An adaptive model hierarchy for data-augmented training of kernel models for reactive flow ⋆

B. Haasdonk * **M. Ohlberger** ** **F. Schindler** **

* *Institute of Applied Analysis and Numerical Simulation, Pfaffenwaldring 57, D-70569 Stuttgart (e-mail:* haasdonk@mathematik.uni-stuttgart.de*).*
** *Mathematics Münster, Westfälische Wilhelms-Universität Münster, Einsteinstr. 62, D-48149 Münster (e-mail:* {mario.ohlberger,felix.schindler}@uni-muenster.de*)*

1. REFERENCE MODEL

We are interested in constructing efficient and accurate models to approximate time-dependent quantities of interest (QoI) $f \in L^2\big(\mathcal{P}; L^2([0,T])\big)$ in the context of reactive flow, with $T > 0$ and where $\mathcal{P} \subset \mathbb{R}^p$ for $p \geq 1$ denotes the set of possible input parameters. As a class of QoI functions, we consider those obtained by applying linear functionals $s_\mu \in V'$ to solution trajectories $c_\mu \in L^2(0,T;V)$ of, e.g., parametric parabolic partial differential equations. Thus, $f(\mu;t) := s_\mu(c_\mu(t))$, where for each parameter $\mu \in \mathcal{P}$, the concentration c_μ with $\partial_t c_\mu \in L^2(0,T;V')$ and initial condition $c_0 \in V$ is the unique weak solution of

$$\langle \partial_t c_\mu, v\rangle + a_\mu(c_\mu, v) = l_\mu(v) \quad \forall\, v \in V,\ c_\mu(0) = c_0. \quad (1)$$

Here, $V \subset H^1(\Omega) \subset L^2(\Omega) \subset V'$ denotes a Gelfand triple of Hilbert-spaces associated with a spatial Lipschitz-domain Ω and, for $\mu \in \mathcal{P}$, $l_\mu \in V'$ denotes a continuous linear functional and $a_\mu : V \times V \to \mathbb{R}$ a continuous coercive bilinear form.

As a basic model for reactive flow in catalytic filters, (1) could stem from a single-phase one-dimensional linear advection-diffusion-reaction problem with Dammköhler- and Péclet-numbers as input (thus $p = 2$), where c models the dimensionless molar concentration of a species and the break-through curve s measures the concentration at the outflow, as detailed in Gavrilenko et al. (2022).

Since direct evaluations of f are not available, we resort to a full order model (FOM) as reference model, yielding

$$f_h : \mathcal{P} \to \mathbb{R}^{N_T} \text{ for } N_T \geq 1, \quad f_h(\mu;t) := s_\mu(c_{h,\mu}(t)), \quad (2)$$

which we assume to be a sufficiently accurate approximation of the QoI. For simplicity, we consider a P^1-conforming Finite Element space $V_h \subset V$ and obtain the FOM solution trajectory $c_{h,\mu} \in L^2(0,T;V_h)$ by Galerkin projection of (1) onto V_h and an implicit Euler approximation of the temporal derivative.

⋆ Funded by BMBF under contracts 05M20PMA and 05M20VSA. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under contracts OH 98/11-1 and SCHI 1493/1-1, as well as under Germany's Excellence Strategy EXC 2044 390685587, Mathematics Münster: Dynamics – Geometry – Structure, and EXC 2075 390740016, Stuttgart Center for Simulation Science (SimTech).

2. SURROGATE MODELS

The evaluation of (2) may be arbitrarily costly, in particular in multi- or large-scale scenarios where $\dim V_h \gg 1$, but also if $N_T \gg 1$ due to long-time integration or when a high resolution of f_h is required. We thus seek to build a machine learning (ML) based surrogate model

$$f_{\mathrm{ml}} : \mathcal{P} \to \mathbb{R}^{N_T},\ f_{\mathrm{ml}}(\mu;t_n) \approx f_h(\mu;t_n)\ \forall 1 \leq n \leq N_T, \quad (3)$$

to predict all values $\{f_{\mathrm{ml}}(\mu;t_n)\}_{n=1}^{N_T}$ at once, without time-integration. Such models based on Neural Networks or Kernels typically rely on a large amount of training data

$$\big\{\big(\mu, f_h(\mu)\big) \,\big|\, \mu \in \mathcal{P}_{\mathrm{train}}\big\}, \quad \mathcal{P}_{\mathrm{train}} \subset \mathcal{P}, \quad |\mathcal{P}_{\mathrm{train}}| \gg 1, \quad (4)$$

rendering their training prohibitively expensive in the aforementioned scenarios; we refer to Gavrilenko et al. (2022) and the references therein and in particular to Santin and Haasdonk (2021). In Gavrilenko et al. (2022) we thus seek to employ an intermediate surrogate to generate sufficient training data.

2.1 Structure preserving Reduced Basis models

The idea of projection-based model order reduction by Reduced Basis (RB) methods is to approximate the state c_h in a low-dimensional subspace $V_{\mathrm{rb}} \subset V_h$ and to obtain online-efficient approximations of f_h by Galerkin projection of the FOM detailed in Section 1 onto V_{rb} and a precomputation of all quantities involving V_h in a possibly expensive offline-computation; we refer to Milk et al. (2016) and the references therein. Using such structure preserving reduced order models (ROM)s we obtain RB trajectories $c_{\mathrm{rb},\mu} \in L^2(0,T;V_{\mathrm{rb}})$ and a RB model

$$f_{\mathrm{rb}} : \mathcal{P} \to \mathbb{R}^{N_T}, \qquad f_{\mathrm{rb}}(\mu;t) := s_\mu(c_{\mathrm{rb},\mu}(t)), \quad (5)$$

with a computational complexity independent of $\dim V_h$, the solution of which, however, still requires time-integration.

The quality and efficiency of RB models hinges on the problem adapted RB space V_{rb} which could be constructed in an iterative manner steered by a posteriori error estimates using the POD-greedy algorithm from Haasdonk (2013). Instead, we obtain by the method of snapshots

$$V_{\mathrm{rb}} := \langle \texttt{POD}(\{c_{h,\mu} \,|\, \mu \in \mathcal{P}_{\mathrm{rb}}\rangle\,, \qquad \text{with } \mathcal{P}_{\mathrm{rb}} \subset \mathcal{P} \quad (6)$$

consisting of only few a priori selected parameters (e.g. the outermost four points in $\mathcal{P}$), where we use the hierarchic approximate POD from Himpe et al. (2018) for $N_T \gg 1$ to avoid computing the SVD of a dense snapshot Gramian of size N_T^2.

2.2 Kernel models

While still requiring time-integration, we can afford to use RB ROMs to generate a sufficient amount of training data $X_{\text{train}} = \{(\mu, f_{\text{rb}}(\mu)) \,|\, \mu \in \mathcal{P}_{\text{ml}}\} \cup \{(\mu, f_h(\mu) \,|\, \mu \in \mathcal{P}_{\text{rb}})\}$, augmented by the FOM-data available as a side-effect from generating V_{rb}. Using this data, we obtain the ML model f_{ml} from (3) using the vectorial greedy orthogonal kernel algorithm from Santin and Haasdonk (2021).

While resulting in substantial computational gains, the presented approach from Gavrilenko et al. (2022) still relies on the traditional offline/online splitting of the computational process to train the RB ROM as well as the ML model to be valid for all of $\mathcal{P}$, requiring a priori choices regarding $\mathcal{P}_{\text{rb}}$ and $\mathcal{P}_{\text{ml}}$ with a significant impact on the overall performance and applicability of these models.

3. AN ADAPTIVE MODEL HIERARCHY

Keil et al. (2021) introduced an approach beyond the classical offline/online splitting where a RB ROM is adaptively enriched based on rigorous a posteriori error estimates, following the path of an optimization procedure through the parameter space. Similarly, we propose an adaptive enrichment yielding a hierarchy of FOM, RB ROM and ML models, based on the standard residual-based a posteriori estimate on the RB output error, $\|f_h(\mu) - f_{\text{rb}}(\mu)\|_{L^2([0,T])} \leq \Delta_{\text{rb}}(\mu)$, for which we refer to the references in Milk et al. (2016).

Algorithm 1 Adaptive QoI model generation

Require: ROM tolerance $\varepsilon > 0$, ML trust/train criteria
$X_{\text{train}} = \emptyset$, $\Phi_{\text{RB}} = \{\}$, $V_{\text{rb}} := \langle \Phi_{\text{rb}} \rangle$, $f_{\text{ml}} := 0$
for all $\mu \in \mathcal{P}$ selected by outer loop **do**
  if ML model is trustworthy **then return** $f_{\text{ml}}(\mu)$
  else
    compute $f_{\text{rb}}(\mu)$, $\Delta_{\text{rb}}(\mu)$
    if $\Delta_{\text{rb}}(\mu) \leq \varepsilon$ **then**
      collect $X_{\text{train}} = X_{\text{train}} \cup \{(\mu, f_{\text{rb}}(\mu))\}$
      (optionally) fit ML model, **return** $f_{\text{rb}}(\mu)$
    else $\Pi_{\Phi_{\text{rb}}}$: orth. proj. onto $\langle \Phi_{\text{rb}} \rangle$
      compute $f_h(\mu)$
      enrich $\Phi_{\text{rb}} = \Phi_{\text{rb}} \cup \texttt{POD}(c_h(\mu) - \Pi_{\Phi_{\text{rb}}}[c_h(\mu)])$
      update RB ROM
      collect $X_{\text{train}} = X_{\text{train}} \cup \{(\mu, f_h(\mu))\}$
      (optionally) fit ML model, **return** $f_h(\mu)$

As a means to judge if a ML model is trustworthy, we propose a manual validation using the following a posteriori error estimate on the ML QoI error. While not as cheaply computable as f_{ml}, it still allows to validate the ML model without computing f_h.

Proposition 1. (ML model a posteriori error estimate). Let $f_{\text{rb}}(\mu)$, $f_{\text{ml}}(\mu) \in \mathbb{R}^{N_T}$ denote the RB ROM and ML model approximations of $f_h(\mu)$, respectively, and let $\Delta_{\text{rb}}(\mu)$ denote an upper bound on the RB-output error. We then have by triangle inequality for all $\mu \in \mathcal{P}$

$$\|f_h(\mu) - f_{\text{ml}}(\mu)\|_{L^2([0,T])} \leq \Delta_{\text{rb}}(\mu) + \|f_{\text{rb}}(\mu) - f_{\text{ml}}(\mu)\|_{L^2([0,T])}, \tag{7}$$

where the right hand side is computable with a computational complexity independent of $\dim V_h$.

Applying Algorithm 1 to the example of one-dimensional single-phase reactive flow from the last row of Table 1 in Gavrilenko et al. (2022), with $\dim \mathcal{P} = 2$, $N_T = 24576$ time steps, $\dim V_h = 65537$ gives the behaviour shown in Figure 1, where we set $\varepsilon = 10^{-2}$, retrain the ML model every 10 collected samples and unconditionally trust the ML model as soon as $|X_{\text{train}}| \geq 50$.[1] For the considered diffusion dominated regime, we only require a single evaluation of f_h (yielding a $\dim V_{\text{rb}} = 15$-dimensional RB ROM), which results in even further computational savings, compared to the results obtained in Gavrilenko et al. (2022).

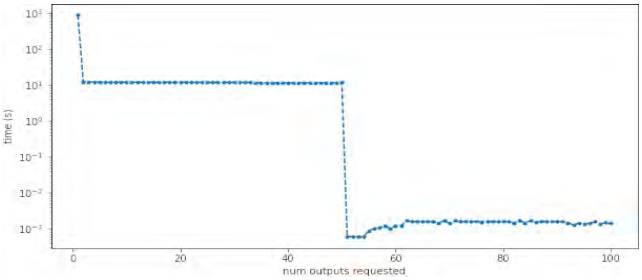

Fig. 1. Each dot correspond to the input-to-output query time of the adaptive model from Algorithm 1 applied to Gavrilenko et al. (2022).

REFERENCES

Gavrilenko, P., Haasdonk, B., Iliev, O., Ohlberger, M., Schindler, F., Toktaliev, P., Wenzel, T., and Youssef, M. (2022). A full order, reduced order and machine learning model pipeline for efficient prediction of reactive flows. In *Large-Scale Scientific Computing*, 378–386. Springer International Publishing.

Haasdonk, B. (2013). Convergence rates of the POD-greedy method. *ESAIM Math. Model. Numer. Anal.*, 47(3), 859–873.

Himpe, C., Leibner, T., and Rave, S. (2018). Hierarchical Approximate Proper Orthogonal Decomposition. *SIAM Journal on Scientific Computing*, 40(5), A3267–A3292.

Keil, T., Mechelli, L., Ohlberger, M., Schindler, F., and Volkwein, S. (2021). A non-conforming dual approach for adaptive trust-region reduced basis approximation of PDE-constrained parameter optimization. *ESAIM Math. Model. Numer. Anal.*, 55(3), 1239–1269.

Milk, R., Rave, S., and Schindler, F. (2016). pyMOR – generic algorithms and interfaces for model order reduction. *SIAM J. Sci. Comput.*, 38(5), S194–S216.

Santin, G. and Haasdonk, B. (2021). Kernel methods for surrogate modeling. In P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L.M. Silveira (eds.), *Model Order Reduction*, volume 2, 311–353. De Gruyter.

[1] The experiments were performed using `pyMOR` from Milk et al. (2016) and `dune-gdt` from `https://docs.dune-gdt.org/`.