast scale and further, tolerances must be defined to control the approximation quality all problems that are involved in the multiscale approach. If one or both of the scales are described by partial differential equations it is also necessary to control the spatial discretization parameters.

Here, we will derive and discuss an a posteriori error estimator based on the dual weighted residual method [3] for estimating functional errors that cover all error contributions coming from discretization and multiscale approximation. By splitting and localizing the error estimator to the various components, an adaptive multiscale scheme is realized that allows to optimally balance the different error contributions. The concept of goal oriented error estimation is chosen, since it allows for a uniform handling of temporal [25] and spatial discretization errors [2] but also of truncation errors coming from the violation of conformity by the non-exact solution of sub problems [15] and finally it also allows to include the multiscale error which can be considered as a kind of model error [4]. Since the design of the error estimator is complex and involves a staggered approach with adjoint and tangent problems on both scales, we restrict the presentation to a system of ordinary differential equations to keep the notation and discussion as brief as possible.

In the next section we describe the problem under consideration and we briefly summarize the multiscale approximation scheme as introduced in [12]. Then, Section 3 casts the multiscale scheme into a temporal Galerkin formulation that will act as basis for the error estimator derived in Section 4. Numerical examples are discussed in Section 5 and finally, we summarize in a short conclusion.

2 Model problem and multiscale approximation

We consider a system of ordinary differential equations.

Problem 1 (Model Problem). *On $I = [0, T]$ find $y : I \rightarrow \mathbb{R}^c$ and $u : I \rightarrow \mathbb{R}^d$ with*

$c, d \in \mathbb{N}$ such that

$$y'(t) = \epsilon f(y(t), u(t)), \quad u'(t) = g(t, y(t), u(t)) \quad (2.1)$$

with $y(0) = y_0 \in \mathbb{R}^c$ and $u(0) = u_0 \in \mathbb{R}^d$ and the scale separation parameter $0 < \epsilon \ll 1$. We will call $y(t)$ the slow component and $u(t)$ the fast component of the problem.

Let the following assumptions hold.

Assumption 2 (Slow Scale). *Let f be continuous on $I \times \mathbb{R}^c \times \mathbb{R}^d$, bounded*

$$\|f(y, u)\| \leq C_{A2.2} \quad \forall (y, u) \in \mathbb{R}^c \times \mathbb{R}^d, \quad (2.2)$$

and Lipschitz and differentiable with respect to y and u , such that

$$\begin{aligned} \|f(y, u) - f(Y, u)\| &\leq C_{A2.3} \|y - Y\| \quad \forall y, Y \in \mathbb{R}^c, \quad \forall u \in \mathbb{R}^d \\ \|f(y, u) - f(y, U)\| &\leq C_{A2.3} \|u - U\| \quad \forall u, U \in \mathbb{R}^d, \quad \forall y \in \mathbb{R}^c. \end{aligned} \quad (2.3)$$

Remark 3 (Separation of temporal scales). *By the boundedness of f , it holds*

$$\|y'(t)\| \leq C_{A2.2} \cdot \epsilon, \quad \|y(t)\| \leq C_{A2.4} \text{ for } t \in [0, T], \quad (2.4)$$

with $C_{A2.4} = C_{A2.2} \cdot \epsilon \cdot T + \|y_0\|$. We define

$$X^c := \{y \in \mathbb{R}^c, \|y\| \leq C_{A2.4}\}.$$

On the fast scale problem we impose the following assumptions.

Assumption 4 (Fast Scale). *Let g be continuous on $I \times \mathbb{R}^c \times \mathbb{R}^d$ and Lipschitz*

$$\|g(t, y_1, u) - g(t, y_2, u)\| \leq C_{A2.5} \|y_1 - y_2\| \text{ for all } y_1, y_2 \in X^c, \quad (2.5)$$

uniform in $t \in I$ and $u \in \mathbb{R}^d$. We assume that g is differentiable with respect to $y \in \mathbb{R}^c$ and $u \in \mathbb{R}^d$ with a Jacobian $\nabla_u g$ whose eigenvalues all have negative real part

$$-\lambda_{A2.6} := \sup\{\operatorname{Re}(\lambda), \lambda \text{ is Eigenvalue of } \nabla_u g(t, y, u)\} < 0, \quad (2.6)$$

uniform in $t \in [0, T]$ and $y \in X^c$. Finally, let g be 1-periodic in time

$$g(t, y, u) = g(t + 1, y, u) \quad \forall (t, y, u) \in \mathbb{R} \times \mathbb{R}^c \times \mathbb{R}^d \quad (2.7)$$

and we assume that for each $y \in X^c$ there exists a unique periodic-in-time solution to

$$u'_y(t) = g(t, y, u_y(t)) \quad t \in I^P := [0, 1], \quad u_y(0) = u_y(1), \quad (2.8)$$

which is bound by

$$\max_{t \in I^P} \|u_y\| \leq C_{A2.9}, \quad (2.9)$$

with a constant $C_{A2.9} = C(C_{A2.4})$. Finally, we assume that for all $y_1, y_2 \in X^c$ and the corresponding periodic-in-time solutions $u_{y_1}, u_{y_2} : [0, 1] \rightarrow \mathbb{R}^d$ it holds

$$\sup_{t \in I^P} \|u_{y_1}(t) - u_{y_2}(t)\| \leq C_{A2.10} \|y_1 - y_2\|. \quad (2.10)$$

Remark 5 (Scope of applications). *In Section 5 we will discuss specific systems of differential equations that meet this assumptions. The examples considered in [12] and [18] also fit into the frame of assumptions given above. In particular for nonlinear equations, like the Navier-Stokes equations which are considered in [18], the existence of time-periodic solutions can only be shown under very strict assumptions on the problem data, with limitations to small Reynolds numbers [14]. Hence, in particular (2.8)-(2.10) will be difficult to validate in most application cases. Also, (2.6) calls for a damping behavior of the fast scale problem and is here used to obtain uniqueness of solutions. Considering the Navier-Stokes equations this also calls for bounds on the problem data. In Section 5 we will present a simple test case where all assumptions can be validated.*

The multiscale scheme for the efficient approximation of Problem 1 is based on [12] and on the following averaged multiscale problem.

Problem 6 (Averaged Multiscale Problem). *On $I = [0, T]$ find $Y : I \rightarrow \mathbb{R}^c$ such that*

$$Y'(t) = \epsilon \int_t^{t+1} f(Y(t), u_{Y(t)}(s)) ds, \quad Y(0) = y_0, \quad (2.11)$$

and where $u_{Y(t)}(\cdot)$, for each $Y(t) \in \mathbb{R}^c$, is given as the solution to the periodic-in-time micro problem

$$\frac{d}{ds} u_{Y(t)}(s) = g(s, Y(t), u_{Y(t)}(s)) \text{ in } I^P = [0, 1], \quad u_{Y(t)}(1) = u_{Y(t)}(0). \quad (2.12)$$

Defining a *fast scale feedback operator* $\mathcal{F} : \mathbb{R}^c \rightarrow \mathbb{R}^c$ by

$$\mathcal{F}(Y) := \int_t^{t+1} f(Y(t), u_{Y(t)}(s)) ds, \quad (2.13)$$

with $u_{Y(t)}$ given by (2.12), the averaged multiscale problem can be written as

$$Y'(t) = \epsilon \mathcal{F}(Y(t)), \quad Y(0) = y_0, \quad (2.14)$$

and the fast scale problem is formally removed. This problem notation is basis for the numerical multiscale method to be described in Section 3. We will discretize (2.14) with very large time steps, and, in each time step $Y_{n-1} \mapsto Y_n$, the fast scale feedback operator $\mathcal{F}(\cdot)$ must be evaluated in the approximations Y_{n-1} and Y_n . The clue in the design of this multiscale approach is the introduction of locally periodic-in-time micro solutions $u_{Y(t)}(\cdot)$. This allows to formally decouple the microscale from the macroscale. $\mathcal{F}(Y_n)$ can be approximated without requiring any initial values that might have to be obtained from the previous macro step solution Y_{n-1} .

In the following we will show that this averaged multiscale problem has a solution and we will show that this solution $Y(t)$ is close to the original solution $y(t)$.

2.1 Analysis of the multiscale error

Given the above listed assumptions on the slow and the fast scale, we can show that the solution to the multiscale problem, Problem 6 is close to the resolved original solution. The following theorem is a generalization of [12, Lemma 10] to a more general class of equations.

Theorem 7 (Multiscale error). *Let Assumptions 2 and 4 hold. There exists a unique solution $Y(t)$ to Problem 6. Let $(y(t), u(t))$ be the solution to Problem 1. It holds*

$$\max_{t \in [0, T]} \|y(t) - Y(t)\| = \mathcal{O}(\epsilon).$$

Proof. Let $Y_1, Y_2 : I \rightarrow X^c$ be given. For the right hand side of Problem 6, given in the form (2.14) it holds

$$\begin{aligned} \|\mathcal{F}(Y_1(t)) - \mathcal{F}(Y_2(t))\| &\leq \int_t^{t+1} \|f(Y_1(t), u_{Y_1(t)}(s)) - f(Y_2(t), u_{Y_2(t)}(s))\| \, ds \\ &\leq C_{A2.3} \int_t^{t+1} \|Y_1(t) - Y_2(t)\| + \|u_{Y_1(t)}(s) - u_{Y_2(t)}(s)\| \, ds \\ &\leq C_{A2.3}(1 + C_{A2.10})\|Y_1(t) - Y_2(t)\|. \end{aligned}$$

Hence, the problem is Lipschitz and a unique solution exists. The bound $\|Y(t)\| \leq C_{A2.2} \cdot \epsilon \cdot t$ shows that this solution exists on all $I = [0, T]$.

To start with, we introduce the average

$$\bar{Y}(t) := \int_t^{t+1} y(s) \, ds.$$

For this average it holds

$$\|y(t) - \bar{Y}(t)\| \leq \int_t^{t+1} \|y(t) - y(s)\| \, ds = \int_t^{t+1} \left\| \int_t^s y'(r) \, dr \right\| \, ds \leq \frac{C_{A2.2}}{2} \epsilon. \quad (2.15)$$

The remaining error $w(t) := \bar{Y}(t) - Y(t)$ is governed by

$$w'(t) = \epsilon \int_t^{t+1} f(y(s), u(s)) - f(Y(t), u_{Y(t)}(s)) \, ds, \quad w(0) = \int_0^1 y(s) \, ds - y_0, \quad (2.16)$$

where the initial value is estimated with help of (2.15)

$$\|w(0)\| \leq \int_0^1 \|y(s) - y(0)\| \, ds \leq \frac{C_{A2.2}}{2} \epsilon. \quad (2.17)$$

To estimate $w(t)$ we split the right hand side of (2.16) into

$$\begin{aligned} &f(y(s), u(s)) - f(Y(t), u_{Y(t)}(s)) \\ &= \left(f(y(s), u(s)) - f(Y(t), u(s)) \right) + \left(f(Y(t), u(s)) - f(Y(t), u_{Y(t)}(s)) \right). \end{aligned} \quad (2.18)$$

Then, Lipschitz continuity of $f(\cdot)$ gives

$$\|w'(t)\| \leq C_{A2.3}\epsilon \int_t^{t+1} \left(\|y(s) - Y(t)\| + \|u_{Y(t)}(s) - u(s)\| \right) ds. \quad (2.19)$$

The first term is bounded by introducing $\pm \bar{Y}(t)$

$$\begin{aligned} \int_t^{t+1} \|y(s) - Y(t)\| ds &\leq \int_t^{t+1} \|y(s) - \bar{Y}(t)\| ds + \|\bar{Y}(t) - Y(t)\| \\ &\leq \int_t^{t+1} \int_t^{t+1} \|y(s) - y(r)\| dr ds + \|w(t)\| \\ &= \int_t^{t+1} \int_t^{t+1} \left\| \int_r^s y'(t) dt \right\| dr ds + \|w(t)\| \leq \frac{C_{A2.2}}{3}\epsilon + \|w(t)\| \end{aligned} \quad (2.20)$$

To bound the second term in (2.19) we introduce $\pm u_{y(s)}(s)$

$$\int_t^{t+1} \|u_{Y(t)}(s) - u(s)\| ds \leq \int_t^{t+1} \|u_{Y(t)}(s) - u_{y(s)}(s)\| + \|u_{y(s)}(s) - u(s)\| ds \quad (2.21)$$

Here, the first term is bounded with help of (2.10) and (2.15)

$$\begin{aligned} \int_t^{t+1} \|u_{Y(t)}(s) - u_{y(s)}(s)\| ds &\leq C_{A2.10} \int_t^{t+1} \|Y(t) - y(s)\| ds \\ &\leq C_{A2.10} \left(\|Y(t) - \bar{Y}(t)\| + \|\bar{Y}(t) - y(t)\| + \int_t^{t+1} \|y(t) - y(s)\| ds \right) \\ &\leq C_{A2.10} \left(\|w(t)\| + C_{A2.2}\epsilon \right). \end{aligned} \quad (2.22)$$

The second term in (2.21), $\|u_{y(s)}(s) - u(s)\|$, is more subtle to estimate. It measures the difference between the fully dynamic solution $u(s)$, evolving around the slow scale $y(s)$ to the isolated periodic-in-time solutions for each $y(s)$ fixed. For better readability we move the estimate of this term to the separate Lemma 8 which shows

$$\int_t^{t+1} \|u_{y(s)}(s) - u(s)\| ds \leq C_{A2.27} \cdot \epsilon. \quad (2.23)$$

Combining this with (2.17)-(2.22) gives

$$\|w'(t)\| \leq C_{A2.3}\epsilon \left((1 + C_{A2.10})\|w(t)\| + \left(\frac{C_{A2.2}}{3} + C_{A2.10} + C_{A2.27} \right) \epsilon \right), \quad (2.24)$$

where $\|w(0)\| \leq \frac{C_{A2.2}}{2}\epsilon$. Using [7, Sec. 3] we can estimate $\|w(t)\|$ by the solution to the differential equation with the corresponding right hand side

$$\|w(t)\| \leq \left(1 + \exp(C_{A2.3}(1 + C_{A2.10})\epsilon \cdot t) \right) (C_{A2.2} + C_{A2.10} + C_{A2.27})\epsilon, \quad (2.25)$$

which for $t = \mathcal{O}(\epsilon^{-1})$ is bound by

$$\|\bar{Y}(t) - Y(t)\| \leq C_{A2.26}\epsilon, \quad (2.26)$$

with a constant $C_{A2.26} = C(C_{A2.2}, C_{A2.3}, C_{A2.10}, C_{A2.27})$. Finally, the claim of the theorem follows by combining (2.15) and (2.26)

$$\|y(t) - Y(t)\| \leq \|y(t) - \bar{Y}(t)\| + \|\bar{Y}(t) - Y(t)\| \leq \frac{C_{A2.3}}{2}\epsilon + C_{A2.26}\epsilon.$$

□

It remains to prove estimate (2.23), the difference between the dynamically evolving fast scale solution and the local periodic-in-time solutions. This lemma is close to [12, Lemma 9].

Lemma 8. *Let $y \in C^1(I; X^c)$ be given such that $y(0) = y_0$ and $\|y'(t)\| \leq C_{A2.2} \cdot \epsilon$. Let $u_{y(t)}(s)$ be the family of periodic-in-time fast scale solutions for $y(t)$ fixed and let $u(t)$ be the dynamic solution to*

$$u'(t) = g(t, y(t), u(t)), \quad u(0) = u_{y(0)}(0).$$

There exists a constant $C_{A2.27} := C(C_{A2.2}, C_{A2.10}, \lambda_{A2.6})$ such that

$$\|u(t) - u_{y(t)}(t)\| \leq C_{A2.27} \cdot \epsilon. \quad (2.27)$$

Proof. For the periodic-in-time solutions it holds

$$\frac{d}{dt}u_{y(t)}(t) = u'_{y(t)}(t) + \underbrace{\frac{du_{y(t)}}{dy(t)}(t)}_{=: Du_{y(t)}(t)} y'(t),$$

and the difference $v(t) := u(t) - u_{y(t)}(t)$ is governed by

$$v'(t) = g(t, y(t), u(t)) - g(t, y(t), u_{y(t)}(t)) - Du_{y(t)}(t)y'(t),$$

and, since g is differentiable, it holds

$$v'(t) = \nabla_y g(t, y(t), \gamma(t))v(t) - Du_{y(t)}(t)y'(t),$$

where $\gamma(t) \in \mathbb{R}^u$ is an intermediate between $u(t)$ and $u_{y(t)}(t)$. Since $v(0) = u_{y(0)}(0) - u_{y(0)}(0) = 0$ it holds

$$v(t) = \int_0^t \Phi(s) Du_{y(s)}(s) y'(s) ds, \quad \Phi(t) = \exp \left(\nabla_y g(t, y(t), \gamma(t)) \cdot t \right),$$

such that

$$\|v(t)\| \leq C_{A2.2}\epsilon \int_0^t \Phi(s) ds \cdot \sup_{s \in [0, t]} |Du_{y(s)}(s)| \leq \sup_{s \in [0, t]} |Du_{y(s)}(s)| \cdot \frac{C_{A2.2}}{\lambda_{A2.6}}\epsilon, \quad (2.28)$$

Using that all eigenvalues of $\nabla_y g$ have a negative real part. To estimate $Du_y(t)$ we first note that g is differentiable with respect to $y \in \mathbb{R}^c$ such that

$$Du'_y(t) = \nabla_u g(t, y, u_y(t)) Du_y(t) + \nabla_y g(t, y, u_y(t)).$$

This equation is linear in $Du_y(t)$ such that a unique solution exists. Given two periodic solutions u_{Y_1} and u_{Y_2} and using (2.10) we can bound the derivative $Du_y(t)$ by estimating

$$\frac{|u_{Y_1}(t) - u_{Y_2}(t)|}{|Y_1 - Y_2|} \leq C_{A2.10}, \quad (2.29)$$

uniform in $Y_1, Y_2 \in \mathbb{R}^d$. Hence, combining (2.28) and (2.29)

$$\|u(t) - u_{y(t)}(t)\| \leq \frac{C_{A2.2} C_{A2.10}}{\lambda_{A2.6}} \epsilon. \quad (2.30)$$

□

Theorem 7 is the analytical basis for the following numerical approximation scheme. We have shown that the solution to the averaged multiscale problem, Problem 6 is close to the solution of the original model problem. The numerical scheme will be based on the approximation of the averaged equation in Problem 6 using large time steps K , which are by far larger than the micro scale. For evaluating the right hand side, we will have to evaluate the transfer operator $\mathcal{F}(Y(t))$. This will require the solution of a periodic-in-time micro problem. Here a small time step $k \ll K$ must be employed. In [12] we have further given an a priori error estimator for the numerical discretization of the averaged multiscale method and, in [12, Theorem 18] we have shown that the multiscale scheme converges with second order, $\mathcal{O}(\epsilon^2 K^2 + k^2)$, if both averaged macro scale problem and periodic micro problem are discretized with second order time stepping schemes.

2.2 Variational formulation

We conclude by defining the variational formulation of the averaged multiscale problem which will be the basis for the Galerkin discretization and the a posteriori error estimator.

Problem 9 (Variational formulation of the multiscale problem). *Find $Y \in \mathcal{X}$ such that*

$$A(Y, \Phi) = 0 \quad \forall \Phi \in \mathcal{Y},$$

$$A(Y, \Phi) := \int_0^T (Y'(t) - \epsilon \mathcal{F}(Y(t))) \cdot \Phi(t) dt, \quad \mathcal{F}(Y) := \int_0^1 f(Y, u_Y(s)) ds \quad (2.31)$$

where

$$\mathcal{X} := \{\Phi \in H^1(I; \mathbb{R}^c), \Phi(0) = y_0\}, \quad \mathcal{Y} := L^2(I; \mathbb{R}^c), \quad (2.32)$$

and where, for a fixed value $Y \in \mathbb{R}^c$, the periodic fast scale solutions u_Y are defined on $I^P = [0, 1]$ by

$$u_Y \in \mathcal{V}^\pi : \quad B(Y; u_Y; \phi) : \int_0^1 (u_Y'(t) - g(t, Y, u_Y(t))) \phi(t) dt = 0 \quad \forall \phi \in \mathcal{W}, \quad (2.33)$$

with test and trial spaces defined as

$$\mathcal{V}^\pi := \{\phi \in H^1(I^P; \mathbb{R}^d), \phi(0) = \phi(1)\}, \quad \mathcal{W} := L^2(I^P; \mathbb{R}^d). \quad (2.34)$$

3 Discretization and Solution

Discretization of (2.14) is based on a temporal Galerkin scheme of Problem 9. For general literature on temporal Galerkin formulations we refer to [27, 11]. Both long term and short term problem are discretized with continuous and piecewise linear functions using piecewise constant test functions with possible discontinuities at the discrete time steps. This results in a time stepping scheme which is of second order and which is, up to numerical quadrature, equivalent to the trapezoidal rule. We introduce the partitioning I_K of $I = [0, T]$ by

$$0 = T_0 < T_1 < \dots < T_N = T, \quad K_n := T_n - T_{n-1}, \quad I_n := (T_{n-1}, T_n] \quad (3.1)$$

and define the discrete subspaces

$$\begin{aligned} \mathcal{X}_K &:= \{\Psi \in C(\bar{I})^c : \Psi|_{I_n} \in P^1(I_n; \mathbb{R}^c), \ n = 1, \dots, N, \ \Psi(0) = y_0\} \subset \mathcal{X} \\ \mathcal{Y}_K &:= \{\Phi \in L^2(I; \mathbb{R}^c) : \Phi|_{I_n} \in P^0(I_n; \mathbb{R}^c), \ n = 1, \dots, N\} \subset \mathcal{Y}, \end{aligned} \quad (3.2)$$

where we denote by $P^r(I) = \text{span}\{1, t, \dots, t^r\}$ the space of polynomials up to degree $r \in \mathbb{N}$. Likewise, for discretization of the micro problems (2.33) we introduce partitionings I_n^P of $I^P = [0, 1]$ by defining

$$0 = t_n^0 < t_n^1 < \dots < t_n^{M_n} = 1, \quad k_n := t_n^m - t_n^{m-1}, \quad I_{n,m}^P := (t_n^{m-1}, t_n^m]. \quad (3.3)$$

While we allow for different micro discretizations in each macro step $n = 1, \dots, N$, we assume that each of them is uniform with step size k_n . We introduce

$$\begin{aligned} \mathcal{V}_{k;n}^\pi &:= \{\phi \in C(I^P; \mathbb{R}^d) : \phi|_{I_{n,m}^P} \in P^1(I_{n,m}^P)^d, \ m = 1, \dots, M_n, \ \phi(1) = \phi(0)\} \subset \mathcal{V}^\pi, \\ \mathcal{W}_{k;n} &:= \{\phi \in L^2(I^P; \mathbb{R}^d) : \phi|_{I_{n,m}^P} \in P^0(I_{n,m}^P)^d, \ m = 1, \dots, M_n\} \subset \mathcal{W}. \end{aligned} \quad (3.4)$$

Mostly, we will skip the index n if we refer to these micro spaces. Discretization is accomplished by restricting trial and test functions to the discrete function spaces $\mathcal{X}_K, \mathcal{Y}_K$ and $\mathcal{V}_{k;n}^\pi, \mathcal{W}_{k;n}$, respectively.

For general right hand sides $f(\cdot)$ and $g(\cdot)$, the integrals appearing in the variational formulations (2.31) and (2.33) cannot be evaluated exactly, since the trial spaces \mathcal{X}_K and $\mathcal{V}_{k;n}^\pi$ are piecewise linear and these functions are potentially nonlinear. Instead, we numerically approximate them with a summed two-point Gaussian quadrature rule, which we define by

$$\int_{I_K} f(t) dt := \sum_{n=1}^N \frac{K_n}{2} \left(f\left(\bar{T}_n - \frac{K_n}{\sqrt{12}}\right) + f\left(\bar{T}_n + \frac{K_n}{\sqrt{12}}\right) \right), \quad \bar{T}_n := \frac{T_{n-1} + T_n}{2}. \quad (3.5)$$

Integration on each micro partitioning I_n^P is defined in the same spirit. This quadrature rule is of fourth order, see [26], and it guarantees the additional higher order consistency error $\mathcal{O}(\epsilon K^4 + k_n^4)$ in contrast to all other error terms which are of order two. Hence

from here on we will neglect the conformity error coming from numerical quadrature on both scales.

Altogether, the fully discrete multiscale solution is described by the following problem formulation:

Problem 10 (Discretized variational formulation of the multiscale problem).

Find $Y_{K,k} \in \mathcal{X}_K$ such that

$$A_k(Y_{K,k}, \Phi) = 0 \quad \forall \Phi \in \mathcal{Y}_K, \quad (3.6)$$

$$A_k(Y, \Phi) := \int_0^T (Y'(t) - \epsilon \mathcal{F}_k(Y(t))) \cdot \Phi(t) dt, \quad \mathcal{F}_k(Y) := \int_0^1 f(Y, u_{k;Y}(s)) ds$$

where $\mathcal{X}_K, \mathcal{Y}_K$ are given in (3.2). For $t \in I_n$ and $Y = Y(t) \in \mathbb{R}^c$ fixed, the discrete periodic fast scale solution $u_{k;Y} \in \mathcal{V}_{k;n}^\pi$ is defined by

$$u_{k;Y} \in \mathcal{V}_{k;n}^\pi : B(Y; u_{k;Y}; \phi_k) = 0 \quad \forall \phi_k \in \mathcal{W}_{k;n}, \quad (3.7)$$

where the function spaces $\mathcal{V}_{k;n}^\pi$ and $\mathcal{W}_{k;n}$ are given in (3.4) and $B(\cdot)$ in (2.33).

Remark 11 (Efficiency of Galerkin discretizations). *The approximation of Galerkin time discretizations with high order quadrature rules (the approach that we describe) causes additional effort since at least two evaluations of all nonlinear operators and functions are required in each step. The closely related trapezoidal rule would only require one evaluation. However, the consistency error between the Galerkin approach and the trapezoidal rule is of the same order such that both approaches must be considered as separate discretization schemes. In [16, 17] we have demonstrated how the more efficient trapezoidal rule can be used for solving the problem while including the consistency error within the error estimator. It shows that the quadrature error is indeed of the same (or even higher) order than further contributions to the error estimator.*

We conclude by summarizing the algorithmic realization of the multiscale process. The discrete variational formulation (3.6) decouples into separate time steps on $I_n = [T_{n-1}, T_n]$. On each interval it remains to solve a nonlinear problem for $Y_n \in P^1(I_n; \mathbb{R}^c)$

$$Y_n(T_n) \cdot \Phi_n - \epsilon \int_{T_{n-1}}^{T_n} \mathcal{F}_k(Y_n(t)) dt \cdot \Phi_n = Y_{n-1}(T_{n-1}) \cdot \Phi_n \quad \forall \Phi_n \in P^0(I_n; \mathbb{R}^c). \quad (3.8)$$

Here, the integral is approximated by a two point Gaussian quadrature rule. In each quadrature point it is necessary to solve the periodic in time micro problem for $Y_{n,q} = Y_n(\chi_{n,q})$, $\chi_{n,q} \in I_n$ being, for $q = 1, 2$, the two Gauss points. The corresponding solutions $u_{Y_{n,q}}(t)$ are given by (3.7) and their approximation requires the solution of a periodic problem, see Remark 12.

To solve (3.8), we use an approximate Newton scheme. We approximate the Jacobian by neglecting the derivatives of the transfer operator \mathcal{F} with respect to the micro solution, i.e. we approximate

$$\tilde{\nabla}_Y \mathcal{F}(Y)(\delta Y) := \int_0^1 \nabla_Y f(Y, u_Y(s)) \cdot \delta Y ds,$$

and do not consider the partial derivative $\nabla_u f(Y, u_Y) \nabla_Y u_y$ which would require the approximation of a further periodic in time tangent problem. Numerical tests have shown that this approximation does not strongly worsen the convergence rate of the Newton scheme.

Remark 12 (Approximation of periodic solutions). *In each Gaussian quadrature point $\chi_{n,q} \in [T_{n-1}, T_n]$, periodic-in-time micro problems $u_{k;Y_{n,q}}(t)$ must be computed for the fixed slow scale variable $Y_{n,q} := Y_K(\chi_{n,q})$. These are approximated until the periodicity mismatch $\|u_{k;Y_{n,q}}(1) - u_{k;Y_{n,q}}(0)\| < \text{tol}_P$. This can either be done by simply letting the dynamic problem run into a cyclic state or by using different acceleration schemes, see [24, 22].*

4 Error estimation

We follow the framework of the dual weighted residual estimator (DWR) introduced in [2, 3]. We are interested in functional outputs $J : \mathcal{X} \rightarrow \mathbb{R}$ of the long scale problem. We aim at estimating the functional error $J(y) - J(Y_{K,k})$ between the analytic solution $y(t)$ given by (2.1)-(5.2) and the fully discrete multiscale approximation defined in Problem 10. In between, we must consider several approximation steps:

1. The averaging error **(EA)** introduced by deriving the averaged model problem, Problem 6

$$J(y) - J(Y_{K,k}) = \underbrace{(J(y) - J(Y))}_{\textbf{(EA)}} + \underbrace{(J(Y) - J(Y_{K,k}))}_{\textbf{(ED)}},$$

and the remaining discretization error **(ED)**, which is further split.

2. The error from Galerkin discretization **(EG)** of the averaged long term problem

$$J(y) - J(Y_{K,k}) = \underbrace{(J(y) - J(Y))}_{\textbf{(EA)}} + \underbrace{(J(Y) - J(Y_K))}_{\textbf{(EG)}} + \underbrace{(J(Y_K) - J(Y_{K,k}))}_{\textbf{(EF)}}$$

which also reveals **(EF)**, the error coming from discretizing the fast scale problem.

By Y_K we define the solution to the semidiscrete problem, which is discrete in terms of the long scale, e.g. $Y_K \in \mathcal{X}_K$, but which is based on the analytic transfer operator \mathcal{F} . This intermediate solution will enter the estimate as an analytical tool only.

While the averaging error **(EA)** is bound by the a priori estimate in Theorem 7, the remaining errors **(ED) = (EG) + (EF)** can be formulated as residual errors of a non conforming Galerkin formulation. Non conformity comes from the approximation of the transfer operator \mathcal{F} by \mathcal{F}_k . As outlined above, we have neglected the error coming from Gaussian quadrature since it is negligible.

The general framework of the dual weighted residual error estimator for such a non conforming discretization is discussed in [3, Section 2.3] or [21, Theorem 8.7]. An application to the multiscale scheme will require a nested application of the DWR method

to also take care of the error coming from approximating the transfer operator \mathcal{F} which implicitly depends on the fast scale contributions. We state the main result.

Theorem 13 (DWR estimator for the long term problem). *Let $I = [0, T]$ and let $y \in C^1(I; \mathbb{R}^c)$ be the solution to (2.1)-(5.2) and $Y_{K,k} \in \mathcal{X}_K$ be the fully discrete solution to Problem 10. Let $\text{tol}_P > 0$ be the tolerance for the approximation of the temporal periodicity in all micro problems, i.e. $\|u_Y(1) - u_Y(0)\| < \text{tol}_P$. Let $J : \mathcal{X} \rightarrow \mathbb{R}$ be three times differentiable. It holds*

$$\begin{aligned} J(y) - J(Y_{K,k}) &= -\frac{1}{2}A_k(Y_{K,k}, Z - i_Y Z) \\ &\quad + \frac{1}{2} \left(J'(Y_{K,k})(Y - i_X Y) - A'_k(Y_{K,k})(Z_{K,k}, Y - i_X Y) \right) \\ &\quad + \frac{1}{2} \epsilon \int_0^T \eta^\pi(Y_{K,k}(t)) \cdot (Z(t) + Z_{K,k}(t)) \, ds \\ &\quad + \mathcal{O}(\text{tol}_P) + \mathcal{O}(\epsilon) + \mathcal{O}(\epsilon k^2 + k^4 + K^4) + \mathcal{R}_K^{(3)} + \mathcal{R}_k^{(3)}, \end{aligned} \quad (4.1)$$

where $i_X : \mathcal{X} \rightarrow \mathcal{X}_K$ is the nodal interpolation into the space of piecewise linear polynomials and $i_Y : \mathcal{Y} \rightarrow \mathcal{Y}_K$ is the projection to the piecewise constants, $Z \in \mathcal{Y}$ and $Z_{K,k} \in \mathcal{Y}_K$ are the adjoint solutions to $A'(Y)(\Phi, Z) = J'(Y)(\Phi)$ for all $\Phi \in \mathcal{X}^0$ and $\Phi_K \in \mathcal{X}_K^0$, respectively. \mathcal{X}^0 and \mathcal{X}_K^0 differ from \mathcal{X} and \mathcal{X}_K in the sense that homogeneous initial values are realized. The fast scale error $\eta^\pi(Y_{K,k})$ is given by

$$\begin{aligned} \eta^\pi(Y) &:= \mathcal{O}(\text{tol}_P) + \frac{1}{2} \left(G(z_Y - i_W z_Y) - B(u_{k;Y}, z_Y - i_W z_Y) \right) \\ &\quad + \frac{1}{2} \left(J^{\pi'}(u_{k;Y})(u - i_V u) - B'(u_{k;Y})(u - i_V u, z_{k;Y}) \right) \end{aligned} \quad (4.2)$$

and the adjoint micro scale solutions $z_Y \in \mathcal{W}$ and $z_{k;Y} \in \mathcal{W}_k$ are defined for each fixed Y by $B'(u_Y)(\phi, z_Y) = \int_0^1 \nabla_u f(Y, u_Y(s)) \phi(s) \, ds$ for all $\phi \in \mathcal{V}^\pi$ and $\phi_k \in \mathcal{V}_k^\pi$, respectively. $i_V : \mathcal{V}^\pi \rightarrow \mathcal{V}_k^\pi$ and $i_W : \mathcal{W} \rightarrow \mathcal{W}_k$ are interpolation operators. By $\mathcal{R}_K^{(3)}$ and $\mathcal{R}_k^{(3)}$ we denote remainders which are of third order in the error.

Proof. The proof follows by combining Theorem 7, Lemma 14, 15 and Remark 16. Details on the adjoint problems are given in Section 4.2. \square

4.1 Derivation of the error estimator

The averaging error **(EA)** $J(y) - J(Y)$ is estimated by a priori arguments. Given a differentiable functional $J(y)$ it holds with Theorem 7 that

$$|(\mathbf{EA})| = |J(y) - J(Y)| = |J'(\zeta)(y - Y)| \leq \|J'(\zeta)\| \cdot \|y - Y\| = \mathcal{O}(\epsilon),$$

where ζ is an intermediate between y and Y . We turn our attention to the Galerkin error **(EG)** estimating $J(Y) - J(Y_{K,k})$.

Lemma 14 (DWR estimator of the averaged long term problem).

Let $Y \in \mathcal{X}$ be the solution to Problem 9 and $Y_{K,k} \in \mathcal{X}_K$ be the solution to Problem (10), $Z \in \mathcal{Y}$ and $Z_{K,k} \in \mathcal{Y}_K$ the adjoint solutions to (4.11). It holds

$$\begin{aligned} J(Y) - J(Y_{K,k}) &= \mathcal{R}_K^{(3)}(Y - Y_{K,k}, Z - Z_{K,k}) \\ &+ \frac{1}{2} \left(J'(Y_{K,k})(Y - i_{\mathcal{X}}Y) - A'_k(Y_{K,k})(Z_{K,k}, Y - i_{\mathcal{X}}Y) - A_k(Y_{K,k}, Z - i_{\mathcal{Y}}Z) \right) \\ &+ \frac{1}{2} \left([A'_k - A'](Y_{K,k})(Z_{K,k}, Y - Y_{K,k}) + [A_k - A](Y_{K,k}, Z + Z_{K,k}) \right), \end{aligned} \quad (4.3)$$

where $\mathcal{R}_K^{(3)}$ is of third order in the primal and adjoint discretization error.

Proof. We introduce one Lagrangian for the continuous and the semidiscrete model and one for the fully discrete model

$$L(Y, Z) := J(Y) - A(Y, Z), \quad L_k(Y, Z) := J(Y) - A_k(Y, Z).$$

For the solutions $Y \in \mathcal{X}$ and $Y_{K,k} \in \mathcal{X}_K$ it holds for all $Z \in \mathcal{Y}$ and $Z_K \in \mathcal{Y}_K$ that

$$\begin{aligned} J(Y) - J(Y_{K,k}) &= L(Y, Z) - L_k(Y_{K,k}, Z_K) \\ &= \underbrace{L(Y, Z) - L(Y_{K,k}, Z_K)}_{=(\mathbf{Ea})} + \underbrace{L(Y_{K,k}, Z_K) - L_k(Y_{K,k}, Z_K)}_{=(\mathbf{Eb})}. \end{aligned} \quad (4.4)$$

The first part **(Ea)** is a standard DWR term, which by defining $\mathbf{x} := (Y, Z)$ and $\mathbf{x}_{K,k} := (Y_{K,k}, Z_K)$, is approximated by writing the difference as integral over its derivative and by approximation with the trapezoidal rule, compare [3, Proposition 2.1]

$$(\mathbf{Ea}) = L(\mathbf{x}) - L(\mathbf{x}_{K,k}) = \frac{1}{2} \left(L'(\mathbf{x})(\mathbf{x} - \mathbf{x}_{K,k}) + L'(\mathbf{x}_{K,k})(\mathbf{x} - \mathbf{x}_{K,k}) \right) + \mathcal{R}_K^{(3)}(\mathbf{x} - \mathbf{x}_{K,k}),$$

where the remainder $\mathcal{R}_K^{(3)}(\mathbf{x} - \mathbf{x}_{K,k})$ is of third order in the error. The relation

$$L'(\mathbf{x})(\delta \mathbf{x}) = J'(Y)(\delta Y) - A'(Y, Z)(\delta Y) - A(Y, \delta Z)$$

shows that it holds $L'(\mathbf{x})(\delta \mathbf{x}) = L'(Y, Z)(\delta Y, \delta Z) = 0$ for the analytical solutions $Y, Z \in \mathcal{X} \times \mathcal{Y}$ and for all $\delta \mathbf{x} = (\delta Y, \delta Z) \in \mathcal{X} \times \mathcal{Y}$. We neglect the remainder $\mathcal{R}^{(3)}$ and approximate

$$(\mathbf{Ea}) \approx \frac{1}{2} \left(J'(Y_{K,k})(Y - Y_{K,k}) - A'(Y_{K,k})(Z_{K,k}, Y - Y_{K,k}) - A(Y_{K,k}, Z - Z_K) \right). \quad (4.5)$$

The forms $A(\cdot, \cdot)$ and $A'(\cdot)(\cdot, \cdot)$ are based on the non-discrete transfer operator \mathcal{F} . The discrete primal and adjoint solutions $Y_{K,k} \in \mathcal{X}_K$ and $Z_K \in \mathcal{Y}_K$ are however defined by using the discrete form $A_k(\cdot, \cdot)$. We insert $\pm A_k$ and $\pm A'_k$ in (4.5) such that we can apply Galerkin orthogonality to introduce interpolations $i_{\mathcal{X}} : \mathcal{X} \rightarrow \mathcal{X}_K$ and $i_{\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{Y}_K$

$$\begin{aligned} (\mathbf{Ea}) &\approx \frac{1}{2} \left(J'(Y_{K,k})(Y - i_{\mathcal{X}}Y) - A'_k(Y_{K,k})(Z_{K,k}, Y - i_{\mathcal{X}}Y) - A_k(Y_{K,k}, Z - i_{\mathcal{Y}}Z) \right) \\ &+ \frac{1}{2} \left([A'_k - A'](Y_{K,k})(Z_{K,k}, Y - Y_{K,k}) + [A_k - A](Y_{K,k}, Z - Z_K) \right). \end{aligned} \quad (4.6)$$

The notation $[A_k - A](Y, Z) := A_k(Y, Z) - A(Y, Z)$ is introduced for brevity. In (4.4), the second error component **(Eb)** is a conformity error and given as

$$\mathbf{(Eb)} = [A_k - A](Y_{K,k}, Z_K).$$

Together with (4.6) we obtain the postulated result. \square

The second line of (4.3) is the standard residual representation of the DWR error estimator. Given a reconstruction of the *weights* $Y - i_{\mathcal{X}}Y$ and $Z - i_{\mathcal{Y}}Z$ it can be evaluated numerically, we refer to Section 4.2 for details. The third line in (4.3) combines two conformity errors coming from the replacement of the transfer operator \mathcal{F} by its discrete counterpart \mathcal{F}_k . These terms will be discussed in the following paragraphs.

Lemma 15 (Primal conformity error). *Let the assumptions of Theorem 13 hold. For the primal conformity error it holds*

$$[A_k - A](Y_{K,k}, Z + Z_{K,k}) = \epsilon \int_0^T \eta^\pi(Y_{K,k}(t)) \cdot (Z(t) + Z_{K,k}(t)) dt + \epsilon \int_0^T \mathcal{R}_k^{(3)}(Y_{K,j}(t)) dt,$$

where $\eta^\pi(Y)$ is defined in (4.2) and where $\mathcal{R}_k^{(3)}$ is a remainder of third order in the error.

Proof. For ease of notation we introduce $\tilde{Z} := Z + Z_{K,k}$. This term does not carry any convergence properties and the full order of convergence must be reconstructed from the difference of the two forms $[A_k - A]$. Subtracting (2.31) from (3.6) gives

$$[A_k - A](Y_{K,k}, \tilde{Z}) = \epsilon \sum_{n=1}^N \int_{T_{n-1}}^{T_n} \left(\mathcal{F}(Y_{K,k}(t)) - \mathcal{F}_k(Y_{K,k}(t)) \right) \tilde{Z}(t) dt. \quad (4.7)$$

For the evaluation of this term we use a nested application of the DWR estimator, since the difference between the fast scale influences u_Y and $u_{k;Y}$ enters implicitly.¹

Now, let $t \in [0, T]$ be fixed and $Y := Y_{K,k}(t)$. We introduce

$$J^\pi(u_Y) := \int_0^1 f(Y, u_Y(s)) ds \quad (4.8)$$

such that $\mathcal{F}(Y_{K,k}(t)) - \mathcal{F}_k(Y_{K,k}(t)) = J^\pi(u_Y) - J^\pi(u_{k;Y})$, where by u_Y we denote the continuous periodic micro solution to (2.33) and by $u_{k;Y}$ the discrete solution to (3.7), which satisfies the periodicity approximately, i.e. $\|u_{k;Y}(1) - u_{k;Y}(0)\| < \text{tol}_P$. With the solution $z_{k;Y} \in \mathcal{W}_k$ to the adjoint micro problem

$$B'(u_{k;Y})(\psi_k, z_{k;Y}) = J^{\pi'}(u_Y)(\psi_k) \quad \forall \psi_k \in \mathcal{V}_k^\pi$$

¹A practical evaluation of this error term will require numerical quadrature of the integrals on the right hand side, e.g. by the 2-point Gauss rule. The error term $\mathcal{F}(Y_{K,k}(t)) - \mathcal{F}_k(Y_{K,k}(t))$ must hence be approximated in two points in each time step $[T_{n-1}, T_n]$.

we estimate in the usual DWR way, see Lemma [3, Section 2.3],

$$\begin{aligned} J^\pi(u_Y) - J^\pi(u_{k;Y}) &= \mathcal{R}_k^{(3)}(u_{k;Y})(u_Y - u_{k;Y}, z_Y - z_{k;Y}) - \frac{1}{2}B(u_{k;Y}, z_Y - i_{\mathcal{W}}z_Y) \\ &\quad + \frac{1}{2}\left(J^{\pi'}(u_{k;Y})(u_Y - i_{\mathcal{V}}u_Y) - B'(u_{k;Y})(u_Y - i_{\mathcal{V}}u_Y, z_{k;Y})\right) + \mathcal{O}(\text{tol}_P), \end{aligned} \quad (4.9)$$

where the tol_P -term arises from the disturbed Galerkin orthogonality. To clarify the impact of the approximated tolerance we give a sketch: assume that $u_{k;Y}^\pi(t)$ is the fully periodic solution, strictly satisfying $u_{k;Y}^\pi(1) = u_{k;Y}^\pi(0)$. Then,

$$\begin{aligned} B(u_{k;Y}; \phi_{K,k}) &= \underbrace{B(u_{k;Y}^\pi, \phi_{K,k})}_{=0} + B(u_{k;Y} - u_{k;Y}^\pi, \phi_{K,k}) \\ \Rightarrow |B(u_{k;Y}, \phi_{K,k})| &\leq c \|u_{k;Y} - u_{k;Y}^\pi\|_{L^\infty(I^P)} \|\phi_{K,k}\|_{L^\infty(I^P)} \leq c \text{tol}_P \|\phi_{K,k}\|_{L^\infty(I^P)}, \end{aligned}$$

where we estimated the periodicity error $\|u_{k;Y}^\pi - u_{k;Y}\| \leq \text{tol}_P$ by the imposed threshold.

Details on the adjoint solution $z_Y = z_Y \in \mathcal{W}$ and its discretization $z_{k;Y} \in \mathcal{W}_k$ entering (4.9) are discussed in Section 4.2. \square

Remark 16. The adjoint consistency error arising in Lemma 14 can be estimated as

$$\begin{aligned} &\left| [A'_k - A'](Y_{K,k})(Y - Y_{K,k}, Z_{K,k}) \right| \\ &\leq \epsilon \int_0^T \left| \left(\mathcal{F}'_k(Y_{K,k}(t)) - \mathcal{F}'(Y_{K,k}(t)) \right) \cdot (Y(t) - Y_{K,k}(t)) \cdot Z_{K,k}(t) \right| dt \\ &\leq \epsilon T \|Z_{K,k}\|_{L^\infty(I)} \|\mathcal{F}'_k(Y_{K,k}) - \mathcal{F}'(Y_{K,k})\|_{L^\infty(I)} \|Y - Y_{K,k}\|_{L^\infty(I)} \end{aligned} \quad (4.10)$$

In [12] we have shown a second order error estimator for the primal error in a comparable multiscale setting

$$\|Y - Y_{K,k}\|_{L^\infty(I)} = \mathcal{O}\left(\epsilon + \epsilon^2 K^2 + k^2\right).$$

The first term in (4.10) is bounded, since the adjoint problem $A'(Y)(\Phi, Z) = J'(Y)(\Phi)$, going backward in time, is equivalent to

$$-Z'(t) - \epsilon \mathcal{F}'(Y(t))Z(t) = 0, \quad Z(T) = 1 \quad \Rightarrow \quad Z(t) = \exp\left(\epsilon \int_T^t \mathcal{F}'(Y(s)) ds\right),$$

which is bounded for $0 \leq \epsilon t \leq \epsilon T = \mathcal{O}(1)$, since the adjoint transfer operator $\mathcal{F}'(Y)$ is bounded. The remaining term in (4.10) measures the discretization error in the adjoint fast scale problem. With arguments similar to those used in the proof to Lemma 15, second order convergence in k can be shown. Overall, the adjoint consistency error is of higher order $\left| [A'_k - A'](Y_{K,k})(Y - Y_{K,k}, Z_{K,k}) \right| = \mathcal{O}\left(\epsilon k^2 + \epsilon^2 k^2 K^2 + k^4\right)$.

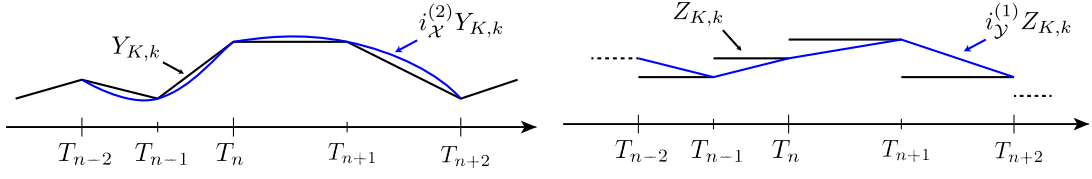


Figure 1: Reconstructing a higher order approximation from the discrete solutions. By $i_{\mathcal{X}}^{(2)}$ (left) we piecewise quadratic reconstruction on the mesh with twice the mesh spacing and by $i_{\mathcal{Y}}^{(1)}$ (right) the linear reconstruction on the same mesh.

4.2 Adjoint problems and evaluation of the error estimator

The error estimator (4.1) depends on the adjoint solution $Z \in \mathcal{Y}$ and also on the adjoint micro scale solutions $z_Y \in \mathcal{W}$. We shortly sketch the steps required to approximate these adjoint solutions, as the multiscale framework will require a nested approach. For $Y \in \mathcal{X}$ given, $Z \in \mathcal{Y}$ is defined as solution to

$$\int_0^T (\Psi'(t) - \epsilon \nabla_Y \mathcal{F}(Y(t))(\Psi(t))) \cdot Z(t) dt = J'(Y)(\Psi) \quad \forall \Psi \in \mathcal{X}. \quad (4.11)$$

The derivative of the transfer operator is given by

$$\nabla_Y \mathcal{F}(Y)(\Psi) = \int_0^1 \left(\nabla_Y f(Y(t), u_Y(s)) + \nabla_u f(Y(t), u_Y(s)) (D_Y u_Y(s)) \right) \Psi(t) ds. \quad (4.12)$$

While the first part involving the derivative in direction of Y is directly accessible, evaluation of the second term requires a further tangent solution $D_Y u_Y \in \mathcal{V}^\pi$, the derivative of the time periodic solution $u_Y(s)$ with respect to Y . It is given as the solution to

$$B'(u_Y)(D_Y u_Y, \phi) = 0 \quad \forall \phi \in \mathcal{W}. \quad (4.13)$$

Finally, to estimate the conformity error introduced by replacing the transfer operator \mathcal{F} by its discrete counterpart a further adjoint micro scale solution $z_Y \in \mathcal{W}$, given by the following periodic in time problem must be solved

$$B'(u_Y)(\psi, z_Y) = J^{\pi'}(u_Y)(\psi) \quad \forall \psi \in \mathcal{V}^\pi, \quad (4.14)$$

where $J^\pi(u_Y) = \int_0^1 f(Y, u_Y(s)) ds$.

The a posteriori error estimator presented in Theorem 13 cannot be evaluated exactly since it depends on the unknown exact solutions $Y \in \mathcal{X}$ and $Z \in \mathcal{Y}$. Further, several higher order remainders appear, which are simply omitted. To approximate primal and dual residuals weights $Y - i_{\mathcal{X}} Y$ and $Z - i_{\mathcal{Y}} Z$ and also to approximate the sum of continuous and discrete adjoint solution $Z + Z_{K,k}$ we use the usual reconstruction mechanism that is based on computing $Y_{K,k} \in \mathcal{X}_K$ and $Z_{K,k} \in \mathcal{Y}_K$ and applying a higher order interpolation by reinterpreting the piecewise linear function $Y_{K,k}$ as piecewise quadratic

and the piecewise constant function $Z_{K,k}$ as piecewise linear. Fig. 1 illustrates this procedure. We ensure that all macro meshes have a patch structure: two adjacent intervals I_{2n-1} and I_{2n} each have the size $K_{2n-1} = K_{2n}$. The micro meshes are uniform.

This reconstruction of the weights must be considered a computational tool for approximating the functional error. It cannot however give rigorous upper and lower bounds. For general details on this reconstruction we refer to [3, 23] and in particular to [16] in the context of temporal Galerkin schemes.

5 Numerical examples

Problem 17. On $I = [0, T]$ with $T = 6 \cdot 10^5$ find $y \in C^1(I)$ and $u \in C^2(I)$ such that

$$\begin{aligned} y'(t) &= \epsilon f(y(t), u(t)), & y(0) &= 0, \\ u''(t) + \frac{1}{2}u'(t) + \gamma(y(t))u(t) &= \sin(2\pi t), & u(0) &= u_0, \quad u'(0) = u'_0, \end{aligned} \quad (5.1)$$

with the scale separation parameter $\epsilon = 10^{-6}$ and

$$f(y, u) := \frac{1}{(1+y)(1+64u^2)}, \quad \gamma(y) := (4\pi^2 + 32(y-1)). \quad (5.2)$$

As functional of interest we consider the slow scale component at final time T

$$J(y) = y(T).$$

We produce reference values for the functional output $J(y) = y(T)$ by resolved simulations based on a direct discretization of Problem 17 with the trapezoidal rule using a small time step size k over the full period of time $I = [0, T]$. Extrapolating $k \rightarrow 0$ shows the experimental order of convergence $\mathcal{O}(k^{2.0015})$ and for all further comparisons we set the reference value to

$$J(y_{ref}) := 1.08704164. \quad (5.3)$$

We first show that this problem fits into the framework introduced in Section 2.

Lemma 18. Problem 17 satisfies Assumptions 2 and 4.

Proof. For $f(t, y)$ it holds $|f(t, y)| \leq 1$ and, for $y, Y \in \mathbb{R}^c$

$$|f(y, u) - f(Y, u)| = \frac{|Y - y|}{|1+y| \cdot |1+Y| \cdot |1+64u^2|} \leq |y - Y|,$$

as well as

$$|f(y, u) - f(y, U)| = \frac{64|u+U| \cdot |u-U|}{|1+y| \cdot |1+64u^2| \cdot |1+64U^2|} \leq 8|u - U|,$$

which shows boundedness and Lipschitz continuity with $C_{A2.2} = 1$ and $C_{A2.3} = 8$.

Next we reformulate the microscale problem as a first order system in $v(t) := (u_1(t), u_2(t))$, with $u_1(t) = u(t)$ and $u_2(t) = u'(t)$

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}'(t) = \begin{pmatrix} 0 & 1 \\ -\gamma(y) & -\frac{3}{5} \end{pmatrix} \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ \sin(2\pi t) \end{pmatrix} =: G(t, y, v(t)) \quad (5.4)$$

The Jacobian $\nabla_u G$ has the eigenvalues $\lambda_{1/2} = -\frac{3}{10} \pm \frac{\sqrt{9-100\gamma(y)}}{10}$, which for the specific choice of $\gamma(y)$ have strictly negative real part with $\lambda_{A2.6} = \frac{3}{10}$. This shows that there exists a unique solution which is bound by the initial values and the right hand side $g(t) = \sin(2\pi t)$

$$\|v(t)\| \leq C \left(\|v(0)\| + \frac{1}{\lambda_{A2.6}} \right), \quad (5.5)$$

with a constant $C = C(y)$ which depends on the eigenvectors of $\nabla_u G$, hence on $y \in X^c$. Furthermore, for $y \in X^c$ fixed there exists a unique periodic solution, since, for each initial value the difference $v(t+1) - v(t)$ will decay to zero. We will denote such a periodic solution by $v_y(t)$. And since this solution is also reached for the initial $v(0) = 0$, estimate (5.5) gives

$$\|v_y(t)\| \leq C \frac{1}{\lambda_{A2.6}}. \quad (5.6)$$

Having two such periodic solution $v_1(t)$ and $v_2(t)$ belonging to the parameters $y_1, y_2 \in X^c$ satisfy

$$v_1'(t) - v_2'(t) = \begin{pmatrix} 0 & 1 \\ -\gamma(y_1) & -\frac{3}{5} \end{pmatrix} (v_1(t) - v_2(t)) + \begin{pmatrix} 0 & 0 \\ \gamma(y_1) - \gamma(y_2) & 0 \end{pmatrix} v_2(t).$$

Considered as equations for the difference $v_1(t) - v_2(t)$ this corresponds to problem (5.4) with $g(t)$ replaced by the periodic function $(\gamma(y_1) - \gamma(y_2))v_2(t)$ such that (5.6) yields

$$\|v_1(t) - v_2(t)\| \leq \frac{C}{\lambda_{A2.6}} \sup_{t \in I^P} \|v_2(t)\| \cdot \|\gamma(y_1) - \gamma(y_2)\| \leq \frac{32C^2}{\lambda_{A2.6}^2} \|y_1 - y_2\|.$$

□

5.1 Convergence of the multiscale algorithm

We start by analyzing the convergence of the multiscale scheme by running simulations with different but uniform time step sizes for K and k specified by

$$K_i = 100\,000 \cdot 2^{-i}, \quad k_j = 0.1 \cdot 2^{-j}, \quad i, j \in \{0, 1, 2, 3, 4, 5\}. \quad (5.7)$$

In Fig. 2 we show convergence with respect to the small time step size k (left) and with respect to the large time step size K (right). In both cases second order convergence is obtained as long as the step size under investigation is dominant. Furthermore, the results show that the range of chosen step sizes (5.7) is balanced with a slight dominance of the small scale error depending on k . The raw data is also given in Table 1.

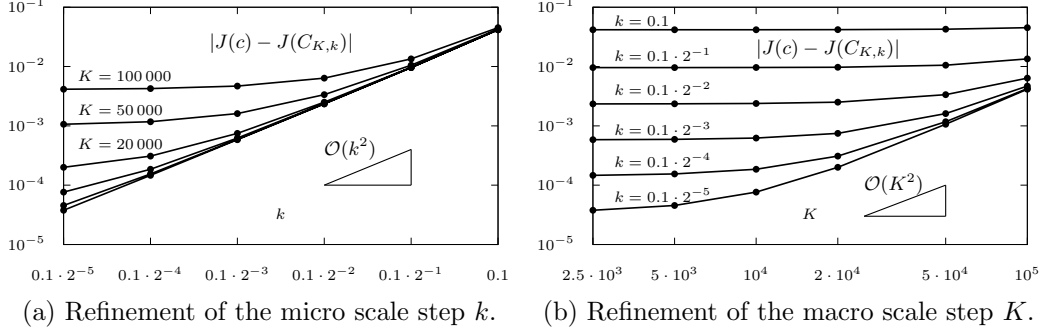


Figure 2: Error with respect to the small step size k and the large step size K . Left: each line represents a fixed value of K . Right: each line takes k fixed.

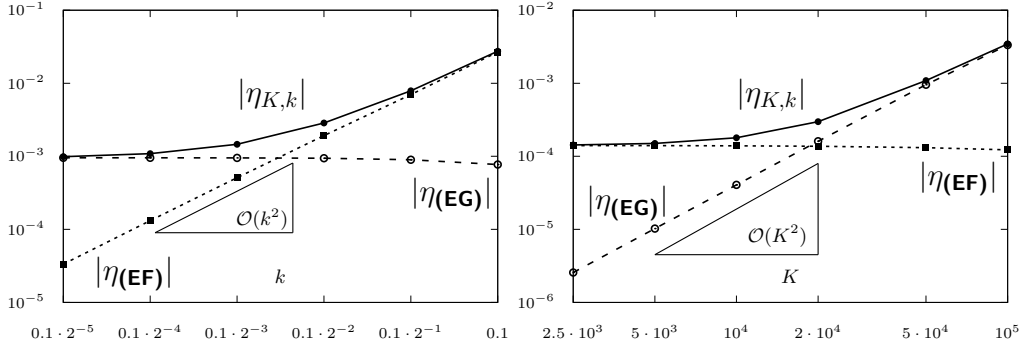


Figure 3: Error estimator $\eta_{K,k}$ and the parts $\eta(\text{EG})$ and $\eta(\text{EF})$ that make it up. Left: fixed macro step size $K = 50\,000$, and right: fixed micro step size $k = 0.00625$.

5.2 Evaluation of the error estimator

Next, we analyze the quality of the a posteriori error estimator derived in the previous section. We will show that this error estimator is accurate in predicting $J(y) - J(Y_{K,k})$ for the complete range of step sizes shown in (5.7). The smallest step sizes reach $K_{\min} = 2\,500$ and $k_{\min} = 0.003125$ such that the discretization error $\mathcal{O}(\epsilon^2 K^2 + k^2)$ will still dominate the averaging error $\mathcal{O}(\epsilon)$. The tolerance for reaching periodicity is set to $\text{tol}_p = 10^{-9}$. The raw values indicating $J(y_{\text{ref}}) - J(Y_{K,k})$, overall error estimator $\eta_{K,k}$ and its splitting into discretization error $\eta(\text{EG})$, primal conformity error (fine scale error) $\eta(\text{EF})$ adjoint conformity error (which is of higher order) $\eta(\text{EF})'$ are shown in Table 1. Finally, we also present the efficiency of the error estimator by indicating the effectivity index

$$\text{eff}_{K,k} = 100\% \cdot \frac{\eta_{K,k}}{J(y_{\text{ref}}) - J(Y_{K,k})}. \quad (5.8)$$

Values above 100% show an overestimation of the error, values below 100% an underestimation. The results collected in Table 1 however show a highly robust estimation for all combinations of small and large step sizes.

k	K	$J(y_{ref}) - J(Y_{K,k})$	$\eta_{K,k}$	$\eta(\text{EG})$	$\eta(\text{EF})$	$\eta(\text{EF})'$	eff $_{K,k}$
0.1	100000	$4.52 \cdot 10^{-2}$	$2.81 \cdot 10^{-2}$	$2.74 \cdot 10^{-3}$	$2.55 \cdot 10^{-2}$	$-1.77 \cdot 10^{-4}$	62.1%
	50000	$4.26 \cdot 10^{-2}$	$2.75 \cdot 10^{-2}$	$7.71 \cdot 10^{-4}$	$2.68 \cdot 10^{-2}$	$-3.36 \cdot 10^{-5}$	64.6%
	20000	$4.19 \cdot 10^{-2}$	$2.77 \cdot 10^{-2}$	$1.30 \cdot 10^{-4}$	$2.76 \cdot 10^{-2}$	$-5.30 \cdot 10^{-6}$	66.3%
	10000	$4.18 \cdot 10^{-2}$	$2.79 \cdot 10^{-2}$	$3.29 \cdot 10^{-5}$	$2.79 \cdot 10^{-2}$	$-1.23 \cdot 10^{-6}$	66.9%
	5000	$4.17 \cdot 10^{-2}$	$2.81 \cdot 10^{-2}$	$8.28 \cdot 10^{-6}$	$2.81 \cdot 10^{-2}$	$-3.07 \cdot 10^{-7}$	67.2%
	2500	$4.17 \cdot 10^{-2}$	$2.81 \cdot 10^{-2}$	$2.08 \cdot 10^{-6}$	$2.81 \cdot 10^{-2}$	$-7.67 \cdot 10^{-8}$	67.4%
0.05	100000	$1.34 \cdot 10^{-2}$	$9.64 \cdot 10^{-3}$	$3.12 \cdot 10^{-3}$	$6.52 \cdot 10^{-3}$	$-9.50 \cdot 10^{-6}$	71.7%
	50000	$1.06 \cdot 10^{-2}$	$7.90 \cdot 10^{-3}$	$8.99 \cdot 10^{-4}$	$7.00 \cdot 10^{-3}$	$-2.56 \cdot 10^{-6}$	74.8%
	20000	$9.74 \cdot 10^{-3}$	$7.45 \cdot 10^{-3}$	$1.51 \cdot 10^{-4}$	$7.30 \cdot 10^{-3}$	$-2.70 \cdot 10^{-7}$	76.5%
	10000	$9.63 \cdot 10^{-3}$	$7.44 \cdot 10^{-3}$	$3.83 \cdot 10^{-5}$	$7.41 \cdot 10^{-3}$	$-6.62 \cdot 10^{-8}$	77.3%
	5000	$9.60 \cdot 10^{-3}$	$7.47 \cdot 10^{-3}$	$9.64 \cdot 10^{-6}$	$7.46 \cdot 10^{-3}$	$-1.63 \cdot 10^{-8}$	77.8%
	2500	$9.59 \cdot 10^{-3}$	$7.48 \cdot 10^{-3}$	$2.42 \cdot 10^{-6}$	$7.48 \cdot 10^{-3}$	$-4.05 \cdot 10^{-9}$	78.0%
0.025	100000	$6.39 \cdot 10^{-3}$	$5.07 \cdot 10^{-3}$	$3.28 \cdot 10^{-3}$	$1.78 \cdot 10^{-3}$	$2.80 \cdot 10^{-6}$	79.3%
	50000	$3.36 \cdot 10^{-3}$	$2.86 \cdot 10^{-3}$	$9.41 \cdot 10^{-4}$	$1.92 \cdot 10^{-3}$	$7.57 \cdot 10^{-8}$	85.2%
	20000	$2.51 \cdot 10^{-3}$	$2.16 \cdot 10^{-3}$	$1.58 \cdot 10^{-4}$	$2.00 \cdot 10^{-3}$	$3.70 \cdot 10^{-8}$	86.3%
	10000	$2.38 \cdot 10^{-3}$	$2.07 \cdot 10^{-3}$	$4.01 \cdot 10^{-5}$	$2.03 \cdot 10^{-3}$	$1.04 \cdot 10^{-8}$	87.0%
	5000	$2.35 \cdot 10^{-3}$	$2.06 \cdot 10^{-3}$	$1.01 \cdot 10^{-5}$	$2.05 \cdot 10^{-3}$	$2.67 \cdot 10^{-9}$	87.5%
	2500	$2.35 \cdot 10^{-3}$	$2.06 \cdot 10^{-3}$	$2.53 \cdot 10^{-6}$	$2.05 \cdot 10^{-3}$	$6.71 \cdot 10^{-10}$	87.7%
0.0125	100000	$4.68 \cdot 10^{-3}$	$3.80 \cdot 10^{-3}$	$3.32 \cdot 10^{-3}$	$4.74 \cdot 10^{-4}$	$1.12 \cdot 10^{-6}$	81.1%
	50000	$1.61 \cdot 10^{-3}$	$1.46 \cdot 10^{-3}$	$9.53 \cdot 10^{-4}$	$5.09 \cdot 10^{-4}$	$9.70 \cdot 10^{-8}$	90.9%
	20000	$7.46 \cdot 10^{-4}$	$6.92 \cdot 10^{-4}$	$1.60 \cdot 10^{-4}$	$5.32 \cdot 10^{-4}$	$2.22 \cdot 10^{-8}$	92.7%
	10000	$6.23 \cdot 10^{-4}$	$5.79 \cdot 10^{-4}$	$4.06 \cdot 10^{-5}$	$5.39 \cdot 10^{-4}$	$5.86 \cdot 10^{-9}$	93.1%
	5000	$5.92 \cdot 10^{-4}$	$5.53 \cdot 10^{-4}$	$1.02 \cdot 10^{-5}$	$5.43 \cdot 10^{-4}$	$1.48 \cdot 10^{-9}$	93.4%
	2500	$5.84 \cdot 10^{-4}$	$5.47 \cdot 10^{-4}$	$2.56 \cdot 10^{-6}$	$5.44 \cdot 10^{-4}$	$3.72 \cdot 10^{-10}$	93.7%
0.00625	100000	$4.26 \cdot 10^{-3}$	$3.46 \cdot 10^{-3}$	$3.33 \cdot 10^{-3}$	$1.22 \cdot 10^{-4}$	$3.16 \cdot 10^{-7}$	81.2%
	50000	$1.17 \cdot 10^{-3}$	$1.09 \cdot 10^{-3}$	$9.55 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$	$3.37 \cdot 10^{-8}$	92.5%
	20000	$3.09 \cdot 10^{-4}$	$2.98 \cdot 10^{-4}$	$1.60 \cdot 10^{-4}$	$1.37 \cdot 10^{-4}$	$7.13 \cdot 10^{-9}$	96.2%
	10000	$1.86 \cdot 10^{-4}$	$1.80 \cdot 10^{-4}$	$4.07 \cdot 10^{-5}$	$1.39 \cdot 10^{-4}$	$1.86 \cdot 10^{-9}$	96.9%
	5000	$1.55 \cdot 10^{-4}$	$1.50 \cdot 10^{-4}$	$1.02 \cdot 10^{-5}$	$1.40 \cdot 10^{-4}$	$4.70 \cdot 10^{-10}$	97.2%
	2500	$1.47 \cdot 10^{-4}$	$1.43 \cdot 10^{-4}$	$2.57 \cdot 10^{-6}$	$1.40 \cdot 10^{-4}$	$1.18 \cdot 10^{-10}$	97.4%
0.003125	100000	$4.15 \cdot 10^{-3}$	$3.37 \cdot 10^{-3}$	$3.34 \cdot 10^{-3}$	$3.11 \cdot 10^{-5}$	$8.27 \cdot 10^{-8}$	81.2%
	50000	$1.07 \cdot 10^{-3}$	$9.90 \cdot 10^{-4}$	$9.56 \cdot 10^{-4}$	$3.34 \cdot 10^{-5}$	$9.58 \cdot 10^{-9}$	92.8%
	20000	$2.00 \cdot 10^{-4}$	$1.95 \cdot 10^{-4}$	$1.61 \cdot 10^{-4}$	$3.48 \cdot 10^{-5}$	$1.98 \cdot 10^{-9}$	97.6%
	10000	$7.64 \cdot 10^{-5}$	$7.60 \cdot 10^{-5}$	$4.07 \cdot 10^{-5}$	$3.53 \cdot 10^{-5}$	$5.14 \cdot 10^{-10}$	99.5%
	5000	$4.55 \cdot 10^{-5}$	$4.58 \cdot 10^{-5}$	$1.03 \cdot 10^{-5}$	$3.55 \cdot 10^{-5}$	$1.30 \cdot 10^{-10}$	100.7%
	2500	$3.77 \cdot 10^{-5}$	$3.82 \cdot 10^{-5}$	$2.57 \cdot 10^{-6}$	$3.57 \cdot 10^{-5}$	$3.25 \cdot 10^{-11}$	101.3%

Table 1: Functional error $J(Y_{K,k}) - J(y_{ref})$ and error estimator $\eta_{K,k}$ for different step sizes for the large scale and small scale problem. $\eta(\text{EG})$, $\eta(\text{EF})$ and $\eta(\text{EF})'$ are the contributions of the estimator $\text{eff}_{K,k}$ the effectivity, compare (5.8). All values are rounded to the first three relevant digits.

The analysis of the different contribution shows that $\eta_{(\mathbf{EG})}$ indicates the long term error depending mostly on K and $\eta_{(\mathbf{EF})}$ indicates the short term error depending on k , both converging with order two. The adjoint consistency error $\eta_{(\mathbf{EF})'}$ shows higher order convergence $O(\epsilon k^2 + \epsilon^2 k^2 K^2 + k^4)$ as stated in Remark 16 and hence, it can be neglected.

For $K = 50\,000$ fixed and varying k and for $k = 0.00625$ fixed and varying K respectively, Figure 3 shows the separation of the error estimator into long term and short term influences each. These results motivate to use $\eta_{(\mathbf{EF})}$ and $\eta_{(\mathbf{EG})}$ for controlling an adaptive procedure to find an optimally balanced discretization K, k .

5.3 Adaptive control

We write the error estimator as a sum over the subdivisions of the long time horizon I . By doing this we can quantify the error contribution of each subdivision. For the error contribution of element $I_n = (T_{n-1}, T_n]$, the error contribution is comprised of two parts. The error discretization of the averaged long term problem

$$\begin{aligned} \eta_{(\mathbf{EG})}^n := & -\frac{1}{2} A_{k|I_n}(Y_{K,k}, Z - i_Y Z) \\ & + \frac{1}{2} \left(J'_{|I_n}(Y_{K,k})(Y - i_X Y) - A'_{k|I_n}(Y_{K,k})(Z_{K,k}, Y - i_X Y) \right) \end{aligned} \quad (5.9)$$

and the error discretization of the fast scale problem.

$$\eta_{(\mathbf{EF})}^n := \frac{1}{2} \epsilon \int_{I_n} \eta^\pi(Y_{K,k}(t)) \cdot (Z(t) + Z_{K,k}(t)) \, ds. \quad (5.10)$$

We introduce the following method for refining macro scale and micro scale:

Algorithm 19 (Adaptive refinement). *Let an initial subdivision $I_K^{(1)}$ into $N^{(1)}$ macro steps be given with uniform but possibly distinct partitions $I_n^{P,(1)}$ for each subdivision $n = 1, \dots, N^{(1)}$. Let $\beta \in \mathbb{R}$ with $\beta \approx 1$. Iterate for $l = 1, 2, \dots$*

1. *Compute $\eta_{(\mathbf{EG})}^n$ and $\eta_{(\mathbf{EF})}^n$ for each $n = 1, \dots, N^{(l)}$.*

2. *Calculate the average*

$$\bar{\eta}^{(l)} := \frac{1}{N^{(l)}} \sum_{n=1}^{N^{(l)}} \left(|\eta_{(\mathbf{EG})}^n| + |\eta_{(\mathbf{EF})}^n| \right) \quad (5.11)$$

3. *For each $n = 1, \dots, N^{(l)}$: if $|\eta_{(\mathbf{EG})}^n| + |\eta_{(\mathbf{EF})}^n| > \beta \cdot \bar{\eta}^{(l)}$, we refine this cell:*

- a) *If $|\eta_{(\mathbf{EG})}^n| > \beta |\eta_{(\mathbf{EF})}^n|$ we refine $I_n = (T_{n-1}, T_n]$ into two intervals $(T_{n-1}, T_n^*]$ and $(T_n^*, T_n]$ where T_n^* is the midpoint of I_n . $I_n^{P,(l)}$ is kept for both new steps.*
- b) *If $|\eta_{(\mathbf{EF})}^n| > \beta |\eta_{(\mathbf{EG})}^n|$ refine the subdivision $I_n^{P,(l)}$ by cutting the step size in half.*
- c) *Otherwise refine I_n and I_n^P according to 3.a) and 3.b).*

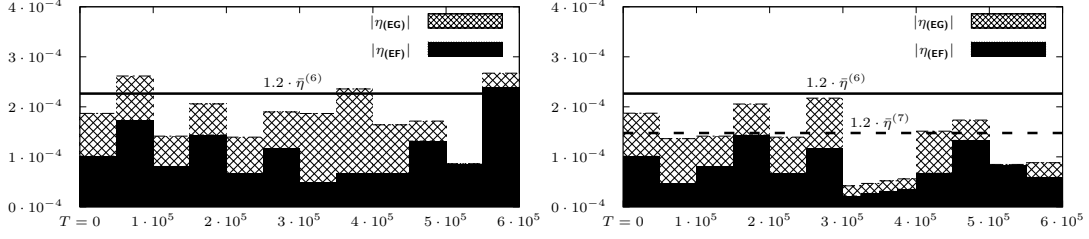


Figure 4: Absolute values of the error estimator $|\eta_{(\text{EG})}^n| + |\eta_{(\text{EF})}^n|$ for all macro steps before (left) and after (right) the sixth step of adaptive refinement. The bold line indicates the refinement threshold for iteration 6, the dashed line indicates iteration 7.

We illustrate the functionality of Algorithm 19 starting with $I_K^{(1)}$ with $K = 50\,000$ and $k = 0.05$ on each $I_n^{P,(1)}$.

In Figure 4 we discuss the sixth refinement step of Algorithm 19 in detail. The upper figure shows the error estimator $\eta_{K,k}^{(6)}$ and its partitioning into $\eta_{(\text{EG})}$ and $\eta_{(\text{EF})}$ for each of the 12 macro steps (there has been no refinement of K in the first 5 iterations). The bold line indicates the tolerance for refinement, i.e. $\beta \cdot \bar{\eta}^{(6)}$ for $\beta = 1.2$. Three steps exceed this limit and will be refined. In $I_2^{(6)}$ and $I_{12}^{(6)}$ the micro scale error is dominating and *Step 3.b*) is applied, in $I_8^{(6)}$ the dominance of the macro scale error leads to a refinement on the K -scale according to *Step 3.a*). To keep the patch structure of the macro mesh we also refine $I_7^{(6)}$. The resulting discretization and the error estimator in the next step is shown in the lower plot.

The adaptive algorithm roughly balances the error contributions coming from macro error and micro error over the first couple of steps, see Fig. 5c for details. In Fig. 5a we further plot the effectivity index (5.8) on this sequence of adaptively refined meshes and show that the error estimator still gains accuracy for increased resolution in k and K .

Refinement in Algorithm 19 is based on the absolute values of the local error contributions $|\eta_{(\text{EF})}^n|$ and $|\eta_{(\text{EG})}^n|$ and we introduce the indicator index

$$\text{ind}_{K,k} := \sum_{n=1}^N \frac{|\eta_{(\text{EG})}^n| + |\eta_{(\text{EF})}^n|}{|J(y_{\text{ref}}) - J(Y_{K,k})|}. \quad (5.12)$$

Figs. 5b and 5d show values close to one and suggest no significant overestimation, neither in the complete error or in the single parts.

To measure the computational effort on locally refined discretizations we count the overall number of time steps to be computed in the macro and the micro problem:

$$E_{K,k}^{(l)} := \sum_{n=1}^{N^{(l)}} \left(1 + \frac{1}{k_n^{(l)}} \right). \quad (5.13)$$

We do not take into account that multiple iterations are required within the Newton solver and that multiple cycles must be repeated for finding a periodic solution. The

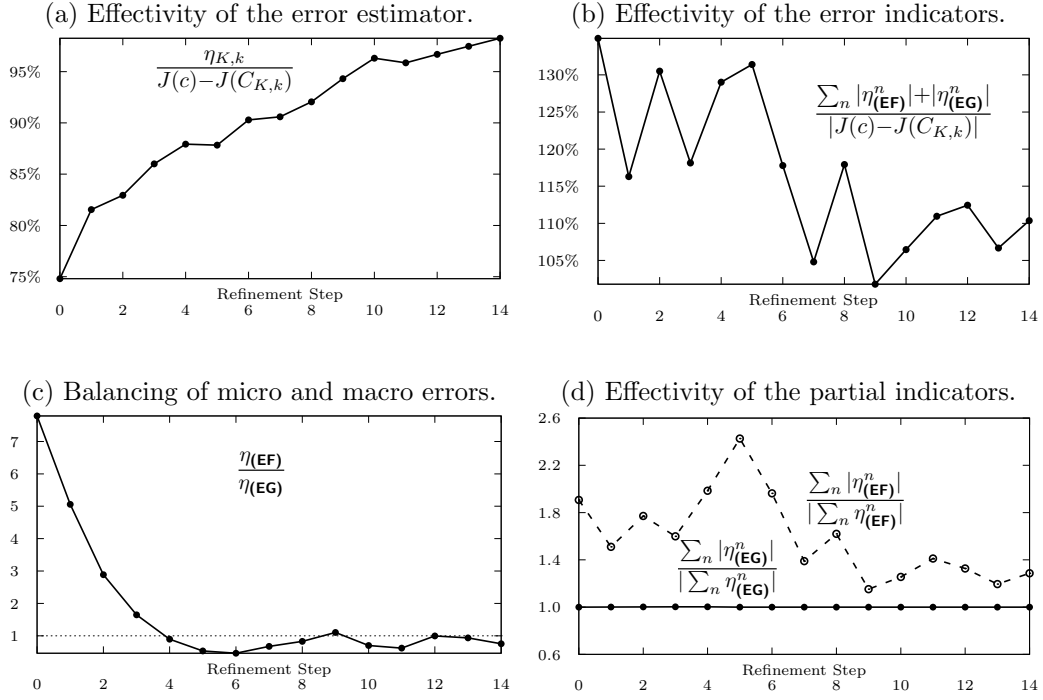


Figure 5: Performance on adaptive meshes: Effectivity (5.8) (top/left), indicator-effectivity (5.12) (right) and trend towards balancing error contributions (bottom/right).

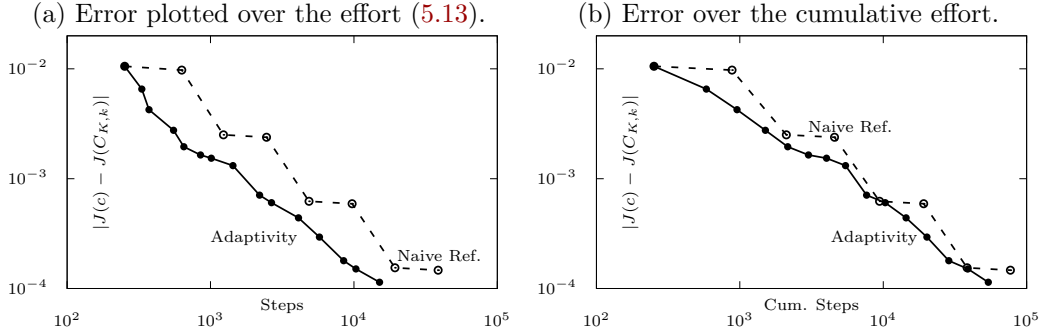


Figure 6: Comparison of the accuracy on uniform mesh refinement (alternately in micro and macro problem) and adaptive meshes.

computation of the dual solution requires roughly the same effort, since the scheme runs backwards in time and also calls for the solution of periodic in time micro problem. With these solutions one can compute the error estimator. Figure 6 shows the error $J(y_{ref}) - J(Y_{K,k})$ on sequences of adaptive and uniform meshes plotted over the effort (5.13).

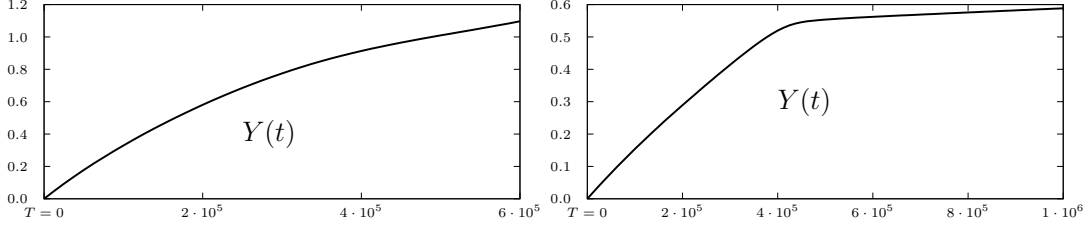


Figure 7: Solution $Y(t)$ of Problem 17 (left) and Problem 20 (right).

Adaptivity gives a slight advantage for the adaptive discretization. Since the regularity of the solution is very high, significant local effects cannot be expected.

5.4 Test case with pronounced local behavior

As a second test case we consider the following slightly modified problem that shows a more pronounced dependency of the slow scale on the fast scale solution. Here, we expect a larger benefit of local mesh adaptivity. Fig. 7 shows the averaged slow scale solution $Y(t)$ for both test cases.

Problem 20. On $I = [0, 10^6]$ find $y \in C^1(I)$ and $u \in C^2(I)$ such that

$$\begin{aligned} y'(t) &= \epsilon f(y(t), u(t)), & y(0) &= 0, \\ u''(t) + \frac{1}{2}u'(t) + \gamma(y(t))u(t) &= \sin(2\pi t), & u(0) &= u_0, \quad u'(0) = u'_0, \end{aligned}$$

with the scale separation parameter $\epsilon = 10^{-6}$ and

$$f(y(t), u(t)) := \frac{\tanh(500u(t)^2 - 5) + 1.01}{1 + y(t)}, \quad \gamma(y) := 20 \tanh(-10y(t) + 6) + 21.$$

Similar to Lemma 18 we can show that this problem also falls into the general framework discussed in this paper. The proof follows that of Lemma 18 line by line.

Lemma 21. Problem 20 satisfies Assumptions 2 and 4.

Again we produce reference values for the functional output $J(y) = y(10^6)$ by resolved simulations based on a direct discretization of Problem 20 with the trapezoidal rule using a small time step size k over the full period of time $I = [0, 10^6]$. Extrapolating $k \rightarrow 0$ shows the experimental order of convergence $\mathcal{O}(k^{2.0014})$ as expected and for further comparisons we set the reference value to

$$J(y_{ref}) := 0.59223654. \tag{5.14}$$

The solution to this problem alters its character at $t \approx 4.5 \cdot 10^5$, see Fig. 7. Due to the sigmoid nature of $f(y, u)$ in the second argument we have a sudden decay in the rate of change of Y . The contributions to the error are also concentrated around this point of

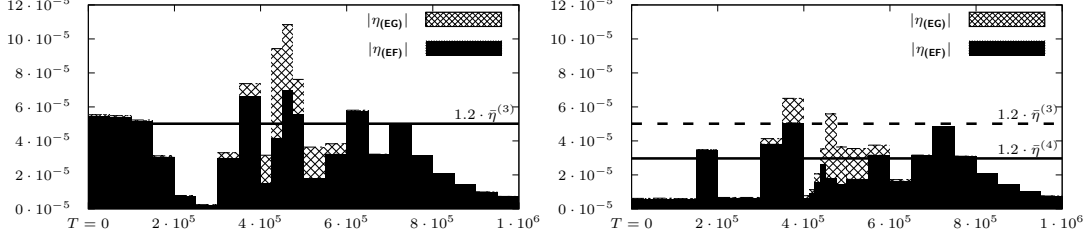


Figure 8: Absolute values of the error estimator $|\eta_{(\text{EG})}^n| + |\eta_{(\text{EF})}^n|$ for all macro steps before (left) and after (right) the third step of adaptive refinement. The bold line indicates the refinement threshold for iteration 3, the dashed line indicates iteration 4.

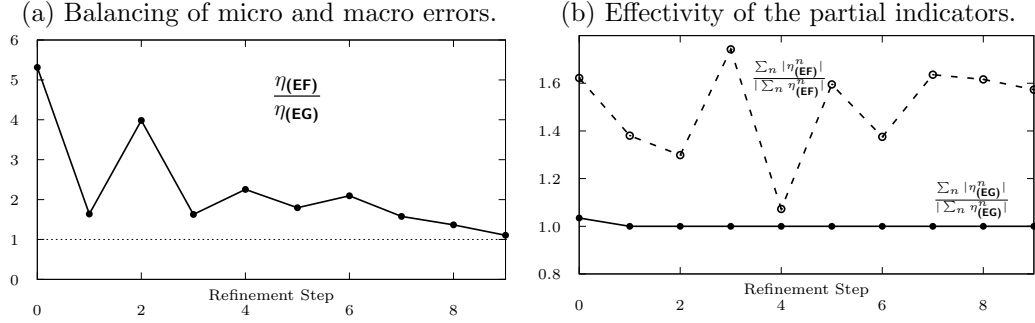


Figure 9: Performance on adaptive meshes: Indicator-effectivity (5.12) (right) and trend towards balancing contributions (left).

interest. For this case we illustrate the functionality of Algorithm 19 starting with $I_K^{(1)}$ with $K = 50\,000$ and $k = 0.05$ on each $I_n^{P,(1)}$.

In Fig. 8 the third refinement step is shown. In the upper graph we see the contributions to the error estimator from the subdivisions of I after two refinement steps. The error still concentrates around the point of interest but due to the refinement of the adjacent patches, the estimator values are already better balanced. As in Fig. 4, the bold line indicates the tolerance for refinement and for all subdivisions where the total contribution to the error estimator is above this threshold, either k or K is refined. The results of refinement are presented in the right graph. Comparing the bold line to the dotted line shows that there is an improvement in the value for the error.

We can see that the error parts $\eta_{(\text{EG})}$ and $\eta_{(\text{EF})}$ tend to balance, see Fig. 9a. This is intuitive since refinement is done on the dominant error term of a subdivision according to Algorithm 19. Figure 9b shows that the partial indicators reasonably approximate the error estimator in this case even though the sign of the estimator can and does change on different subdivisions and during refinement. This shows that contributions of one sign dominate in this case but this is not always guaranteed. In the rare case of the positive and negative contributions to the error estimator cancelling each other out, the

error estimator can fail to give an accurate representation of the error.

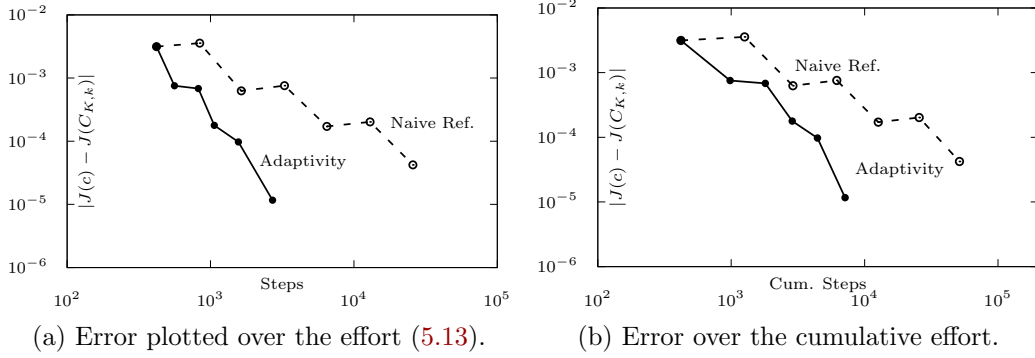


Figure 10: Comparison of the accuracy on uniform and adaptive meshes.

For this problem our algorithm speeds up performance significantly. According to Fig. 10a, to achieve a similar error of around $5 \cdot 10^{-5}$, we need only around one tenth of the effort. The cumulative effort, see Fig. 10b, is also of similar magnitude. This can be explained by the unequal distribution of $\eta(\mathbf{EG})$ and especially $\eta(\mathbf{EF})$, see Fig. 8. By uniform refinement of I and I^P we cannot alleviate this phenomenon. However our error estimator is able to identify sections of I that need to be treated on a finer scale and also sections, where I^P needs to be subdivided with a smaller k compared to other sections. Local refinement then gives a better distribution of the computing resources to calculate the solution $Y(t)$. Since the dependency on the slow scale is not uniform in this problem, we are able to gain efficiency with adaptive refinement unlike in the first numerical example.

To conclude we compare in Table 2 the effort, measured in time-steps of the micro-problem to be solved, for the fully resolved simulation of the original problem with the multiscale approach based on uniform meshes and on adaptive meshes. To measure the effort we must multiply the cumulative effort listed above by 2, to account for the additional effort for solving the adjoint problem and also by 5, which accounts for the number of cycles required to find the periodic state. In each case we show the results for the choice of discretization parameters where the error is below $5 \cdot 10^{-5}$. By using the multiscale scheme the effort is reduced by about 1 : 400, although we must solve adjoint solutions and although the approximation of the local periodic solutions requires multiple cycles in each macro step. If we further employ adaptive mesh refinement the effort is reduced by 1 : 2800. The adaptive multiscale scheme has the further advantage of giving an estimate on the discretization error.

6 Conclusion

We have presented an a posteriori error estimator for a temporal multiscale scheme of HMM type that has recently been introduced [12]. This multiscale scheme is based on separating micro and macro scale by replacing the micro scale influences by localized

Approach	Error	k	K	Micro-steps
Resolved simulation	$3.43 \cdot 10^{-5}$	$\frac{1}{200}$	–	200 000 000
Multiscale (uniform)	$4.24 \cdot 10^{-5}$	$\frac{1}{160}$	6 250	513 800
Multiscale (adaptive)	$1.16 \cdot 10^{-5}$	$\frac{1}{320}$	12 500	71 360

Table 2: Comparison of the effort for the resolved simulation, the multiscale scheme based on uniform and adaptive discretizations measured in the number of overall steps of the micro problem to be solved.

periodic in time solutions. The resulting scheme calls for the solution of one such periodic micro problem in each macro step.

The error estimator is based on the dual weighted residual method for estimating errors in goal functionals. The adjoint problem entering the error estimator has a structure similar to the primal one: each adjoint macro time step requires the solution of a periodic micro problem. In addition, to incorporate the error of the periodic in time micro scale problems, a further adjoint micro problem must be solved in each macro step. The resulting error estimator allows for a splitting of the local error contributions into micro scale and macro scale influences. We have shown very good efficiency of the estimator for a wide range of discretization parameters.

Based on the splitting into micro scale errors and macro scale errors an adaptive refinement loop is presented that allows to optimally balance all discretization parameters.

For the future it remains to extend this setting to temporal multiscale problems involving partial differential equations as discussed in [12, 18] that will add the further complexity of finding optimal spatial discretization parameters for macro and micro problems.

Acknowledgements The work of both authors has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the GRK 2297 MathCore - 314838170, as well within the project 411046898. Further, the work of TR has been supported by the Federal Ministry of Education and Research of Germany, grant number 05M16NMA.

References

- [1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden. The heterogeneous multi-scale method. *Acta Numerica*, pages 1–87, 2012.
- [2] R. Becker and R. Rannacher. Weighted a posteriori error control in FE methods. In et al. H. G. Bock, editor, *ENUMATH’97*. World Sci. Publ., Singapore, 1995.
- [3] R. Becker and R. Rannacher. An optimal control approach to a posteriori error

- estimation in finite element methods. In A. Iserles, editor, *Acta Numerica 2001*, volume 37, pages 1–202. Cambridge University Press, 2001.
- [4] M. Braack and A. Ern. A posteriori control of modeling errors and discretization errors. *Multiscale Model. Simul.*, 1(2):221–238, 2003.
 - [5] M. Běhounková, G. Tobie, G. Choblet, and O. Čadek. Coupling mantle convection and tidal dissipation: Applications to enceladus and earth-like planets. *J. Geophys. Research*, 115(E9), 2010.
 - [6] Doina Cioranescu and Patrizia Donato. *An Introduction to Homogenization*. Oxford Lecture Series in Mathematics and its applications. Oxford University Press, 1999.
 - [7] R. Conti. Sulla prolungabilità delle soluzioni di un sistema di equazioni differenziali ordinarie. *Bollettino dell’Unione Matematica Italiana, Serie 3*, 11(4):510–514, 1956.
 - [8] W. E. *Principles of Multiscale Modeling*. Cambridge University Press, 2011.
 - [9] W. E and B. Engquist. The heterogeneous multiscale method. *Comm. Math. Sci.*, 1(1):87–132, 2003.
 - [10] B. Engquist and Y.-H. Tsai. Heterogeneous multiscale methods for stiff ordinary differential equations. *Math. Comp.*, 74(252):1707–1742, 2005.
 - [11] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. Introduction to adaptive methods for differential equations. In A. Iserles, editor, *Acta Numerica 1995*, pages 105–158. Cambridge University Press., 1995.
 - [12] S. Frei and T. Richter. Efficient approximation of flow problems with multiple scales in time. *SIAM Multiscale Modeling and Simulation*, 2020.
 - [13] S. Frei, T. Richter, and T. Wick. Long-term simulation of large deformation, mechano-chemical fluid-structure interactions in ALE and fully Eulerian coordinates. *J. Comp. Phys.*, 321:874 – 891, 2016.
 - [14] G. P. Galdi and M. Kyed. Time-periodic solutions to the Navier-Stokes equations in three-dimensional whole-space with a non-zero drift term: Asymptotic profile at spatial infinity. arXiv:1610.00677v1, 2016.
 - [15] D. Meidner, R. Rannacher, and J. Vihharev. Goal-oriented error control of the iterative solution of finite element equations. *J. Num. Anal.*, 17(2), 2009.
 - [16] D. Meidner and T. Richter. Goal-oriented error estimation for the fractional step theta scheme. *Comp. Meth. Appl. Math.*, 14:203–230, 2014.
 - [17] D. Meidner and T. Richter. A posteriori error estimation for the fractional step theta discretization of the incompressible Navier-Stokes equations. *Comp. Meth. Appl. Mech. Engrg.*, 288:45–59, 2015.

- [18] J. Mizerski and T. Richter. The candy wrapper problem - a temporal multiscale approach for pde/pde systems. In *Numerical Mathematics and Advanced Applications - Enumath 2019*, Lecture Notes in Computational Science and Engineering. Springer, 2020.
- [19] S. Murakami. *Continuum Damage Mechanics*. Springer, 2012.
- [20] O.A. Oleinik, A.S. Shamaev, and G.A. Yosifian. *Mathematical Problems in Elasticity and Homogenization*, volume 26 of *Studies in Mathematics and its Applications*. North-Holland, 1992.
- [21] T. Richter. *Fluid-structure Interactions. Models, Analysis and Finite Elements*, volume 118 of *Lecture notes in computational science and engineering*. Springer, 2017.
- [22] T. Richter. An averaging scheme for the efficient approximation of time-periodic flow problems. *Computers and Fluids*, 214, 2021. <https://arxiv.org/abs/1806.00906>.
- [23] T. Richter and T. Wick. Variational localizations of the dual weighted residual method. *J. Comput. Appl. Math.*, 279:192–208, 2015.
- [24] T. Richter and W. Wollner. Efficient computation of time-periodic solutions of partial differential equations. *Viet. J. Math.*, 46(4), 2018.
- [25] M. Schmich and B. Vexler. Adaptivity with dynamic meshes for space-time finite element discretizations of parabolic equations. *SIAM J. Sci. Comp.*, 30(1):369–393, 2008.
- [26] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer, 2002.
- [27] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Number 25 in Springer Series in Computational Mathematics. Springer, 1997.