Modal Decomposition of Flow Data via Gradient-Based Transport Optimization

Felix Black[1(✉)], Philipp Schulze[1], and Benjamin Unger[2]

[1] Institute of Mathematics, Technische Universität Berlin, Berlin, Germany
{black,pschulze}@math.tu-berlin.de
[2] Stuttgart Center for Simulation Science, Universität Stuttgart, Stuttgart, Germany
benjamin.unger@simtech.uni-stuttgart.de

Abstract. In the context of model reduction, we study an optimization problem related to the approximation of given data by a linear combination of transformed modes, called *transformed proper orthogonal decomposition* (tPOD). In the simplest case, the optimization problem reduces to a minimization problem well-studied in the context of proper orthogonal decomposition. Allowing transformed modes in the approximation renders this approach particularly useful to compress data with transported quantities, which are prevalent in many flow applications. We prove the existence of a solution to the infinite-dimensional optimization problem. Towards a numerical implementation, we compute the gradient of the cost functional and derive a suitable discretization in time and space. We demonstrate the theoretical findings with three numerical examples using a periodic shift operator as transformation.

Keywords: Nonlinear model order reduction · Transport-dominated phenomena · Transformed modes · Gradient-based optimization

1 Introduction

Projection-based model order reduction (MOR) typically relies on the fact that the solution manifold of a (parametrized) differential equation can be approximately embedded in a low-dimensional linear subspace. The best subspace of a given dimension, where best is understood as the minimal worst-case approximation error, is characterized by the Kolmogorov n-widths [14]. In practice, the minimizing subspace for the n-widths is difficult to compute. Instead, one relies on the proper orthogonal decomposition (POD) [11], which is typically combined with a greedy-search within the parameter domain, to get an approximate solution. In more detail, for given parameters $\mu_\sigma \in \mathbb{M}$ $(\sigma = 1, \dots, \ell)$, associated data samples $z(t, x; \mu_\sigma)$ with time variable $t \in \mathbb{T} := [0, T]$, space variable $x \in \Omega \subseteq \mathbb{R}^d$, and desired dimension $r \in \mathbb{N}$ of the low-dimensional subspace, POD determines orthonormal basis functions of a low-dimensional subspace solving the minimization problem

R. King and D. Peitsch (Eds.): AFCC 2021, NNFM 152, pp. 203–224, 2022.
https://doi.org/10.1007/978-3-030-90727-3_13

$$\begin{cases} \min \frac{1}{2} \sum_{\sigma=1}^{\ell} \int_0^T \left\| z(t,x;\mu_\sigma) - \sum_{i=1}^{r} \alpha_i(t;\mu_\sigma)\varphi_i(x) \right\|^2 \mathrm{d}t \\ \text{with } \alpha_i(t;\mu_\sigma) := \langle z(t,x;\mu_\sigma), \varphi_i(x) \rangle \quad \text{for } i = 1,\dots,r,\ \sigma = 1,\dots,\ell, \\ \text{s.t. } \langle \varphi_i, \varphi_j \rangle = \delta_{ij} \quad \text{for } i,j = 1,\dots,r. \end{cases} \tag{1}$$

If the n-widths, respectively the Hankel singular values for linear dynamical systems [33], decay fast, then one can expect to construct an effective reduced-order model (ROM) able to approximate the full dynamics with a small approximation error. Although one can show exponential decay for a large class of problems [18], it is well-known, see for instance [7,10], that the decay of the n-widths for flow problems is typically slow, thus conspiring against MOR. The main reason for slowly decaying n-widths is that flow problems often exhibit a strong space-time coupling, thus conspiring against the separation of space and time inherent to the definition of the Kolmogorov n-widths and POD approximation. Consequently, POD is often not able to produce accurate ROMs with a small number of modes.

To remedy this issue prevalent in transport-dominated phenomena, several strategies have been proposed in the literature. We refer to [7,12,15,22–24,27,31,32,34] to name just a few. One promising approach, introduced in [16,26,30] and formalized in [3,6], is to replace the POD minimization problem with

$$\min \frac{1}{2} \sum_{\sigma=1}^{\ell} \int_0^T \left\| z(t,x;\mu_\sigma) - \sum_{i=1}^{r} \alpha_i(t;\mu_\sigma)\varphi_i(x - p_i(t;\mu_\sigma)) \right\|^2 \mathrm{d}t, \tag{2}$$

thus accounting explicitly for the transportation of quantities throughout the spatial domain. Let us emphasize that in the above formulation we may have $x - p_i(t;\mu_\sigma) \notin \Omega$. This may be resolved by a periodic domain, by an extrapolation approach [5], or by defining the modes on an extended domain $\widetilde{\Omega}$ [3, Sec. 7.2].

Consequently, the linear subspace in the Kolmogorov n-widths is replaced with a subspace able to adapt itself to the solution over time, hence rendering this a nonlinear approach, which we refer to as *transformed POD* (tPOD). Note that in contrast to the POD minimization problem (1), we do not require the modes to be orthogonal to each other. This is due to the fact that in the setting of (2), we would need orthogonality of $\varphi_i(x - p_i(t;\mu_\sigma))$ and $\varphi_j(x - p_j(t;\mu_\sigma))$ for all $i \neq j$, $t \in [0,T]$, and $\sigma = 1,\dots,\ell$, which is in general not a reasonable assumption, cf. [3, Ex. 4.4] for an illustrative example.

In the past years, there have been some attempts of solving discretized versions of (2) or related minimization problems. In [26], the authors propose a heuristic iterative method for computing a decomposition of a given snapshot matrix by an approximation ansatz as in (2). The numerical experiments indicate promising results, but it is not clear in which situations the proposed method actually determines an optimal solution. Another heuristic has been recently proposed in [5, sec. 5.2.1] and it is applied to snapshot data of a wildland fire simulation. This method is based on a decomposition of the snapshot matrix and involves a small number of singular value decompositions without requiring an iterative procedure. The numerical results presented in [5] demonstrate

the effectiveness of this approach, but it is in general not optimal in the sense of the minimization problem (2). In contrast, the method introduced in [30] directly solves a fully discretized version of (2) by determining optimal modes φ and coefficients α, but assumes the paths p to be given or determined in a pre-processing step. A similar optimization problem has been proposed in [25] and aims at approximating the snapshot matrix by a sum of matrices representing different reference frames while achieving a fast singular value decay in each of the reference frames. The corresponding cost function is shown to be an upper bound for a fully discretized version of (2) and the cost functions coincide for the special case that only one reference frame is considered, i.e., if the same shift is applied to all modes in (2). Again, the paths are not considered as part of the optimization problem, but instead determined in a pre-processing step via peak or front tracking. On the contrary, the authors in [20] focus on determining optimal paths, whereas the determination of optimal ansatz functions and coefficients is not addressed. Moreover, to simplify the optimization problem, the paths are sought within a low-dimensional subspace consisting of predefined time-dependent library functions. As in the case of the other mentioned works, also the authors in [20] consider a fully discrete optimization problem.

We conclude that a gradient-based algorithm for the full optimization problem (2) is currently not available. Besides, a rigorous proof showing that (2) has a solution is missing in the literature. A notable exception is provided in [3, Thm. 4.6], albeit under the assumption that the path variables $p_i(t)$ are known a-priori. In this paper we aim to close this gap. Our main contributions are the following:

1. We show in Theorem 3 the existence of a minimizing solution for the optimization problem (6), which generalizes the minimization problem (2). Afterward, we reformulate the constrained minimization problem (6) as unconstrained problem (9) by adding appropriate penalty terms and conclude from Theorem 3 that also the unconstrained problem has a solution, cf. Corollary 1. In addition, Theorem 5 details that the solution of the unconstrained problem converges to the solution of the constraint problem.
2. We compute the gradient of the unconstrained problem in Theorem 7, which enables the use of gradient-based methods to solve (2). In this context, a remarkable finding is that the paths have to be sufficiently smooth (e.g. in $H^1(0,T)$), since otherwise some directional derivatives of the cost functional with respect to the paths may not exist, cf. Example 2.
3. We discuss the discretization of the gradient in space and time in Sect. 4 and explicitly compute the path-dependent inner products for the shift operator with periodic boundary conditions in Example 3. Finally, the effectiveness of gradient-based optimization is demonstrated for several examples in Sect. 5.

Notation. We denote the space of real $m \times n$ matrices by $\mathbb{R}^{m\times n}$ and the transpose of a matrix A is written as $A^\top$. Furthermore, for a vector with n entries

all equal to one we use the symbol $\mathbf{1}_n$. Besides, for abbreviating diagonal and blockdiagonal matrices we use

$$\operatorname{diag}(a_1, \dots, a_n) := \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix}, \quad \operatorname{blkdiag}(A_1, \dots, A_n) := \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_n \end{bmatrix},$$

respectively, where $a_1, \dots, a_n$ are scalars and $A_1, \dots, A_n$ matrices of arbitrary size. For the Kronecker product of two matrices A and B we write $A \otimes B$. The space of square-integrable functions mapping from an interval (a, b) to a Banach space $\mathscr{X}$ is denoted by $L^2(a, b; \mathscr{X})$ and, similarly, the space of essentially bounded measurable functions by $L^\infty(a, b; \mathscr{X})$. Furthermore, we use $H^1(a, b; \mathscr{X})$ for the Sobolev subspace of functions in $L^2(a, b; \mathscr{X})$ possessing also a weak derivative in $L^2(a, b; \mathscr{X})$. The corresponding subspace consisting of $H^1(a, b; \mathscr{X})$ functions whose values at the boundaries a and b coincide is denoted by $H^1_{\mathrm{per}}(a, b; \mathscr{X})$. Besides, for the space of continuous functions from $[a, b]$ to $\mathscr{X}$ we use the symbol $C([a, b]; \mathscr{X})$. For the special case $\mathscr{X} = \mathbb{R}$, we omit the last argument, i.e., we write, for instance, $L^2(a, b)$ instead of $L^2(a, b; \mathbb{R})$.

2 Preliminaries and Problem Formulation

To formalize the optimization problem (2), we introduce the following spaces and notation. Consider a real Hilbert space $(\mathscr{X}, \langle \cdot, \cdot \rangle_{\mathscr{X}})$ with induced norm $\| \cdot \|_{\mathscr{X}}$, and let $\mathscr{Y}$ denote a dense subspace of $\mathscr{X}$ that itself is a reflexive Banach space with norm $\| \cdot \|_{\mathscr{Y}}$. Our standing assumption is that we are minimizing the mean-squared distance to the data $z \in L^2(0, T; \mathscr{Y})$ in the Bochner space $L^2(0, T; \mathscr{X})$ with the additional requirement that the modes φ_i are elements of $\mathscr{Y}$.

To formalize the meaning of $\varphi_i(x - p_i(t))$ in (2), we follow the notation in [3] and introduce a family of linear and bounded operators $\mathcal{T}_i \colon \mathscr{P}_i \to \mathscr{B}(\mathscr{X})$ with real, finite-dimensional vector space $\mathscr{P}_i$, for which we postulate the following properties, taken from [3, Ass. 4.1].

Assumption 1. *For every $i = 1, \dots, r$, every $\varphi_i \in \mathscr{Y}$, and every $p_i \in \mathscr{P}_i$, the operator $\mathcal{T}_i(p_i)$ is $\mathscr{Y}$-invariant, i.e., $\mathcal{T}_i(p_i)\mathscr{Y} \subseteq \mathscr{Y}$, and the mapping*

$$\mathscr{P}_i \to \mathscr{X}, \qquad p_i \mapsto \mathcal{T}_i(p_i)\varphi_i$$

is continuous.

A particular example for such a family of operators is given by the shift operator with periodic boundary conditions, see for instance [3, Ex. 5.2]. For further examples we refer to [2,15].

For the ease of presentation, we restrict ourselves to the case $\mathscr{P}_i = \mathbb{R}$, and emphasize that all results can be generalized to $\mathscr{P}_i = \mathbb{R}^{m_i}$ for some $m_i \in \mathbb{N}$. For

$$\mathscr{Z} := L^2(0, T; \mathbb{R}^r) \times H^1(0, T; \mathbb{R}^r) \times \mathscr{Y}^r \tag{3}$$

let us define the cost functional

$$J\colon \mathscr{Z} \to \mathbb{R}, \qquad (\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) \mapsto \frac{1}{2} \left\| z - \sum_{i=1}^{r} \alpha_i \mathcal{T}_i\left(p_i\right) \varphi_i \right\|^2_{L^2(0,T;\mathscr{X})} \tag{4}$$

and for $C > 0$ the space

$$\mathcal{A}_C := \left\{ (\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) \in \mathscr{Z} \;\middle|\; \max\left\{ \|\varphi_i\|_{\mathscr{Y}}, \|\alpha_i\|_{L^2(0,T)}, \|p_i\|_{H^1(0,T)} \right\} \leq C \right\}, \tag{5}$$

where we use the notation $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$ to denote the coefficients of $\boldsymbol{\alpha}$ and analogously for $\boldsymbol{p}$ and $\boldsymbol{\varphi}$. To ensure that the norm in (4) is defined, we invoke the following assumption, which is for instance satisfied if the family of operators $\mathcal{T}_i(\cdot)$ is uniformly bounded, cf. [3, Lem 4.2].

Assumption 2. *For every $i = 1, \ldots, r$, $\alpha_i \in L^2(0,T)$, $p_i \in H^1(0,T)$, and every $\varphi_i \in \mathscr{Y}$, we assume*

$$\alpha_i \mathcal{T}_i(p_i) \varphi_i \in L^2(0,T;\mathscr{X}).$$

With these preparations, the constrained minimization problem that we are interested in takes the form

$$\min_{(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})} J\left(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}\right), \quad \text{s. t. } (\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) \in \mathcal{A}_C. \tag{6}$$

Note that we have set $\ell = 1$ in (2) to simplify the notation. We emphasize that it is straightforward to generalize all results to $\ell > 1$.

3 Main Results

As first main result, we discuss the existence of a solution for the optimization problem (6), thus generalizing [3, Thm. 4.6] to include the path variables.

Theorem 3. *Assume that the reflexive Banach space $\mathscr{Y}$ is compactly embedded into $\mathscr{X}$, and let $z \in L^2(0,T;\mathscr{Y})$. Furthermore, let the family of transformation operators satisfy Assumptions 1 and 2. Then the constraint minimization problem (6) has a solution for every $C > 0$.*

Proof. The proof follows along the lines of the proof of [3, Thm. 4.6], with slight modifications to account for the optimization with respect to the path variables. Let $C > 0$. We first observe that the optimization problem possesses a finite infimum $J^\star \geq 0$. This follows directly from $J \geq 0$ and $(0,0,0) \in \mathcal{A}_C$. We may thus choose a sequence $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k)_{k\in\mathbb{N}} \in \mathcal{A}_C$ satisfying

$$\lim_{k\to\infty} J(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k) = J^\star.$$

Additionally, we have

$$\begin{aligned} &\left\|(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k)\right\|^2_{L^2(0,T;\mathbb{R}^r)\times H^1(0,T;\mathbb{R}^r)\times \mathscr{Y}^r} \\ &\qquad = \|\boldsymbol{\alpha}^k\|^2_{L^2(0,T;\mathbb{R}^r)} + \|\boldsymbol{p}^k\|^2_{H^1(0,T;\mathbb{R}^r)} + \|\boldsymbol{\varphi}^k\|^2_{\mathscr{Y}^r} \leq 3rC^2 \end{aligned}$$

for all $k \in \mathbb{N}$, such that the Eberlein-Šmuljan theorem [35, Thm. 21.D] ensures the existence of a weakly convergent subsequence $(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n})_{n\in\mathbb{N}} \subseteq \mathcal{A}_C$ with weak limit $(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star) \in \mathcal{A}_C$, cf. [35, Prop. 21.23 (c)]. Due to the compact embeddings $\mathscr{Y} \hookrightarrow \mathscr{X}$ and $H^1(0,T;\mathbb{R}^r) \hookrightarrow L^2(0,T;\mathbb{R}^r)$, we conclude that $(\boldsymbol{\varphi}^{k_n})_{n\in\mathbb{N}}$ and $(\boldsymbol{p}^{k_n})_{n\in\mathbb{N}}$ converge strongly in $\mathscr{X}$ and $L^2(0,T;\mathbb{R}^r)$ to $\boldsymbol{\varphi}^\star$ and $\boldsymbol{p}^\star$, respectively, cf. [35, Prop. 21.35]. Using [29, Thm. 3.12], we conclude the existence of a subsequence, for which we use the same indexing, such that $(\boldsymbol{p}^{k_n})_{n\in\mathbb{N}}$ converges pointwise to $\boldsymbol{p}^\star$ for almost all $t \in (0,T)$.

For the next part of the proof, we introduce the mapping

$$\beta\colon \mathscr{Z} \to L^2(0,T;\mathscr{X}), \qquad (\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) \mapsto \sum_{i=1}^{r} \alpha_i \mathcal{T}_i(p_i)\varphi_i \tag{7}$$

with $\mathscr{Z}$ as defined in (3) and notice

$$J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) = \frac{1}{2}\|z - \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})}.$$

If $\beta(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n})$ converges weakly to $\beta(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star)$, then the weak sequential lower semicontinuity of the norm, see for instance [35, Prop. 21.23 (c)], implies that $(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star)$ is a minimizer of J. It thus remains to show that $\beta(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n})$ converges weakly to $\beta(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star)$.

To this end, we observe that

$$\begin{aligned}
&\|\mathcal{T}_i(p_i^{k_n}(t))\varphi_i^{k_n} - \mathcal{T}_i(p_i^\star(t))\varphi^\star\|_{\mathscr{X}} \\
&\le \|\mathcal{T}_i(p_i^{k_n}(t))\varphi_i^{k_n} - \mathcal{T}_i(p_i^\star(t))\varphi_i^{k_n}\|_{\mathscr{X}} + \|\mathcal{T}_i(p_i^\star(t))\varphi_i^{k_n} - \mathcal{T}_i(p_i^\star(t))\varphi_i^\star\|_{\mathscr{X}}
\end{aligned}$$

together with Assumption 1 and the strong convergence of $(\varphi_i^{k_n})_{n\in\mathbb{N}}$ in $\mathscr{X}$ implies

$$\|\mathcal{T}_i(p_i^{k_n}(t))\varphi_i^{k_n} - \mathcal{T}_i(p_i^\star(t))\varphi^\star\|_{\mathscr{X}} \to 0 \qquad \text{for } n \to \infty$$

for $i = 1,\dots,r$ and almost all $t \in (0,T)$. Let $f \in L^2(0,T;\mathscr{X})$. Then clearly

$$\langle f(t), \mathcal{T}_i(p_i^{k_n}(t))\varphi_i^{k_n}\rangle_{\mathscr{X}} \to \langle f(t), \mathcal{T}_i(p_i^\star(t))\varphi_i^\star\rangle_{\mathscr{X}} \qquad \text{for } n \to \infty$$

for $i = 1,\dots,r$ and almost all $t \in (0,T)$ such that [35, Prop 21.23 (j)] implies

$$\sum_{i=1}^{r} \left\langle \alpha_i^{k_n}, \langle f, \mathcal{T}_i(p_i^{k_n})\varphi_i^{k_n}\rangle_{\mathscr{X}} \right\rangle_{L^2(0,T)} \to \sum_{i=1}^{r} \langle \alpha_i^\star, \langle f, \mathcal{T}_i(p_i^\star)\varphi_i^\star\rangle_{\mathscr{X}}\rangle_{L^2(0,T)}$$

for $n \to \infty$ and thus

$$\beta(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n}) \rightharpoonup \beta(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star) \qquad \text{for } n \to \infty,$$

which completes the proof. □

For numerical methods, it may be easier to work with unconstrained optimization problems. To this end, we use a penalty method, see for instance [17, Cha. 13.1], i.e., we add the constraint equation with a penalty parameter to the cost functional. In more detail, we assume for $C > 0$ a penalty functional

$$\Lambda_C \colon \mathscr{Z} \to \mathbb{R} \tag{8}$$

with the following properties available.

Assumption 4. *The penalty functional (8) is continuous, weakly sequentially lower semicontinuous, non-negative and has the following properties:*

- *We have $\Lambda_C(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) = 0$ if, and only if, $(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) \in \mathcal{A}_C$.*
- *For any sequence $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k)$ with*

$$\max\{\|\boldsymbol{\alpha}^k\|_{L^2(0,T;\mathbb{R}^r)}, \|\boldsymbol{p}^k\|_{H^1(0,T;\mathbb{R}^r)}, \|\boldsymbol{\varphi}^k\|_{\mathscr{Y}^r}\} \to \infty \qquad \text{for } k \to \infty,$$

we have $\Lambda_C(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k) \to \infty$ for $k \to \infty$.

Example 1. The penalty functional

$$\Lambda_C(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \sum_{i=1}^{r} \max\{0, \max\{\|\alpha_i\|_{L^2(0,T)}, \|p_i\|_{H^1(0,T)}, \|\varphi_i\|_{\mathscr{Y}}\} - C\}$$

satisfies Assumption 4.

The penalized cost functional is then given as

$$\widetilde{J}_C(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}, \lambda) := J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) + \lambda \Lambda_C(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})$$

with penalty coefficient $\lambda > 0$. The associated (unconstrained) minimization problem is thus given by

$$\min_{(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) \in \mathscr{Z}} \widetilde{J}_C(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}, \lambda) \tag{9}$$

with $\mathscr{Z}$ as defined in (3) and given $\lambda > 0$.

Corollary 1. *Let the assumptions of Theorem 3 and Assumption 4 be satisfied. Then for any $C > 0$ and any $\lambda > 0$ the optimization problem (9) has a solution.*

Proof. Similarly as in the proof of Theorem 3, we conclude the existence of a finite infimum, such that we can choose a minimizing sequence $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k) \in \mathscr{Z}$. Due to Assumption 4, we deduce that $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k)_{k \in \mathbb{N}}$ is bounded in $\mathscr{Z}$, i.e., there exists some $\widetilde{C} > 0$ such that $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k) \in \mathcal{A}_{\widetilde{C}}$ for all $k \in \mathbb{N}$. The remaining proof thus follows along the lines of the proof of Theorem 3. □

Theorem 5. *Let $(\lambda^k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ denote a non-decreasing sequence of positive numbers with $\lim_{k \to \infty} \lambda^k = \infty$, and let $C > 0$. For $k \in \mathbb{N}$, let $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k) \in \mathscr{Z}$ denote a solution of (9) with penalty parameter λ^k. If the assumptions of Corollary 1 are satisfied, then any limit point of $(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k)_{k \in \mathbb{N}}$ is a solution of (6).*

Proof. The proof follows along the lines of the proof of the main theorem in [17, Cha. 13.1]. Let $(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star) \in \mathcal{A}_C$ denote a minimizer of (6) with minimum $J^\star$. Then for every $k \in \mathbb{N}$ we have

$$\begin{aligned}\widetilde{J}_C(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k, \lambda^k) &\leq \widetilde{J}_C(\boldsymbol{\alpha}^{k+1}, \boldsymbol{p}^{k+1}, \boldsymbol{\varphi}^{k+1}, \lambda^k) \\ &\leq \widetilde{J}_C(\boldsymbol{\alpha}^{k+1}, \boldsymbol{p}^{k+1}, \boldsymbol{\varphi}^{k+1}, \lambda^{k+1})\end{aligned}$$

and

$$J(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k) \leq \widetilde{J}_C(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k, \lambda^k) \leq \widetilde{J}_C(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star, \lambda^k) = J(\boldsymbol{\alpha}^\star, \boldsymbol{p}^\star, \boldsymbol{\varphi}^\star) = J^\star.$$

Thus $(\widetilde{J}_C(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k, \lambda^k))_{k\in\mathbb{N}}$ is a monotone sequence bounded above by $J^\star$. We thus set

$$\widetilde{J}_C^\star := \lim_{k\to\infty} \widetilde{J}_C(\boldsymbol{\alpha}^k, \boldsymbol{p}^k, \boldsymbol{\varphi}^k, \lambda^k) \leq J^\star. \tag{10}$$

Let $(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n})$ denote a convergent subsequence with limit $(\boldsymbol{\alpha}^\dagger, \boldsymbol{p}^\dagger, \boldsymbol{\varphi}^\dagger)$ and set

$$J^\dagger := \lim_{n\to\infty} J(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n}) = J(\boldsymbol{\alpha}^\dagger, \boldsymbol{p}^\dagger, \boldsymbol{\varphi}^\dagger), \tag{11}$$

using the continuity of J. Subtracting (10) from (11) yields

$$\lim_{n\to\infty} \lambda^{k_n} \Lambda_C(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n}) = \widetilde{J}_C^\star - J^\dagger.$$

Assumption 4 and $\lambda^{k_n} \to \infty$ for $n \to \infty$ together with the continuity of Λ_C thus implies

$$\Lambda_C(\boldsymbol{\alpha}^\dagger, \boldsymbol{p}^\dagger, \boldsymbol{\varphi}^\dagger) = \lim_{n\to\infty} \Lambda_C(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n}) = 0,$$

showing $(\boldsymbol{\alpha}^\dagger, \boldsymbol{p}^\dagger, \boldsymbol{\varphi}^\dagger) \in \mathcal{A}_C$. We conclude

$$J^\dagger = \lim_{n\to\infty} J(\boldsymbol{\alpha}^{k_n}, \boldsymbol{p}^{k_n}, \boldsymbol{\varphi}^{k_n}) \leq J^\star,$$

which completes the proof. □

Although (9) is an unconstrained optimization problem, we still have to choose a suitable constant $C > 0$ for the admissible set. Let us emphasize that the proofs of Theorem 3 and Corollary 1 heavily depend on the fact that we have bounded sequences, which is the main reason for the constant $C > 0$ in the admissible set (5). However, we observed faster convergence in our numerical experiments when considering the unconstrained minimization problem without penalization. For this reason and the sake of a concise presentation, we consider in the following only the unconstrained optimization problem (9) with penalty parameter $\lambda = 0$. Nevertheless, we emphasize that adding the derivatives of the penalty terms to the gradient formulas is straightforward as long as the partial Fréchet derivatives of Λ_C are available.

To solve the optimization problem (9) with penalty parameter $\lambda = 0$ numerically, we employ a gradient-based algorithm and thus have to compute the gradient of the objective function (4). It is easy to see that the directional derivatives

of J with respect to the coefficient function $\boldsymbol{\alpha} \in L^2(0,T;\mathbb{R}^r)$ and the modes $\boldsymbol{\varphi} \in \mathscr{Y}^r$ in directions $\boldsymbol{d} \in L^2(0,T;\mathbb{R}^r)$ and $\boldsymbol{h} \in \mathscr{Y}^r$, respectively, are given by

$$\partial_{\boldsymbol{\alpha},\boldsymbol{d}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) = \sum_{i=1}^{r} \left\langle \sum_{j=1}^{r} \alpha_j \mathcal{T}_j(p_j)\varphi_j - z, d_i \mathcal{T}_i(p_i)\varphi_i \right\rangle_{L^2(0,T;\mathscr{X})}, \tag{12a}$$

$$\partial_{\boldsymbol{\varphi},\boldsymbol{h}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) = \sum_{i=1}^{r} \left\langle \sum_{j=1}^{r} \alpha_j \mathcal{T}_j(p_j)\varphi_j - z, \alpha_i \mathcal{T}_i(p_i)h_i \right\rangle_{L^2(0,T;\mathscr{X})}. \tag{12b}$$

The situation is slightly different for the partial derivative with respect to the path variable. First of all, we have to ensure that the transformed modes are differentiable (with respect to the path variable), i.e., we have to evoke the following assumption.

Assumption 6. *For every $\varphi_i \in \mathscr{Y}$ and every $i = 1,\dots,r$, the mapping*

$$\mathbb{R} \to \mathscr{X}, \qquad p_i \mapsto \mathcal{T}_i(p_i)\varphi_i,$$

is continuously differentiable with derivatives in $\mathscr{X}$. For $p_i \in \mathbb{R}$ we denote the derivative by $\frac{\partial}{\partial p_i}\mathcal{T}_i(p_i)\,\varphi_i \in \mathscr{X}$ and assume $\alpha_i \frac{\partial}{\partial p_i}\mathcal{T}_i(p_i)\varphi_i \in L^2(0,T;\mathscr{X})$ for all $\alpha_i \in L^2(0,T)$ and all $p_i \in H^1(0,T)$.

In this case, the directional derivative in direction $\boldsymbol{g} \in H^1(0,T;\mathbb{R}^r)$ is given as

$$\partial_{\boldsymbol{p},\boldsymbol{g}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) = \sum_{i=1}^{r} \left\langle \sum_{j=1}^{r} \alpha_j \mathcal{T}_j(p_j)\varphi_j - z, \alpha_i \left[\tfrac{\partial}{\partial p_i}\mathcal{T}_i(p_i)\varphi_i\right] g_i \right\rangle_{L^2(0,T;\mathscr{X})}. \tag{12c}$$

Note that the Sobolev embedding theorems, see for instance [35, Thm. 21.A.(d)], imply $g_i \in C([0,T]) \subseteq L^\infty(0,T)$, such that (12c) is defined.

Theorem 7. *Let the transformation operators satisfy Assumptions 1, 2, and 6. Let $(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) \in \mathscr{Z}$ and assume*

$$\mathcal{T}_i(p_i)\varphi_i \in L^\infty(0,T;\mathscr{X}), \tag{13a}$$

$$\alpha_i \|\mathcal{T}_i(p_i)\| \in L^2(0,T) \tag{13b}$$

for $i = 1,\dots,r$, then the partial Fréchet derivatives of the cost functional J (defined in (4)) with respect to the coefficients, paths, and modes at $(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) \in \mathscr{Z}$ are given by

$$\partial_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})(\boldsymbol{d}) := \partial_{\boldsymbol{\alpha},\boldsymbol{d}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \qquad \forall \boldsymbol{d} \in L^2(0,T;\mathbb{R}^r), \tag{14a}$$

$$\partial_{\boldsymbol{p}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})(\boldsymbol{g}) := \partial_{\boldsymbol{p},\boldsymbol{g}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \qquad \forall \boldsymbol{g} \in H^1(0,T;\mathbb{R}^r), \tag{14b}$$

$$\partial_{\boldsymbol{\varphi}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})(\boldsymbol{h}) := \partial_{\boldsymbol{\varphi},\boldsymbol{h}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \qquad \forall \boldsymbol{h} \in \mathscr{Y}^r, \tag{14c}$$

with directional derivatives as defined in (12).

Proof. It suffices to show that J is partially Fréchet differentiable with respect to the coefficients, paths, and modes. Let $(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}), (\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{h}) \in \mathscr{Z}$. Using (13a) we obtain

$$\begin{aligned} J(\boldsymbol{\alpha}+\boldsymbol{d}, \boldsymbol{p}, \boldsymbol{\varphi}) - J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) - \partial_{\boldsymbol{\alpha},\boldsymbol{d}} J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) &= \frac{1}{2} \left\| \sum_{i=1}^{r} d_i \mathcal{T}_i(p_i)\varphi_i \right\|^2_{L^2(0,T;\mathscr{X})} \\ &\leq \frac{1}{2} \max_{i=1,\dots,r} \|\mathcal{T}_i(p_i)\varphi_i\|^2_{L^\infty(0,T;\mathscr{X})} \left(\sum_{i=1}^{r} \|d_i\|_{L^2(0,T)} \right)^2 \\ &\leq \frac{r^2}{2} \max_{i=1,\dots,r} \|\mathcal{T}_i(p_i)\varphi_i\|^2_{L^\infty(0,T;\mathscr{X})} \|\boldsymbol{d}\|^2_{L^2(0,T;\mathbb{R}^r)} \end{aligned}$$

and thus

$$\lim_{\|\boldsymbol{d}\|_{L^2(0,T;\mathbb{R}^r)} \to 0} \frac{|J(\boldsymbol{\alpha}+\boldsymbol{d}, \boldsymbol{p}, \boldsymbol{\varphi}) - J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) - \partial_{\boldsymbol{\alpha},\boldsymbol{d}} J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})|}{\|\boldsymbol{d}\|_{L^2(0,T;\mathbb{R}^r)}} = 0.$$

We conclude that J is Fréchet differentiable with respect to the coefficients with Fréchet derivative as in (14a). For the partial derivative with respect to the modes we obtain

$$\begin{aligned} J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}+\boldsymbol{h}) - J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) - \partial_{\boldsymbol{\varphi},\boldsymbol{h}} J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) &= \frac{1}{2} \left\| \sum_{i=1}^{r} \alpha_i \mathcal{T}_i(p_i) h_i \right\|^2_{L^2(0,T;\mathscr{X})} \\ &\leq \frac{1}{2} \int_0^T \left(\sum_{i=1}^{r} |\alpha_i(t)| \|\mathcal{T}_i(p_i(t))\| \|h_i\|_{\mathscr{X}} \right)^2 \mathrm{d}t \\ &\leq \frac{\|\boldsymbol{h}\|^2_{\mathscr{X}^r}}{2} \int_0^T \left(\sum_{i=1}^{r} |\alpha_i(t)| \|\mathcal{T}_i(p_i(t))\| \right)^2 \mathrm{d}t. \end{aligned}$$

Using (13b), we observe that the integral is finite. Similarly as before, we thus conclude that J is Fréchet differentiable with respect to the modes with Fréchet derivative as in (14c). We conclude our proof for the partial derivative with respect to the path variable. Note that the Sobolev embedding theorem [1, Thm 4.12, Part I, Case A] implies that the Sobolev space $H^1(0,T)$ is continuously embedded into the space $L^\infty(0,T)$, i.e., there exists a constant $\gamma > 0$ independent of g_i, such that $\|g_i\|_{L^\infty(0,T)} \leq \gamma \|g_i\|_{H^1(0,T)}$. We define

$$f_i(p_i, \varphi_i, g_i) := \mathcal{T}_i(p_i + g_i)\varphi_i - \mathcal{T}_i(p_i)\varphi_i - \left[\tfrac{\partial}{\partial p_i} \mathcal{T}_i(p_i) \right] g_i.$$

For $g_i \equiv 0$ we have $f_i(p_i, \varphi_i, g_i) = 0$ for almost all $t \in (0,T)$. For $\|g_i\|_{H^1(0,T)} \neq 0$, let us define $\widehat{\mathbb{T}}_i := \{t \in (0,T) \mid g_i(t) \neq 0\}$. Then

$$\frac{\langle z, \alpha_i f_i(p_i, \varphi_i, g_i) \rangle_{L^2(0,T;\mathscr{X})}}{\|g_i\|_{H^1(0,T)}} \leq \gamma \int_{\widehat{\mathbb{T}}_i} \alpha_i(t) \left\langle z(t), \frac{f_i(p_i(t), \varphi_i, g_i(t))}{|g_i(t)|} \right\rangle_{\mathscr{X}} \mathrm{d}t.$$

From Assumption 6 we conclude

$$\lim_{\|g_i\|_{H^1(0,T)}\to 0} \frac{\langle z, \alpha_i f_i(p_i,\varphi_i,g_i)\rangle_{L^2(0,T;\mathscr{X})}}{\|g_i\|_{H^1(0,T)}} = 0,$$

and thus

$$\lim_{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}\to 0} \frac{\sum_{i=1}^r \langle z, \alpha_i f_i(p_i,\varphi_i,g_i)\rangle_{L^2(0,T;\mathscr{X})}}{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}} = 0.$$

Furthermore, using β as defined in (7), we obtain

$$\begin{aligned}
&\tfrac{1}{2}\|\beta(\boldsymbol{\alpha},\boldsymbol{p}+\boldsymbol{g},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})} - \tfrac{1}{2}\|\beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})} \\
&\qquad - \sum_{j=1}^r \left\langle \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \alpha_j \left[\tfrac{\partial}{\partial p_j}\mathcal{T}_j(p_j)\varphi_j\right] g_j \right\rangle_{L^2(0,T;\mathscr{X})} \\
&= \tfrac{1}{2}\|\beta(\boldsymbol{\alpha},\boldsymbol{p}+\boldsymbol{g},\boldsymbol{\varphi}) - \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) + \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})} \\
&\qquad - \tfrac{1}{2}\|\beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})} \\
&\qquad - \sum_{j=1}^r \left\langle \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \alpha_j \left[\tfrac{\partial}{\partial p_j}\mathcal{T}_j(p_j)\varphi_j\right] g_j \right\rangle_{L^2(0,T;\mathscr{X})} \\
&= \tfrac{1}{2}\|\beta(\boldsymbol{\alpha},\boldsymbol{p}+\boldsymbol{g},\boldsymbol{\varphi}) - \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})} \\
&\qquad + \sum_{j=1}^r \left\langle \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \alpha_j f_j(p_j,\varphi_j,g_j)\right\rangle_{L^2(0,T;\mathscr{X})}.
\end{aligned}$$

Similarly as before, we obtain

$$\begin{aligned}
\lim_{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}\to 0} \frac{1}{2}\frac{\|\beta(\boldsymbol{\alpha},\boldsymbol{p}+\boldsymbol{g},\boldsymbol{\varphi}) - \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})\|^2_{L^2(0,T;\mathscr{X})}}{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}} &= 0, \\
\lim_{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}\to 0} \frac{\left\langle \beta(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}), \sum_{j=1}^r \alpha_j f_j(p_j,\varphi_j,g_j)\right\rangle_{L^2(0,T;\mathscr{X})}}{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}} &= 0.
\end{aligned}$$

Combining the previous results, we thus infer

$$\lim_{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}\to 0} \frac{|J(\boldsymbol{\alpha},\boldsymbol{p}+\boldsymbol{g},\boldsymbol{\varphi}) - J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) - \partial_{\boldsymbol{p},\boldsymbol{g}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi})|}{\|\boldsymbol{g}\|_{H^1(0,T;\mathbb{R}^r)}} = 0,$$

which concludes the proof. □

Remark 1. If the family of transformation operators is uniformly bounded, i.e., there exists some $\overline{C} > 0$ such that

$$\|\mathcal{T}_i(p_i)\| \le \overline{C} \qquad \text{for all } p_i \in \mathbb{R},$$

then it is easy to see that condition (13) is satisfied. Note that in this case Assumption 2 is also satisfied, cf. [3, Lem 4.2]. An example for such a family of operators is (again) the periodic shift operator.

Let us emphasize that it is essential for the directional derivative $\partial_{\boldsymbol{p},\boldsymbol{g}} J$ to have the path variable and associated directions in $H^1(0,T;\mathbb{R}^r)$. The following example details that if we take a direction in $L^2(0,T;\mathbb{R}^r)$, then the directional derivative may not be finite.

Example 2. Consider the shift operator $\mathcal{T}(p)\varphi = \varphi(\cdot - p)$ with periodic embedding into the spaces $\mathscr{X} := L^2(0,2\pi)$ and $\mathscr{Y} := H^1_{\mathrm{per}}(0,2\pi)$, cf. [3, Ex. 4.3 and 5.12]. It is well-known, that the shift operator is a semi-group with generator $-\frac{\partial}{\partial x}$, see for instance [9, Sec. II.2.10]. Let $z(t,x) = t^{-1/3}\cos(x)$, $r = 1$, $p_1 \equiv 0$, and $\varphi_1(x) = \sin(x)$. Then for any $\alpha_1, g_1 \in L^2(0,T)$, we obtain

$$\begin{aligned}\partial_{\boldsymbol{p},\boldsymbol{g}} J(\alpha_1, p_1, \varphi_1) &= -\left\langle \alpha_1\varphi_1 - z, \alpha_1 \tfrac{\partial}{\partial x}\varphi_1 g_1 \right\rangle_{L^2(0,T;\mathscr{X})} \\ &= \langle z, \alpha_1 \tfrac{\partial}{\partial x}\varphi_1 g_1\rangle_{L^2(0,T;\mathscr{X})} = \|\cos(\cdot)\|^2_{\mathscr{X}} \int_0^T t^{-1/3}\alpha_1(t)g_1(t)\,\mathrm{d}t.\end{aligned}$$

We notice that for $\alpha_1(t) = g_1(t) = t^{-1/3}$ we have $\alpha_1, g_1 \in L^2(0,T)$ but the product $t^{-1/3}\alpha_1 g_1$ is not in $L^1(0,T)$. We conclude $\partial_{\boldsymbol{p},\boldsymbol{g}} J(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) \notin \mathbb{R}$.

Remark 2. To ensure $\boldsymbol{p} \in H^1(0,T;\mathbb{R}^r)$ during a (numerical) optimization, we may choose a suitable low-dimensional subspace with continuously differentiable basis functions, such as the space of polynomials with given maximal degree. The associated gradient is easily computed from Theorem 7 via the chain rule. Besides the reduced computational cost, such an approach yields an interpretable representation for the wave speeds. We refer to [20] for a similar idea in a fully discretized setting.

4 Discretization

Towards a numerical implementation, we derive discretized versions of the partial derivatives from Theorem 7. To shorten notation, we introduce for $(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) \in \mathscr{Z}$ and $\boldsymbol{h} \in \mathscr{Y}^r$ the quantities

$$v_i(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) := \left\langle \sum_{j=1}^r \alpha_j \mathcal{T}_j(p_j)\varphi_j - z, \mathcal{T}_i(p_i)\varphi_i \right\rangle_{\mathscr{X}}, \tag{15a}$$

$$\xi_i(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi}) := \left\langle \sum_{j=1}^r \alpha_j \mathcal{T}_j(p_j)\varphi_j - z, \alpha_i \left[\tfrac{\partial}{\partial p_i}\mathcal{T}_i(p_i)\varphi_i\right] \right\rangle_{\mathscr{X}}, \tag{15b}$$

$$\mu_i(\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{\varphi},\boldsymbol{h}) := \left\langle \sum_{j=1}^r \alpha_j \mathcal{T}_j(p_j)\varphi_j - z, \alpha_i \mathcal{T}_i(p_i)h_i \right\rangle_{\mathscr{X}}, \tag{15c}$$

for $i = 1,\ldots,r$.

We start our exposition with the discretization with respect to time. To this end, consider a time grid $0 = t_0 < t_1 < \ldots < t_m = T$ and associated quadrature

rule defined by weights $\omega_\ell \geq 0$ for $\ell = 0, \ldots, m$. The approximation of the directional derivative of J with respect to $\boldsymbol{\alpha}$ is thus given by

$$\begin{aligned}\partial_{\boldsymbol{\alpha},\boldsymbol{d}} J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) &= \sum_{i=1}^{r} \int_0^T v_i(\boldsymbol{\alpha}(t), \boldsymbol{p}(t), \boldsymbol{\varphi}) d_i(t) \,\mathrm{d}t \\ &\approx \sum_{i=1}^{r} \sum_{k=0}^{m} w_k v_i(\boldsymbol{\alpha}(t_k), \boldsymbol{p}(t_k), \boldsymbol{\varphi}) d_i(t_k) \\ &= \left(\boldsymbol{v}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})\right)^\top (I_r \otimes W)\, \boldsymbol{d}^m,\end{aligned}$$

where $\otimes$ is the Kronecker product and

$$\begin{aligned} W &:= \operatorname{diag}(w_0, \ldots, w_m) \in \mathbb{R}^{(m+1)\times(m+1)}, \\ \boldsymbol{v}_i^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) &:= \left[v_i(\boldsymbol{\alpha}(t_0), \boldsymbol{p}(t_0), \boldsymbol{\varphi}) \cdots v_i(\boldsymbol{\alpha}(t_m), \boldsymbol{p}(t_m), \boldsymbol{\varphi})\right]^\top \in \mathbb{R}^{m+1}, \\ \boldsymbol{v}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) &:= \left[\boldsymbol{v}_1^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})^\top \cdots \boldsymbol{v}_r^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})^\top\right]^\top \in \mathbb{R}^{r(m+1)}, \\ \boldsymbol{d}_i^m &:= \left[d_i(t_0) \cdots d_i(t_m)\right]^\top \in \mathbb{R}^{m+1}, \\ \boldsymbol{d}^m &:= \left[(\boldsymbol{d}_1^m)^\top \cdots (\boldsymbol{d}_r^m)^\top\right]^\top \in \mathbb{R}^{r(m+1)}. \end{aligned}$$

The time-discrete approximation of the partial derivative is thus given by

$$\partial_{\boldsymbol{\alpha}} J^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \left(\boldsymbol{v}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})\right)^\top (I_r \otimes W) \in \mathbb{R}^{1\times r(m+1)}.$$

Analogously, the time-discrete approximation of the partial derivative of J with respect to the path variables, i.e., $\partial_{\boldsymbol{p}} J(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})$, is given by

$$\partial_{\boldsymbol{p}} J^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \left(\boldsymbol{\xi}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})\right)^\top (I_r \otimes W) \in \mathbb{R}^{1\times r(m+1)},$$

with $\boldsymbol{\xi}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})$ defined analogously as $\boldsymbol{v}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi})$. In the same fashion, we obtain the time-discrete approximation for the directional derivative with respect to the mode variables as

$$\partial_{\boldsymbol{\varphi},\boldsymbol{h}} J^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \left(\boldsymbol{\mu}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}, \boldsymbol{h})\right)^\top (I_r \otimes W)\, \mathbf{1}_{r(m+1)} \in \mathbb{R},$$

where we denote by $\mathbf{1}_{r(m+1)} \in \mathbb{R}^{r(m+1)}$ the vector with all entries equal to 1, and $\boldsymbol{\mu}^m$ defined analogously as $\boldsymbol{v}^m$.

For the spatial discretization, let $\mathscr{Y}_n$ denote an n-dimensional subspace of $\mathscr{Y}$ with basis functions $\psi_1, \ldots, \psi_n \in \mathscr{Y}$. Let us define for $i, j = 1, \ldots, r$ and $p_i, p_j \in \mathbb{R}$ the matrices $M_{i,j}, N_{i,j}, F_i, G_i \in \mathbb{R}^{n\times n}$ via

$$[M_{i,j}(p_i, p_j)]_{k,\ell} := \langle \mathcal{T}_j(p_j)\psi_k, \mathcal{T}_i(p_i)\psi_\ell \rangle_{\mathscr{X}}, \tag{16a}$$

$$[N_{i,j}(p_i, p_j)]_{k,\ell} := \langle \mathcal{T}_j(p_j)\psi_k, \tfrac{\partial}{\partial p_i}\mathcal{T}_i(p_i)\psi_\ell \rangle_{\mathscr{X}}, \tag{16b}$$

$$[F_i(p_i)]_{k,\ell} := \langle \psi_k, \mathcal{T}_i(p_i)\psi_\ell \rangle_{\mathscr{X}}, \tag{16c}$$

$$[G_i(p_i)]_{k,\ell} := \langle \psi_k, \tfrac{\partial}{\partial p_i}\mathcal{T}_i(p_i)\psi_\ell \rangle_{\mathscr{X}}, \tag{16d}$$

for $k, \ell = 1, \dots, r$. For the data $z \in L^2(0,T;\mathscr{Y})$, the modes $\varphi_i \in \mathscr{Y}$, and directions $h_i \in \mathscr{Y}$, we consider the approximations

$$z(t) \approx \sum_{\ell=1}^{n} \widehat{z}_\ell(t)\psi_\ell, \qquad \varphi_i \approx \sum_{\ell=1}^{n} \widehat{\varphi}_{i,\ell}\psi_\ell, \qquad h_i \approx \sum_{\ell=1}^{n} \widehat{h}_{i,\ell}\psi_\ell,$$

with

$$\widehat{\boldsymbol{z}}(t) := [\widehat{z}_1(t) \ \cdots \ \widehat{z}_n(t)]^\top \in \mathbb{R}^n, \qquad \widehat{\boldsymbol{\varphi}}_i := [\widehat{\varphi}_{i,1} \ \cdots \ \widehat{\varphi}_{i,n}]^\top \in \mathbb{R}^n,$$
$$\widehat{\boldsymbol{h}}_i := [\widehat{h}_{i,1} \ \cdots \ \widehat{h}_{i,n}]^\top \in \mathbb{R}^n.$$

With these preparations, we obtain the spatial approximation of the inner products in (15) as

$$\widehat{v}_i(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \left(\sum_{j=1}^{r} \alpha_j \widehat{\boldsymbol{\varphi}}_j^\top M_{i,j}(p_i, p_j) - \widehat{\boldsymbol{z}}^\top F_i(p_i) \right) \widehat{\boldsymbol{\varphi}}_i \in \mathbb{R},$$

$$\widehat{\xi}_i(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \alpha_i \left(\sum_{j=1}^{r} \widehat{\boldsymbol{\varphi}}_j^\top N_{i,j}(p_i, p_j) - \widehat{\boldsymbol{z}}^\top G_i(p_i) \right) \widehat{\boldsymbol{\varphi}}_i \in \mathbb{R},$$

$$\widehat{\mu}_i(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \alpha_i \left(\sum_{j=1}^{r} \alpha_j \widehat{\boldsymbol{\varphi}}_j^\top M_{i,j}(p_i, p_j) - \widehat{\boldsymbol{z}}^\top F_i(p_i) \right) \in \mathbb{R}^{1\times n}.$$

We thus obtain the space- and time-discretized partial derivatives as

$$\partial_{\boldsymbol{\alpha}} \widehat{J}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := (\widehat{\boldsymbol{v}}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}))^\top (I_r \otimes W) \in \mathbb{R}^{1\times(m+1)r},$$
$$\partial_{\boldsymbol{p}} \widehat{J}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := (\widehat{\boldsymbol{\xi}}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}))^\top (I_r \otimes W) \in \mathbb{R}^{1\times(m+1)r},$$
$$\partial_{\boldsymbol{\varphi}} \widehat{J}^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}) := \mathbf{1}_{r(m+1)}^\top (I_r \otimes W)\, \mathcal{M} \in \mathbb{R}^{1\times nr},$$

with $\mathcal{M} := \operatorname{blkdiag}(\widehat{\mu}_1^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}), \dots, \widehat{\mu}_r^m(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\varphi}))$.

We conclude this section with a specific computation of the quantities depending on the inner products for the periodic shift operator and P_1 finite elements.

Example 3. Let us assume we have a one-dimensional domain $\Omega = (0,1)$ and a corresponding equidistant grid of step size $h := \frac{1}{n}$. We discretize $\mathscr{Y} = H^1_{\mathrm{per}}(\Omega)$ via periodic P_1 finite element functions. For $\mathscr{X} = L^2(0,1)$ and shift operator with periodic embedding, we observe

$$M_{i,j}(p_i, p_j) = F_i(p_i - p_j) \qquad \text{and} \qquad N_{i,j}(p_i, p_j) = G_i(p_i - p_j).$$

For $p_i = qh + \tilde{p}_i$ with $q \in \mathbb{Z}$ and $\tilde{p}_i \in [0, h)$ we obtain

$$\langle \psi_k, \mathcal{T}_i(p_i)\psi_\ell \rangle = \begin{cases} \frac{1}{h^2}\left(\frac{2}{3}(h-\tilde{p}_i)^3 + \tilde{p}_i(h-\tilde{p}_i)^2 + \tilde{p}_i h(h-\tilde{p}_i) + \frac{1}{6}\tilde{p}_i^3\right), & \text{if } \ell = k-q, \\ \frac{1}{6h^2}(h-\tilde{p}_i)^3, & \text{if } \ell = k-q+1, \\ \frac{1}{h^2}\left(\frac{1}{6}(h-\tilde{p}_i)^3 - \frac{1}{3}\tilde{p}_i^3 + h^2\tilde{p}_i\right), & \text{if } \ell = k-q-1, \\ \frac{1}{6h^2}\tilde{p}_i^3, & \text{if } \ell = k-q-2, \\ 0, & \text{otherwise.} \end{cases}$$

For further details, including the computation of $\langle \psi_k, \frac{\partial}{\partial p_i} \mathcal{T}_i(p_i)\psi_\ell \rangle_{\mathscr{X}}$, we refer to the preprint version of this manuscript [4].

5 Numerical Examples

For our numerical examples, we use an equidistant time grid $t_i := i\tau$ with step size $\tau > 0$. The weights for the time integration are chosen based on the trapezoidal rule. For the discretization in space, we also use an equidistant grid and follow Example 3 for approximating the inner products occurring in the cost functional and the gradient. In particular, we use the periodic shift operator for all numerical examples. The optimization itself is carried out with the MATLAB® package GRANSO with default settings, see [8], which is based on a quasi-Newton solver.

For notational convenience, we assumed so far that there is exactly one mode per transformation operator. In practice, it is often more reasonable to cluster the modes into different reference frames, see, for instance, [3, sec. 7.1]. Thus, we use the clustered approximation ansatz

$$z \approx \sum_{\rho=1}^{q} \sum_{i=1}^{r_\rho} \alpha_{\rho,i} \mathcal{T}_\rho\left(p_\rho\right) \varphi_{\rho,i} \tag{18}$$

for the following numerical experiments and emphasize that this only requires a minor and straightforward modification of the gradient. We denote the approximation based on our optimization results with tPOD. Furthermore, we use dashed lines in the plots to display the (optimized) path variables. Finally, note that even though the data for the numerical examples stems from partial differential equations, our main concern in this work is the approximation and compression of any given data. We thus forego a thorough treatment of the actual values in the following pseudocolor plots.

To ensure reproducibility of the conducted experiments, the code for the numerical examples is publicly available under https://doi.org/10.5281/zenodo.5471404.

5.1 Viscous Burgers' Equation

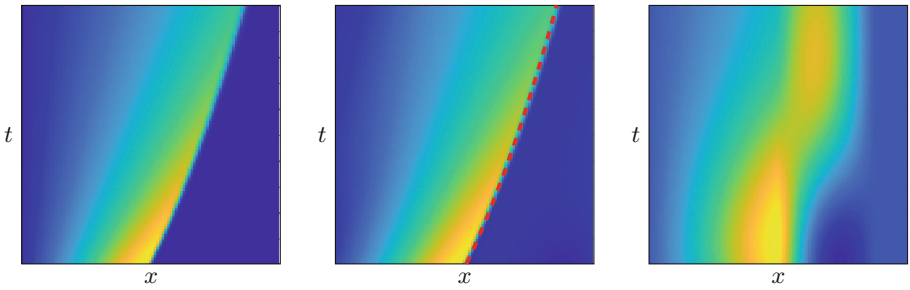
We consider the one-dimensional viscous Burgers' equation

$$\frac{\partial}{\partial t} z(t,x) = \frac{1}{\mathrm{Re}} \frac{\partial^2}{\partial x^2} z(t,x) - z(t,x) \frac{\partial}{\partial x} z(t,x), \qquad (t,x) \in (0,2) \times (0,1), \tag{19}$$

and, following [19], use the analytical solution

$$z(t,x) = \frac{x}{t+1} \left(1 + \sqrt{\frac{t+1}{\exp\left(\frac{\mathrm{Re}}{8}\right)}} \exp\left(\mathrm{Re} \frac{x^2}{4t+4} \right) \right)^{-1} \tag{20}$$

with Reynolds number $\mathrm{Re} = 1000$ for our experiment. Let us emphasize that (20) is not periodic in x. Nevertheless, the shock front (as depicted in Fig. 1a) stays



(a) snapshot data (b) tPOD $r = 2$ (c) POD $r = 2$

Fig. 1. Burgers' equation – comparison of original data (a), tPOD approximation (b), and POD approximation (c). The red dashed line in (b) denotes the optimized path variable.

within the computational domain and the solution is approximately zero at the boundaries, allowing us to treat this problem as quasi-periodic. We test our algorithm with data obtained from the analytical solution, evaluated on a grid with 100 equidistant intervals in space and time, respectively. We use a single frame, i.e., $q = 1$ in (18), and $r_1 = r = 2$ modes, and supply the first snapshot and the zero vector as starting values for the modes. The corresponding coefficients are initialized as a constant function with value 1. For the path, we start with a straight line given by $p(t) = \frac{37}{200}t$. The results are depicted in Fig. 1, detailing that already with $r = 2$, an accurate approximation with a relative L^2 error of less than 3% can be achieved, while the POD approximation is not able to reproduce the shock front. Besides, we compare the relative L^2 errors of the approximations for different mode numbers in Table 1, detailing the superior approximation capabilities of tPOD for this test case.

Table 1. Burgers' equation – comparison of relative L^2 errors

r	tPOD	POD
1	1.224×10^{-1}	4.510×10^{-1}
2	2.910×10^{-2}	2.863×10^{-1}
3	1.328×10^{-2}	2.111×10^{-1}
4	8.453×10^{-3}	1.662×10^{-1}
5	6.821×10^{-3}	1.354×10^{-1}

5.2 Nonlinear Schrödinger Equation

In this section, we consider the nonlinear Schrödinger equation

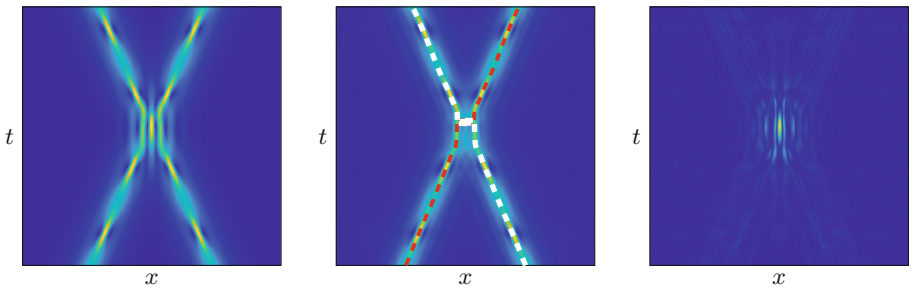
$$\begin{aligned} \mathrm{i}\tfrac{\partial}{\partial t} z(t,x) &= -\tfrac{1}{2}\tfrac{\partial^2}{\partial x^2} z(t,x) + \kappa |z(t,x)|^2 z(t,x), \\ z(t,0) &= 2\mathrm{sech}(x+7)\exp(2\mathrm{i}x) + 2\mathrm{sech}(x-7)\exp(-2\mathrm{i}x), \end{aligned}$$

as presented in [20, Example 4]. We compute a solution using the code from [20] on a uniform grid of 501×1024 points on the domain $\mathbb{T} \times \Omega = [0, 2\pi] \times [-15, 15]$. The absolute value of the numerical solution is presented in Fig. 2a.

We initialize our algorithm by assuming an approximation with two frames, each with two modes. As starting value for the modes, we use

$$\varphi_{\rho,1}(x) = 2\mathrm{sech}(x - (-1)^\rho 7)\exp(-(-1)^\rho 2\mathrm{i}x) \qquad \text{and} \qquad \varphi_{\rho,2} \equiv 0$$

for $\rho = 1, 2$. The coefficients are initialized as constants, with value 1 at each time point. For the initial paths, we use $p_1(t) = 2t$ and $p_2(t) = -2t$. The corresponding approximation and the absolute error are presented in Figs. 2b and c. We observe that the error results mainly from the complicated wave dynamics in the middle of the spatial and time domain, whereas the transported wave profiles are captured accurately. Let us emphasize that the error is very localized such that it can be captured with only a few additional POD modes. We notice that the optimizer does not keep the linear path over the whole time domain. Instead, as



(a) snapshot data (b) tPOD $r = 4$ (c) absolute error

Fig. 2. Nonlinear Schrödinger equation – original data (a), tPOD approximation with $r = 4$ total modes (b), and absolute error (c) for initial paths $p_1(t) = 2t$ and $p_2(t) = -2t$. The red and white dashed lines in (b) denote the optimization results for the first and second path variable, respectively.

depicted in Fig. 2b, in the middle of the computational domain, the paths jump between the wavefronts. Inspecting the snapshot matrix in the co-moving frame along the path $p_1(t) = 2t$ in Fig. 3 provides a possible explanation: the vertical wavefront features an offset after the two waves have crossed. The optimizer needs to account for this offset, which explains the jump. Let us emphasize that with a different initialization for the path variables, the optimizer finds another local minimum with a similar approximation quality. Using piecewise linear paths as depicted in Fig. 4a, we observe that the resulting optimized path smoothes out the edges of the initial path in the middle of the domain (cf. Fig. 4c) and

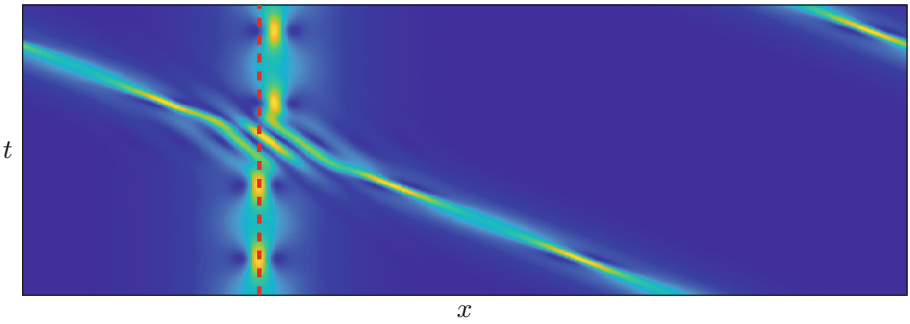


Fig. 3. Nonlinear Schrödinger equation – transformation of Fig. 2a to the co-moving frame along the path $p_1(t) = 2t$, i.e., applying the shift operator along the path $-p_1(t)$ to the original snapshot data in Fig. 2a, such that the wave front originally traveling to the right becomes stationary. For illustration purposes, the image is stretched to clearly present the offset in the vertical wave profile after the two waves have crossed.

does not feature any jumps. It is smooth and tracks the wavefronts as if the waves reflect off each other.

5.3 FitzHugh–Nagumo Wave Train

We follow [13] and consider the FitzHugh–Nagumo model given by

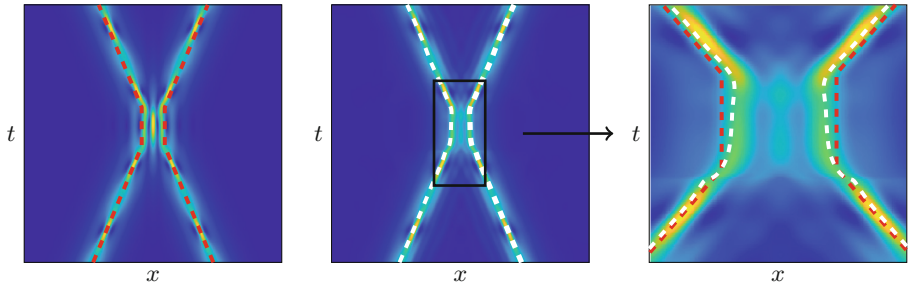
$$\begin{aligned} \tfrac{\partial}{\partial t} u_1(t,x) &= \nu \tfrac{\partial^2}{\partial x^2} u_1(t,x) - u_2(t,x) + u_1(t,x)(1-u_1(t,x))(u_1(t,x)-a), \\ \tfrac{\partial}{\partial t} u_2(t,x) &= \epsilon(b u_1(t,x) - u_2(t,x)), \end{aligned} \tag{21}$$

with spatial domain $(0, 500)$ and time interval $(0, 1000)$. The partial differential equation (21) is closed by periodic boundary conditions and the initial condition

$$u_1(0,x) = \tfrac{1}{2}\left(1 + \sin\left(\tfrac{\pi}{50}x\right)\right), \quad u_2(0,x) = \tfrac{1}{2}\left(1 + \cos\left(\tfrac{\pi}{50}x\right)\right).$$

For the parameter values, we choose $\nu = 1$, $a = -0.1$, $\epsilon = 0.05$, and $b = 0.3$. The spatial discretization of (21) is performed via a central sixth-order finite-difference scheme with mesh width $h = 0.5$ and for the time integration we use MATLAB®'s `ode45` function based on a time grid with step size 1. The corresponding numerical solution for the variable u_1 is depicted in Fig. 5a.

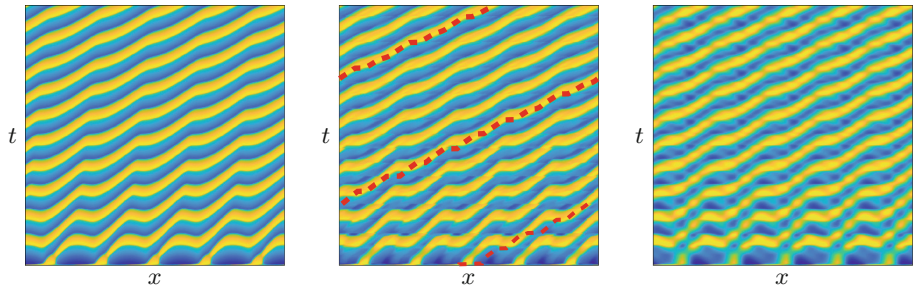
For the optimization we consider only the data of the variable u_1 and use an approximation with one reference frame to account for the traveling wave train. Furthermore, we reduce the computational complexity by considering the optimization problem only in terms of the path, whereas the coefficients and modes are computed in each iteration via a truncated singular value decomposition of the snapshot matrix shifted into the co-moving reference frame. Here we exploit that the periodic shift operator is isometric, such that we can solve the optimization problem via classical POD with transformed data, cf. [3, Thm. 4.8]. As starting value for the path, we choose a linear function in t with slope



(a) initial paths (b) tPOD $r = 4$ (c) zoom: initial vs. optimized path

Fig. 4. Nonlinear Schrödinger equation – comparison of original data (a) with piecewise linear initialization of the path variables (red dashed lines) and tPOD approximation (b) with $r = 4$ total modes and optimized path variables (white dashed lines). Figure 4c shows a zoom into Fig. 4b and illustrates that the optimized path variables (white dashed lines) differ from the initial path variables (red dashed lines).

1.04, which we determined by inspecting the first and the last snapshot of the original data. The corresponding approximation obtained from the optimization procedure is depicted in Fig. 5b. As reference approximation, we consider a POD



(a) snapshot data (b) tPOD $r = 4$ (c) POD $r = 4$

Fig. 5. FitzHugh–Nagumo model – comparison of original data for u_1 (a), tPOD approximation with $r = 4$ modes (b), and POD approximation with $r = 4$ modes (c). The red dashed line in (b) denotes the optimized path variable.

approximation with the same number of modes in Fig. 5c. The corresponding total relative errors are 15% for the approximation based on shifted modes and 31% for the POD approximation. We note that in contrast to the Burgers test case considered in Sect. 5.1, the traveling wave train can be better approximated by POD due to the lack of a traveling shock wave. Correspondingly, the difference between tPOD and POD approximation for the considered FitzHugh–Nagumo test case is less striking than the one observed for the example in Sect. 5.1.

6 Summary

In this paper, we analyze the problem of determining an optimal approximation of given snapshot data by a linear combination of dynamically transformed modes. This data compression can, for instance, be used for model order reduction of transport-dominated systems [3]. As optimization parameters, we consider the modes, the corresponding coefficients or amplitudes, and the so-called path variables, which parameterize the coordinate transforms applied to the modes. We first show that the considered infinite-dimensional optimization problem possesses a minimizing solution if the admissible set is constrained such that the optimization parameters are norm bounded. Afterward, we derive a corresponding unconstrained optimization problem by adding an appropriate penalization term and show that the unconstrained problem also has a solution. Furthermore, we demonstrate that if the penalization coefficient tends to infinity, each limit point of the corresponding sequence of minimizers is a solution to the original constrained optimization problem. To derive a gradient-based optimization procedure, we compute the partial Fréchet derivatives of the unconstrained cost functional and discuss their space and time discretization. Finally, we apply the optimization procedure to some numerical test cases and observe that the optimized decompositions are significantly more accurate than corresponding approximations obtained by the classical POD with the same number of modes.

After full discretization, the optimization problem still features a large number of optimization parameters scaling with the number of grid points in space and time. Thus, an interesting future research direction is to investigate approaches for reducing the computational complexity of the optimization procedure, for instance, by using multigrid optimization techniques [21], or by making use of low-dimensional parametrizations of the optimization parameters, cf. Remark 2. Such a parametrization seems to be especially promising for reducing the complexity in the path variables since our numerical experiments revealed that the optimization procedure is sensitive with respect to the paths. Furthermore, let us emphasize that, although we have only discussed applications in a one-dimensional spatial domain with a periodic shift operator, our framework is not restricted to this case. Thus, another promising direction for the future is to explore the applicability to problems with higher-dimensional spatial domains using different transformation operators, see e.g. [15,28,31] for some contributions in this direction.

Acknowledgments. The work of F. Black is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Collaborative Research Center (CRC) 1029 *Substantial efficiency increase in gas turbines through direct use of coupled unsteady combustion and flow dynamics*, project number 200291049. P. Schulze acknowledges funding by the DFG CRC Transregio 154 *Mathematical Modelling, Simulation and Optimization Using the Example of Gas Networks*, project number 239904186. B. Unger acknowledges funding from the DFG under Germany's Excellence Strategy – EXC 2075 – 390740016 and is thankful for support by the Stuttgart Center for Simulation Science (SimTech).

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces, 2nd edn. Elsevier, Oxford (2003)
2. Beyn, W.J., Thümmler, V.: Freezing solutions of equivariant evolution equations. SIAM J. Appl. Dyn. Syst. **3**(2), 85–116 (2004)
3. Black, F., Schulze, P., Unger, B.: Projection-based model reduction with dynamically transformed modes. ESAIM Math. Model. Numer. Anal. **54**(6), 2011–2043 (2020)
4. Black, F., Schulze, P., Unger, B.: Decomposition of flow data via gradient-based transport optimization. ArXiv preprint 2107.03481, ArXiv (2021)
5. Black, F., Schulze, P., Unger, B.: Efficient wildland fire simulation via nonlinear model order reduction. Fluids **6**(8), 280 (2021)
6. Black, F., Schulze, P., Unger, B.: Model order reduction with dynamically transformed modes for the wave equation. PAMM **20**(1), e202000321 (2021)
7. Cagniart, N., Maday, Y., Stamm, B.: Model order reduction for problems with large convection effects. In: Chetverushkin, B.N., Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmäki, P., Periaux, J., Pironneau, O. (eds.) Contributions to Partial Differential Equations and Applications, pp. 131–150. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-78325-3_10
8. Curtis, F.E., Mitchell, T., Overton, M.L.: A BFGS-SQP method for nonsmooth, nonconvex, constrained optimization and its evaluation using relative minimization profiles. Optim. Methods Softw. **32**(1), 148–181 (2017)
9. Engel, K.J., Nagel, R.: One-Parameter Semigroups for Linear Evolution Equations. Springer, New York (2000). https://doi.org/10.1007/b97696
10. Greif, C., Urban, K.: Decay of the Kolmogorov n-width for wave problems. Appl. Math. Lett. **96**, 216–222 (2019)
11. Gubisch, M., Volkwein, S.: Chapter 1: Proper orthogonal decomposition for linear-quadratic optimal control. In: Benner, P., Cohen, A., Ohlberger, M., Willcox, K. (eds.) Model Reduction and Approximation, pp. 3–63. SIAM, Philadelphia (2017)
12. Iollo, A., Lombardi, D.: Advection modes by optimal mass transfer. Phys. Rev. E **89**(2), 022923 (2014)
13. Koch, J.: Data-driven modeling of nonlinear traveling waves. Chaos **31**(4), 043128 (2021)
14. Kolmogoroff, A.: Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. Ann. Math. **37**(1), 107–110 (1936)
15. Krah, P., Sroka, M., Reiss, J.: Model order reduction of combustion processes with complex front dynamics. In: Vermolen, F.J., Vuik, C. (eds.) Numerical Mathematics and Advanced Applications ENUMATH 2019, pp. 803–811. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-55874-1_79
16. Lee, K., Carlberg, K.T.: Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. J. Comput. Phys. **404**, 108973 (2020)
17. Luenberger, D.G., Ye, Y.: Linear and Nonlinear Programming. ISORMS, vol. 116. Springer, New York (2008). https://doi.org/10.1007/978-0-387-74503-9_15
18. Maday, Y., Patera, A.T., Turinici, G.: A priori convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations. J. Sci. Comput. **17**(1), 437–446 (2002)
19. Mauli, R., Lusch, B., Balaprakash, P.: Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. Phys. Fluids **33**, 037106 (2021)

20. Mendible, A., Brunton, S.L., Aravkin, A.Y., Lowrie, W., Kutz, J.N.: Dimensionality reduction and reduced-order modeling for traveling wave physics. Theoret. Comput. Fluid Dyn. **34**(4), 385–400 (2020). https://doi.org/10.1007/s00162-020-00529-9
21. Nash, S.G.: A multigrid approach to discretized optimization problems. Optim. Methods Softw. **14**(1–2), 99–116 (2000)
22. Nonino, M., Ballarin, F., Rozza, G., Maday, Y.: Overcoming slowly decaying Kolmogorov n-width by transport maps: application to model order reduction of fluid dynamics and fluid structure interaction problems. ArXiv preprint 1911.06598 (2019)
23. Ohlberger, M., Rave, S.: Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing. C. R. Acad. Sci. Paris **351**(23–24), 901–906 (2013)
24. Peherstorfer, B.: Model reduction for transport-dominated problems via online adaptive bases and adaptive sampling. SIAM J. Sci. Comput. **42**(5), A2803–A2836 (2020)
25. Reiss, J.: Optimization-based modal decomposition for systems with multiple transports. SIAM J. Sci. Comput. **43**(3), A2079–A2101 (2021)
26. Reiss, J., Schulze, P., Sesterhenn, J., Mehrmann, V.: The shifted proper orthogonal decomposition: a mode decomposition for multiple transport phenomena. SIAM J. Sci. Comput. **40**(3), A1322–A1344 (2018)
27. Rim, D., Moe, S., LeVeque, R.J.: Transport reversal for model reduction of hyperbolic partial differential equations. SIAM/ASA J. Uncertain. Quantif. **6**(1), 118–150 (2018)
28. Rim, D., Peherstorfer, B., Mandli, K.T.: Manifold approximations via transported subspaces: model reduction for transport-dominated problems. ArXiv preprint 1912.13024v2 (2020)
29. Rudin, W.: Real and Complex Analysis. ISORMS, Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-2886-2_4
30. Schulze, P., Reiss, J., Mehrmann, V.: Model reduction for a pulsed detonation combuster via shifted proper orthogonal decomposition. In: King, R. (ed.) Active Flow and Combustion Control 2018, pp. 271–286. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-98177-2_17
31. Taddei, T.: A registration method for model order reduction: data compression and geometry reduction. SIAM J. Sci. Comput. **42**(2), A997–A1027 (2020)
32. Taddei, T., Perotto, S., Quarteroni, A.: Reduced basis techniques for nonlinear conservation laws. ESAIM Math. Model. Numer. Anal. **49**(3), 787–814 (2015)
33. Unger, B., Gugercin, S.: Kolmogorov n-widths for linear dynamical systems. Adv. Comput. Math. **45**(5–6), 2273–2286 (2019)
34. Welper, G.: Transformed snapshot interpolation with high resolution transforms. SIAM J. Sci. Comput. **42**(4), A2037–A2061 (2020)
35. Zeidler, E.: Nonlinear Functional Analysis and its Applications IIA: Linear Monotone Operators. Springer, New York (1990). https://doi.org/10.1007/978-1-4612-0985-0