

# RANDOMIZED QUASI-OPTIMAL LOCAL APPROXIMATION SPACES IN TIME

JULIA SCHLEUSS, KATHRIN SMETANA, AND LUKAS TER MAAT

**ABSTRACT.** We target time-dependent partial differential equations (PDEs) with heterogeneous coefficients in space and time. To tackle these problems, we construct reduced basis/ multiscale ansatz functions defined in space that can be combined with time stepping schemes within model order reduction or multiscale methods. To that end, we propose to perform several simulations of the PDE for few time steps in parallel starting at different, randomly drawn start points, prescribing random initial conditions; applying a singular value decomposition to a subset of the so obtained snapshots yields the reduced basis/ multiscale ansatz functions. This facilitates constructing the reduced basis/ multiscale ansatz functions in an embarrassingly parallel manner. In detail, we suggest using a data-dependent probability distribution based on the data functions of the PDE to select the start points. Each local in time simulation of the PDE with random initial conditions approximates a local approximation space in one time point that is optimal in the sense of Kolmogorov. The derivation of these optimal local approximation spaces which are spanned by the left singular vectors of a compact transfer operator that maps arbitrary initial conditions to the solution of the PDE in a later point of time, is one other main contribution of this paper. By solving the PDE locally in time with random initial conditions, we construct local ansatz spaces in time that converge provably at a quasi-optimal rate and allow for local error control. Numerical experiments demonstrate that the proposed method can outperform existing methods like the proper orthogonal decomposition even in a sequential setting and is well capable of approximating advection-dominated problems.

## 1. INTRODUCTION

Applications that require repeated simulations for different parameters or a real-time simulation response of complex systems of partial differential equations (PDEs) or dynamical systems are ubiquitous. Moreover, heterogeneous problems that exhibit multiscale features or include rough data functions are particularly challenging. A direct numerical simulation using standard techniques such as the finite element (FE) method can be prohibitively expensive for such tasks. Well-known strategies to tackle these (heterogeneous) problems comprise multiscale methods which are based on local ansatz functions that incorporate the local behavior of the (numerical) solution of the

---

*Date:* October 4, 2022.

*2010 Mathematics Subject Classification.* 65C20, 65M12, 65M15, 65M55, 65M60, 65M75.

*Key words and phrases.* multiscale methods, model order reduction, randomized numerical linear algebra, Kolmogorov n-width.

The work of Julia Schleuß was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure.

PDE and model order reduction methods that exploit a carefully chosen set of problem-adapted basis functions to reduce the high-dimensional problem.

In this paper, we consider heterogeneous time-dependent PDEs and propose reduced basis/ multiscale ansatz functions defined in space that can be combined with time stepping schemes within model order reduction or multiscale methods. We provide one of the first contributions that facilitates constructing the reduced basis/ multiscale ansatz functions in an embarrassingly parallel manner in time. As a major new contribution in this paper, we select important points in time and only perform local simulations of the PDE on the corresponding local time intervals instead of decomposing the global time interval into consecutive subintervals [10, 31]. As the numerical experiments show, this can result in a reduced total number of computed time steps, whereas the existing approaches [10, 31] require (local) computations everywhere in the entire time interval. To choose relevant points in time, we employ data-dependent sampling strategies from randomized numerical linear algebra (NLA) [16, 12] in a completely new context since they are usually used to construct low-rank matrix decompositions (cf., e.g., [34]). As another key contribution we derive for the first time local approximation spaces in time that are optimal in the sense of Kolmogorov. Moreover, we provide for the first time a rigorous local a priori error analysis in time as one major new contribution.

A well-established tool for compressing and reducing time trajectories is the proper orthogonal decomposition (POD) [3, 27, 56], which is based on a singular value decomposition (SVD) of the functions evaluated in the time grid points and allows for error control. However, in order to perform a POD on simulation data, the (global) solution trajectory of the considered problem has to be computed sequentially prior to reducing.

In contrast, the approach we propose in this paper enables, as one major contribution, to construct reduced basis/ multiscale ansatz functions in parallel in time. To facilitate a time-parallel procedure, we propose to perform several simulations of the PDE for only few time steps in parallel. To this end, we start the simulations at different start time points that are randomly drawn from a data-dependent sampling distribution and prescribe random initial conditions. Subsequently, we apply an SVD to a subset of the computed snapshots to obtain the reduced basis/ multiscale ansatz functions. The proposed method is thus well-suited to be used on modern computer architectures allowing for many parallel computations and on each single compute unit a simulation for only few time steps has to be performed. Moreover, as another major contribution, the approach is especially tailored to time-dependent problems with heterogeneous time-dependent data functions. To draw start time points for the temporally local PDE simulations, we employ uniform, squared norm [20], or leverage score [17] sampling, which are standard sampling techniques from randomized NLA [12, 16]. In particular, both squared norm and leverage score sampling take into account the time-dependent data functions of the PDE and are commonly used in a variety of applications [12], for instance, to construct CUR<sup>1</sup> or similar matrix decompositions by approximating the matrix via its columns or rows. As one major contribution of this paper, we employ these methods in a completely new context for the purpose of time point selection.

---

<sup>1</sup>A CUR decomposition of a matrix  $A$  consists of three matrices  $C$ ,  $U$ , and  $R$ , where  $C$  ( $R$ ) contains columns (rows) of  $A$  and  $U$  is constructed such that the product  $CUR$  approximates  $A$ . As the matrices  $C$  and  $R$  are constructed from actual elements of  $A$ , the decomposition is usually more interpretable with respect to the original data compared to, e.g., a truncated SVD (cf., e.g., [34]).

To this end, the time-dependent data functions are discretized and represented by a matrix, where each column of the matrix corresponds to one time point. Moreover, other randomized subset selection technique, as proposed, for instance, in [1, 14, 13, 15], may also be used to choose time points.

The key observation, motivating a localized construction in time, is that for certain time-dependent problems the solution exhibits a very rapid, exponential decay of energy in time. To detect the functions that still persist at a point of time and are thus relevant for approximation, we introduce as a key new contribution a compact transfer operator in time that maps arbitrary initial conditions to the solution of the PDE in a later point of time. Spanning the local space by the leading left singular vectors of the transfer operator results in an approximation space that is optimal in the sense of Kolmogorov [26] and hence minimizes the approximation error among all spaces of the same dimension. While there are many methods that exploit localization in space [2, 22, 36, 35, 40, 41, 42, 57, 18, 24, 25, 33], we provide with this paper one of the first contributions that exploit localization in time [10, 31]. Whereas existing methods [10, 31] decompose the entire time interval into consecutive local subintervals, we select relevant time points in a data-dependent manner and only perform local computations on the corresponding local time intervals. Moreover, we provide for the first time an a priori error bound for the local approximation error in time.

As a direct calculation of the leading left singular vectors of the transfer operator can become computationally expensive, we employ random sampling as proposed in [6] for elliptic PDEs to efficiently approximate the optimal local spaces and facilitate an embarrassingly parallel construction of the reduced ansatz functions even for a single point of time. To this end, we solve the PDE locally in time with random initial conditions. We show that the resulting local space yields an approximation that converges at a quasi-optimal rate, allowing for local error control. While we only provide a local a priori error bound in this paper, we conjecture that it might be possible to also derive a global error bound.

Preliminary numerical experiments can be found in one of the authors bachelor thesis [59], where the proposed algorithm has been tested for problems with time-dependent source terms by uniformly sampling time points.

Optimal spatially local approximation spaces have been introduced for elliptic [2, 32, 57] and parabolic [49] problems, and random sampling has been employed to efficiently approximate the optimal local spaces in the elliptic setting in [6, 8]. Further spatially localizable multiscale methods for parabolic problems have been proposed in [10, 35, 40, 41, 42]. In [31] a space-time multiscale method for the linear heat equation is introduced, where for each coarse space-time node a corrector function that is localized in both space and time is computed to capture (local) fine-scale features. While in [31] a global a posteriori error bound is proved assuming that certain localization parameters are large enough to guarantee sufficiently small localization errors, but no rigorous a priori error analysis is performed, we provide in this paper for the first time a local a priori error bound in time. Moreover, we refer to [5] for an overview of methods to construct local reduced spaces.

In system and control theory balanced truncation is a well-known method to reduce the complexity of input-output systems [38, 45]. Balanced truncation for systems including time-dependent data functions has been introduced in [54, 62] and its error analysis has been first studied in [28, 47]. Nevertheless, solving matrix differential

equations or matrix inequalities is required which is prohibitively expensive for high dimensional problems. In [7] the authors propose an iterative procedure that is computationally more appealing as it is based on operations which exploit sparsity of the model. However, the required computations are global in time and have to be carried out in a sequential manner, leading to a complexity that depends on the global time discretization. In contrast, the approach we propose here requires only local computations in time that are in addition parallelizable.

In [61] it is shown that for linear time-invariant systems the concepts of Kolmogorov  $n$ -widths and Hankel singular values are directly connected and that the right singular vectors of the Hankel operator restricted to the unit ball span the optimal reduced input space in the sense of Kolmogorov and can be linked to active subspaces. We suggest that the leading right singular vectors of the transfer operator introduced here span the optimal reduced input space and can thus be used to regularize inverse problems and data assimilation procedures.

Furthermore, dynamic mode decomposition [50, 60] fits an operator that maps the solution from one time point to the next to simulation data. The fitted operator is thus similar to the transfer operator we introduce in this paper. However, similar to the POD and in contrast to the approach we propose here, access to the (global) solution trajectory is required to compress the entire dynamics. Recently, probabilistic numerical methods that yield a probability distribution over the (unknown) solution of an ordinary or partial differential equation have been proposed, for instance, in [51, 52, 11, 39]. Moreover, randomized subset selection techniques are used in [46] for the purpose of hyperreduction. In comparison, in this paper data-dependent probability distributions and randomized subset selection techniques are exploited to select both start time points and initial conditions for the temporally local PDE simulations.

The remainder of this paper is organized as follows. In [section 2](#) we introduce the general time-dependent model problem together with an exemplary test case and its numerical approximation. Subsequently, we first sketch the key new contributions of this paper along with some motivation in [section 3](#) and then develop the main contributions in [sections 4](#) and [5](#). We propose optimal local spaces in time in [section 4](#) and address their approximation via random sampling in [section 5](#). Moreover, in [section 5](#) we propose a randomized algorithm to construct one reduced space for the global approximation by solving several local problems in time in parallel. We discuss both its basic properties and the choice of the probability distribution used for drawing time points. Finally, we present numerical experiments in [section 6](#) to demonstrate the approximation properties of the proposed algorithm and draw conclusions in [section 7](#).

## 2. PROBLEM SETTING

We consider a time-dependent linear PDE, say, an advection-diffusion-reaction problem, that may include heterogeneous time-dependent coefficients. To that end, let  $D \subseteq \mathbb{R}^n$  denote a bounded Lipschitz domain of dimension  $n \in \{1, 2, 3\}$  and let  $I = (0, T) \subset \mathbb{R}$  be a time interval with  $0 < T < \infty$ . We assume that on  $I \times D$  a Bochner space  $\mathcal{V}$  and a reflexive Bochner space  $\mathcal{W}$  with dual space  $\mathcal{W}^*$  are given that will serve as ansatz and test space, respectively. Furthermore, we denote by  $\partial_t + \mathcal{A} : \mathcal{V} \rightarrow \mathcal{W}^*$  a linear, continuous, surjective, and inf-sup stable operator. Let  $\mathcal{F} \in \mathcal{W}^*$  be a bounded linear functional. We assume that  $\mathcal{F}$  accounts for both source terms and boundary data. Moreover, we denote the initial values by  $u_0 \in L^2(D)$ . Here, we assume that for

any function  $u \in \mathcal{V}$  we have that  $u(t, \cdot) \in L^2(D)$  for every point in time  $t \in I$ . Then, we consider the following variational problem: Find  $u \in \mathcal{V}$  such that  $u(0, \cdot) = u_0$  in  $L^2(D)$  and

$$(2.1) \quad \partial_t u + \mathcal{A}u = \mathcal{F} \quad \text{in } \mathcal{W}^*.$$

The Banach-Nečas-Babuška theorem (e.g. [19, Theorem 2.6]) and the assumptions above guarantee the existence of a unique solution of (2.1).

**Exemplary model problem: advection-diffusion-reaction problem.** We consider an advection-diffusion-reaction problem as a representative model problem of (2.1). To this end, we assume that  $\partial D = \Sigma_D \cup \Sigma_N$  with  $|\Sigma_D| > 0$  and denote by  $f : I \times D \rightarrow \mathbb{R}$  source terms,  $u_0 : D \times \mathbb{R}$  initial conditions,  $g_D : I \times \Sigma_D \rightarrow \mathbb{R}$  Dirichlet boundary conditions,  $g_N : I \times \Sigma_N \rightarrow \mathbb{R}$  Neumann boundary conditions, and  $n$  the outer unit normal, respectively. In its weak form the conductivity coefficient is assumed to satisfy  $\kappa \in L^\infty(I \times D)^{n \times n}$  with  $\kappa_0(t, x)|v|^2 \leq v^\top \kappa(t, x)v \leq \kappa_1(t, x)|v|^2$  for every  $v \in \mathbb{R}^n$ ,  $0 < \kappa_0 \leq \kappa_0(t, x) \leq \kappa_1(t, x) \leq \kappa_1 < \infty$  for almost every  $(t, x) \in I \times D$ , where we do not consider the case  $\kappa_0 \rightarrow 0$  such that we can assume that a weak solution with spatial regularity of  $H^1(D)$  exists. Furthermore, the advection field and the reaction coefficient are assumed to satisfy  $b \in L^\infty(I \times D)^n$ ,  $\nabla \cdot b \in L^\infty(I \times D)$ , and  $c \in L^\infty(I \times D)$  with  $|b(t, x) \cdot v|^2 \leq b_1(t, x)|v|^2$  for every  $v \in \mathbb{R}^n$ ,  $0 \leq b_1(t, x) \leq b_1 < \infty$  for almost every  $(t, x) \in I \times D$ , and  $|c(t, x)| \leq c_1 < \infty$  for almost every  $(t, x) \in I \times D$ , respectively. To ensure well-posedness, we moreover assume that  $c(t, x) - \frac{1}{2}\nabla \cdot b(t, x) \geq 0$  for almost every  $(t, x) \in I \times D$ . We then seek the solution  $u : I \times D \rightarrow \mathbb{R}$  such that

$$(2.2) \quad \begin{aligned} u_t(t, x) - \operatorname{div}(\kappa(t, x)\nabla u(t, x)) + b(t, x) \cdot \nabla u(t, x) + c(t, x)u(t, x) &= f(t, x) \\ &\text{for every } (t, x) \in I \times D, \\ u(t, x) &= g_D(t, x) \quad \text{for every } (t, x) \in I \times \Sigma_D, \\ \kappa(t, x)\nabla u(t, x) \cdot n(x) &= g_N(t, x) \quad \text{for every } (t, x) \in I \times \Sigma_N, \\ u(0, x) &= u_0(x) \quad \text{for every } x \in D. \end{aligned}$$

**Discretization.** Here, we consider  $f \in L^2(I, L^2(D))$  and  $g_N \in L^2(I, L^2(\Sigma_N))$ . To simplify notation, we assume that  $f$  also accounts for Dirichlet boundary conditions  $g_D \in L^2(I, H^{1/2}(\Sigma_D))$ . For the numerical approximation of (2.2), we employ the implicit Euler method. To this end, we assume that the time interval  $I$  is partitioned via  $N_I$  equidistant time points  $0 = t_0 < t_1 < \dots < t_{N_I-1} = T$  of distance  $\Delta_T = T/(N_I - 1)$ . However, the approximation can be done completely analogously for other time stepping schemes. Moreover, we consider a piecewise linear conforming FE space  $X$  of dimension  $N_D \in \mathbb{N}$  with basis functions  $\phi_i \in X$ ,  $i = 1, \dots, N_D$ . Then, the mass and stiffness matrices  $\mathbf{M}, \mathbf{A}_l \in \mathbb{R}^{N_D \times N_D}$  and the right-hand side vectors  $\mathbf{F}_l \in \mathbb{R}^{N_D}$  are given as

$$(2.3) \quad \begin{aligned} \mathbf{M}_{ij} &:= (\phi_j, \phi_i)_{L^2(D)}, \quad 1 \leq i, j \leq N_D, \\ (\mathbf{A}_l)_{ij} &:= (\kappa(t_l)\nabla \phi_j, \nabla \phi_i)_{L^2(D)} + (b(t_l) \cdot \nabla \phi_j, \phi_i)_{L^2(D)} + (c(t_l)\phi_j, \phi_i)_{L^2(D)}, \\ &\quad 1 \leq i, j \leq N_D, \quad 1 \leq l \leq N_I - 1, \\ (\mathbf{F}_l)_i &:= (f(t_l), \phi_i)_{L^2(D)} + (g_N(t_l), \phi_i)_{L^2(\Sigma_N)}, \quad 1 \leq i \leq N_D, \quad 1 \leq l \leq N_I - 1. \end{aligned}$$

To simplify the presentation, we assume here that this discretization is stable, meaning that the advection should not be too dominant. Given a discrete representation  $\mathbf{u}_0 \in \mathbb{R}^{N_D}$  of the initial values  $u_0$ , we approximate the solution of (2.2) by computing  $\mathbf{u}_l \in$

$\mathbb{R}^{N_D}$  at time point  $t_l$  for  $l = 1, \dots, N_I - 1$  via

$$(2.4) \quad (\mathbf{M} + \Delta_T \mathbf{A}_l) \mathbf{u}_l = \Delta_T \mathbf{F}_l + \mathbf{M} \mathbf{u}_{l-1}.$$

In the following, we assume that the discretization in space and time is chosen sufficiently fine such that the discretization error between the exact solution and the discrete solution of (2.4) is negligibly small compared to the multiscale or model order reduction error. To reduce the possibly very high-dimensional discrete problem (2.4) (due to, for instance, fine-scale features in the coefficient functions that need to be well resolved) we apply multiscale or model order reduction methods.

**Reduced approximation.** We assume for now that suitable reduced ansatz functions  $\varphi_1, \dots, \varphi_N \in X$  are given that will be determined below (cf. section 5). The matrix  $\mathbf{U}_{\text{red}} = [\varphi_1 \dots \varphi_N] \in \mathbb{R}^{N_D \times N}$  contains the corresponding FE coefficient vectors. We then compute a reduced approximation of (2.4) via Galerkin projection of the FE space onto the space spanned by the reduced basis: For  $l = 1, \dots, N_I - 1$  find  $\mathbf{u}_{\text{red},l} \in \mathbb{R}^N$  such that

$$(2.5) \quad \mathbf{u}_{\text{red},l} = (\mathbf{M}_{\text{red}} + \Delta_T \mathbf{A}_{\text{red},l})^{-1} (\Delta_T \mathbf{F}_{\text{red},l} + \mathbf{M}_{\text{red}} \mathbf{u}_{\text{red},l-1}),$$

where  $\mathbf{M}_{\text{red}} = \mathbf{U}_{\text{red}}^\top \mathbf{M} \mathbf{U}_{\text{red}}$ ,  $\mathbf{A}_{\text{red},l} = \mathbf{U}_{\text{red}}^\top \mathbf{A}_l \mathbf{U}_{\text{red}}$ ,  $\mathbf{F}_{\text{red},l} = \mathbf{U}_{\text{red}}^\top \mathbf{F}_l$ , and  $\mathbf{u}_{\text{red},0} = \mathbf{M}_{\text{red}}^{-1} \mathbf{U}_{\text{red}}^\top \mathbf{M} \mathbf{u}_0$ .

### 3. MOTIVATION AND KEY NEW CONTRIBUTIONS OF THIS PAPER

To construct reduced ansatz functions  $\varphi_1, \dots, \varphi_N \in X$  as given above, a well-established strategy is to perform a POD on (the first part of) the global solution trajectory. For this purpose, prior to reducing, the global solution in time has to be computed in a sequential manner (cf. Fig. 1 (top row)). In contrast, we propose in this paper, as a major new contribution, to generate a reduced basis in an embarrassingly time-parallel manner which enables to split the computational budget and distribute it over the entire time interval (cf. Fig. 1 (middle row)). The approach is thus well-suited to be used on modern computer architectures allowing for many parallel computations. Moreover, we conjecture that the proposed ideas may in general contribute to reduce computational costs and exploit parallelization in applications where the computation of classical full order solutions is extremely expensive, for instance, by combining the proposed approach with the parareal methodology [30].

To motivate the proposed approximation strategy, we consider the toy model problem visualized in Fig. 2 and observe that the time-dependent source terms  $f_1$  and  $f_2$  clearly determine the behavior of the corresponding solution in time. Therefore, we suggest that the time-dependent data functions may help to determine time points that are relevant for approximation. To this end, we represent the time-dependent data functions as matrices, where each column corresponds to one time point in the time grid, see Fig. 2 (top right). To detect and select significant time points, we then employ column subset selection techniques from randomized NLA [12, 16]. These methods are generally used to construct low-rank matrix decompositions (cf., e.g., [34]) by approximating (the range of) the matrix from selected columns or rows. The decompositions are thus more interpretable with respect to the original data compared to an SVD and error bounds are available (cf., e.g., [16, 34]). As one key contribution, we exploit column subset selection techniques in a completely new context for the purpose of time point selection.

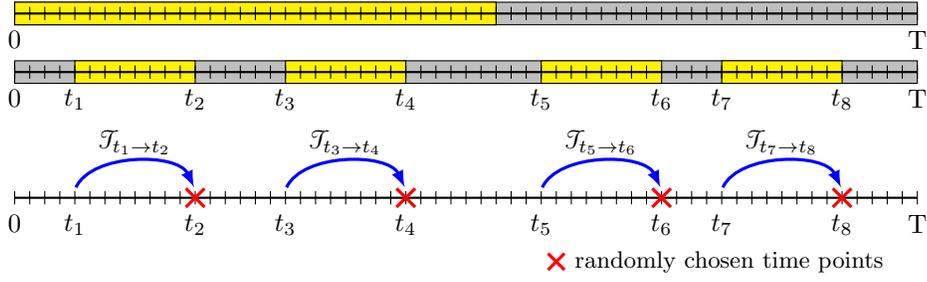


FIGURE 1. Computational budget in time: Sequential for POD (top row) vs. split and distributed for randomized approach proposed in this paper (middle row). Transfer operators corresponding to randomly chosen time points (bottom row).

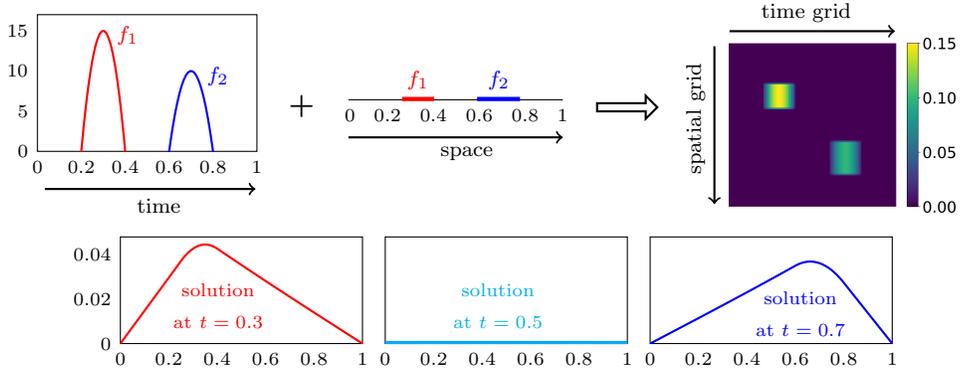


FIGURE 2. Data-dependent sampling based on  $f$  helps to determine time points for detecting shapes of the solution. Right-hand side data functions  $f_1$  and  $f_2$  in time (top left) corresponding to spatially disjoint sources in one spatial dimension (top middle) and associated right-hand side data matrix (top right). Solution of heat equation for  $f_1$ ,  $f_2$ , homogeneous initial and Dirichlet boundary conditions evaluated at different points in time (bottom).

Subsequently, we generate reduced ansatz functions corresponding to the selected time points in an embarrassingly parallel manner by solving independent local problems in time (cf. Fig. 1 (middle row)). To motivate a localized construction of the reduced ansatz functions in time, we recall the following well-known property that many time-dependent problems share via the example of the linear heat equation: If  $\mathcal{F} \equiv 0$  it is straightforward to show that  $\|u(t, \cdot)\|_{L^2(D)} \leq e^{-Ct} \|u_0\|_{L^2(D)}$  for any  $t \in I$ , where the constant  $C$  depends only on the shape of  $D$  and the heat conductivity coefficient. To detect the functions that still persist at a selected time point  $t$  and are thus relevant for approximation purposes, we introduce, as a key new contribution, a transfer operator  $\mathcal{T}_{s \rightarrow t}$  in time that takes arbitrary initial conditions in  $L^2(D)$  at time  $s \in I$  with  $s < t$ , solves the PDE locally in  $(s, t)$ , and evaluates the solution at target time  $t$ ; see Fig. 1 (bottom row) for an illustration. In this way, the transfer operator captures the decay behavior of solutions of the PDE in time.

After discretization, say, with the FE method, the transfer operator can be represented by a matrix. According to the well-known Eckart-Young theorem the range of this matrix can be optimally approximated by its leading left singular vectors. While

the *discrete* transfer operator is trivially compact thanks to its finite rank, in the *continuous* setting we need to prove compactness of the transfer operator to facilitate its singular value decomposition (SVD) via the Hilbert-Schmidt theorem [44, Theorem 8.94]. Then, the space spanned by the leading left singular vectors provides an optimal approximation in the sense of Kolmogorov, meaning that it minimizes the approximation error among all linear spaces of the same dimension (see [43, Theorem 2.2 in Chapter 4]). Moreover, in the elliptic setting it has been shown that the optimal local approximation spaces outperform other approaches (based on, for instance, Legendre-type functions or empirical modes) numerically (see [57]).

To further facilitate an efficient parallel computation, we approximate the optimal spaces via random sampling [6, 23], i.e. we solve the PDE locally in time with random initial conditions, which results in a provably nearly optimal local approximation.

#### 4. OPTIMAL LOCAL APPROXIMATION SPACES IN TIME

First, we introduce in subsection 4.1, as a key new contribution, a transfer operator in time for which one can prove compactness and thus obtain local ansatz spaces in time, which are optimal in the sense of Kolmogorov for the approximation of the solution of (2.1) in a point of time. Subsequently, we show how to compute an approximation of the optimal local spaces and discuss its practical realization via Krylov subspace methods and random sampling in subsection 4.2.

**4.1. Constructing optimal local approximation spaces via a transfer operator.** We first observe that for any local subinterval  $(s, t) \subseteq I$  of the global time interval the solution  $u$  of (2.1) solves the PDE locally in time with (unknown) initial conditions given by  $u(s, \cdot) \in L^2(D)$ . Therefore, we consider all local solutions  $u_{\text{loc}} \in \mathcal{V}|_{(s,t) \times D}$  with arbitrary initial conditions  $u_{\text{loc}}(s, \cdot) \in L^2(D)$  that satisfy

$$(4.1) \quad \partial_t u_{\text{loc}} + \mathcal{A}_{\text{loc}} u_{\text{loc}} = \mathcal{F}_{\text{loc}} \quad \text{in } (\mathcal{W}|_{(s,t) \times D})^*.$$

Here,  $\mathcal{A}_{\text{loc}}$  and  $\mathcal{F}_{\text{loc}}$  denote the respective local operator and functional associated with  $(s, t) \times D$ . All solutions of (4.1) can be split into a function that solves (4.1) for  $\mathcal{F}_{\text{loc}}$  and homogeneous initial conditions and a function that solves (4.1) for  $\mathcal{F}_{\text{loc}} \equiv 0$  and arbitrary initial conditions. In the following, we will first address the case where  $\mathcal{F}_{\text{loc}} \equiv 0$  and discuss the general case at the end of the subsection.

As we want to approximate the evaluation of the global solution at time point  $t$ , we consider the space  $\mathcal{H}_t$  of all solutions of (4.1) evaluated at time  $t$ :

$$(4.2) \quad \mathcal{H}_t := \{w(t, \cdot) \in L^2(D) \mid w \in \mathcal{V}|_{(s,t) \times D} \text{ solves (4.1), } w(s, \cdot) \in L^2(D), \mathcal{F}_{\text{loc}} \equiv 0\}.$$

We equip the space  $\mathcal{H}_t$  with the  $L^2(D)$ -inner product and -norm.

As one major contribution of this paper, we next introduce a transfer operator in time that can be proven to be compact and thus facilitates to derive optimal local ansatz spaces for the approximation of the solution space at time  $t$  via its SVD (see Fig. 3 for an illustration).

**Definition 4.1.** For  $s < t$ ,  $s, t \in I$ , the transfer operator  $\mathcal{T}_{s \rightarrow t} : L^2(D) \rightarrow \mathcal{H}_t$  is given by

$$(4.3) \quad \mathcal{T}_{s \rightarrow t} w(s, \cdot) = w(t, \cdot)$$

for  $w \in \mathcal{V}|_{(s,t) \times D}$  that solves (4.1) for  $\mathcal{F}_{\text{loc}} \equiv 0$ .

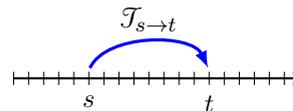


FIGURE 3. Transfer operator  $\mathcal{T}_{s \rightarrow t}$ .

**Assumption 1.** We assume that the transfer operator  $\mathcal{T}_{s \rightarrow t}$  introduced in (4.3) is compact for arbitrary  $0 \leq s < t \leq T$ .

**Remark 4.2.** In [Appendix A](#) we exemplarily prove compactness of the transfer operator for an advection-diffusion-reaction problem as introduced in [section 2](#).

Compactness of the transfer operator guarantees the existence of its SVD via the Hilbert-Schmidt theorem [44, Theorem 8.94]. Similar to [2, 49, 57] it can then be shown that the leading left singular vectors of  $\mathcal{T}_{s \rightarrow t}$  span an optimal approximation space in the local solution space  $\mathcal{H}_t$ . Here, we use the concept of optimality in the sense of Kolmogorov [26]: A subspace  $\mathcal{H}_t^n \subset \mathcal{H}_t$  of dimension at most  $n$  for which holds  $d_n(\mathcal{T}_{s \rightarrow t}(L^2(D)); \mathcal{H}_t) = \|\mathcal{T}_{s \rightarrow t} - \mathcal{P}_{\mathcal{H}_t^n} \mathcal{T}_{s \rightarrow t}\|$  is called an optimal subspace for  $d_n(\mathcal{T}_{s \rightarrow t}(L^2(D)); \mathcal{H}_t)$ , where the Kolmogorov  $n$ -width  $d_n(\mathcal{T}_{s \rightarrow t}(L^2(D)); \mathcal{H}_t)$  is defined as  $d_n(\mathcal{T}_{s \rightarrow t}(L^2(D)); \mathcal{H}_t) := \inf_{\mathcal{H}_t^n \subset \mathcal{H}_t; \dim(\mathcal{H}_t^n) = n} \|\mathcal{T}_{s \rightarrow t} - \mathcal{P}_{\mathcal{H}_t^n} \mathcal{T}_{s \rightarrow t}\|$  and  $\mathcal{P}_{\mathcal{H}_t^n}$  denotes the orthogonal projection onto  $\mathcal{H}_t^n$ .

**Theorem 4.3** (Optimal local approximation spaces in time). *Let  $\sigma_{s \rightarrow t}^{(i)} \in \mathbb{R}^+$  and  $\varphi_{s \rightarrow t}^{(i)} \in \mathcal{H}_t$ ,  $i = 1, \dots, \infty$ , denote the singular values and left singular vectors of the transfer operator  $\mathcal{T}_{s \rightarrow t}$  defined in (4.3). Then the optimal approximation space for  $d_n(\mathcal{T}_{s \rightarrow t}(L^2(D)); \mathcal{H}_t)$  is given by*

$$(4.4) \quad \mathcal{H}_t^n := \text{span}\{\varphi_{s \rightarrow t}^{(1)}, \dots, \varphi_{s \rightarrow t}^{(n)}\}$$

and the Kolmogorov  $n$ -width satisfies

$$\begin{aligned} & d_n(\mathcal{T}_{s \rightarrow t}(L^2(D)); \mathcal{H}_t) \\ &= \sup_{\psi \in L^2(D)} \inf_{\zeta \in \mathcal{H}_t^n} \frac{\|\mathcal{T}_{s \rightarrow t}\psi - \zeta\|_{L^2(D)}}{\|\psi\|_{L^2(D)}} = \|\mathcal{T}_{s \rightarrow t} - \mathcal{P}_{\mathcal{H}_t^n} \mathcal{T}_{s \rightarrow t}\| = \sigma_{s \rightarrow t}^{(n+1)}. \end{aligned}$$

*Proof.* The assertion directly follows from the Hilbert-Schmidt theorem [44, Theorem 8.94] and [43, Theorem 2.2 in Chapter 4].  $\square$

**Remark 4.4** (Discussion of [Theorem 4.3](#)). *While a compact transfer operator has already been used to construct optimal spatially local approximation spaces for elliptic [2, 32, 57] and parabolic [49] problems, the key new contribution in this paper is the introduction of a compact transfer operator in time that enables to generate an optimal ansatz space for the approximation of the solution space at a point of time. Moreover, we provide, to the best of our knowledge for the first time, a priori error analysis for the local approximation error in time (cf. also [Proposition 5.1](#)). We conjecture that it is also possible to derive a global error bound by using ideas as employed in [4].*

To address non-homogeneous data  $\mathcal{F}_{\text{oc}}$ , we define  $u_{s \rightarrow t}^f \in \mathcal{V}|_{(s,t) \times D}$  as the solution of (4.1) with homogeneous initial conditions at time  $s$ . Finally, the optimal local approximation space at time  $t$  is given by

$$(4.5) \quad \mathcal{H}_t^{n, \text{data}} := \text{span}\{\varphi_{s \rightarrow t}^{(1)}, \dots, \varphi_{s \rightarrow t}^{(n)}, u_{s \rightarrow t}^f(t, \cdot)\}.$$

**4.2. Approximation of the optimal local approximation spaces in time.** In this subsection we describe how to compute an approximation of the optimal local space  $\mathcal{H}_t^{n, \text{data}}$  in (4.5) for the example of the advection-diffusion-reaction problem (cf. [section 2](#)). In the following, we use the notation introduced in [section 2](#) (see e.g. (2.3)). Assuming that  $s = t_i$  and  $t = t_j$  for  $0 \leq i < j \leq N_I - 1$  and given a discrete version

$\mathbf{u}_{\text{loc},i} \in \mathbb{R}^{N_D}$  of the arbitrary local initial values  $u_{\text{loc}}(s, \cdot)$ , we compute a local solution  $\mathbf{u}_{\text{loc},l} \in \mathbb{R}^{N_D}$  at time point  $t_l$  for  $l = i + 1, \dots, j$  via (cf. (2.4) and (4.1))

$$(4.6) \quad \mathbf{u}_{\text{loc},l} = (\mathbf{M} + \Delta_T \mathbf{A}_l)^{-1} (\Delta_T \mathbf{F}_l + \mathbf{M} \mathbf{u}_{\text{loc},l-1}).$$

As the discrete transfer operator  $T_{t_i \rightarrow t_j}$  acts on the space of local solutions with  $\mathbf{F}_l = 0$  ( $i \leq l \leq j$ ), the matrix version  $\mathbf{T}_{t_i \rightarrow t_j} \in \mathbb{R}^{N_D \times N_D}$  of  $T_{t_i \rightarrow t_j}$  is given by (cf. (4.3))

$$(4.7) \quad \mathbf{T}_{t_i \rightarrow t_j} \boldsymbol{\xi} = [(\mathbf{M} + \Delta_T \mathbf{A}_j)^{-1} \mathbf{M}] [(\mathbf{M} + \Delta_T \mathbf{A}_{j-1})^{-1} \mathbf{M}] \dots [(\mathbf{M} + \Delta_T \mathbf{A}_{i+1})^{-1} \mathbf{M}] \boldsymbol{\xi}.$$

Finally, we compute the  $n$  leading left singular vectors  $\boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(1)}, \dots, \boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(n)} \in \mathbb{R}^{N_D}$  of  $\mathbf{T}_{t_i \rightarrow t_j}$  to approximate the optimal local space  $\mathcal{H}_{t_j}^n$  (cf. (4.4)) and define

$$(4.8) \quad H_{t_j}^n := \text{span}\{\boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(1)}, \dots, \boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(n)}\},$$

where  $\boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(k)}$  is the FE function corresponding to the coefficient vector  $\boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(k)}$  for  $1 \leq k \leq n$ . Consequently, we have that  $\|T_{t_i \rightarrow t_j} - P_{H_{t_j}^n} T_{t_i \rightarrow t_j}\| = \sigma_{t_i \rightarrow t_j}^{(n+1)}$  (Eckart-Young theorem e.g. in [21]), where  $\sigma_{t_i \rightarrow t_j}^{(n+1)}$  is the  $n+1$ -st singular value of  $T_{t_i \rightarrow t_j}$  (listed in non-increasing order of magnitude) and  $P_{H_{t_j}^n}$  denotes the orthogonal projection onto  $H_{t_j}^n$  (cf. Theorem 4.3). We use the same notation for continuous and discrete singular values and vectors expecting that the respective meaning is clear from the context.

To address non-homogeneous data  $\mathbf{F}_l$  ( $l = i + 1, \dots, j$ ), we compute the solution of (4.6) for homogeneous initial conditions  $\mathbf{u}_{\text{loc},i}^{\mathbf{F}} \equiv 0$ , add the resulting solution  $\mathbf{u}_{\text{loc},j}^{\mathbf{F}}$  at time  $t_j$  to the FE basis, and define  $H_{t_j}^{n,\text{data}}$  as the span of the FE functions associated with the coefficient vectors  $\boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(1)}, \dots, \boldsymbol{\varphi}_{t_i \rightarrow t_j}^{(n)}, \mathbf{u}_{\text{loc},j}^{\mathbf{F}}$ .

**Remark 4.5** (Comparison of computational approaches to approximate the optimal local spaces). *In practice, the left singular vectors of  $\mathbf{T}_{t_i \rightarrow t_j}$  can be computed via the eigenvectors of  $\mathbf{T}_{t_i \rightarrow t_j}^{\top} \mathbf{T}_{t_i \rightarrow t_j}$ . A direct computation of the optimal local space would therefore require to compute the evaluation of local solutions at time  $t_j$  for all  $N_D$  basis functions that span the local solution space at time  $t_i$ , thus evaluate the transfer operator  $N_D$  times, and solve a dense generalized eigenproblem of dimension  $N_D \times N_D$ . As this becomes infeasible for large  $N_D$ , one would in general use Krylov subspace or randomized methods for the approximation of the optimal local spaces [6, 23, 37].*

*In Krylov subspace methods, the application of the transfer operator (4.7) would be implicitly passed to the eigenvalue solver. To calculate the  $m$  leading eigenvectors of  $\mathbf{T}_{t_i \rightarrow t_j}^{\top} \mathbf{T}_{t_i \rightarrow t_j}$  using, for instance, the implicitly restarted Arnoldi method (IRAM) from [29],  $\mathcal{O}(m)$  evaluations of  $\mathbf{T}_{t_i \rightarrow t_j}$  and  $\mathbf{T}_{t_i \rightarrow t_j}^{\top}$  are required in every iteration. While Krylov subspace methods can lead to more accurate approximations especially for slowly decaying singular values, randomized methods have the main advantage that they are inherently stable and amenable to parallelization [23, 64].*

*To approximate the space spanned by the  $m$  leading left singular vectors of the transfer operator via random sampling as described in subsection 5.1 in more detail,  $m + s$  evaluations of the transfer operator are required. As randomized methods can outperform Krylov subspace methods even in the sequential setting (see, e.g., [6]), they are thus an appealing choice for the approximation of the optimal local spaces. For a more in-depth comparison of Krylov subspace and randomized methods, we refer, for instance, to [23, section 6].*

**Remark 4.6** (Computational complexity). *The computational complexity of the local basis construction is clearly dominated by the evaluation of  $\mathbf{T}_{t_i \rightarrow t_j}$  or  $\mathbf{T}_{t_i \rightarrow t_j}^\top$  and thus the numerical solution of the local PDE, where we employ a sparse direct solver. For a standard FE discretization in two or three dimensions the factorizations of  $(\mathbf{M} + \Delta_T \mathbf{A}_{i+1}), \dots, (\mathbf{M} + \Delta_T \mathbf{A}_j) \in \mathbb{R}^{N_D \times N_D}$  can be computed in  $\mathcal{O}((j-i)N_D^{3/2})$  or  $\mathcal{O}((j-i)N_D^2)$  work [21]. After factorizing, the computational complexity for each local solution trajectory of the PDE is  $\mathcal{O}((j-i)N_D \log(N_D))$  or  $\mathcal{O}((j-i)N_D^{4/3})$  [21].*

## 5. GENERATING REDUCED ANSATZ FUNCTIONS IN PARALLEL IN TIME

In this section we derive a randomized algorithm that provides an approximation to the *discrete* solution of problem (2.1) by exploiting techniques from randomized NLA [6, 17, 20, 12, 16, 23] (see also motivation in section 3). For this purpose, we first sketch in subsection 5.1 how a nearly optimal ansatz space for the approximation of the discrete solution in a single point of time can be generated in an efficient and parallel manner via random sampling. Subsequently, we present in subsection 5.2 a randomized algorithm that constructs a reduced basis by performing several simulations of the PDE for only few time steps in parallel. The local simulations start at different time points that are randomly drawn from a data-dependent probability distribution whose choice is discussed in subsection 5.3; random initial conditions are prescribed. The proposed algorithm thus enables to split and distribute the available computational budget over the entire time interval, facilitates an embarrassingly parallel computation, and is therefore well-suited to be used on modern computer architectures.

### 5.1. Approximating the range of one transfer operator via random sampling.

To motivate the random sampling strategy (cf. [23]), suppose that we want to approximate the range of a large matrix of rank  $m$ . By multiplying the matrix with  $m$  random vectors we draw  $m$  random samples from the range of the matrix at high probability. As these  $m$  samples are likely linearly independent thanks to the randomness, they span the range of the matrix with high probability. In cases where the rank of the matrix is unknown and to compensate for the fact that any of the drawn random vectors might lie in the null space of the matrix, we draw a small number  $s$  of additional random vectors to ensure that the resulting  $m+s$  samples most likely span the targeted  $m$ -dimensional subspace.

To construct a suitable approximation of the *discrete* optimal local ansatz space  $H_{t_j}^m$  (4.8), we thus prescribe  $n = m + s$  random initial conditions at the local starting time point  $t_i < t_j$ , where the coefficient vectors of the corresponding FE functions are mutually independent normal random vectors  $\mathbf{r}_1, \dots, \mathbf{r}_n \sim N(\mathbf{0}, (\mathbf{A}_i^\top \mathbf{A}_i)^{-1})$ . Recall that  $\mathbf{A}_i$  denotes the stiffness matrix corresponding to time point  $t_i$  (2.3). Moreover, the oversampling parameter  $s$  is typically not greater than 2 or 3 (cf. [23, 6]).

Then, the matrix version of the transfer operator  $\mathbf{T}_{t_i \rightarrow t_j}$  (4.7) is applied to the random vectors, meaning we solve the PDE locally on the time interval  $(t_i, t_j)$  with initial conditions given by  $\mathbf{r}_1, \dots, \mathbf{r}_n$  and evaluate the solutions at the local end time  $t_j$ . We highlight that the computation of the  $n$  local solutions is embarrassingly parallel. The space  $H_{t_j, \text{rand}}^n$  is then spanned by the  $n$  resulting local solutions evaluated at time  $t_j$ .

The following probabilistic a priori error bound shows that the space  $H_{t_j, \text{rand}}^n$  yields an approximation that converges at a nearly optimal rate which is only slightly worse than the rate  $\sigma_{t_i \rightarrow t_j}^{(m+1)}$  achieved by the optimal space  $H_{t_j}^m$  (cf. subsection 4.2).

**Proposition 5.1.** (Probabilistic a priori error bound, [6, Proposition 3.2] based on [23, Theorem 10.6]). *Let  $\lambda_{\min}^{\mathbf{M}}$  and  $\lambda_{\max}^{\mathbf{M}}$  denote the smallest and largest eigenvalue of  $\mathbf{M}$  (cf. (2.3)). Moreover, we denote by  $\sigma_{\min}^{\mathbf{A}_i}$  and  $\sigma_{\max}^{\mathbf{A}_i}$  the smallest and largest singular value of  $\mathbf{A}_i$ . Then, for  $n \geq 4$  it holds that*

$$(5.1) \quad \mathbb{E}(\|T_{t_i \rightarrow t_j} - PH_{t_j, \text{rand}}^n T_{t_i \rightarrow t_j}\|) \leq \frac{\sigma_{\max}^{\mathbf{A}_i} \lambda_{\max}^{\mathbf{M}}}{\sigma_{\min}^{\mathbf{A}_i} \lambda_{\min}^{\mathbf{M}}} \min_{\substack{m+s=n \\ m \geq 2, s \geq 2}} \left[ \left( 1 + \left( \frac{m}{s-1} \right)^{1/2} \right) \sigma_{t_i \rightarrow t_j}^{(m+1)} + \frac{e\sqrt{n}}{s} \left( \sum_{l>m} (\sigma_{t_i \rightarrow t_j}^{(l)})^2 \right)^{1/2} \right].$$

*Proof.* By applying the Courant minimax principle the result follows directly from [6, Proposition 3.2], which is based on [23, Theorem 10.6].  $\square$

As we can hope that the transfer operator has fast decaying singular values, the square root of the sum of squared singular values in the last term of (5.1) behaves often roughly as  $\sigma_{t_i \rightarrow t_j}^{(m+1)}$ . Hence, the error bound in (5.1) decays for increasing  $n$  if the singular values decay faster than  $n^{-1/2}$ . This is a valid assumption as the singular values often decay exponentially as can be seen in the numerical experiments (see, for instance, Fig. 8). For further details we refer to [6] where methods from randomized linear algebra [23] have been used to approximate the optimal local approximation spaces in the elliptic setting.

**5.2. Randomized reduced basis generation algorithm.** Algorithm 1 summarizes the embarrassingly parallel randomized basis generation. To provide a good approximation of the discrete solution of problem (2.1), we propose to randomly choose  $n_{\text{rand}} \in \mathbb{N}$  time points in  $\{t_0, \dots, t_{N_I-1}\}$  according to the probability distribution  $p$  that may be based on the time-dependent data functions of the PDE (see subsection 5.3 for details on the choice of  $p$ ). Recall that the time interval  $I$  is discretized via  $N_I \in \mathbb{N}$  time points  $0 = t_0 \leq \dots \leq t_{N_I-1} = T$ . To construct suitable ansatz functions in the chosen time points, we then apply the corresponding transfer operators to random initial conditions (cf. subsection 5.1) as illustrated in Fig. 4.

In detail, for each starting time point  $t_i$ , we draw in line 8 a Gaussian random vector  $\mathbf{rand\_u}_0 \sim N(\mathbf{0}, (\mathbf{A}_i^\top \mathbf{A}_i)^{-1})$ , where  $\mathbf{A}_i$  is the stiffness matrix introduced in (2.3). In line 9, we then compute the local solution of the PDE with initial condition  $\mathbf{rand\_u}_0$  for the respective local time interval employing the time stepping scheme  $\mathbf{t\_steps}$  that is suitably chosen by the user.

Subsequently, we add the resulting local solution trajectories evaluated at the last  $n_t - k + 1$  ( $k \leq n_t$ ) time points to the snapshot matrix  $\mathbf{S}$  in line 10. By choosing the parameter  $k$  smaller than  $n_t$ , we sample from the ranges of multiple transfer operators simultaneously (see Fig. 4 for an illustration), which often results in an improved approximation accuracy as the numerical experiments in section 6 show. For some guidance on the choice of  $k$  and  $n_t$  we refer to subsection 6.4. Moreover, we highlight that the computations in lines 7 to 11 are embarrassingly parallel.

**Algorithm 1:** Reduced basis generation via random sampling and SVD

---

```

1 Function RandomizedReducedBasisGeneration( $n_{\text{rand}}, n_t, k, p, \text{tol}$ ):
   Input : number of randomly chosen time points  $n_{\text{rand}}$ , number of time
           steps for local PDE simulations  $n_t$ , starting time step for
           collecting snapshots  $k$  ( $\leq n_t$ ), (data-dependent) probability
           distribution  $p$ , time stepping scheme  $\mathbf{t\_steps}$ , tolerance  $\text{tol}$ 
   Output: reduced basis  $\mathbf{U}_{\text{red}}$  chosen according to tolerance  $\text{tol}$ 
2  $\text{rand\_ints} \leftarrow$  draw  $n_{\text{rand}}$  integers in  $\{0, \dots, N_I - 1\}$  according to  $p$ 
3 if  $\text{rand\_ints}[i] \leq n_t$  then
4   | remove entry  $\text{rand\_ints}[i]$  from  $\text{rand\_ints}$ 
5  $\text{endpoints} \leftarrow \text{timegrid}[\text{rand\_ints}]$ ,  $\text{startpoints} \leftarrow \text{timegrid}[\text{rand\_ints} - n_t]$ 
   // initialize snapshot matrix
6  $\mathbf{S} \leftarrow \emptyset$ 
7 for  $i = 1, \dots, \#\text{startpoints}$  do
   // draw random initial condition
8    $\text{rand\_u}_0 \leftarrow \text{Gaussian}(\text{size} = N_D)$ 
   // solve locally and store solution at time steps  $k$  to  $n_t$ 
9    $\text{snapshots} \leftarrow \mathbf{t\_steps}(\text{startpoints}[i], \text{endpoints}[i], n_t, \text{rand\_u}_0)[:, k:n_t]$ 
10  |  $\mathbf{S} \leftarrow [\mathbf{S}, \text{snapshots}]$ 
   // add representation of  $\mathbf{u}_0$  for first  $n_t$  time steps
11  $\text{snapshots\_u}_0 \leftarrow \mathbf{t\_steps}(0, \text{timegrid}[n_t], n_t, \mathbf{u}_0)$ 
12  $\mathbf{S} \leftarrow [\mathbf{S}, \text{snapshots\_u}_0]$ 
   // compute SVD of collected snapshots and cut using  $\text{tol}$ 
13  $\mathbf{U}_{\text{red}}, -, - \leftarrow \text{svd}(\mathbf{S}, \text{tol})$ 

```

---

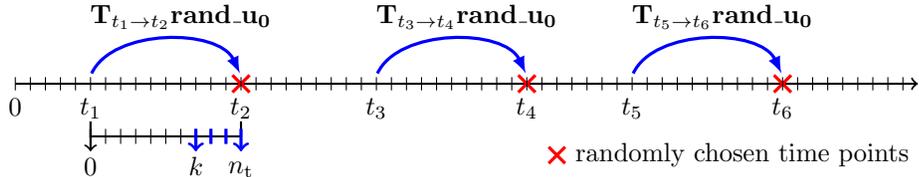


FIGURE 4. Evaluation of multiple transfer operators at randomly chosen points in time for random initial conditions. Sketch of parameters  $k$  and  $n_t$  in Algorithm 1.

As the local PDEs are solved for  $n_t (\ll N_I)$  time steps, we discard any chosen time point smaller than or equal to  $t_{n_t}$  in line 4 and add the evolution of the initial conditions  $\mathbf{u}_0$  for the first  $n_t$  time steps to the snapshot matrix in line 12.

Finally, we compress all collected snapshots via an SVD in line 13 to construct the reduced approximation space.

**Remark 5.2** (Choice of  $n_{\text{rand}}$ ). In Algorithm 1 the user determines the number of random initial conditions  $n_{\text{rand}}$  a priori, for instance, based on the knowledge that the employed computer architecture provides  $n_{\text{rand}}$  parallel compute units. While we compress the snapshot matrix via an SVD, which is advantageous in cases where no error estimator is available or is very costly to evaluate, we conjecture that it might

also be possible to develop an adaptive randomized algorithm that adaptively augments the reduced basis relying on a probabilistic a posteriori error estimator.

**Remark 5.3** (Number of random initial conditions per time point). *In lines 8-9 of Algorithm 1 we propose to draw one random initial condition and thus compute one local basis function per chosen time point. In light of the probabilistic a priori error bound (5.1), we conjecture that for transfer operators with fast decaying singular values this is sufficient in order to obtain a good approximation accuracy and we also observe this in the numerical experiments at least for the considered test cases.*

*However, for transfer operators with more slowly decaying singular values, we suggest to draw multiple random initial conditions per chosen time point to enhance the quality of the approximation. In that case, the computation in line 9 has to be split into (multiply) solving for random initial conditions and homogeneous right-hand side and (once) solving for homogeneous initial conditions and right-hand side.*

**Remark 5.4** (Computational complexity of Algorithm 1). *Assuming that we employ the discretization introduced in section 2 and subsection 4.2, the complexity for computing the  $n_{\text{rand}} + 1$  solution trajectories in lines 9 and 11 of Algorithm 1 is  $\mathcal{O}((n_{\text{rand}} + 1)n_t(N_D^{3/2} + N_D \log(N_D)))$  in two or  $\mathcal{O}((n_{\text{rand}} + 1)n_t(N_D^2 + N_D^{4/3}))$  in three spatial dimensions (cf. Remark 4.6). We highlight that the computations in lines 9 and 11 are embarrassingly parallel as the local solution trajectories can be computed completely independently from each other. Algorithm 1 is thus well-suited to be used on modern computer architectures allowing for many parallel computations.*

*Moreover, the computational complexity for compressing the collected snapshots via an SVD in line 13 is  $\mathcal{O}((n_t - k + 1)n_{\text{rand}} + n_t)N_D^2$ . Alternatively, we can approximate the SVD of  $\mathbf{S}$  via a randomized SVD [23]. For details on the computational complexity of assembling the probability distribution  $p$  we refer to subsection 5.3.*

**5.3. Choice of probability distribution.** In this subsection we discuss how to choose the probability distribution  $p$  that is used in Algorithm 1 for sampling points in time. As the behavior of the solution in time is influenced by the (time-dependent) data functions of the PDE, we guide the time point selection by the behavior of the data functions in time. To this end, we represent the time-dependent data functions as matrices, where each column corresponds to one time point and employ column subset selection techniques from randomized NLA [12, 16]. In detail, we focus on the following standard sampling strategies: uniform, squared norm, and leverage score sampling. While these methods are often used to construct low-rank matrix decompositions (cf., e.g., [34]), we exploit them, as a key new contribution, for the purpose of time point selection.

In the following, we assume that the matrix  $\mathbf{B} \in \mathbb{R}^{N_D \times N_I}$  represents the time-dependent data encoded in  $\mathcal{A}$  or  $\mathcal{F}$  (for instance, a coefficient function or source terms and boundary data that vary in time). In case of a time-dependent coefficient function, one option is to choose the entries of  $\mathbf{B}$  equal to the value of the function in the respective space-time nodes. For a time-dependent  $\mathcal{F}$ , one could set the columns of  $\mathbf{B}$  equal to the right-hand side vectors  $\mathbf{F}_l$  for  $0 \leq l \leq N_I - 1$  (cf. (2.3)).

Next, we present the probability distributions (see, e.g., [12, 16] for an overview).

**Uniform sampling.** In the case of uniform sampling, we sample a point of time  $t_i$  with probability  $p_i = 1/N_I$  for  $i = 0, \dots, N_I - 1$ , where  $N_I$  denotes the number of time points determined by the partition  $0 = t_0 \leq \dots \leq t_{N_I-1} = T$  of the time interval  $I$ . The computational costs for constructing the probability distribution are zero.

**Squared norm sampling.** If we employ squared norm sampling [20], we select a point of time  $t_i \in \{t_0, \dots, t_{N_I-1}\}$  with probability  $p_i = \|\mathbf{B}[:, i]\|_2^2 / \|\mathbf{B}\|_F^2$ , where  $\|\cdot\|_F^2$  denotes the Frobenius norm and  $\mathbf{B}[:, i]$  is the  $i$ -th column of  $\mathbf{B}$ . The computation of the probabilities  $p_i$  can be carried out in parallel and its complexity scales linearly in the number of non-zero entries of  $\mathbf{B}$ , e.g. for a dense matrix  $\mathbf{B}$  the computational complexity is  $\mathcal{O}(N_I N_D)$ .

**Leverage score sampling.** The leverage score sampling approach [17] captures the statistical leverage of the columns of  $\mathbf{B}$  on its best rank- $r$  approximation and preferably chooses columns which have a large influence on the best rank- $r$  fit of  $\mathbf{B}$  [34]. To this end, one computes the leading  $r$  right singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$  of  $\mathbf{B}$  and selects a column  $i$  of  $\mathbf{B}$  (a time point  $t_i \in \{t_0, \dots, t_{N_I-1}\}$ ) with probability  $p_i = 1/r \sum_{j=1}^r \mathbf{v}_j[i]^2$ , where  $\mathbf{v}_j[i]$  is the  $i$ -th entry of  $\mathbf{v}_j$ . The complexity for computing the SVD of  $\mathbf{B}$  is  $\mathcal{O}(r N_I N_D)$ . Alternatively, we can approximate the SVD of  $\mathbf{B}$  via a randomized SVD [23].

5.3.1. *Discussion and comparison of probability distributions.* In the following, we discuss and compare the probability distributions introduced in subsection 5.3 with respect to their capability of detecting relevant time points of (heterogeneous) time-dependent data functions and computational costs. For a more general comparison, we refer to [16].

As no data-dependent information is incorporated, the uniform sampling approach might not detect relevant problem-specific features in time unless a large number of time points is drawn and can thus lead to poor results. For instance, a data matrix  $\mathbf{B}$  with only one non-zero column would require to draw  $\mathcal{O}(N_I)$  columns (time points) in order to detect the non-zero data at the single point in time with high probability. In contrast to uniform sampling, there exist error bounds for both the squared norm and the leverage score sampling approach (cf. Appendix B).

As illustrated for two different time-dependent data functions<sup>2</sup> in Fig. 5, we can infer that the squared norm sampling approach might not detect parts of the data that has values on smaller scales or smaller temporal scales. In contrast, the leverage scores weight the heterogeneous parts of the data equally with the same expectation. As a result, the leverage score sampling approach more likely detects all dominant features of the heterogeneous data and might lead to a better approximation accuracy than the squared norm approach. Moreover, as can be expected from the definition of leverage scores, we observe from numerical experiments not included in this paper that leverage scores are capable of detecting repetitions in the data functions.

Nevertheless, the costs for computing the rank- $r$  leverage scores are  $r$  times the costs of computing the squared norm sampling distribution (for a dense matrix  $\mathbf{B}$ ). To reduce the computational complexity, one could employ a randomized SVD and parallelize computations. However, this potentially still leads to costs that are not negligible compared to the uniform and squared norm sampling approach.

Consequently, uniform (and squared norm) sampling can be advantageous compared to leverage score sampling if many parallel compute units are available and the data is, for instance, spread relatively uniform over the whole time interval. In that case,

---

<sup>2</sup>To ensure reproducibility, the data functions and discretization parameters for Example 1 are listed in Appendix C.

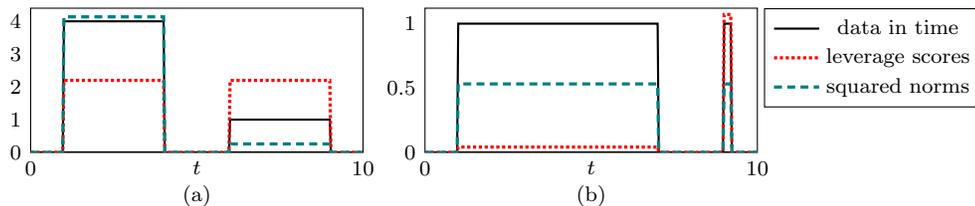


FIGURE 5. Example 1: Rank-2 leverage score (LS) and squared norm (SN) probability distributions for two (spatially disjoint) signals with different values (a) or different temporal scales (b). Values of LS and SN are scaled with a factor of 400 (a) or 15 (LS) and 100 (SN) (b).

drawing a large number of time points will likely yield a good approximation accuracy without having to compute (or approximate) the SVD of a large matrix.

5.3.2. *Sampling from multiple probability distributions.* If both  $\mathcal{A}$  and  $\mathcal{F}$  encode time-dependent data due to, for instance, a coefficient function and source terms that vary in time, one might want to include both data matrices  $\mathbf{B}_{\mathcal{A}}$  and  $\mathbf{B}_{\mathcal{F}}$  in the time point selection process.

To this end, one option is to attach  $\mathbf{B}_{\mathcal{A}}$  to  $\mathbf{B}_{\mathcal{F}}$  and sample from the probability distribution computed from  $[\mathbf{B}_{\mathcal{A}}\mathbf{B}_{\mathcal{F}}] \in \mathbb{R}^{N_D \times 2N_I}$ . However, if the data encoded in  $\mathcal{F}$ , for instance, has values on smaller scales compared to  $\mathcal{A}$  (due to, e.g., high conductivity channels), employing the squared norm sampling approach one might sample solely from the part of the probability distribution that is associated with  $\mathcal{A}$  and neglect the information encoded in  $\mathcal{F}$  (cf. Fig. 5 and the discussion in subsection 5.3.1). As leverage score sampling is based on the SVD of  $[\mathbf{B}_{\mathcal{A}}\mathbf{B}_{\mathcal{F}}]$ , the approach more likely detects the dominant modes encoded in both  $\mathcal{A}$  and  $\mathcal{F}$ . Nevertheless, computing the SVD of the large matrix  $[\mathbf{B}_{\mathcal{A}}\mathbf{B}_{\mathcal{F}}] \in \mathbb{R}^{N_D \times 2N_I}$  is more costly than computing the SVDs of  $\mathbf{B}_{\mathcal{A}} \in \mathbb{R}^{N_D \times N_I}$  and  $\mathbf{B}_{\mathcal{F}} \in \mathbb{R}^{N_D \times N_I}$  separately. Therefore, we propose to assemble the probability distributions associated with  $\mathbf{B}_{\mathcal{A}}$  and  $\mathbf{B}_{\mathcal{F}}$  separately and draw from both distributions simultaneously. Moreover, we note that in certain cases it might be necessary to also sample from data functions that are constant in time to achieve a good quality of approximation (e.g. in case of a constant advection field, cf. Experiment 3 in subsection 6.2).

## 6. NUMERICAL EXPERIMENTS

In this section we demonstrate the excellent approximation properties of the reduced basis generated via Algorithm 1. In subsections 6.1 and 6.2 we first comprehensively test how the results depend on various parameters such as the number of chosen time points  $n_{\text{rand}}$ , the local oversampling size  $n_t$ , the number of collected snapshots for the SVD determined via  $k$ , or the probability distribution for drawing points in time. For this purpose, we consider both the linear heat equation and an advection-diffusion problem for time-dependent source terms. In particular, we demonstrate that the proposed method is able to tackle higher values of advection. Subsequently, we show in subsection 6.3 that the randomized approach is well capable of approximating a problem with a time-dependent permeability coefficient that is rough with respect to both space and time using real-world data taken from the SPE10 benchmark problem [9]. Moreover, we demonstrate in subsections 6.2 and 6.3 how to sample from multiple probability distributions simultaneously.

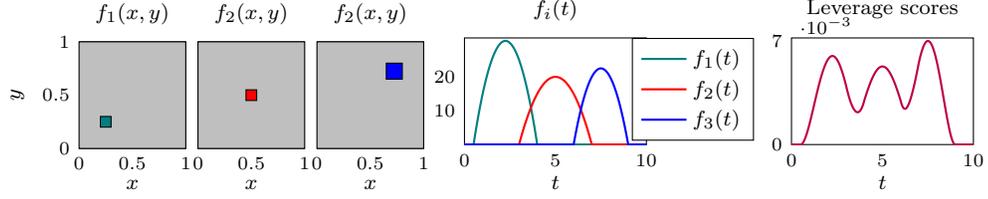


FIGURE 6. Example 2: Source term  $f(t, x, y) = \sum_{i=1}^3 f_i(t)f_i(x, y)$  and corresponding rank-3 leverage score probability distribution. Gray equates to 0, while green, red, and blue equate to 1 (left).

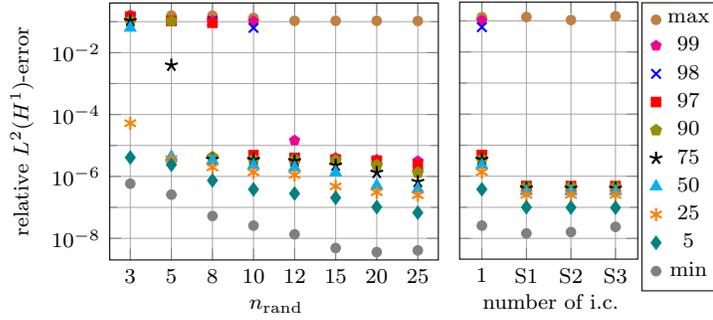


FIGURE 7. Example 2: Quantiles of relative  $L^2(I, H^1(D))$ -error for  $n_t = 15$ ,  $k = 13$ ,  $\text{tol} = 10^{-8}$ , and 100.000 realizations for varying numbers of  $n_{\text{rand}}$  (left) or  $n_{\text{rand}} = 10$  and 1, 2, or 3 random initial conditions (i.c.) per time point (right). Here, S indicates that local computations are performed separately for right-hand side and initial conditions (cf. Remark 5.3).

For the experiments, we employ the discretization introduced in section 2, use Algorithm 1 to generate the randomized reduced basis, and construct the reduced approximation via Galerkin projection as described in section 2. Hence, in what follows the term *error* always refers to the error between the solution of (2.4) and its reduced approximation determined by solving the reduced problem (2.5). Moreover, we prescribe homogeneous Dirichlet boundary conditions on  $I \times \Sigma_D$  in all experiments. The source code to reproduce all results shown in this section is provided in [48].

**6.1. Stove problem.** In this subsection, we consider the heat equation ((2.2) with  $b \equiv c \equiv 0$ ) and investigate the following numerical experiment, which we refer to as Example 2: We choose  $I = (0, 10)$ ,  $D = (0, 1)^2$ ,  $\Sigma_N = \emptyset$ , and discretize the spatial domain  $D$  with a regular quadrilateral mesh with mesh size  $1/100$  in both directions. For the implicit Euler method, we use an equidistant time step size of  $1/30$ . Furthermore, we choose the initial condition  $u_0(x, y) = \sum_{i=1}^3 \sin(i\pi x) \sin(i\pi y)$ , the coefficient  $\kappa \equiv 1$ , and the source term  $f(t, x, y) = \sum_{i=1}^3 f_i(t)f_i(x, y)$  involving three spatially disjoint heat sources (stoves) that are turned on and off in time as illustrated in Fig. 6 (left and middle). Fig. 6 (right) shows the rank-3 leverage score probability distribution computed from the right-hand side matrix  $\mathbf{F}$ , whose columns are the right hand-side vectors  $\mathbf{F}_l$  for  $l = 0, \dots, 300$  (2.3). Unless stated otherwise, we use the rank-3 leverage score probability distribution to draw time points in Algorithm 1 for this example.

First, we investigate the influence of the number of drawn time points  $n_{\text{rand}}$  on the approximation accuracy of the reduced basis. As the rank of the right-hand side is three, we choose  $n_{\text{rand}} \geq 3$ . In Fig. 7 (left) we observe that for  $n_{\text{rand}} = 3$  in 50% of

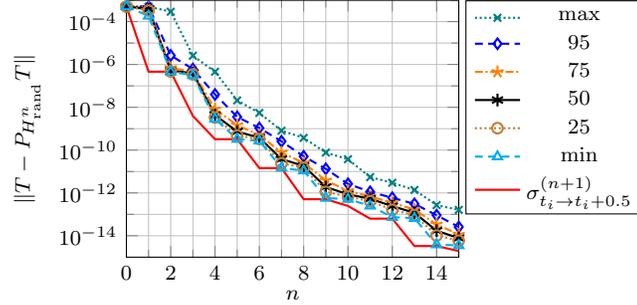


FIGURE 8. Example 2: Singular value decay of transfer operator  $T_{t_i \rightarrow t_i + 15/30}$  ( $0 \leq i \leq 285$ ) and quantiles of projection error  $\|T_{t_i \rightarrow t_i + 0.5} - P_{H_{t_i + 0.5, \text{rand}}^n} T_{t_i \rightarrow t_i + 0.5}\|$  over basis size  $n$  for 2.000 realizations.

cases the error is of the order of  $10^{-1}$  and only in 5% of cases the error is below  $10^{-5}$ , while for  $n_{\text{rand}} = 8$  or  $n_{\text{rand}} = 10$  in 90% or 97% of cases the error is below  $10^{-5}$ . We thus infer that only a small amount of oversampling is necessary to detect all three stoves with high probability. If we choose  $n_{\text{rand}} > 10$  we see in Fig. 7 (left) that the approximation accuracy still slightly improves compared to smaller  $n_{\text{rand}}$ . Nevertheless, we recall that also the computational costs increase with increasing  $n_{\text{rand}}$ . For the following tests in this subsection we therefore choose  $n_{\text{rand}} = 10$  as a good trade-off between approximation accuracy and computational costs.

Next, we test how the approximation accuracy depends on the number of random initial conditions per drawn time point. We observe that for S1-S3 99% of the realizations yield an approximation with a relative  $L^2(I, H^1(D))$ -error below  $10^{-6}$  and the approximation accuracy does not improve if we choose more than one random initial condition. This can be traced back to the very fast, exponential decay of the singular values of the transfer operator shown in Fig. 8 (cf. Proposition 5.1). If we draw only one random initial condition per chosen time point and do not separate the local computation for right-hand side and random initial condition, we observe in Fig. 7 (right) that 97% of the realizations yield an approximation with a relative  $L^2(I, H^1(D))$ -error below  $10^{-5}$ . As the approximation accuracy is thus very good and the costs for computing local solutions are only half as much as in the separated approach (S1), we choose to not separate local computations for right-hand side and random initial condition in all other tests in this subsection.

We observe in Fig. 8 that while the randomized spaces often provide an approximation that converges nearly with the optimal rate  $\sigma_{t_i \rightarrow t_i + 0.5}^{(n+1)}$ , in some cases two or three basis vectors more are needed to guarantee the same approximation accuracy. This is in line with the predictions by theory (cf. Proposition 5.1).

Moreover, we investigate how the number of collected snapshots determined via the parameter  $k$  influences the approximation accuracy of the reduced basis. In Fig. 9 (left) we observe that the approximation quality significantly improves if we collect not only solution snapshots at local end time points ( $k = 15$ ), but at the (locally) last two or three time points ( $k = 14$  or  $k = 13$ ) as, for instance, 75% of realizations have a relative  $L^2(I, H^1(D))$ -error below  $10^{-5}$  for  $k = 15$ , while for  $k = 14$  or  $k = 13$  already 90% or 97% of realizations have a relative  $L^2(I, H^1(D))$ -error below  $10^{-5}$ . Fig. 10 illustrates an explanation for this phenomenon. In the particular realization only time points in the first and last stove are drawn from the probability distribution. However, by

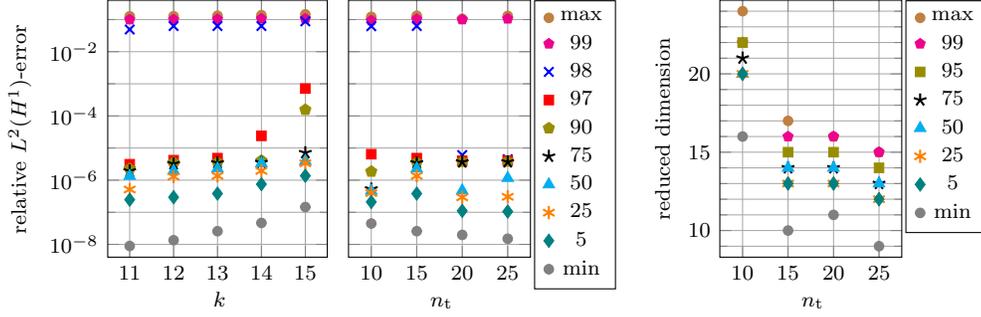


FIGURE 9. Example 2: Quantiles of relative  $L^2(I, H^1(D))$ -error for  $n_t = 15$ , varying numbers of  $k$ ,  $n_{\text{rand}} = 10$ ,  $\text{tol} = 10^{-8}$ , and 100.000 realizations (left). Quantiles of relative  $L^2(I, H^1(D))$ -error (middle) and reduced dimension (right) for varying numbers of  $n_t$ ,  $k = n_t - 2$ ,  $n_{\text{rand}} = 10$ ,  $\text{tol} = 10^{-8}$ , and 100.000 realizations.

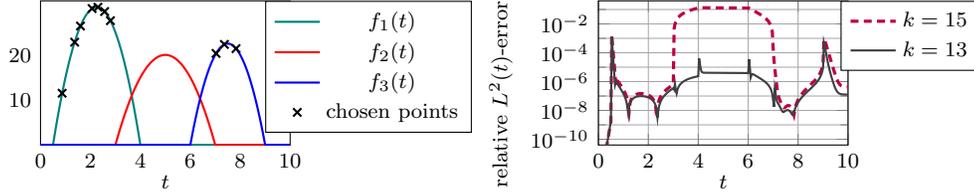


FIGURE 10. Example 2:  $n_{\text{rand}} = 10$  randomly chosen points (left) and relative  $L^2(t)$ -error of corresponding reduced approximation for  $n_t = 15$ ,  $\text{tol} = 10^{-8}$  and  $k = 15$  vs.  $k = 13$  (right).

choosing  $k = 13$  we include a solution snapshot that also detects the second stove and can thus significantly decrease the error in the time interval  $(3, 7)$  for this particular realization compared to  $k = 15$  as shown in Fig. 10 (right). For  $k = 12$  or  $k = 11$  we observe in Fig. 9 (left) only very slight improvements in the approximation quality compared to  $k = 13$ . We therefore choose  $k = n_t - 2$  in all other tests in this subsection as a trade-off between approximation quality and size of the snapshot matrix and thus computational costs for its SVD.

Subsequently, we test how the approximation accuracy depends on the local oversampling size  $n_t$ . For all tested oversampling sizes, and in particular already for a small oversampling size of  $n_t = 10$ , we observe in Fig. 9 (middle) that 97% of realizations have a relative  $L^2(I, H^1(D))$ -error below  $10^{-5}$  (98% for  $n_t = 20$  and  $n_t = 25$ ). Nevertheless, we see in Fig. 9 (right) that for  $n_t = 10$  the reduced basis is significantly larger compared to  $n_t = 15$  (20, 25) as, for instance, in 95% of cases the reduced dimension is larger than or equal to 20 for  $n_t = 10$ , while for  $n_t = 15, 20, 25$  it is smaller than or equal to 15 in 95% of cases. This can be explained by the fact that for smaller  $n_t$  the randomness of the initial conditions has a larger influence compared to larger  $n_t$  due to the exponential decay behavior of (local) solutions in time. In Fig. 9 (middle) we observe that for  $n_t = 20$  or  $n_t = 25$  the quality of approximation slightly improves compared to  $n_t = 15$ , but we recall that also the computational costs increase with increasing  $n_t$ . For all other tests in this subsection we therefore choose  $n_t = 15$  as a good trade-off between approximation quality, computational costs, and size of the reduced basis.

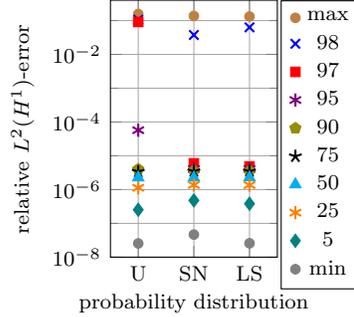


FIGURE 11. Example 2: Quantiles of relative  $L^2(I, H^1(D))$ -error for  $n_{\text{rand}} = 10$ ,  $n_t = 15$ ,  $k = 13$ ,  $\text{tol} = 10^{-8}$ , 100.000 realizations, and uniform (U), squared norm (SN), or leverage score (LS) probability distribution.

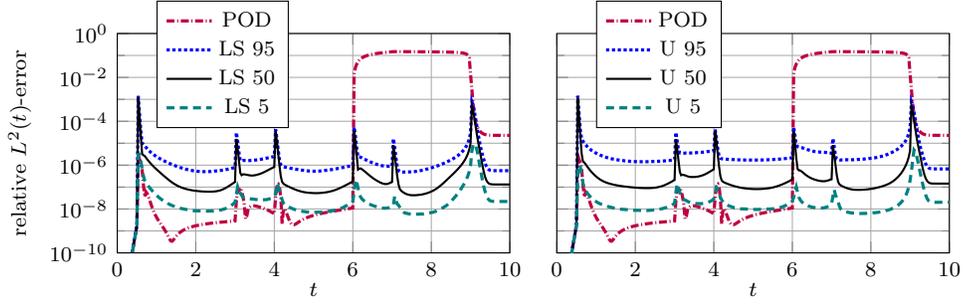


FIGURE 12. Example 2: 5, 50, and 95% quantiles of relative  $L^2(t)$ -error for  $n_{\text{rand}} = 10$ ,  $n_t = 15$ ,  $k = 13$ ,  $\text{tol} = 10^{-8}$ , 100.000 realizations, and leverage score (LS) or uniform (U) probability distribution vs. relative  $L^2(t)$ -error for POD on solution trajectory of first 165 of 300 time steps with tolerance  $10^{-8}$  (POD).

Next, we investigate how the choice of the probability distribution influences the approximation accuracy for the considered test case. To this end, we employ either the uniform, squared norm, or leverage score probability distribution as introduced in subsection 5.3 for drawing time points in Algorithm 1. Both squared norms and leverage scores are computed from the right-hand side matrix  $\mathbf{F}$ . In Fig. 11 we observe that the squared norm and the leverage score probability distribution yield a comparably good approximation quality and in both cases 97% of realizations have a relative  $L^2(I, H^1(D))$ -error below  $10^{-5}$ . The uniform sampling approach also achieves a good approximation quality as, for instance, 95% (90%) of realizations have a relative  $L^2(I, H^1(D))$ -error below  $10^{-4}$  ( $10^{-5}$ ). Therefore, the results shown in Fig. 11 indicate that for data that is spread over almost the whole time interval, aside from the leverage scores, the uniform or squared norm sampling approach can also lead to very good approximation results. Especially if the computer architecture allows for many parallel computations, one can thus save expenses for computing an SVD of a potentially large data matrix and instead draw a large number of time points from the uniform (or squared norm) probability distribution as discussed in subsection 5.3.1.

As the POD is a well-established tool for compressing and reducing time trajectories, we compare in Fig. 12 quantiles of the relative  $L^2(t)$ -approximation error for both the uniform and the leverage score sampling approach with the relative  $L^2(t)$ -error of the

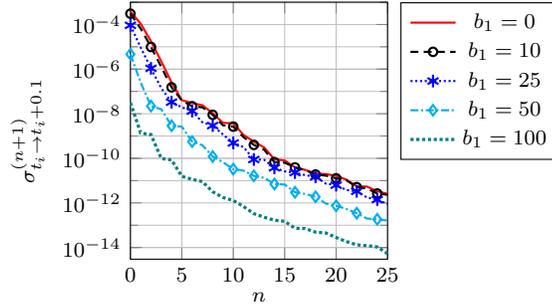


FIGURE 13. Example 3a: Singular value decay of transfer operator  $T_{t_i \rightarrow t_i + 0.1}$  ( $0 \leq i \leq 491$ ) for different values of advection in  $x$ -direction and a constant diffusion of 1.

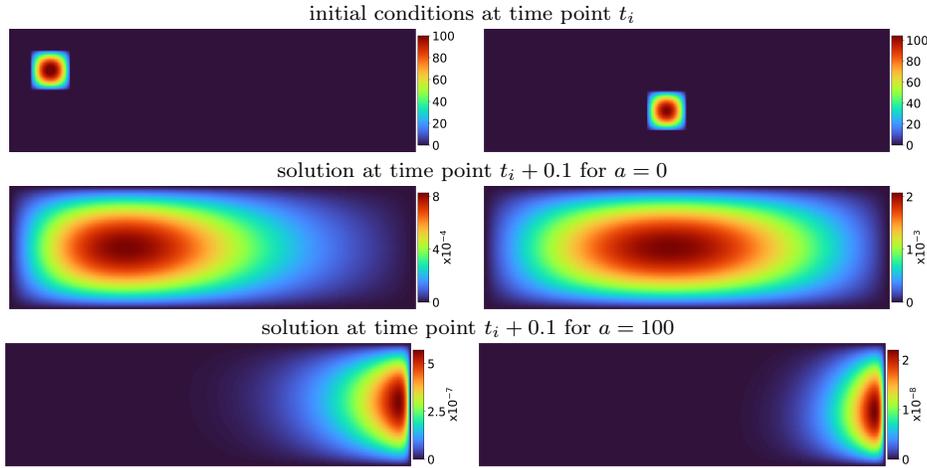


FIGURE 14. Example 3a: Local solutions for different initial conditions and advection  $a = 0$  vs.  $a = 100$ .

approximation via POD on the solution trajectory of the first 165 of 300 time steps. In this way, we compare the standard POD approach with the randomized approach based on the same computational budget of  $(n_{\text{rand}} + 1) \cdot n_t = (10 + 1) \cdot 15 = 165$  time steps. While we also have to compute the SVD of the solution trajectory for the POD and the SVD of the data matrix for the leverage score sampling approach, we here focus on equaling the budget based on the time stepping, which likely dominates the computational costs in complex applications. We observe in Fig. 12 that the POD approach, in contrast to the randomized approach in at least 95% of cases, is not able to detect the third stove and thus yields a much larger relative approximation error in the time interval (6,9) compared to both the uniform and the leverage score sampling approach. Moreover, in the randomized approach the local PDE simulations can be performed in parallel, while for the POD the first 165 time instances of the solution trajectory have to be computed sequentially.

**6.2. Advection-diffusion problem.** Here, we consider an advection-diffusion problem ((2.2) with  $c \equiv 0$ ) and first investigate the following numerical experiment, which we refer to as Example 3a: We choose  $I = (0, 5)$ ,  $D = (0, 1) \times (0, 0.3)$ ,  $\Sigma_N = \emptyset$ , and

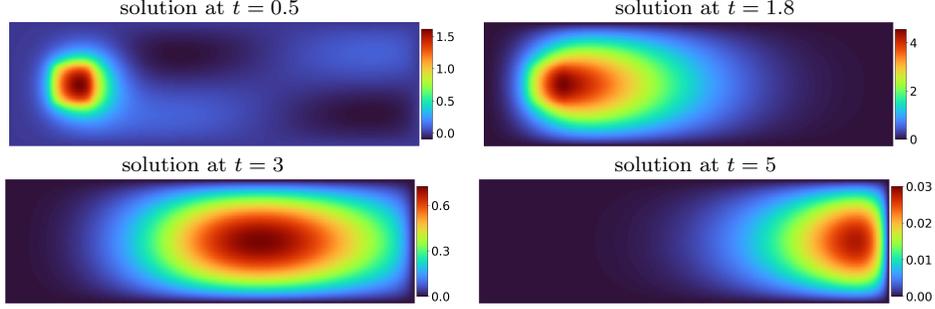


FIGURE 15. Example 3b: Solution evaluated at different points in time. The shades in the middle and right part of the spatial domain at time point  $t = 0.5$  are the vanishing initial conditions.

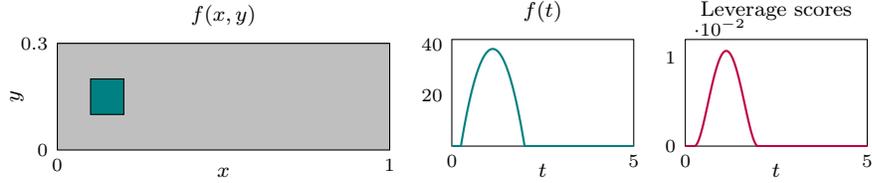


FIGURE 16. Example 3b: Source term  $f(t, x, y) = f(t)f(x, y)$  and corresponding rank-1 leverage score probability distribution (identical to squared norms). Gray equates to 0, green equates to 1 (left).

discretize the spatial domain  $D$  with a regular quadrilateral mesh with mesh size  $1/300$  in both directions. For the implicit Euler method, we use an equidistant time step size of  $1/100$ . Furthermore, we consider the advection field  $b = (b_1, 0)^\top$  for varying constants  $b_1 \in \{0, 10, 25, 50, 100\}$  and choose the conductivity coefficient  $\kappa \equiv 1$ . First, we observe in Fig. 13 a very fast, exponential decay of the singular values of the transfer operators mapping arbitrary initial conditions at time point  $t_i$  to the local solution at time point  $t_i + 0.1$ . Moreover, we see that the singular values are smaller for higher values of advection. This can be traced back to the fact that for a higher value of advection (i.e.  $b_1 = 100$ ) initial conditions at time  $t_i$  move quickly in  $x$ -direction and the corresponding local solutions at time  $t_i + 0.1$  have support only in the right half of the spatial domain close to the right boundary, see Fig. 14 (bottom). In contrast, the plots in Fig. 14 (middle) show that in the absence of advection (i.e.  $b_1 = 0$ ) the local solutions corresponding to different initial conditions differ more from each other compared to the case  $b_1 = 100$  and spread over the whole spatial domain. We can thus infer for this test case that the range of the transfer operator (i.e. the local solution space at a point of time) is extremely low-dimensional for higher values of advection and therefore well amenable to approximation.

Next, we consider an advection-diffusion problem with a solution that is constantly moving in time as shown in Fig. 15, which we refer to as Example 3b. We choose  $I$ ,  $D$ ,  $\Sigma_N$ , and the discretization as in Example 3a. Moreover, we choose the advection field  $b = (0.3, 0)^\top$ , the conductivity coefficient  $\kappa = 0.01$ , the initial condition  $u_0(x, y) = \sum_{i=1}^3 \sin(i\pi x) \sin(i\pi y)$ , and the source term  $f(t, x, y) = f(t)f(x, y)$  as depicted in Fig. 16 (left). For drawing time points in Algorithm 1, we use the rank-1 leverage score probability distribution (see Fig. 16 (right)) computed from the right-hand side matrix  $\mathbf{F}$ , whose columns are the right hand-side vectors  $\mathbf{F}_l$  for  $l = 0, \dots, 500$

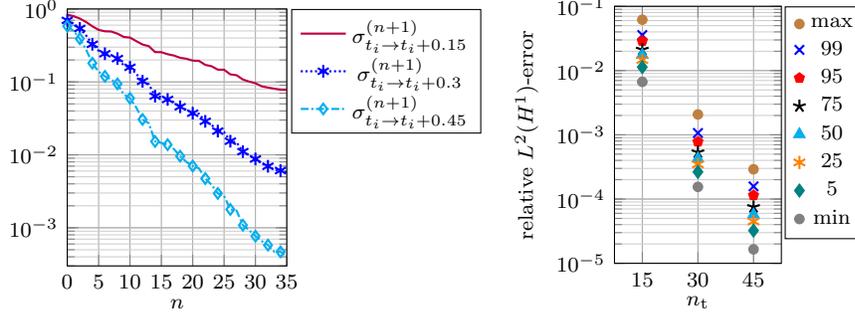


FIGURE 17. Example 3b: Singular value decay of transfer operators  $T_{t_i \rightarrow t_i + 0.15}$ ,  $0 \leq i \leq 486$ ,  $T_{t_i \rightarrow t_i + 0.3}$ ,  $0 \leq i \leq 471$ , and  $T_{t_i \rightarrow t_i + 0.45}$ ,  $0 \leq i \leq 466$ , (left). Quantiles of relative  $L^2(I, H^1(D))$ -error for varying numbers of  $n_t$ ,  $k = n_t - 6$ ,  $n_{\text{rand}}^{\text{rhs}} = 10$ ,  $n_{\text{rand}}^{\text{advec}} = 20$ ,  $\text{tol} = 10^{-8}$ , and 10.000 realizations (right). Singular values decay slower due to lower diffusion compared to Example 2, 3a, and 4.

(2.3). We emphasize that the latter distribution is identical to the squared norm probability distribution for this test case. As the solution is constantly moving in time (cf. Fig. 15), we additionally sample uniformly from the global time grid to capture the global advection in time. We denote the number of chosen time points by  $n_{\text{rand}}^{\text{rhs}}$  and  $n_{\text{rand}}^{\text{advec}}$ , respectively.

First, we observe in Fig. 17 (left) that the singular values of the transfer operator decay exponentially, but more slowly compared to Example 2, 3a, and 4. This is mainly due to the lower diffusion compared to Example 2, 3a, and 4, which results in a slower propagation of (local) solutions in time. However, we see that the singular values decrease faster with an increasing local oversampling size and also observe in Fig. 17 (right) that the approximation accuracy increases with increasing  $n_t$ : While for  $n_t = 15$  95% of realizations yield an approximation with a relative  $L^2(I, H^1(D))$ -error above  $10^{-2}$ , for  $n_t = 30$  or  $n_t = 45$  already 99% of realizations have a relative error below  $10^{-3}$  or  $2 \cdot 10^{-4}$ . For all other tests in this subsection we therefore choose  $n_t = 30$  as a good trade-off between approximation quality and computational costs.

Next, we investigate how the approximation accuracy is influenced by the number of time points selected from the uniform sampling distribution and how many samples are required in order to sufficiently capture the advection in the global problem. In Fig. 18 (left) we see that for  $n_{\text{rand}}^{\text{advec}} = 0$  or  $n_{\text{rand}}^{\text{advec}} = 5$  in 95% of cases the relative  $L^2(I, H^1(D))$ -error is above  $2 \cdot 10^{-2}$  or  $7 \cdot 10^{-3}$ , while for  $n_{\text{rand}}^{\text{advec}} = 20$  the relative error is already below  $10^{-3}$  in 99% of cases. If we choose  $n_t > 20$  we see in Fig. 18 (left) that the approximation accuracy still slightly improves compared to smaller  $n_{\text{rand}}^{\text{advec}}$ . We therefore choose  $n_{\text{rand}}^{\text{advec}} = 20$  in all other tests in this subsection.

Finally, we test how the approximation accuracy depends on the number of collected snapshots determined via  $k$ . In Fig. 18 (right) we observe that the approximation quality significantly improves if we collect not only solution snapshots at local end time points ( $k = 30$ ), but at the (locally) last three, five, or seven time points ( $k = 28, 26, 24$ ) as, for instance, for  $k = 30$  all realizations have a relative  $L^2(I, H^1(D))$ -error above  $3 \cdot 10^{-2}$ , while for  $k = 28$  or  $k = 24$  already 99% of realizations have a relative error below  $4 \cdot 10^{-3}$  or  $10^{-3}$ . This can be explained by the fact that by choosing a smaller  $k$  the propagation of the solution can be captured more accurately. For  $k = 22$  or  $k = 20$

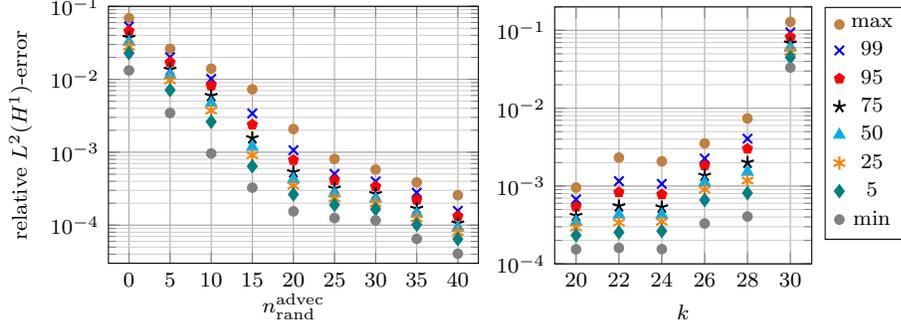


FIGURE 18. Example 3b: Quantiles of relative  $L^2(I, H^1(D))$ -error for varying numbers of  $n_{\text{rand}}^{\text{advec}}$ ,  $n_t = 30$ ,  $k = 24$ ,  $n_{\text{rand}}^{\text{rhs}} = 10$ ,  $\text{tol} = 10^{-8}$ , and 10.000 realizations (left). Quantiles of relative  $L^2(I, H^1(D))$ -error for  $n_t = 30$ , varying numbers of  $k$ ,  $n_{\text{rand}}^{\text{rhs}} = 10$ ,  $n_{\text{rand}}^{\text{advec}} = 20$ ,  $\text{tol} = 10^{-8}$ , and 10.000 realizations (right).

we observe in Fig. 18 (right) only very slight improvements in the approximation quality compared to  $k = 24$ . We therefore choose  $k = n_t - 6$  in all other tests in this subsection.

The computational costs of the randomized approach exceed the costs of the POD for this particular test case. However, we highlight that the computation of the local solution trajectories in Algorithm 1 is embarrassingly parallel while for the POD the global solution trajectory has to be computed in a sequential manner. Depending on the employed computer architecture the randomized approach can thus still lead to a significant speed-up compared to the standard POD approach.

**6.3. Problem with a time-dependent permeability coefficient.** In this subsection we consider a numerical experiment including the real-world permeability coefficient  $\kappa_0$  taken from the SPE10 benchmark problem [9], see Fig. 19. We refer to this experiment as Example 4. In Fig. 19 we observe that the solution trajectory of the problem is quite complex due to different configurations and combinations of permeability and inflow into the domain depicted in Fig. 20.

In detail, we consider the heat equation ((2.2) with  $b \equiv c \equiv 0$ ), choose  $I = (0, 10)$  and  $D = (0, 2.2) \times (0, 0.6)$  with Dirichlet and Neumann boundary as shown in Fig. 20 (left) and discretize the spatial domain  $D$  with a regular quadrilateral mesh with mesh size  $1/100$  in both directions. For the implicit Euler method, we use an equidistant time step size of  $1/50$ . We impose Neumann boundary conditions  $g_N(t, x, y) = g_N(t)$  for  $(x, y) \in (0.4, 1.8) \times \{0.6\}$  modeling a time-dependent inflow as depicted in Fig. 20 (middle) and  $g_N = 0$  elsewhere in  $I \times \Sigma_N$ . The permeability coefficient  $\kappa(t, x, y) = \kappa_0(x, y) + \kappa_1(t) \cdot \kappa_1(x, y) + \kappa_2(t) \cdot \kappa_2(x, y)$  is given by a sum of the permeability field  $\kappa_0$  from [9] shown in Fig. 19 (top, left) and high conductivity channels  $\kappa_1$  and  $\kappa_2$  that are turned on and off in time as depicted in Fig. 20 (middle). Moreover, the initial conditions are given by  $u_0(x, y) = 1$  for  $(x, y) \in (0.5, 0.7) \times (0.3, 0.4)$  and  $u_0(x, y) = 0$  else. As both the permeability  $\kappa$  and the inflow  $g_N$  vary in time, we sample from two probability distributions simultaneously to draw time points in Algorithm 1 (cf. the discussion in subsection 5.3.2). For this purpose, we employ the rank-3 leverage score probability distribution computed from the matrix, whose columns contain the values of  $\kappa(t_l, \cdot, \cdot)$  for all spatial elements at time point  $t_l$  for  $l = 0, \dots, 500$ . In addition, we use the rank-1 leverage score probability distribution computed from the right hand-side

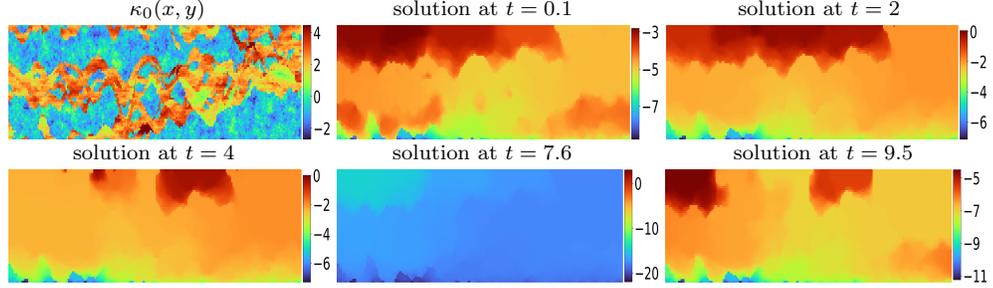


FIGURE 19. Example 4: Permeability field  $\kappa_0$  from [9] and solution evaluated at different points in time, plotted in logarithmic values to the base of 10.

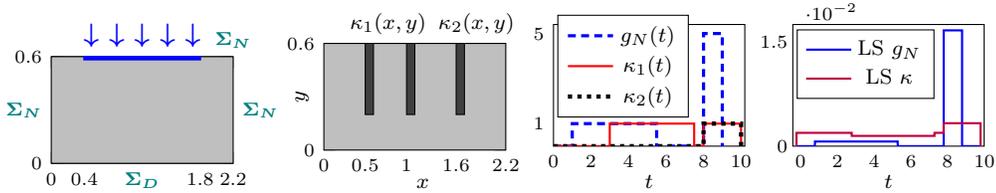


FIGURE 20. Example 4: Dirichlet and Neumann boundary  $\Sigma_D$ ,  $\Sigma_N$  (left), high conductivity channels  $\kappa_1(t) \cdot \kappa_1(x, y) + \kappa_2(t) \cdot \kappa_2(x, y)$ , and Neumann boundary data  $g_N(t, x, y) = g_N(t)$  for  $(x, y) \in (0.4, 1.8) \times \{0.6\}$ ,  $g_N(t, x, y) = 0$  else (middle). Dark gray equates to  $10^3$ , light gray to 0. Rank-1 (LS) associated with  $g_N$  and rank-3 LS corresponding to  $\kappa = \kappa_0 + \kappa_1 + \kappa_2$  (right).

matrix  $\mathbf{F}$ , whose columns are the right hand-side vectors  $\mathbf{F}_l$  for  $l = 0, \dots, 500$  (2.3). Both distributions are depicted in Fig. 20 (right).

First, we test how the approximation accuracy depends on the local oversampling size  $n_t$ . For all tested sizes of  $n_t$ , and especially for a small oversampling size of  $n_t = 10$ , we observe in Fig. 21 (left) that in 88% of cases the relative  $L^2(I, H^1(D))$ -error is below  $2 \cdot 10^{-2}$  and the algorithm succeeds in detecting all different configurations of the time-dependent data functions and thus all different shapes of the solution. Nevertheless, as already observed for Example 2, we see in Fig. 21 (middle) that for  $n_t = 10$  the reduced basis is significantly larger compared to  $n_t = 15$  (20, 25) as, for instance, in 95% of cases the reduced dimension is larger than or equal to 59 for  $n_t = 10$ , while for  $n_t = 15$  (20, 25) it is smaller than or equal to 50 (41, 36) in 95% of cases. The results thus confirm the findings from subsection 6.1 for this test case and in the following we choose  $n_t = 15$ .

Next, we compare the relative  $L^2(t)$ -approximation error for one realization of Algorithm 1 with the relative  $L^2(t)$ -error of the approximation via POD on the solution trajectory of the first 315 of 500 time steps. As already motivated in subsection 6.1, we focus on equaling the computational budget based on the time stepping  $((n_{\text{rand}} + 1) \cdot n_t = (20 + 1) \cdot 15 = 315)$ . In Fig. 21 (right) we observe that the POD yields a much larger error in the time interval (8, 10) compared to the randomized approach and is not able to detect the high conductivity channel  $\kappa_2$ . The randomized approach thus outperforms the POD for this test case even in the sequential setting.

As we observe that the singular values of the transfer operators do not decay as fast as in case of Example 2, see Fig. 22 (left) (and cf. Fig. 8 (right) and the discussion in subsection 6.1), we investigate how the approximation quality depends on the number of random initial conditions per chosen time point. In Fig. 22 (right) we see that the

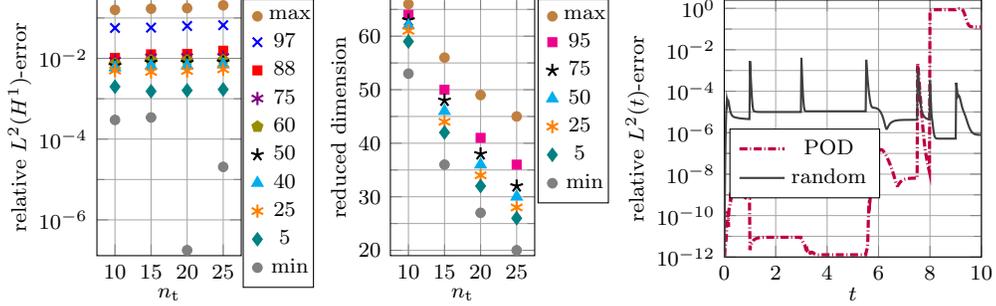


FIGURE 21. Example 4: Quantiles of relative  $L^2(I, H^1(D))$ -error (left) and reduced dimension (middle) for varying numbers of  $n_t$ ,  $k = n_t - 2$ ,  $n_{\text{rand}} = 20(10 + 10)$ ,  $\tau_{\text{ol}} = 10^{-8}$ , and 25.000 realizations. Relative  $L^2(t)$ -error for one realization of Algorithm 1 for  $n_t = 15$  (random) vs. POD on solution trajectory of first 315 of 500 time steps with tolerance  $10^{-8}$  (POD) (right).

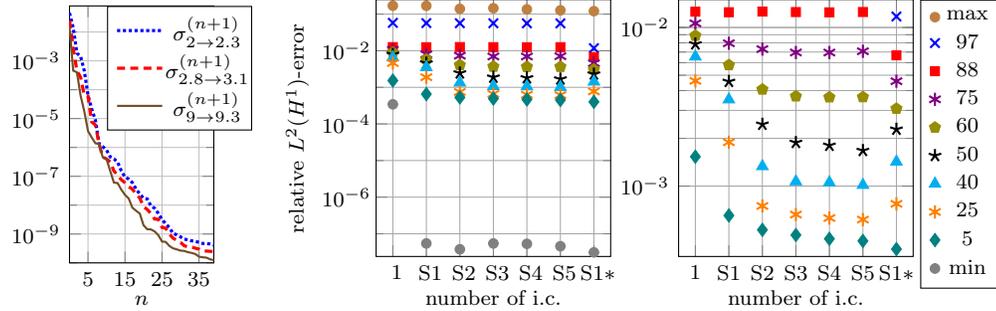


FIGURE 22. Example 4: Singular values of transfer operators (left). Quantiles of relative  $L^2(I, H^1(D))$ -error for  $n_t = 15$ ,  $k = 13$ ,  $n_{\text{rand}} = 20(10 + 10)$ ,  $\tau_{\text{ol}} = 10^{-8}$ , 25.000 realizations, and 1 – 5 random initial conditions (i.c.) per time point (right). S indicates that local computations are performed separately for right-hand side and initial conditions, \* indicates that  $n_{\text{rand}} = 30(15 + 15)$ .

accuracy of the approximation improves if we choose two (S2) instead of one (1, S1) random initial condition per drawn time point. For three or more random initial conditions (S3-S5), the approximation accuracy is at a comparable level with S2. Moreover, we can alternatively improve the approximation quality for this test case by increasing the number of drawn time points  $n_{\text{rand}}$ . If we employ the same computational budget as for S2, but drawn only one random initial condition for 30 instead of 20 drawn time points (S1\*), we observe that in 97% of cases the relative  $L^2(I, H^1(D))$ -error is below  $2 \cdot 10^{-2}$ . Using the same computational budget we thus achieve that the algorithm succeeds to detect all different configurations of the time-dependent data functions in 97% instead of 88% of cases.

**6.4. Choice of parameters in Algorithm 1.** Based on the experiments above we give the following guidance on how to choose the parameters  $n_t$ ,  $k$ , and  $n_{\text{rand}}$ : An appropriate choice for the local oversampling size  $n_t$  seems to be 10 – 15, while for  $n_t = 10$  the resulting reduced basis is larger compared to  $n_t = 15$  and the user can balance computational costs depending on the respective application. In case of coefficients implying a slow spreading or propagation of the solution, a larger  $n_t$ , for instance,  $n_t = 30$ , might be more favorable. In addition, we propose  $k = n_t - 2$  as a reasonable choice for diffusion problems, while for advection-dominated problems a smaller  $k$ , such

as  $k = n_t - 6$ , is more favorable. The choice of drawn time points  $n_{\text{rand}}$  can be based on the number of available parallel compute units. Moreover, in certain cases it is necessary to also sample from data functions that are constant in time (e.g. a constant advection field) to achieve a good approximation accuracy.

## 7. CONCLUSIONS

To tackle time-dependent problems with heterogeneous time-dependent coefficients, we have proposed a randomized algorithm that constructs a reduced approximation space in time by solving several local problems in time in parallel. Based on techniques from randomized NLA [12, 16, 23] points in time are drawn from a data-driven probability distribution and the PDE is solved locally in time using these points as end points with random initial conditions. The approach allows for local error control [6] and the computation of the local basis functions is embarrassingly parallel.

The numerical experiments demonstrate that the proposed algorithm can outperform the POD even in the sequential setting for complex problems with heterogeneous time-dependent data functions and is also well capable of tackling higher values of advection. Moreover, we have observed that leverage scores are capable of detecting multiscale features in the data functions.

### APPENDIX A. COMPACTNESS OF THE TRANSFER OPERATOR IN TIME FOR THE ADVECTION-DIFFUSION-REACTION PROBLEM

In this section we prove a Caccioppoli inequality and compactness of the transfer operator in time (cf. (4.3)) for the advection-diffusion-reaction problem (2.2) introduced in section 2. The Caccioppoli inequality in Proposition A.1 is closely linked to the exponential decay behavior of solutions of the PDE in time (cf. the discussion in section 3) and allows to bound the  $L^2(D)$ -norm of solutions evaluated at a point of time in terms of their  $L^2(I, L^2(D))$ -norm. The second key ingredient that we use to prove compactness of the transfer operator in time in Proposition A.2 is the compactness theorem of Aubin-Lions [55, Corollary 5], which states that the embedding  $\{v \in L^2(I, H_0^1(D)) \mid v_t \in L^2(I, H^{-1}(D))\} \hookrightarrow L^2(I, L^2(D))$  is compact. The combination of a Caccioppoli-type inequality with a suitable compactness theorem is usually used to show compactness of the transfer operator; see also [2, 32, 49, 57, 58].

We may consider the following weak formulation (for a proof of well-posedness see, e.g., [63, 53]): Find the solution  $u \in W^{1,2,2}(I, H_0^1(D), H^{-1}(D)) := \{v \in L^2(I, H_0^1(D)) \mid v_t \in L^2(I, H^{-1}(D))\}$  such that  $u(0) = u_0$  in  $L^2(D)$  and

$$\begin{aligned} & \int_I \langle u_t(t), \psi(t) \rangle_{H_0^1(D)} dt + \int_I (\kappa(t) \nabla u(t), \nabla \psi(t))_{L^2(D)} dt + \int_I (b(t) \cdot \nabla u(t), \psi(t))_{L^2(D)} dt \\ & + \int_I (c(t)u(t), \psi(t))_{L^2(D)} dt = \int_I \langle f(t), \psi(t) \rangle_{H_0^1(D)} dt \quad \forall \psi \in L^2(I, H_0^1(D)). \end{aligned}$$

The data functions are introduced in section 2. In particular, the source terms and initial conditions are given by  $f \in L^2(I, H^{-1}(D))$  and  $u_0 \in L^2(D)$ . To simplify notations, we assume homogeneous Dirichlet boundary conditions on  $I \times \partial D$ . However, the theory analogously applies to non-homogeneous Dirichlet boundary conditions.

We highlight that the embedding  $W^{1,2,2}(I, H_0^1(D), H^{-1}(D)) \hookrightarrow C^0(\bar{I}, L^2(D))$  is not compact unless additional regularity is assumed. In fact, the space

$$W^{1,\infty,r}(I, H_0^1(D), H^{-1}(D)) := \{v \in L^\infty(I, H_0^1(D)) \mid v_t \in L^r(I, H^{-1}(D))\}$$

embeds compactly in  $C^0(\bar{I}, L^2(D))$  for  $r > 1$  [55, Corollary 5]. In contrast, we prove compactness in [Appendix A](#) requiring significantly less regularity. First, the compactness theorem of Aubin-Lions [55, Corollary 5] states that the embedding  $W^{1,2,2}(I, H_0^1(D), H^{-1}(D)) \hookrightarrow L^2(I, L^2(D))$  is compact. Next, the Caccioppoli inequality that we prove in [Proposition A.1](#) is the key ingredient that facilitates the restriction to a point of time as the inequality bounds the  $L^2(D)$ -norm of (local) solutions evaluated at a point of time in terms of their  $L^2(I, L^2(D))$ -norm. Consequently, we show in [Proposition A.2](#) that the space of (local) solutions contained in the generalized Sobolev space  $W^{1,2,2}(I, H_0^1(D), H^{-1}(D))$  embeds compactly in  $C^0(\bar{I}, L^2(D))$ .

In the following, we denote by  $t^* \in I$  the local end time point to simplify notations in the proofs. We then consider the local time interval  $(s, t^*) \subseteq I$  and seek local solutions  $u_{\text{loc}} \in W^{1,2,2}((s, t^*), H_0^1(D), H^{-1}(D))$  with initial conditions  $u_{\text{loc}}(s, \cdot) \in L^2(D)$  such that for all  $\psi \in L^2((s, t^*), H_0^1(D))$

$$(A.1) \quad \int_s^{t^*} \langle (u_{\text{loc}})_t(t), \psi(t) \rangle_{H_0^1(D)} dt + \int_s^{t^*} (\kappa(t) \nabla u_{\text{loc}}(t), \nabla \psi(t))_{L^2(D)} dt \\ + \int_s^{t^*} (b(t) \cdot \nabla u_{\text{loc}}(t), \psi(t))_{L^2(D)} dt + \int_s^{t^*} (c(t) u_{\text{loc}}(t), \psi(t))_{L^2(D)} dt = \int_s^{t^*} \langle f(t), \psi(t) \rangle_{H_0^1(D)} dt.$$

**Proposition A.1** (Caccioppoli inequality in time). *Let  $w$  satisfy (A.1) with  $f \equiv 0$  and arbitrary initial conditions  $w(s, \cdot) \in L^2(D)$ . Then, we have that*

$$(A.2) \quad \|w(t^*, \cdot)\|_{L^2(D)}^2 \leq \frac{2}{(t^* - s)} \|w\|_{L^2((s, t^*), L^2(D))}^2.$$

*Proof.* The first paragraph closely follows the proof of Proposition 3.1 in [49]. Since  $w$  satisfies (A.1), we can choose  $\psi = v\varphi$  for arbitrary  $v \in H_0^1(D)$  and  $\varphi \in C_0^\infty((s, t^*))$  as a test function in (A.1). As  $\varphi \in C_0^\infty((s, t^*))$  is chosen arbitrarily, the fundamental lemma of calculus of variations yields that  $\langle w_t(t), v \rangle_{H_0^1(D)} + (\kappa(t) \nabla w(t), \nabla v)_{L^2(D)} + (b(t) \cdot \nabla w(t), v)_{L^2(D)} + (c(t)w(t), v)_{L^2(D)} = 0$  for all  $v \in H_0^1(D)$  and almost every  $t \in (s, t^*)$ . Next, we introduce a cut-off function  $\eta \in C^1((s, t^*))$  that satisfies  $0 \leq \eta \leq 1$ ,  $\eta(s) = 0$ ,  $\eta(t^*) = 1$ , and  $|\eta_t| \leq \frac{1}{(t^* - s)}$ . In the following, we want to use  $w\eta^2$  as a test function. To enable rearranging the part of the weak formulation that includes the time derivative, we approximate  $w$  by a sequence  $w_n \in C_0^\infty((s, t^*), H_0^1(D))$  such that  $w_n$  converges strongly to  $w$  in  $L^2((s, t^*), H_0^1(D))$ . Then, for almost every  $t \in (s, t^*)$  and each  $n \in \mathbb{N}$  we have that

$$\langle w_t(t), w_n(t)\eta^2(t) \rangle_{H_0^1(D)} + (\kappa(t) \nabla w(t), \nabla w_n(t)\eta^2(t))_{L^2(D)} \\ + (b(t) \cdot \nabla w(t), w_n(t)\eta^2(t))_{L^2(D)} + (c(t)w(t), w_n(t)\eta^2(t))_{L^2(D)} = 0.$$

Integrating over the time interval  $(s, t^*)$  yields

$$(A.3) \quad \int_s^{t^*} \langle w_t(t), w_n(t)\eta^2(t) \rangle_{H_0^1(D)} dt + \int_s^{t^*} (\kappa(t) \nabla w(t), \nabla w_n(t)\eta^2(t))_{L^2(D)} dt \\ + \int_s^{t^*} (b(t) \cdot \nabla w(t), w_n(t)\eta^2(t))_{L^2(D)} dt + \int_s^{t^*} (c(t)w(t), w_n(t)\eta^2(t))_{L^2(D)} dt = 0.$$

Using integration by parts and  $\eta(s) = 0$ , we rewrite the first term in (A.3) as follows:

$$\begin{aligned}
& \int_s^{t^*} \langle w_t(t), w_n(t) \eta^2(t) \rangle_{H_0^1(D)} dt \\
&= - \int_s^{t^*} (w(t), (w_n(t))_t \eta^2(t))_{L^2(D)} dt - \int_s^{t^*} (w(t), w_n(t) 2\eta(t) \eta_t(t))_{L^2(D)} dt \\
&\quad + (w(t^*), w_n(t^*) \eta^2(t^*))_{L^2(D)} \\
&= - \int_s^{t^*} (w(t) \eta(t), (w_n(t))_t \eta(t))_{L^2(D)} dt - 2 \int_s^{t^*} (w(t) \eta(t), w_n(t) \eta_t(t))_{L^2(D)} dt \\
&\quad + (w(t^*) \eta(t^*), w_n(t^*) \eta(t^*))_{L^2(D)} \\
&= \int_s^{t^*} \langle (w(t) \eta(t))_t, w_n(t) \eta(t) \rangle_{H_0^1(D)} dt - \int_s^{t^*} (w(t) \eta(t), w_n(t) \eta_t(t))_{L^2(D)} dt.
\end{aligned}$$

Letting  $n$  go to  $\infty$  thus yields

$$\begin{aligned}
& \int_s^{t^*} \langle (w(t) \eta(t))_t, w(t) \eta(t) \rangle dt - \int_s^{t^*} (w(t) \eta(t), w(t) \eta_t(t))_{L^2(D)} dt \\
\text{(A.4)} \quad & + \int_s^{t^*} (\kappa(t) \nabla w(t), \nabla w(t) \eta^2(t))_{L^2(D)} dt + \int_s^{t^*} (b(t) \cdot \nabla w(t), w(t) \eta^2(t))_{L^2(D)} dt \\
& + \int_s^{t^*} (c(t) w(t), w(t) \eta^2(t))_{L^2(D)} dt = 0.
\end{aligned}$$

Employing the assumption  $c - \frac{1}{2} \nabla \cdot b \geq 0$  and Gauss's theorem we then conclude that

$$\begin{aligned}
& \int_s^{t^*} (b(t) \cdot \nabla w(t), w(t) \eta^2(t))_{L^2(D)} dt + \int_s^{t^*} (c(t) w(t), w(t) \eta^2(t))_{L^2(D)} dt \\
&= \frac{1}{2} \int_s^{t^*} \int_D \nabla \cdot (b(w\eta)^2) dx dt + \int_s^{t^*} \int_D (c - \frac{1}{2} \nabla \cdot b)(w\eta)^2 dx dt \\
&\geq \frac{1}{2} \int_s^{t^*} \int_{\partial D} b(w\eta)^2 n dx dt = 0.
\end{aligned}$$

Finally, we exploit that  $\int_s^{t^*} (\kappa(t) \nabla w(t), \nabla w(t) \eta^2(t))_{L^2(D)} dt \geq 0$ , the properties of the cut-off function  $\eta$  and the identity  $\int_s^{t^*} \langle (w(t) \eta(t))_t, w(t) \eta(t) \rangle dt = \frac{1}{2} \|w(t^*) \eta(t^*)\|_{L^2(D)}^2$  to infer that

$$\|w(t^*, \cdot)\|_{L^2(D)}^2 \leq \frac{2}{(t^* - s)} \|w\|_{L^2((s, t^*), L^2(D))}^2.$$

□

**Proposition A.2.** *The transfer operator  $\mathcal{F}_{s \rightarrow t^*}$  in (4.3) is compact.*

*Proof.* Let  $(\xi_n)_{n \in \mathbb{N}}$  be a bounded sequence in  $L^2(D)$ . We denote by  $(w_n)_{n \in \mathbb{N}} \subset W^{1,2,2}((s, t^*), H_0^1(D), H^{-1}(D))$  the corresponding sequence of solutions of problem (A.1) with initial conditions  $w_n(s) = \xi_n$  in  $L^2(D)$  and right-hand side  $f \equiv 0$ , obtaining  $\|w_n\|_{L^2((s, t^*), H_0^1(D))} \leq C$  for a constant  $0 < C < \infty$ . Then, there exists a subsequence  $(w_{n_l})_{l \in \mathbb{N}}$  and a limit function  $w \in L^2((s, t^*), H_0^1(D))$  such that the subsequence converges weakly to  $w$  in  $L^2((s, t^*), H_0^1(D))$ . Thanks to this weak convergence and the Hahn-Banach theorem we infer that  $w \in W^{1,2,2}((s, t^*), H_0^1(D), H^{-1}(D))$

and that  $w$  solves the local PDE (A.1) for  $f \equiv 0$  (cf. proofs of Lemma 3.2 and Theorem A.1 in [49]) and initial conditions  $w(s, \cdot) \in L^2(D)$  (due to the embedding  $W^{1,2,2}((s, t^*), H_0^1(D), H^{-1}(D)) \hookrightarrow C^0([s, t^*], L^2(D))$ ). Here, we moreover employ that the sequence  $((w_{n_l})_{l \in \mathbb{N}})$  converges weakly-\* to  $w_t$  in  $L^2((s, t^*), H^{-1}(D))$  (cf. proof of Theorem A.1 in [49]). As there also holds that  $(w_{n_l})_{l \in \mathbb{N}} \subseteq W^{1,2,2}((s, t^*), H_0^1(D), H^{-1}(D))$ , the compactness theorem of Aubin-Lions [55, Corollary 5] yields a subsequence  $(w_{n_{l_m}})_{m \in \mathbb{N}}$  which converges strongly to  $w$  in  $L^2((s, t^*), L^2(D))$ . Since the sequence  $e_{n_{l_m}} := w - w_{n_{l_m}}$  thus solves (A.1) with  $f \equiv 0$  and some initial conditions in  $L^2(D)$ , we may invoke Proposition A.1 to infer that

$$\|w(t^*, \cdot) - w_{n_{l_m}}(t^*, \cdot)\|_{L^2(D)}^2 = \|e_{n_{l_m}}(t^*, \cdot)\|_{L^2(D)}^2 \leq \frac{2}{(t^* - s)} \|e_{n_{l_m}}\|_{L^2((s, t^*), L^2(D))}^2 \longrightarrow 0.$$

□

## APPENDIX B. ERROR BOUNDS FOR SQUARED NORM AND LEVERAGE SCORE SAMPLING

In this subsection, we state and briefly discuss error bounds for the squared norm and leverage score sampling approach introduced in subsection 5.3.

The following theorem from [20] gives an additive error bound on the approximation quality of the squared norm sampling approach.

**Theorem B.1.** ([20, Theorem 2]). *Let  $\mathbf{C} \in \mathbb{R}^{N_D \times m}$  be a sample of  $m$  columns of  $\mathbf{B}$  using the squared norm probability distribution and let  $\mathbf{C}^\dagger$  denote the Moore-Penrose inverse of  $\mathbf{C}$ . Then, with probability at least 0.9 it holds that*

$$(B.1) \quad \|\mathbf{B} - \mathbf{C}\mathbf{C}^\dagger\mathbf{B}\|_F^2 \leq \|\mathbf{B} - \mathbf{B}_r\|_F^2 + \frac{10r}{m} \|\mathbf{B}\|_F^2.$$

Here,  $\mathbf{B}_r$  denotes the best rank- $r$  approximation to  $\mathbf{B}$ .

The subsequent theorem from [17] gives a multiplicative error bound on the approximation quality of the leverage score sampling approach.

**Theorem B.2.** ([17, Theorem 3]). *Let  $\varepsilon \in (0, 1]$  and let  $\mathbf{C} \in \mathbb{R}^{N_D \times m}$  be a sample of  $m = 3200r^2/\varepsilon^2$  columns of  $\mathbf{B}$  using the leverage score probability distribution. Then, with probability at least 0.7 it holds that*

$$(B.2) \quad \|\mathbf{B} - \mathbf{C}\mathbf{C}^\dagger\mathbf{B}\|_F \leq (1 + \varepsilon) \|\mathbf{B} - \mathbf{B}_r\|_F.$$

Here,  $\mathbf{C}^\dagger$  denotes the Moore-Penrose inverse of  $\mathbf{C}$  and  $\mathbf{B}_r$  denotes the best rank- $r$  approximation to  $\mathbf{B}$ .

The observations and discussion in subsection 5.3.1 are in line with the error bounds stated in Theorems B.1 and B.2: The rank- $r$  best approximation error in (B.1) and (B.2) is given by  $\|\mathbf{B} - \mathbf{B}_r\|_F = (\sum_{i=r+1}^R \sigma_i^2)^{1/2}$  (Eckart-Young theorem e.g. in [21]), where  $\sigma_1 \geq \sigma_2 \geq \dots$  denote the singular values of  $\mathbf{B}$  and  $R \leq \min\{N_D, N_I\}$  denotes its rank. Moreover, the Frobenius-norm in the last term of (B.1) is determined by  $\|\mathbf{B}\|_F = (\sum_{i=1}^R \sigma_i^2)^{1/2}$ . Consequently, error bound (B.1) indicates that in the squared norm sampling approach the number  $m$  of selected columns needs to be large in order to detect structures that correspond to singular values  $\sigma_i$  for which  $\sigma_i/\sigma_j$  with  $i \in \{r+1, \dots, R\}$  and  $j \in \{1, \dots, r\}$  is small. As  $m$  appears in the denominator in the last term of (B.1), it has to be chosen as  $m \approx 10r(\sum_{i=1}^R \sigma_i^2)/(\sum_{i=r+1}^R \sigma_i^2)$  to achieve

an error that is of the order of the best rank- $r$  approximation to  $B$ . In contrast, the bound for the leverage score sampling approach in (B.2) is of the order  $(\sum_{i=r+1}^R \sigma_i^2)^{1/2}$  for a number of selected columns that is independent of the size of the singular values of  $\mathbf{B}$ .

#### APPENDIX C. DATA FUNCTIONS AND PARAMETERS FOR EXAMPLE 1

To ensure reproducibility, we list here data functions and discretization parameters corresponding to Example 1 in subsection 5.3.1. We choose  $I = (0, 10)$ ,  $D = (0, 1)^2$ ,  $\Sigma_N = \emptyset$ , and discretize the spatial domain  $D$  with a regular quadrilateral mesh with mesh size  $1/50$  in both directions. For the implicit Euler method, we use an equidistant time step size of  $1/30$ . Furthermore, the two source terms are given by  $f_i(t, x, y) = \sum_{j=1}^2 f_{i,j}(t) f_j(x, y)$  with  $f_{1,1}(t) = 4 \mathbb{1}_{(t \geq 1)} \mathbb{1}_{(t \leq 4)}$ ,  $f_{1,2}(t) = \mathbb{1}_{(t \geq 6)} \mathbb{1}_{(t \leq 9)}$ ,  $f_{2,1}(t) = \mathbb{1}_{(t \geq 1)} \mathbb{1}_{(t \leq 7)}$ ,  $f_{2,2}(t) = \mathbb{1}_{(t \geq 9)} \mathbb{1}_{(t \leq 9.2)}$ ,  $f_1(x, y) = \mathbb{1}_{(x \geq 0.2)} \mathbb{1}_{(x \leq 0.3)} \mathbb{1}_{(y \geq 0.2)} \mathbb{1}_{(y \leq 0.3)}$ , and  $f_2(x, y) = \mathbb{1}_{(x \geq 0.7)} \mathbb{1}_{(x \leq 0.8)} \mathbb{1}_{(y \geq 0.7)} \mathbb{1}_{(y \leq 0.8)}$ . For computing the probability distributions, we choose  $\mathbf{B} = \mathbf{F}$ , i.e. the corresponding right-hand side matrix (2.3).

#### ACKNOWLEDGMENTS

The authors thank Dr. Alexander Heinlein for providing us with the data file of the permeability field  $\kappa_0$  used in Example 4. Moreover, we thank Dr. Christian Himpe for discussions regarding system and control theory.

#### REFERENCES

- [1] A. ALAOUI AND M. W. MAHONEY, *Fast randomized kernel ridge regression with statistical guarantees*, Adv. Neural Inf. Process. Syst., 28 (2015).
- [2] I. BABUŠKA AND R. LIPTON, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul., 9 (2011), pp. 373–406.
- [3] G. BERKOOZ, P. HOLMES, AND J. L. LUMLEY, *The proper orthogonal decomposition in the analysis of turbulent flows*, Annu. Rev. Fluid Mech., 25 (1993), pp. 539–575.
- [4] A. BONITO, A. COHEN, R. DEVORE, G. PETROVA, AND G. WELPER, *Diffusion coefficients estimation for elliptic partial differential equations*, SIAM J. Math. Anal., 49 (2017), pp. 1570–1592.
- [5] A. BUHR, L. IAPICHINO, M. OHLBERGER, S. RAVE, F. SCHINDLER, AND K. SMETANA., *Localized model reduction for parameterized problems*, in P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W.H.A. Schilders, L.M. Sileira (eds.). Model Order Reduction, Volume 2, Snapshot-Based Methods and Algorithms., Walter De Gruyter GmbH, Berlin, 2020.
- [6] A. BUHR AND K. SMETANA, *Randomized Local Model Order Reduction*, SIAM J. Sci. Comput., 40 (2018), pp. A2120–A2151.
- [7] Y. CHAHLAOUI AND P. V. DOOREN, *Model reduction of time-varying systems*, in Dimension reduction of large-scale systems, Springer, 2005, pp. 131–148.
- [8] K. CHEN, Q. LI, J. LU, AND S. J. WRIGHT, *Randomized Sampling for Basis Function Construction in Generalized Finite Element Methods*, Multiscale Model. Simul., 18 (2020), pp. 1153–1177.
- [9] M. A. CHRISTIE AND M. J. BLUNT, *Tenth SPE comparative solution project: A comparison of upscaling techniques*, SPE Reservoir Evaluation & Engineering, 4 (2001), pp. 308–317.
- [10] E. T. CHUNG, Y. EFENDIEV, W. T. LEUNG, AND S. YE, *Generalized multiscale finite element methods for space-time heterogeneous parabolic equations*, Comput. Math. Appl., 76 (2018), pp. 419–437.
- [11] P. R. CONRAD, M. GIROLAMI, S. SÄRKKÄ, A. STUART, AND K. ZYGALAKIS, *Statistical analysis of differential equations: introducing probability measures on numerical solutions*, Stat. Comput., 27 (2017), pp. 1065–1082.
- [12] M. DEREZIŃSKI AND M. W. MAHONEY, *Determinantal point processes in randomized numerical linear algebra*, Notices Amer. Math. Soc., 68 (2021), pp. 34–45.

- [13] A. DESHPANDE AND L. RADEMACHER, *Efficient volume sampling for row/column subset selection*, in 2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010, IEEE Computer Soc., Los Alamitos, CA, 2010, pp. 329–338.
- [14] A. DESHPANDE, L. RADEMACHER, S. VEMPALA, AND G. WANG, *Matrix approximation and projective clustering via volume sampling*, *Theory Comput.*, 2 (2006), pp. 225–247.
- [15] P. DRINEAS, M. MAGDON-ISMAIL, M. W. MAHONEY, AND D. P. WOODRUFF, *Fast approximation of matrix coherence and statistical leverage*, *J. Mach. Learn. Res.*, 13 (2012), pp. 3475–3506.
- [16] P. DRINEAS AND M. W. MAHONEY, *RandNLA: Randomized Numerical Linear Algebra*, *Commun. ACM*, 59 (2016), pp. 80–90.
- [17] P. DRINEAS, M. W. MAHONEY, AND S. MUTHUKRISHNAN, *Relative-error CUR matrix decompositions*, *SIAM J. Matrix Anal. Appl.*, 30 (2008), pp. 844–881.
- [18] J. L. EFTANG AND A. T. PATERA, *Port reduction in parametrized component static condensation: approximation and a posteriori error estimation*, *Int. J. Numer. Methods Eng.*, 96 (2013), pp. 269–302.
- [19] A. ERN AND J.-L. GUERMOND, *Theory and Practice of Finite Elements*, vol. 159 of Applied Mathematical Sciences, Springer New York, New York, NY, 2004.
- [20] A. FRIEZE, R. KANNAN, AND S. VEMPALA, *Fast Monte-Carlo algorithms for finding low-rank approximations*, *J. ACM*, 51 (2004), pp. 1025–1041.
- [21] G. H. GOLUB AND C. F. VAN LOAN, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, fourth ed., 2013.
- [22] L. GRASEDYCK, I. GREFF, AND S. SAUTER, *The AL basis for the solution of elliptic problems in heterogeneous media*, *Multiscale Model. Simul.*, 10 (2012), pp. 245–258.
- [23] N. HALKO, P. G. MARTINSSON, AND J. A. TROPP, *Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions*, *SIAM Rev.*, 53 (2011), pp. 217–288.
- [24] D. B. P. HUYNH, D. J. KNEZEVIC, AND A. T. PATERA, *A static condensation reduced basis element method: Approximation and a posteriori error estimation*, *ESAIM. Math. Model. Numer. Anal.*, 47 (2013), pp. 213–251.
- [25] L. IAPICHINO, A. QUARTERONI, AND G. ROZZA, *A reduced basis hybrid method for the coupling of parametrized domains represented by fluidic networks*, *Comput. Methods Appl. Mech. Eng.*, 221/222 (2012), pp. 63–82.
- [26] A. KOLMOGOROFF, *Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse*, *Ann. of Math. (2)*, 37 (1936), pp. 107–110.
- [27] K. KUNISCH AND S. VOLKWEIN, *Galerkin proper orthogonal decomposition methods for parabolic problems*, *Numer. Math.*, 90 (2001), pp. 117–148.
- [28] S. LALL AND C. BECK, *Error-bounds for balanced model-reduction of linear time-varying systems*, *IEEE Trans. Automat. Control*, 48 (2003), pp. 946–956.
- [29] R. B. LEHOUCQ, D. C. SORENSEN, AND C. YANG, *ARPACK Users’ Guide*, (1998).
- [30] J.-L. LIONS, Y. MADAY, AND G. TURINICI, *Résolution d’EDP par un schéma en temps “pararéel”*, *C. R. Acad. Sci. Paris Sér. I Math.*, 332 (2001), pp. 661–668.
- [31] P. LJUNG, R. MAIER, AND A. MÅLQVIST, *A space-time multiscale method for parabolic problems*, 2022.
- [32] C. MA, R. SCHEICHL, AND T. DODWELL, *Novel design and analysis of generalized finite element methods based on locally optimal spectral approximations*, *SIAM J. Numer. Anal.*, 60 (2022), pp. 244–273.
- [33] Y. MADAY AND E. M. RØNQVIST, *A reduced-basis element method*, *J. Sci. Comput.*, 17 (2002), pp. 447–459.
- [34] M. W. MAHONEY AND P. DRINEAS, *CUR matrix decompositions for improved data analysis*, *Proc. Natl. Acad. Sci. U.S.A.*, 106 (2009), pp. 697–702.
- [35] A. MÅLQVIST AND A. PERSSON, *Multiscale techniques for parabolic equations*, *Numer. Math.*, 138 (2018), pp. 191–217.
- [36] A. MÅLQVIST AND D. PETERSEIM, *Localization of elliptic multiscale problems*, *Math. Comp.*, 83 (2014), pp. 2583–2603.
- [37] P.-G. MARTINSSON, V. ROKHLIN, AND M. TYGERT, *A randomized algorithm for the decomposition of matrices*, *Appl. Comput. Harmon. Anal.*, 30 (2011), pp. 47 – 68.

- [38] B. MOORE, *Principal component analysis in linear systems: Controllability, observability, and model reduction*, IEEE Trans. Automat. Contr., 26 (1981), pp. 17–32.
- [39] H. OWHADI, *Bayesian numerical homogenization*, Multiscale Model. Simul., 13 (2015), pp. 812–828.
- [40] H. OWHADI AND L. ZHANG, *Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast*, Multiscale Model. Simul., 9 (2011), pp. 1373–1398.
- [41] H. OWHADI AND L. ZHANG, *Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients*, J. Comput. Phys., 347 (2017), pp. 99–128.
- [42] H. OWHADI, L. ZHANG, AND L. BERLYAND, *Polyharmonic homogenization, rough polyharmonic splines and sparse super-localization*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 517–552.
- [43] A. PINKUS, *n-widths in approximation theory*, vol. 7, Springer-Verlag, Berlin, 1985.
- [44] M. RENARDY AND R. C. ROGERS, *An introduction to partial differential equations*, vol. 13, Springer Science & Business Media, 2004.
- [45] C. W. ROWLEY, *Model reduction for fluids, using balanced proper orthogonal decomposition*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 15 (2005), pp. 997–1013.
- [46] A. K. SAIBABA, *Randomized discrete empirical interpolation method for nonlinear model reduction*, SIAM J. Sci. Comput., 42 (2020), pp. A1582–A1608.
- [47] H. SANDBERG AND A. RANTZER, *Balanced truncation of linear time-varying systems*, IEEE Trans. Automat. Control, 49 (2004), pp. 217–229.
- [48] J. SCHLEUSS, *Source code to “Randomized quasi-optimal local approximation spaces in time”*, 2022, <https://doi.org/10.5281/zenodo.6287484>.
- [49] J. SCHLEUSS AND K. SMETANA, *Optimal local approximation spaces for parabolic problems*, Multiscale Model. Simul., 20 (2022), pp. 551–582.
- [50] P. J. SCHMID, *Dynamic mode decomposition of numerical and experimental data*, J. Fluid Mech., 656 (2010), pp. 5–28.
- [51] M. SCHOBER, D. K. DUVENAUD, AND P. HENNIG, *Probabilistic ODE solvers with Runge-Kutta means*, Adv. Neural Inf. Process. Syst., 27 (2014).
- [52] M. SCHOBER, S. SÄRKKÄ, AND P. HENNIG, *A probabilistic model for the numerical solution of initial value problems*, Stat. Comput., 29 (2019), pp. 99–122.
- [53] C. SCHWAB AND R. STEVENSON, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comput., 78 (2009), pp. 1293–1318.
- [54] S. SHOKOHI, L. M. SILVERMAN, AND P. M. VAN DOOREN, *Linear time-variable systems: balancing and model reduction*, IEEE Trans. Automat. Control, 28 (1983), pp. 810–822.
- [55] J. SIMON, *Compact sets in the space  $L^p(0, T; B)$* , Ann. Mat. Pura Appl. (4), 146 (1987), pp. 65–96.
- [56] L. SIROVICH, *Turbulence and the dynamics of coherent structures. I. Coherent structures*, Quart. Appl. Math., 45 (1987), pp. 561–571.
- [57] K. SMETANA AND A. T. PATERA, *Optimal local approximation spaces for component-based static condensation procedures*, SIAM J. Sci. Comput., 38 (2016), pp. A3318–A3356.
- [58] T. TADDEI AND A. T. PATERA, *A localization strategy for data assimilation; application to state estimation and parameter estimation*, SIAM J. Sci. Comput., 40 (2018), pp. B611–B636.
- [59] L. TER MAAT, *Random initial conditions in model order reduction*, bachelor thesis, University of Twente, 2019.
- [60] J. H. TU, C. W. ROWLEY, D. M. LUCHTENBURG, S. L. BRUNTON, AND J. N. KUTZ, *On dynamic mode decomposition: Theory and applications*, J. Comput. Dyn., 1 (2014), pp. 391–421.
- [61] B. UNGER AND S. GUGERCIN, *Kolmogorov n-widths for linear dynamical systems*, Adv. Comput. Math., 45 (2019), pp. 2273–2286.
- [62] E. I. VERRIEST AND T. KAILATH, *On generalized balanced realizations*, IEEE Trans. Automat. Control, 28 (1983), pp. 833–844.
- [63] J. WLOKA, *Partielle Differentialgleichungen*, B. G. Teubner, Stuttgart, 1982.
- [64] D. P. WOODRUFF, *Sketching as a tool for numerical linear algebra*, Found. Trends Theor. Comput. Sci., 10 (2014), pp. iv+157.

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, UNIVERSITY OF MÜNSTER, EINSTEINSTR. 62,  
48149 MÜNSTER, GERMANY, JULIA.SCHLEUSS@UNI-MUENSTER.DE.

DEPARTMENT OF MATHEMATICAL SCIENCES, STEVENS INSTITUTE OF TECHNOLOGY, 1 CASTLE POINT  
TERRACE, HOBOKEN, NJ 07030, UNITED STATES OF AMERICA, KSMETANA@STEVENS.EDU.

CURRENTLY MASTER STUDENT AT THE UNIVERSITY OF TWENTE, THE NETHERLANDS, PRIVATE AD-  
DRESS: KIEFTSBEEKLAAN 33, 7607TA ALMELO, THE NETHERLANDS, LUKAS.TM@HOTMAIL.NL.