

## TRANSFORMED SNAPSHOT INTERPOLATION WITH HIGH RESOLUTION TRANSFORMS\*

G. WELPER†

**Abstract.** In the last few years, several methods have been developed to deal with jump singularities in parametric or stochastic hyperbolic PDEs. They typically use some alignment of the jump-sets in physical space before performing well-established reduced order modeling techniques such as reduced basis methods, proper orthogonal decomposition, or simply interpolation. In the current literature, the transforms are typically of low resolution in space, mostly low order polynomials, Fourier modes, or constant shifts. In this paper, we discuss higher resolution transforms in one of the recent methods, the transformed snapshot interpolation. We introduce a new discretization of the transforms with an appropriate behavior near singularities and consider their numerical computation via an optimization procedure.

**Key words.** parametric PDEs, shocks, transformations, interpolation, convergence rates, optimization

**AMS subject classifications.** 41A46, 41A25, 35L67, 65M12

**DOI.** 10.1137/19M126356X

**1. Introduction.** One important ingredient in reduced order modeling and PDEs with random coefficients is the numerical approximation of parametric functions  $u(x, \mu)$  with physical variables  $x \in \Omega \subset \mathbb{R}^d$  and deterministic or random parameters  $\mu$  in some parameter space  $\mathcal{P} \subset \mathbb{R}^r$ . The literature provides a rich toolbox for this kind of approximation; see, e.g., [11, 4, 38, 22] for an overview. However, if  $u(x, \mu)$  has parameter dependent jumps or kinks, the vast majority of the available methods suffer from low convergence rates [12, 32, 49]. As a result, problems from parametric hyperbolic PDEs, elliptic PDEs with jumping diffusion (in parameter dependent locations), and parametric level-set methods still pose severe challenges.

Besides a large body of literature on stability, offline/online decompositions, and error estimators for hyperbolic or singularly perturbed problems [9, 21, 20, 30, 35, 37, 46, 27, 16, 50, 36, 33, 15, 14, 2, 24, 5], in the last few years, several groups have addressed the poor regularity of  $u(x, \mu)$  and developed several methods with drastically improved convergence behavior. While some use localization strategies [12, 34], currently the majority relies on additional transforms, shifts, or similar tools in order to alleviate the issue. All these approaches have in common that they try to align the discontinuities or sharp gradients in physical space before applying classical reduced order modeling techniques such as interpolation, proper orthogonal decomposition (POD), or reduced bases. This alignment drastically improves the smoothness [26], Kolmogorov  $n$ -width, or decay rate of the singular values, respectively, which results in considerable efficiency gains in the presence of jumps. Some approaches use characteristics [45], Lagrangian formulations [28], extra equations or conditions [43, 18, 31, 19], or jump tracking [44]. Another approach is to use displacement interpolation or optimal transport [23, 40, 41], which naturally builds on transport along

\*Submitted to the journal's Methods and Algorithms for Scientific Computing section May 23, 2019; accepted for publication (in revised form) April 17, 2020; published electronically July 6, 2020.  
<https://doi.org/10.1137/19M126356X>

**Funding:** This work was supported by the NSF through grant 1912703.

†Department of Mathematics, University of Central Florida, Orlando, FL 32816 (gerrit.welper@ucf.edu).

the physical variables. One of the more common recent approaches uses a composition of a snapshot  $u(x-s, \mu)$  with a shift  $s$  or alternatively a transform  $u(X(x), \mu)$  in order to align the jump locations in parameter, where  $s$  or  $X$  usually depend on the parameter and may or may not depend on the location  $x$ . These shifted or transformed snapshots are then used for interpolation or reduced bases or are further compressed by POD; see, e.g., [49, 39, 8, 7, 29, 48]. Alternatively, in [42], instead of a function composition, the authors shift the coefficients of a discretization.

For the latter class of methods, one needs numerical procedures to find the shifts  $s$  or transforms  $X$ . Since the shifts or transforms generally depend on  $x$  and  $\mu$ , one first chooses a suitable discretization. Then, one minimizes an error formula among the chosen discrete representation during the “offline phase,” either between individual (transformed) snapshots or the full reconstruction error of the method. In the bulk of the current literature the discretizations are of low order in physical space, either constants, low order polynomials, or Fourier modes. One notable exception is optimal transport based methods [23, 40, 41], where the construction of transforms is better understood. Clearly, these limitations on the transform’s resolutions pose severe limitations on their applicability to many practical problems. As we will see in section 3, already quite simple scenarios require transforms with high resolution in physical space.

In the first part of this paper, we discuss possible discretizations of high resolution transforms, which in the most general case are functions  $X(x, \mu)$  depending on the physical variables  $x$  and an arbitrary parameter  $\mu$ . As we will discuss in section 3, these functions share some major properties with  $u$  itself: they may require high resolution in the spatial variables and can show singular behavior in parameter. However, we cannot use the same discretization strategy as for  $u$  without introducing a second layer of transforms. Instead, we will see that these transforms are naturally given as solutions of ODEs and that variants of implicit solvers thereof provide efficient ways for their discretization. Similar to stability analysis of implicit ODE solvers, the resulting transforms include rational functions in the parameter, which are well suited to handle typical singularities. Note that even for hyperbolic problems where characteristics are available, these transform ODEs are typically not characteristics. Whereas the characteristic curves usually lead into shocks, we will see that the transform ODEs ideally move parallel to the shocks.

In principle, one can then use the same optimization strategies as for low resolution transforms to compute them in the offline phase, but as we will see in some experiments below, this results in poor outcomes. Typically, naive implementations of gradient descent optimizers stall after a few steps and the resulting transform is visually indistinguishable from the initial transform. In order to overcome this problem, in section 6, we discuss that the transforms  $X$  are functions of  $x$  and must be considered in appropriate function spaces. This entails that the Fréchet derivative of the objective function to optimize the transform is only a functional in a dual space and must be appropriately “lifted” by a Riesz map to obtain a well-defined gradient descent method. In this paper, we report on two lifting methods, one by inverting a Laplacian and one by using a multilevel frame.

The paper is organized as follows. To keep it self-contained, in section 2, we briefly review the transformed snapshot interpolation (TSI), which is one of the recently introduced transformed based reduced order modeling techniques for problems with jumps. Then, in section 3, we consider an example problem to better understand the requirements for high resolution transforms. In section 4 we discuss the discretization of the transforms and in section 5 consider their stability. In section 6, we introduce

suitable optimizers to practically find the transforms, and finally in section 7 we provide some numerical experiments.

**2. Transformed snapshot interpolation.** In order to keep the paper self-contained, let us first briefly review the TSI introduced in [49]. It approximates a parametric function  $u(x, \mu)$  with physical variables  $x \in \Omega \subset \mathbb{R}^d$  and deterministic or random parameters  $\mu \in \mathcal{P} \subset \mathbb{R}^r$  by

$$(2.1) \quad u(x, \mu) \approx u_m(x, \mu) := u_m(x, \mu; X) := \sum_{\eta \in \mathcal{P}_m} \ell_\eta(\mu) u(X(\eta; \mu, x), \eta),$$

where  $\ell_\eta(\mu)$  are Lagrange basis polynomials with respect to given interpolation points  $\eta \in \mathcal{P}_m \subset \mathcal{P}$ . Throughout this article, we confine ourselves to one-dimensional parameter domains  $r = 1$ ; however, the results can be extended to higher dimensions with the tensor product type construction in [48]. Similar to reduced basis methods and PODs, this interpolation only requires us to know snapshots  $u(\cdot, \eta)$  at finitely many interpolation points in parameter. In addition we need the transforms  $\eta \rightarrow X(\eta; \mu, x) \subset \Omega$  which can be regarded as curves in physical space, together with an initial value  $X(\mu; \mu, x) = x$ . We often write  $X(\eta) = X(\eta; \mu, x)$  if  $\mu$  and  $x$  are understood from context.

As a notational convention, both  $\mu$  and  $\eta$  refer to parameters, where  $\mu$  usually denotes a target parameter where we want to approximate  $u(\cdot, \mu)$  and  $\eta$  is either a continuous auxiliary parameter in  $\mathcal{P}$  as in the curve  $\eta \rightarrow X(\eta)$  or a discrete “source” parameter in  $\mathcal{P}_m$  at locations where we know the snapshots, e.g., as in the definition of the TSI (2.1). The distinction between continuous or discrete should always be clear from context.

Let us now come back to the TSI (2.1). Choosing the trivial transform  $X(\eta; \mu, x) = x$ , the formula reduces to a standard polynomial interpolation. However, this simple approach does not achieve good convergence rates in case  $u(x, \mu)$  has jumps or kinks in parameter dependent locations, because these entail that  $u$  is not even differentiable in the interpolation direction. The extra transform  $X(\eta; \mu, x)$  is used to align these jumps so that they become “invisible” to the interpolation. In other words, for a target parameter  $\mu$  and interpolation point  $\eta$ , the transform  $X(\eta; \mu, x)$  ensures that the *transformed snapshots*  $(x, \eta) \rightarrow u(X(\eta; \mu, x), \eta)$  have jumps in the same locations as the correct solution  $u(\cdot, \mu)$ , independent of  $\eta$ . In particular, for all  $x$ , which are not in the jump-set of  $u(\cdot, \mu)$ , the transformed snapshots are smooth in  $\eta$ . The TSI (2.1) is merely a polynomial interpolation of these transformed snapshots in  $\eta$  and thus provides high order accuracy.

Note that in the TSI (2.1) the transforms  $X(\eta; \mu, x)$  are evaluated only at finitely many interpolation points  $\eta \in \mathcal{P}_m$  and the continuous variables  $x \in \Omega$  and  $\mu \in \mathcal{P}$ . If the transforms are smooth in  $\mu$ , they can be efficiently discretized in  $\mu$  by interpolation,

$$(2.2) \quad X(\eta; \mu, x) = \sum_{\gamma \in \mathcal{P}_m} \ell_\gamma(\mu) X(\eta; \gamma, x),$$

called *low resolution transforms* in the following. This was the original choice in [49] together with low order polynomials for the remaining finitely many functions  $x \rightarrow X(\eta; \gamma, x)$ ,  $\eta, \gamma \in \mathcal{P}_m$ , with zero normal flux as a numerical substitute for the condition  $X(\eta; \mu, x) \in \Omega$ . As we will see, e.g., in section 3, neither the smoothness in  $\mu$  nor the choice of low order polynomials in  $x$  is warranted in some fairly simple

examples. Therefore, we will discuss alternative representations of the transforms in section 4.

Finally, similar to neural network training or greedy algorithms for reduced basis methods, the remaining transforms  $X(\eta; \gamma, x)$ ,  $\eta, \gamma \in \mathcal{P}_m$ , are trained by minimizing the worst case error

$$(2.3) \quad \sup_{\mu \in \mathcal{P}_T} \|u(x, \mu) - u_m(x, \mu; X)\|_{L_1(\Omega)}.$$

Since we do not know  $u(\cdot, \mu)$  for all  $\mu \in \mathcal{P}$ , we cannot calculate the full error and confine ourselves to a training sample  $\mathcal{P}_T \subset \mathcal{P}$ . Of course other  $L_p$ -norms in parameter are also possible. The  $L_1$ -norm for the physical variables is natural for hyperbolic problems and ensures that the objective function of this optimization problem is differentiable almost everywhere, given that  $u$  is of bounded variation; see [49] for more details.

**3. An example problem.** In this section, we first consider a simple example problem that highlights some typical difficulties for the low resolution transforms (2.2). Namely, let us consider the function

$$(3.1) \quad u(x, \mu) = \begin{cases} 1, & x \leq -(1 - \mu), \\ -1, & x \geq 1 - \mu, \\ 0 & \text{else} \end{cases}$$

for  $\mu < 1$  and

$$(3.2) \quad u(x, \mu) = \begin{cases} 1, & x \leq 0, \\ -1, & x > 0, \end{cases}$$

for  $\mu \geq 1$ , depicted in Figure 3.1. It serves as a prototype of problems for which the original TSI as introduced in section 2 is problematic: it starts out with two jumps for  $\mu < 1$  that collide into one single jump for  $\mu \geq 1$ . This poses a problem for the TSI because we cannot properly align a snapshot with a single jump with a snapshot with two jumps.

More formally, the TSI interpolates the transformed snapshots  $\eta \rightarrow u(X(\eta; \mu, x), \eta)$ , which should ideally be smooth in the  $\eta$  variable. Since the curves  $\eta \rightarrow X(\eta; \dots)$  themselves are smooth, this can be true only if  $X(\eta; \mu, x)$  is not contained

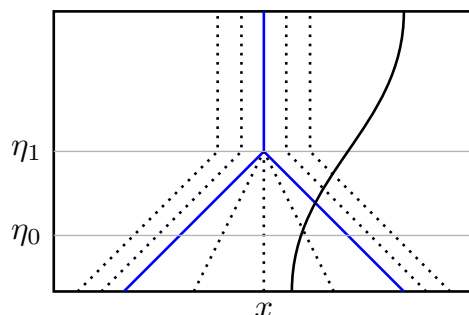


FIG. 3.1. Blue: Jump-set of the example function (3.1), (3.2) in the  $x - \mu$ -plane. Dotted lines: Curves  $\eta \rightarrow X(\eta; \mu, x)$  defined in (3.3). Black: Inappropriate transform  $\eta \rightarrow X(\eta; \mu, x)$  crossing a jump.

in the jump-set of  $u(\cdot, \eta)$  for all  $\eta \in \mathcal{P}$ . In Figure 3.1 this means that the black dotted transform curves should not intersect the blue jump locations, and as a result, the transform curves are ideally as parallel as possible to the jump curves. Obviously, this is impossible at the collision point or for any curve starting between the two jumps in the lower half of Figure 3.1, as indicated by the black solid line.

Away from the blue jump curves, we have some freedom in the choice of the dotted curves  $\eta \rightarrow X(\eta; \dots)$  and merely ensuring smoothness is sufficient for a good TSI reconstruction error (given that  $u$  is smooth away from the jumps). Indeed, if both  $u$  and  $X$  are locally smooth, the composition  $\eta \rightarrow u(X(\eta; \mu, x), \eta)$  is so as well, resulting in small (TSI) interpolation errors. There are many reasonable choices of transforms  $X$ , which satisfy the given conditions. A simple one is shown by the dotted lines and given by

$$(3.3) \quad X(\eta; \mu, x) = \begin{cases} x - (\mu - \eta), & x \leq -(1 - \mu), \\ x + (\mu - \eta), & x \geq 1 - \mu, \\ x - x \frac{\mu - \eta}{\mu - 1} & \text{else,} \end{cases}$$

although the observations below are also applicable to other choices, as well. In particular, if we confine our parameter range to  $\mu < 1$  or  $\mu \geq 1$ , these transforms never cross jumps, corresponding to the parameter ranges where  $u$  has either one or two jumps. Therefore, we may proceed as follows: we subdivide the parameter domain into two pieces for  $\mu \leq 1$  and apply a TSI on each of them. A simple version of this idea, which finds partitions of  $\mathcal{P}$  via  $h$  or  $hp$  adaption, is discussed in [48]. A more sophisticated algorithm that cuts the parameter domain precisely at the jump collision is in preparation. In this paper, we do not deal with the proper subdivision but with the necessarily singular transforms:

1. For  $\mu = 1$  and  $\eta < 1$ , the transform has a jump at  $x = 0$ .
2. The dependence of the  $\mu$  variable is of the type  $\sim 1/(\mu - 1)$  in the regions between the two jumps and therefore singular at  $\mu = 1$ .

Both problems cannot be handled with the low resolution transforms (2.2). These were low order polynomials in both the  $x$  and  $\mu$  variables. For the  $x$  variable one can simply increase the spatial resolution of the transforms to the same resolution of the snapshots. Since the latter must be stored anyway, the overall degrees of freedom do not blow up significantly. However, this higher resolution requires some extra care with regard to the optimization of the error (2.3), as discussed in section 6.

With regard to the singularity in the  $\mu$  variable, a simple increase of the resolution is more challenging without increasing the number of snapshots and therefore we discuss an alternative discretization of the transform in section 4.

Although the example discussed in this section is quite simple, the arising problems are quite typical for problems where the jump-sets cannot be properly aligned between snapshots. As we will see later in section 4 when we have developed some better tools, singular behavior of the transforms generally occurs when the jump-set topology changes in parameter.

**4. Transforms as solutions of ODEs.** In this section, we consider discretizations of the transforms, which allow the types of singular behavior that we observed in our introductory example in section 3.

**4.1. Motivation of transform ODEs.** We have already chosen a notation for the transforms  $\eta \rightarrow X(\eta) = X(\eta; \mu, x)$  with “initial value”  $X(\mu, \mu, x) = x$  that suggests to define them as a solution of an ODE

$$(4.1) \quad \frac{d}{d\eta} X(\eta) = \Phi(X(\eta), \eta), \quad X(\mu) = x,$$

where we call  $\Phi : \Omega \times \mathcal{P} \rightarrow \mathbb{R}$  the *transport field*. First note that this definition restricts us to one-dimensional parameter domains. Nonetheless, higher dimensional parameters can be treated by a tensor product type construction as shown in [48], where we split dimensions of the parameter space one by one. Even for one-dimensional parameters, defining transforms via ODEs restricts our possible choices since it induces a semigroup property in the dynamic variable of the ODE:  $X(\eta; \mu, x) = X(\eta; \nu, X(\nu; \mu, x))$ . This also implies a symmetry in the source and target variables  $\eta$  and  $\mu$ , i.e.,  $X(\eta; \mu, \cdot) = X(\mu; \eta, \cdot)^{-1}$ . However, this restriction does not seem to be overly severe and has two major benefits: it was argued in [49] that it greatly helps optimizers to find global minima of (2.3), and as we shall see in this section, it provides a way to find discretizations of the transforms in view of singularities.

Let us first review the example in Figure 3.1 again. The example transform (3.3) is defined by the transform field

$$(4.2) \quad \Phi(x, \eta) = \begin{cases} 1, & x \leq -(1 - \eta), \\ -1, & x \geq 1 - \eta, \\ \frac{x}{\eta - 1} & \text{else,} \end{cases}$$

which simply records the slopes of  $\eta \rightarrow X(\eta; \mu, x)$  at every point, or more precisely  $\Phi(x, \eta) = \frac{d}{d\eta} X(\eta, \mu, x)|_{\mu=\eta}$ ; see [48]. Recall that the curves  $\eta \rightarrow X(\eta; \dots)$  should not intersect the jump-set of  $u$  or be parallel to it. Hence, in order to properly align the jumps, the corresponding blue lines in Figure 3.1 must be a solution of the ODE (4.1), defining the transforms. This implies that the trajectories intersect at the jump collision so that the ODE does not have a unique solution. Therefore, at least one of the assumptions of the Picard–Lindelöf theorem cannot be valid, which is typically the Lipschitz continuity of  $\Phi$ . This provides another motivation for dealing with high spatial resolution transforms.

One can easily generalize this observation. If the jump-sets of two snapshots  $u(\cdot, \mu)$  and  $u(\cdot, \eta)$  are not homeomorphic, no ODE, which aligns the jumps, can satisfy all assumptions of the Picard–Lindelöf theorem. For if one would, the solution  $x \rightarrow X(\eta; \mu, x)$  would be a homeomorphism of the respective jump-sets, contradicting the assumption of having a change in topology. For clarification, note that this observation is concerned with the uniqueness of the ODE solutions, given a fixed transform field  $\Phi$ . It is independent of possibly nonunique choices of  $\Phi$  itself, which we already observed in section 3 and discuss again in section 6.3 below.

**4.2. Discretization of the transform ODEs.** With the construction (4.1) of transforms as ODE solutions, we have to address two questions. First we have to compute the transform fields  $\Phi$ . This is achieved by optimizing the training error (2.3), this time with respect to the transform field  $\Phi$  instead of the transform  $X$ . If  $\Phi$  has high spatial resolution, this requires some more careful considerations and is discussed in detail in section 6. Second, we have to solve the ODE (4.1) numerically, which is considered in this section.

To this end, let us assume for the moment that we know  $\Phi(x, \mu)$  for all possible values of  $x$  and  $\mu$ . Then we still have to solve the ODE (4.1) to find  $X(\eta)$  and ultimately the TSI interpolation. The easiest choice that comes to mind is Euler's method

$$X(\eta) \approx x + (\eta - \mu)\Phi(x, \mu),$$

which has the drawback that we have to be able to evaluate  $\Phi$  at every possible target parameter  $\mu$ , which are of course unknown at the time we compute the transforms. If, however, we use the implicit Euler method, we have

$$(4.3) \quad X(\eta) \approx x + (\eta - \mu)\Phi(X(\eta), \eta),$$

where  $\Phi$  is only evaluated at the parameter  $\eta$ , which for the TSI reconstruction is only required at the finitely many snapshot locations  $\mathcal{P}_m$ .

Of course the implicit Euler method limits trajectories  $X(\eta)$  to lines, which limits the TSI reconstruction order in case the jump-sets of  $u$  are not. Therefore, in the remainder of this section, we construct higher order alternatives. First note that we cannot use higher order ODE solvers out of the box: Runge–Kutta methods require knowledge of  $\Phi(\cdot, \eta)$  at intermediate parameters depending on the unknown target  $\mu$  and multistep methods rely on single-step methods to get started.

Given these difficulties, we take an alternative but related route. Assume that we know  $\Phi(\cdot, \eta)$  at the interpolation points  $\eta \in \mathcal{P}_n$ , just as we know the snapshots  $u(\cdot, \eta)$  at all  $\mathcal{P}_m$ . Here  $\mathcal{P}_n$  is another set of finitely many nodes, which can in principle differ from  $\mathcal{P}_m$ . Restricting the ODE (4.1) to the information we have, we obtain

$$(4.4) \quad X'(\eta) = \Phi(X(\eta), \eta) \text{ for } \eta \in \mathcal{P}_n, \quad X(\mu) = x.$$

As we shall see in Proposition 4.1 below, these conditions are sufficient to uniquely determine the curve  $X(\eta)$ , if we restrict them to be polynomials of degree  $n = |\mathcal{P}_n|$  in the variable  $\eta$ , denoted by  $\mathbb{P}^n$ . The resulting transforms  $X(\eta)$  are called *high resolution transforms* in the following.

Let us first verify that this construction contains the implicit Euler method as a special case. For one single interpolation point, the transform of (4.4) is a line and therefore uniquely determined by the initial condition  $X(\mu) = x$  and its slope,  $X'(\eta) = \Phi(X(\eta), \eta)$  at the interpolation point  $\eta$ . This implies that the curve is given by

$$X(\xi) = x + (\xi - \mu)\Phi(X(\eta), \eta)$$

at any point  $\xi \in \mathcal{P}$ , which corresponds to the implicit Euler method for  $\xi = \eta$ .

Similar to ODE theory, the next proposition provides some conditions under which the interpolation problem (4.4) has a unique solution.

**PROPOSITION 4.1.** *Let  $I$  be an interval that contains the interpolation points  $\mathcal{P}_n$  and the initial point  $\mu$ .*

1. *Assume that all  $\Phi(\cdot, \eta) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\eta \in \mathcal{P}_n$ , are bounded and continuous. Then the interpolation problem (4.4) has a solution in  $\mathbb{P}^n$  restricted to  $I$ .*
2. *Assume that all  $\Phi(\cdot, \eta) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\eta \in \mathcal{P}_n$ , are Lipschitz continuous with Lipschitz constant  $L$  and that  $\Lambda_{n-1}L|I| < 1$ , where  $\Lambda_{n-1} := \max_{\mu \in I} \sum_{\eta \in \mathcal{P}_n} |\ell_\eta(\mu)|$  is the Lebesgue constant for the interpolation points  $\mathcal{P}_n$ . Then the interpolation problem (4.4) has a unique solution in  $\mathbb{P}^n$  restricted to  $I$ .*

Note that except for the discrete locations of the jump-set topology changes, the maps  $\Phi(\cdot, \eta)$  can be expected to be continuous. Close to the topology changes, the Lipschitz constant will deteriorate, which is discussed at the end of this section. In addition, we are only interested in  $\Phi(x, \eta)$  with  $x \in \Omega$ , so that the boundedness assumption can be met by suitable modifications of  $\Phi$  outside of  $\Omega$ . Therefore, with rather mild conditions the first part of the proposition ensures the global existence of transforms which satisfy (4.4).

As argued at the beginning of this section, we expect nonunique solutions at the jump-set topology changes for the ODE (4.1). Likewise, we cannot expect global uniqueness for the discrete variant (4.4) either. Nonetheless, locally this is possible by classical arguments as stated by the second part of the proposition.

The proof is essentially a discrete version of the Picard–Lindelöf theorem. To see this more clearly and for later reference, by the uniqueness of polynomial interpolation, the problem (4.4) is equivalent to the following continuous variant: find  $x \in \mathbb{P}_n$  such that

$$(4.5) \quad X'(\eta) = [\mathcal{I}_{n-1}\Phi(X(\cdot), \cdot)](\eta) \text{ for } \eta \in \mathcal{P}, \quad X(\mu) = x,$$

or equivalently

$$(4.6) \quad X(\xi) = x + \int_{\mu}^{\xi} [\mathcal{I}_{n-1}\Phi(X(\cdot), \cdot)](\gamma) d\gamma =: F[X](\xi),$$

where  $\mathcal{I}_{n-1}$  is the interpolation operator on the  $n$  points  $\mathcal{P}_n$ . Note that this is not a standard ODE because the right-hand side depends on  $X(\gamma)$  for all  $\gamma \in \mathcal{P}_n$  instead of  $X(\eta)$  at the single point  $\eta$ . Nonetheless, the Picard–Lindelöf theorem carries over verbatim.

*Proof.* We use the usual fixed point argument applied to the fixed point function  $F$  defined in (4.6) on the space  $\mathbb{P}_n(I)$  of polynomials up to degree  $n$ , restricted to the interval  $I$  and equipped with the norm  $\|X\| := \max_{\gamma \in I} |X(\gamma)|$ . Due to the interpolation  $\mathcal{I}_{n-1}$  in the definition of  $F$  and the subsequent integration  $F$  maps  $\mathbb{P}_n(I)$  to itself.

We first show part 1 of the proposition. Instead of the usual contraction mapping principle, we exploit the finite dimensional nature of  $\mathbb{P}_n$  and use Schauder's fixed point theorem. To this end, note that the continuity of  $\Phi(\cdot, \eta)$  implies the continuity of  $F$ . In addition, we have

$$\begin{aligned} \|F[X]\| &= \max_{\xi \in I} \left| x + \int_{\mu}^{\xi} [\mathcal{I}_{n-1}\Phi(X(\cdot), \cdot)](\gamma) d\gamma \right| \\ &\leq |x| + \Lambda_{n-1} \|\Phi(X(\cdot), \cdot)\| \int_{\mu}^{\xi} d\gamma \\ &\leq |x| + \Lambda_{n-1} M |I|, \end{aligned}$$

where  $M$  is the upper bound of  $|\Phi(\cdot, \eta)|$ ,  $\eta \in \mathcal{P}_n$ , and  $\Lambda_{n-1}$  is the Lebesgue constant for the interpolation  $\mathcal{I}_{n-1}$ . It follows that  $F$  maps the ball of all polynomials  $p \in \mathbb{P}_n$  with  $\|p\| \leq |x| + \Lambda_{n-1} M |I|$  to itself. In conclusion, Schauder's fixed point theorem implies the existence of a fixed point of the integral equation (4.6) and thus the discrete definition (4.4) of the transforms.

Since Schauder's fixed point theorem does not guarantee uniqueness, we continue to prove part 2 of the proposition using the classical contraction mapping principle. To show that  $F$  is a contraction, first note that

$$\begin{aligned} \|\mathcal{I}_{n-1}\Phi(X(\cdot), \cdot) - \mathcal{I}_{n-1}\Phi(Y(\cdot), \cdot)\| &\leq \Lambda_{n-1} \max_{\eta \in \mathcal{P}} |\Phi(X(\eta), \eta) - \Phi(Y(\eta), \eta)| \\ &\leq \Lambda_{n-1} L \|X - Y\|, \end{aligned}$$

where again  $\Lambda_{n-1}$  is the Lebesgue constant and  $L$  the Lipschitz constant of  $\Phi$ . It follows that



$$(4.7) \quad \|F[X] - F[Y]\| \leq \Lambda_{n-1} L \|X - Y\| \int_{\mu}^{\xi} d\gamma \leq \Lambda_{n-1} L |I| \|X - Y\|.$$

Since  $\Lambda_{n-1} L |I| < 1$ , the proposition follows from Banach's fixed point theorem.  $\square$

*Remark 4.2.* Since  $X(\eta)$  is a polynomial, which is uniquely defined on the interval  $I$ , we may extend it to the real line  $\mathbb{R}$  still satisfying the discrete and continuous interpolation problems (4.4) and (4.5). However, beyond the interval  $I$  in Proposition 4.1, part 2, the solution is not necessarily unique in the sense that there can be two solutions  $X$  and  $Y$  of (4.4) with different initial values  $X(\mu) = x$  and  $Y(\mu) = y$  that intersect outside of  $I$ . Such nonuniqueness is expected at jump collisions, just as in the continuous case discussed after (4.2).

Let us compare the new transform  $X(\eta)$  constructed in Proposition 4.1 with the original choice (2.2) defined via polynomial interpolation. The original choice is polynomial in  $\mu$ , whereas the “ideal” transform in the example of section 3 is rational in  $\mu$ . In contrast, both the new transform and the “ideal” one are polynomial in  $\eta$ . In addition, the stability function of the implicit Euler method is a rational function of  $\eta - \mu$  and thus for a fixed interpolation node  $\eta$  and variable target  $\mu$  a rational function of  $\mu$ . Again, this matches the “ideal” transform of the example.

**4.3. Numerical application of the implicit ODE solver.** For a practical application of the new transform, we need an algorithm to solve the interpolation problem (4.4). Since the fixed point iteration for the map  $F : \mathbb{P}_n \rightarrow \mathbb{P}_n$  in the proof of Proposition 4.1 is defined on polynomials, it can be carried out practically and it remains to devise an algorithm to compute  $Y = F[X] \in \mathbb{P}_n$ . This is reminiscent of implicit ODE solvers such as, e.g., the implicit Euler method (4.3), which are also often solved by fixed point methods. We carry out the fixed point iteration in a Newton basis, which easily leads to an algorithm that is sufficient for the first experiments presented in this paper, although more clever approaches might be possible.

Although we integrate polynomials, we cannot readily apply standard quadrature rules because the integration bounds are variable and not known beforehand. As an alternative, we first differentiate the definition (4.6) of  $F$  to obtain

$$(4.8) \quad Y'(\xi) := \frac{d}{d\xi} F[X](\xi) = [\mathcal{I}_{n-1} \Phi(X(\cdot), \cdot)](\xi).$$

The derivative  $Y'(\xi)$  is a polynomial in  $\mathbb{P}_{n-1}$  and thus completely specified by its values at the interpolation points  $\eta \in \mathcal{P}_n$ , which are

$$Y'(\eta) = [\mathcal{I}_{n-1} \Phi(X(\cdot), \cdot)](\eta) = \Phi(X(\eta), \eta).$$

Of course these are the same interpolation conditions as in the original problem (4.4), with the difference that currently  $X$  is not necessarily a fixed point. Now, let us represent  $Y$  in a Newton basis with respect to the interpolation points  $\{\mu\} \cup \mathcal{P}_n = \{\mu, \eta_1, \dots, \eta_n\}$ , i.e.,

$$(4.9) \quad Y(\xi) = \sum_{i=0}^n a_i \omega_i(\xi), \quad \omega_0(\xi) = 1, \quad \omega_i(\xi) = (\xi - \mu) \prod_{j < i} (\xi - \eta_j).$$

First note that choosing  $\xi = \mu$  the definition of  $F$  implies that  $a_0 = x$ . For the remaining  $a_1, \dots, a_n$ , we first differentiate  $Y'(\xi) = \sum_{i=1}^n a_i \omega'_i(\xi)$  and plug in the interpolation points to obtain the system

$$(4.10) \quad Ma := \begin{pmatrix} \omega'_1(\eta_1) & \dots & \omega'_n(\eta_1) \\ \vdots & & \vdots \\ \omega'_1(\eta_n) & \dots & \omega'_n(\eta_n) \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \Phi(X(\eta_1), \eta_1) \\ \vdots \\ \Phi(X(\eta_n), \eta_n) \end{pmatrix} =: B.$$

Note that for Newton interpolation we have  $\omega_j(\eta_i) = 0$  for  $j > i$ , which leads to triangular Vandermonde type matrices for standard interpolation problems. However, this does not imply that also  $\omega'_j(\eta_i) = 0$  for  $j > i$ . Therefore, in our case we deal with a full matrix. Nonetheless, we can compute an LU-decomposition beforehand and solve the system efficiently for every new right-hand side.

Now that we can compute  $Y = F[X]$ , we can numerically solve the interpolation problem (4.4) by the fixed point iteration

$$(4.11) \quad X^{k+1} = F[X^k], \quad X^0 = x.$$

The following corollary to Proposition 4.1 states that this converges to a solution of the interpolation problem.

**COROLLARY 4.3.** *Assume that all assumptions of Proposition 4.1, part 2, are satisfied. Then the fixed point iteration (4.11) converges to a solution of the interpolation problem (4.4).*

*Proof.* The corollary is a direct consequence of the proof of Proposition 4.1, part 2.

**Remark 4.4.** By construction the transforms  $\eta \rightarrow X(\eta; \mu, x)$  are polynomial in the continuous variable  $\eta$ . These curves are determined by the discrete interpolation points  $\eta_1, \dots, \eta_n \in \mathcal{P}_n$ , the target parameter  $\mu$ , and the slopes at the interpolation points  $\Phi(\cdot, \eta_i)$ . Via repeatedly solving (4.10) during the fixed point iteration (4.11), the curves  $X$  are rational functions of the interpolation points and target  $\mu$ .

The fixed point iteration (4.11) provides us with polynomials  $\eta \rightarrow X(\eta; \mu, x)$  for  $\eta \in \mathcal{P}$ . For the TSI (2.1), we only need the transforms at the locations  $\eta \in \mathcal{P}_m$ , which we can easily evaluate even if  $\mathcal{P}_m \neq \mathcal{P}_n$ , i.e., we use different nodes for the transform field than for the snapshots. The algorithm can be summarized as follows:

---

```

1: function TSI( $x \in \mathbb{R}^d, \mu \in \mathbb{R}$ )
2:   Initialize transform locations:  $X = [X_\eta]_{\eta \in \mathcal{P}_n} := [x, \dots, x] \in \mathbb{R}^{n \times r}$ .
3:   Compute the LU-decomposition of the interpolation matrix  $M$  in (4.10).
4:   for # fixed point iterations do
5:     Compute  $B_\eta \leftarrow \Phi(X_\eta; \eta)$ ,  $\eta \in \mathcal{P}_n$ .
6:     Solve (4.10), i.e.,  $Ma = B$  for the Newton basis coefficients  $a \in \mathbb{R}^n$ .
7:     Evaluate the Newton polynomial by (4.9):  $X_\eta \leftarrow \sum_{i=1}^n a_i \omega_i(\eta)$ ,  $\eta \in \mathcal{P}_n$ .
8:   end for
9:   return  $u_n(x, \mu) = \sum_{\eta \in \mathcal{P}_n} \ell_\eta(\mu) u(X_\eta, \eta)$ 
10: end function

```

---

The algorithm requires us to store the scalar snapshots and vector valued transform fields for all interpolation points. For  $n$  snapshots,  $d$  spatial dimensions, and  $N$  degrees of freedom for the high fidelity model, this takes  $(d+1)nN$  degrees of freedom. For a fixed spatial dimension, this is only a constant factor more than the  $nN$  degrees of freedom for simple polynomial interpolation or other reduced order modeling techniques.

To compute the TSI for  $\mathcal{O}(N)$  points, as done for quadrature points during the offline optimization of the training error, we compute an LU-decomposition of the polynomial interpolation matrix (4.10) just once in  $\mathcal{O}(n^3)$  steps. For each fixed point step, we evaluate the transform field, compute the coefficients in (4.10) with the pre-computed LU-decomposition, and evaluate the interpolation polynomial in (4.9) in  $\mathcal{O}(n^2N)$  steps. We repeat this for  $k$  fixed point iterations and obtain the transformed points  $X_\eta$  in  $\mathcal{O}(n^3 + kn^2N)$  steps. Finally, we plug these points into the TSI (2.1), which requires a one-time computation of the Lagrange basis functions in  $\mathcal{O}(n^2)$  and the evaluation of the snapshots at the transformed points in  $\mathcal{O}(N)$  steps. The latter requires us to evaluate the high fidelity model at  $N$  arbitrary points in  $\mathcal{O}(N)$  time. This is easily possible on structured grids. For hierarchical or fully unstructured grids, one may exploit a neighboring structure of grid cells together with the observation that even for singular transforms in most parts of the physical domain the transform is Lipschitz. Therefore, once one has found the grid cell of a transformed point  $X(\eta; \mu, x)$  by continuity one can expect to find a nearby transform  $X(\eta; \mu, y)$  for  $y \approx x$  in nearby cells. The development of concrete algorithms is left for future research. Adding up the run-times above, we obtain

$$\mathcal{O}(n^3 + (1 + kn^2)N)$$

steps for a full evaluation of the TSI. In particular, as for most reduced order models this is linear in the degrees of freedom of the high fidelity model.

As discussed earlier, we expect the Lipschitz constant  $L$  to deteriorate near jump-set topology changes so that a plain fixed point iteration as in the last corollary is generally insufficient. Nonetheless, we can counterbalance this with a small interval length  $I$  in order to meet the assumptions of Proposition 4.1. The obvious choice is to add more snapshots close to the singularity, but that clearly defeats the purpose of this section. Alternatively, we use this observation to create better initial values for the fixed point iteration. To this end, note that neither the definition of  $F[X]$  nor the assumptions of Proposition 4.1 need  $\Phi(\cdot, \eta)$  for the full continuum of parameters  $\eta$ , but only for the few discrete  $\eta \in \mathcal{P}_n$ . For clarity, let us call them  $\Phi_0, \dots, \Phi_n$ . Once we have fixed these functions, the interval length  $I$  only enters the equations via the Newton basis (4.9) for the interpolation points  $\{\mu\} \cup \mathcal{P}_n$ . Therefore, in order to obtain a shorter interval  $I$ , we rescale these interpolation points by

$$(4.12) \quad \{\mu\} \cup \{\mu + s(\eta - \mu) : \eta \in \mathcal{P}_n\},$$

which match the original choice with  $s = 1$  and shrink to  $\{\mu\}$  for  $s = 0$ . For  $s$  sufficiently small, the fixed point iteration (4.11) will then be successful for the unchanged  $\Phi_0, \dots, \Phi_n$ . We use the outcome as an improved initial value for a new fixed point iteration with larger scaling  $s$ , although this case may no longer be covered by Corollary 4.3. In the numerical experiments, we choose an increasing sequence of scaling factors  $s_i$  and perform fixed point iterations (4.11) for the corresponding scaled parameters, each using the last iterate from the previous scaling as initial value.

**5. A note about stability.** Since we are dealing with jumps in  $u$  and possibly (near) singular jump-sets, we have to be careful about the stability of the TSI reconstruction with respect to perturbations of the transforms. Such perturbations are generally inevitable because of numerical errors due to iterative optimizers for (2.3) and discretizations of the transport fields  $\Phi$ . A basic analysis for simple transforms is already contained in [49], which, however, does not cover the near singular transforms considered here.

Let us first consider a heuristic motivation in one dimension. Assume we have two transforms  $X_i(x) := X_i(\eta; \mu, x)$ ,  $i = 0, 1$ . Ignoring higher order terms, and using the substitution  $y = X_0(x)$  we can informally estimate the perturbation error between two transformed snapshots  $v \circ X_i(x) := u(X_i(x), \eta)$  by

$$\begin{aligned}
 \|v \circ X_1 - v \circ X_0\|_{L_1(\Omega)} &\approx \int_{\Omega} |u'(X_0(x))| |X_0(x) - X_1(x)| dx \\
 &= \int_{\Omega} |u'(y)| |y - X_1(X_0^{-1}(y))| \frac{1}{|X_0'(x)|} dy \\
 (5.1) \quad &\leq \sup_{y \in \Omega} \left[ \frac{|y - X_1(X_0^{-1}(y))|}{|X_0'(x)|} \right] \int_{\Omega} |u'(y)| dy \\
 &= \sup_{x \in \Omega} \left[ \frac{|X_0(x) - X_1(x)|}{|X_0'(x)|} \right] \|u\|_{BV(\Omega)}.
 \end{aligned}$$

If  $X_0'$  is bounded from above and below, this gives us an estimate of the type

$$\|v \circ X_1 - v \circ X_0\|_{L_1(\Omega)} \leq C \|X_0(x) - X_1(x)\|_{L_{\infty}(\Omega)} \|u\|_{BV(\Omega)}$$

for some constant  $C$ . However, this argument is not fully satisfactory for two reasons:

1. In the vicinity of jump-set topology changes  $X_0'(x)$  is typically not uniformly bounded from above and below.
2. If  $X_i$  has a jump or sharp gradient, reasonable perturbations may not be close in the  $L_{\infty}$ -norm.

If we are more careful, we see that wherever  $X_0$  has a large gradient so that we possibly have a large  $L_{\infty}$  error of the transforms, the denominator of the right-hand side of (5.1) is also large. The argument does not carry over to sharp gradients of  $X_1$ , but with an analogous argument, we could have had  $X_1'$  in the denominator instead. Therefore, in the following, we consider the question of whether a judicious choice of the denominator can at least in principle provide some control of the perturbation error in the face of sharp gradients.

The choice of the denominator, now for arbitrary spatial dimensions, is done by selecting a homotopy, i.e., a continuous transition,  $X_s(\eta; \mu, x)$ ,  $0 \leq s \leq 1$ , between the two transforms so that the boundary cases  $X_0(\eta; \mu, x)$  and  $X_1(\eta; \mu, x)$  are our original two perturbed transforms. A simple candidate is a convex combination  $X_s = (1-s)X_0 + sX_1$ . The following stability result is a rigorous version of our heuristic argument above and a slight refinement from a proposition in [49].

PROPOSITION 5.1. *Assume that  $u \in BV(\Omega)$  and that*

$$\int_{X_s(\eta; \mu, \cdot)^{-1}(A)} dx \leq B|A|$$

for some  $B \geq 0$ , all measurable sets  $A$ , all  $\eta \in \mathcal{P}_n$ ,  $\mu \in \mathcal{P}$ , and  $s \in \{0, 1\}$ . Using  $\dot{X} := \frac{d}{ds} X_s$ , define

$$(5.2) \quad S(\mu, \eta) := \sup_A \frac{1}{|A|} \int_0^1 \int_{X_s(\eta; \mu, \cdot)^{-1}(A)} |\dot{X}_s(\eta; \mu, x)| dx ds.$$

Then, we have

$$\|u_m(\cdot, \mu; X_0) - u_m(\cdot, \mu; X_1)\|_{L_1(\Omega)} \leq \Lambda_n \max_{\eta \in \mathcal{P}_m} [S(\mu, \eta) \|u(\cdot, \eta)\|_{BV(\Omega)}],$$

independent of  $B$  and where  $\Lambda_m$  is the Lebesgue constant.

The proof is essentially the same as in [49] and given in Appendix A. For sufficiently regular transforms  $X_s$ , we can simplify the stability factor  $S(\mu, \eta)$ : with  $X_s(x) = X_s(\eta; \mu, x)$  we have

$$\begin{aligned}
 S(\mu, \eta) &= \sup_A \frac{1}{|A|} \int_0^1 \int_{X_s^{-1}(A)} |\dot{X}_s(x)| dx ds \\
 &= \sup_A \int_0^1 \frac{1}{|A|} \int_A |\dot{X}_s(X_s^{-1}(y))| |\det DX_s^{-1}(y)| dy ds \\
 (5.3) \quad &= \int_0^1 \sup_{y \in \Omega} |\dot{X}_s(X_s^{-1}(y))| |\det DX_s^{-1}(y)| ds \\
 &= \int_0^1 \sup_{x \in \Omega} \frac{|\dot{X}_s(x)|}{|\det DX_s(x)|} ds.
 \end{aligned}$$

Note the similarity to our informal calculation (5.1). For the nominator, we have  $|\dot{X}_s(x)| \approx |X_0(x) - X_1(x)|$  and in one dimension the denominator simplifies to  $|X'_s(x)|$ . In addition, because of the  $s$  dependence, we now have some explicit control over the denominator by choosing a suitable homotopy.

In order to better understand the stability factor  $S(\mu, \eta)$ , let us first consider the integral bound  $X_s(\eta; \mu, \cdot)^{-1}(A)$  in (5.2), or equivalently the denominator in (5.3). To this end, we consider a set  $A \subset \Omega$  and observe how its volume changes when it is transformed along the flow  $\eta \rightarrow A_\eta := X_s(\eta; \mu, A)$ . We expect the stability factor to be small if the volume  $|A_\mu| = |X_s(\eta; \mu, \cdot)^{-1}(A_\eta)|$  is not much larger than the volume  $|A_\eta|$ . This implies that contracting sets from  $\eta \rightarrow \mu$  is expected to be stable, while growing them is probably unstable.

To clarify this point, let us consider the example in Figure 3.1 again. Say we have two parameters  $\mu_c$ , where the two jumps collide, and  $\mu_2 < \mu_c$  for which we have two separate jumps. Let us first try to approximate  $u(\cdot, \mu_c)$  from a transformation of  $u(\cdot, \mu_2)$ . If we choose any set  $A_{\mu_2}$  between the two jumps, the transform indicated in the figure shrinks it to zero volume. By our analysis this is stable and intuitively all perturbations in the set  $A_{\mu_2}$  are squashed to measure zero and thus irrelevant. Conversely, using a transform of the snapshot  $u(\cdot, \mu_c)$  to approximate  $u(\cdot, \mu_2)$  is problematic. Our stability bound becomes large, matching the intuition that the snapshot does not provide any information on how to fill the void between the two jumps of the target  $u(\cdot, \mu_2)$ .

Let us next consider the problem with the  $L_\infty$ -norm for a simple example problem. We choose  $X_s(\eta; \mu, x) = X_0(\eta; \mu, x - s\epsilon v)$  for some direction  $v$ , so that the perturbation  $X_1$  is a shift of  $X_0$ , which generally has large  $L_\infty$  error if  $X_0$  has sharp gradients or jumps. Note that with this definition we also have chosen a homotopy between the two transforms. Next, we assume that  $\|DX_s(\eta; \mu, \cdot)^{-1}\|_2 \leq c$  for some constant  $c$ . This allows the transform  $X_s(\eta; \mu, \cdot)$  to have sharp gradients and places us in the situation where the sets  $A_\eta$  grow at most moderately for which we expect stable behavior by the discussion above. In the example of Figure 3.1, this includes the case of approximating  $u(\cdot, \mu_c)$  from  $u(\cdot, \mu_2)$  using a perturbed transform that misses the correct collision location in physical space by  $\epsilon$ .

Computing the stability factor (5.3) yields

$$S(\mu, \lambda) = \int_0^1 \sup_{x \in \Omega} \frac{|\dot{X}_s(x)|}{|\det DX_s(x)|} ds \leq \int_0^1 \sup_{x \in \Omega} s\epsilon \frac{|DX_s(x)|_2 |v|}{|\det DX_s(x)|} ds,$$

where  $|\cdot|_2$  denotes the  $\ell_2(\mathbb{R}^d)$  matrix norm. With the singular values  $\sigma_1^s(x) \geq \dots \geq \sigma_d^s(x) \geq c$  of  $DX_s$ , this simplifies to

$$S(\mu, \lambda) = \int_0^1 s\epsilon|v| \sup_{x \in \Omega} \frac{\sigma_1^s(x)}{\prod_{i=1}^d \sigma_i^s(x)} ds \leq \epsilon|v|c^{1-d} \int_0^1 s ds = \frac{1}{2}\epsilon|v|c^{1-d}.$$

Thus, even if the  $L_\infty$  error of the perturbed transforms is large, we may nonetheless have a small TSI error. Of course the result relies on a carefully chosen homotopy  $X_s$ . How to choose this in the general case is beyond the scope of this paper.

## 6. Optimizing singular transforms.

**6.1. Difficulties with the optimization of the transform.** In section 4, we discussed singularities of the transforms at jump-set topology changes and efficient ways to approximate them. However, we still need a way to automatically calculate these transforms from the available data, i.e., the snapshots and training snapshots. As in (2.3), we use an optimization problem

$$(6.1) \quad \min_{\Phi \in \mathcal{T}} \sigma(\Phi), \quad \sigma(\Phi) := \|u(\cdot, \mu) - u_m(\cdot, \mu; \Phi)\|_{L_1(\Omega)},$$

where the notation  $u_m(x, \mu; \Phi) = u_m(x, \mu)$  is used to emphasize the dependence of the TSI on the choice of the transport field  $\Phi$  and  $\mathcal{T} \subset \mathcal{B}$  is an admissible set of transforms in a Banach space  $\mathcal{B}$  as discussed below. Note that the constraints  $\mathcal{T}$  can be used to ensure that the resulting curves  $X(\eta)$  do not leave the spatial domain  $\Omega$  and that we only need to know  $\Phi(\cdot, \eta)$  for finitely many  $\eta \in \mathcal{P}_n$ . We therefore only need a spatial discretization of  $\Phi$  by selecting a basis, or more generally a dictionary,  $\{\psi_i\}_{i \in \mathcal{I}}$  for some index set  $\mathcal{I}$ , with  $\text{span}\{\psi_i\}_{i \in \mathcal{I}} = \mathcal{B}_n \subset \mathcal{B}$ , and optimize

$$(6.2) \quad \min_{c \in \mathbb{R}^N} \sigma(c), \quad \sigma(c) := \sigma\left(\sum_{i=1}^N c_i \psi_i\right).$$

Instead of strictly enforcing that  $x \rightarrow X(\eta; \mu, x)$  maps  $\Omega$  to itself, we usually assume that  $\nu \cdot \Phi(\cdot, \eta) = 0$ , where  $\nu$  is the outward normal, so that locally the curves  $X(\eta; \mu, x)$  do not flow through the boundary. As we have seen above, already simple problems require sharp spatial gradients so that we want to choose  $\mathcal{B}_n$  with resolutions matching the high fidelity model of the snapshots. In this section, we discuss the implications for the optimization of (6.1).

It has been shown in [49] that we can expect the map  $\sigma$  to be Lipschitz continuous. Although more sophisticated algorithms for such nonsmooth optimization problems are available [25, 6], we start with a simple gradient descent method:

$$(6.3) \quad c^{n+1} = c^n - \alpha \sigma'(c^n).$$

Unfortunately, this simple choice does not provide satisfactory results. For example, Figure 7.1 shows the convergence behavior for a very simple one-dimensional test case: the gradient descent optimizer reduces the error a little bit but then stalls way before any reasonable alignment has been achieved. Since the example has no singularities, the figure also shows the results for a coarse discretization of  $X(\eta; \mu, x)$  by second order polynomials in  $x$ . As we see, this coarse resolution does not pose a severe obstacle to the optimizer.

**6.2. Optimization in Banach spaces.** To better understand the issue of the higher resolution, we first go all the way to infinite resolution and optimize (6.1) over the full function space  $\mathcal{B}$ . We solve the optimization problem  $\min_{\Phi \in \mathcal{B}} \sigma(\Phi)$  by a descent method so that

$$\Phi^{n+1} = \Phi^n - \alpha d^n$$

for some descent direction  $d^n$ . Usually, we choose the gradient  $d^n = \sigma'(\Phi)$ , but this time we have to be more careful: Since we optimize in a Banach space,  $\sigma'(\Phi)$  is a Fréchet derivative and thus an element of the dual space  $\mathcal{B}'$  of  $\mathcal{B}$ . If  $\mathcal{B} = \mathbb{R}^n$ , we can easily identify  $(\mathbb{R}^n)' = \mathbb{R}^n$ , but for a generic Banach space  $\mathcal{B}$  this is not always possible. This implies that a term  $\Phi^n - \alpha \sigma'(\Phi)$  does not necessarily make sense.

Thus, the descent direction  $d^n$  depends on the Banach space  $\mathcal{B}$ . A good choice of  $\mathcal{B}$  should depend, e.g., on the existence of minimizers or their stability, which is beyond the scope of this paper. Nonetheless, one can easily gain some insight about the space  $\mathcal{B}'$ . Formally, the variation of the error in the direction  $v \in \mathcal{B}$  is given by

$$\partial_v \sigma(\Phi) = \int_{\Omega} \text{sign} [u(x, \mu) - u_m(x, \mu; \Phi)] \partial_v u_m(x, \mu; \Phi) dx,$$

where using the definition of the TSI, we have

$$\partial_v u_m(x, \mu; \Phi) = \sum_{\eta \in \mathcal{P}_m} \ell_{\eta}(\mu) Du(X(\eta; \mu, x), \eta) DX(\eta; \mu, x) v.$$

Note in particular that this derivative contains the derivative  $Du(x, \eta)$  with respect to  $x$ . Since  $u$  has jumps we may assume that  $u$  is of bounded variation so that  $Du$  is a measure. At the jumps this measure has a singular component, e.g., a Dirac delta in  $d = 1$  or generally a measure supported on a  $d - 1$  dimensional manifold. This implies that the dual space  $\mathcal{B}'$  must be rich enough to contain these singular measures, which makes an identification of  $\mathcal{B}$  and  $\mathcal{B}'$  difficult.

Of course the optimization in Banach spaces is well understood and a descent direction  $d^n$  can be constructed along the following lines; see, e.g., [47]. First observe that the value of the objective function after one step is given by

$$\sigma(\Phi^{n+1}) = \sigma(\Phi^n) - \alpha \langle \sigma'(\Phi^n), d^n \rangle + o(\alpha)$$

so that we obtain an error reduction for sufficiently small  $\alpha$  if  $\langle \sigma'(\Phi^n), d^n \rangle > 0$ . For example, we can choose the Riesz representation of the Fréchet derivative defined by

$$(6.4) \quad d^n = \operatorname{argmax}_{v \in \mathcal{B}} \frac{\langle \sigma'(\Phi^n), v \rangle}{\|v\|_{\mathcal{B}}}, \quad \|d^n\|_{\mathcal{B}} = \|\sigma'(\Phi^n)\|_{\mathcal{B}'},$$

so that  $\langle \sigma'(\Phi^n), d^n \rangle = \|\sigma'(\Phi^n)\|_{\mathcal{B}'}^2$ , and we obtain

$$\sigma(\Phi^{n+1}) = \sigma(\Phi^n) - \alpha \|\sigma'(\Phi^n)\|_{\mathcal{B}'}^2 + o(\alpha).$$

Recall that in our case the dual space  $\mathcal{B}$  should contain singular measures supported on  $d - 1$  dimensional manifolds. Via the trace theorem, measures of such type are contained in the Sobolev space  $H^{-1}(\Omega)$ , which is a particularly simple choice because it is a Hilbert space and the computation of the Riesz representer  $d^n$  reduces to solving the Poisson equation. (We use standard notation for Sobolev spaces as, e.g., in [17].)

This can be easily seen as follows. First, for each transform field  $\Phi(\cdot, \eta)$  with  $\eta \in \mathcal{P}_n$ , we choose the Sobolev space  $B_{\eta} = \{v \in H^1(\Omega)^d : \nu \cdot v = 0 \text{ on } \partial\Omega\}$ , with

outward unit normal vector  $\nu$  and slip boundary conditions, which ensure that locally the resulting curves  $X(\eta)$  do not leave  $\Omega$ . We combine all of these into one space  $\mathcal{B} = \prod_{\eta \in \mathcal{P}_m} \mathcal{B}_\eta$ . For every  $f \in \mathcal{B}'_\eta$ , we define the inverse Laplacian by the usual weak formulation

$$(\nabla(\Delta^{-1})f, \nabla v) = \langle f, v \rangle, \quad v \in \mathcal{B}.$$

Choosing the energy norm  $\|\cdot\|_{\mathcal{B}_\eta} = \|\nabla \cdot\|_{L_2(\Omega)}$ , this implies that

$$\|f\|_{\mathcal{B}'_\eta} = \sup_{v \in \mathcal{B}_\eta} \frac{\langle f, v \rangle}{\|v\|_{\mathcal{B}_\eta}} = \sup_{v \in \mathcal{B}_\eta} \frac{(\nabla[-\Delta^{-1}f], \nabla v)}{\|v\|_{\mathcal{B}_\eta}} \leq \|-\Delta^{-1}f\|_{\mathcal{B}_\eta}$$

with equality if and only if  $v = -\Delta^{-1}f$ . Therefore,  $v = -\Delta^{-1}f$  is also the maximizer in the suprema in the last formula. Comparing this with the definition of the Riesz representation (6.4) yields

$$d^n = \bigtimes_{\eta \in \mathcal{P}_n} -\Delta^{-1}[D_{\Phi^n(\cdot, \eta)}\sigma(\Phi^n)], \quad \Phi^n = \bigtimes_{\eta \in \mathcal{P}_n} \Phi^n(\cdot, \eta),$$

where  $D_{\Phi(\cdot, \eta)}\sigma(\Phi^n)$  is the Fréchet derivative of the error with respect to the transform  $\Phi(\cdot, \eta)$ . Hence, with a slight abuse of notation, we obtain the optimization scheme

$$(6.5) \quad \Phi^{n+1} = \Phi^n + \alpha \Delta^{-1} \sigma'(\Phi^n).$$

This is discretized by a Galerkin method in the obvious way. The resulting linear system can be numerically solved in  $\mathcal{O}(N)$  steps to the accuracy warranted by the high fidelity model. Therefore, the extra Laplace inversion only inflates the cost per gradient descent step by a constant. We will refer to this method as *Laplace smoothing* in the following.

As an alternative to inverting a Laplacian, we can carefully choose the dictionaries  $\{\psi_i\}_{i \in \mathcal{I}}$  for the discretization, similar to, e.g., wavelet methods for elliptic equations [13, 10]. To this end, let us assume that the  $\psi_i$  are a frame for  $\mathcal{B}'$ , so that

$$A\|\ell\|_{\mathcal{B}'}^2 \leq \sum_{i \in \mathcal{I}} |\langle \ell, \psi_i \rangle| \leq B\|\ell\|_{\mathcal{B}'}^2$$

for all  $\ell \in \mathcal{B}'$  and constants  $A, B \geq 0$ . Then, there is a dual frame  $\{\bar{\psi}_i\}_{i \in \mathcal{I}}$  such that

$$B^{-1}\|\Phi\|_{\mathcal{B}}^2 \leq \sum_{i \in \mathcal{I}} |\langle \Phi, \bar{\psi}_i \rangle| \leq A^{-1}\|\Phi\|_{\mathcal{B}'}^2$$

for all  $\Phi \in \mathcal{B}$  and

$$\Phi = \sum_{i \in \mathcal{I}} \langle \Phi, \bar{\psi}_i \rangle \psi_i.$$

Note that  $\Phi \in \mathcal{B}$  if and only if the frame coefficients  $c_i = \langle \Phi, \bar{\psi}_i \rangle$  are in  $\ell_2(\mathbb{R}^{\mathcal{I}})$  so that in the original discretization (6.2) we now optimize over  $\mathbf{c} \in \ell_2(\mathbb{R}^{\mathcal{I}})$  for which we easily can identify the space with its dual. Indeed, for the resulting iteration

$$(6.6) \quad \mathbf{c}^{n+1} = \mathbf{c}^n - \alpha \mathbf{d}^n, \quad \mathbf{d}_i^n = \partial_i \sigma(\mathbf{c}) = \left\langle \sigma' \left( \sum_{i=1}^{\infty} c_i \psi_i \right), \psi_i \right\rangle$$

we have the error reduction



$$\begin{aligned}\sigma(\mathbf{c}^{n+1}) &= \sigma(\mathbf{c}^n) - \alpha \sum_{i=1}^{\infty} \left| \left\langle \sigma' \left( \sum_{i=1}^{\infty} c_i \psi_i \right), \psi_i \right\rangle \right|^2 + o(\alpha) \\ &\leq \sigma(\mathbf{c}^n) - \alpha A \left\| \sigma' \left( \sum_{i=1}^{\infty} c_i \right) \right\|_{\mathcal{B}'}^2 + o(\alpha)\end{aligned}$$

for  $\alpha$  sufficiently small.

For simplicity, in the numerical experiments, we do not explicitly construct a frame but rather define the transform as  $X(\eta; \mu, x) = x + \sum_{\ell} (X^{\ell}(\eta; \mu, x) - x)$ , where  $X^{\ell}(\dots)$  are curves generated by transform fields  $\Phi^{\ell}$  with spatial resolution level  $\ell$  and  $X^{\ell}(\dots) - x$  the updates from the respective levels. This is referred to as *multilevel smoothing* in the following.

**6.3. Uniqueness.** The training error is insensitive to smooth changes of the transform away from the jump sets or in higher spatial dimensions tangential to the jump sets. Therefore, a wide selection of transforms may achieve small training errors. Indeed, in our simple example from section 3 any perturbation of the transform in regions where  $u$  is constant change neither the TSI nor the training error.

However, the TSI reconstruction error is the major quantity of interest and as long as it is small we have a sensible approximation. Nonetheless, one may consider this question more carefully in future research: among the “good” transforms some may require more or less degrees of freedom for their discrete representation. In addition, practically we only minimize the error on a training sample  $\mathcal{P}_T$  and different transforms may have better or worse generalization errors.

In optimal transport based methods [23, 40, 41], this nonuniqueness is eliminated by choosing a transform that minimizes a transport cost. Applied to the method of this paper, it is currently not understood how cost and training error minimization interact.

**7. Numerical experiments.** In section 7.1, we first choose a simple problem without singularity to see how the optimizers of section 6 perform in comparison to the original method recalled in section 2. Then, in section 7.2, we consider a test with a singularity where the original low resolution transforms fail. In section 7.3, we compare different orders of the transforms and in section 7.4 we consider a two-dimensional example.

In all examples, the finite element discretizations of the snapshots and transform fields are handled by the Deal.II library [3] and all TSI components are then put together in Tensorflow [1]. This setup allows all gradients of the error to be automatically computed by the back-propagation algorithm in Tensorflow.

**7.1. Testing the optimizers.** In this section, we first consider a very simple example in order to see how the optimizers of section 6 applied to high resolution transforms of section 4 perform in comparison to a simple gradient descent method and low resolution transforms as described in [49] or section 2. To this end, we consider

$$u(x, \mu) = \begin{cases} 1, & x \leq \mu, \\ -1, & x > \mu, \end{cases} \quad \mathcal{P}_m = \mathcal{P}_n = \{-0.2, 0.2\}, \quad \Omega = [-1.5, 1.5].$$

$$\mathcal{P}_T = \{0\},$$

The snapshots and high resolution transforms are discretized by piecewise linear functions on 32 cells and the low resolution transforms are second order polynomials. We use five fixed point iterations (4.11) and no scaling (4.12).

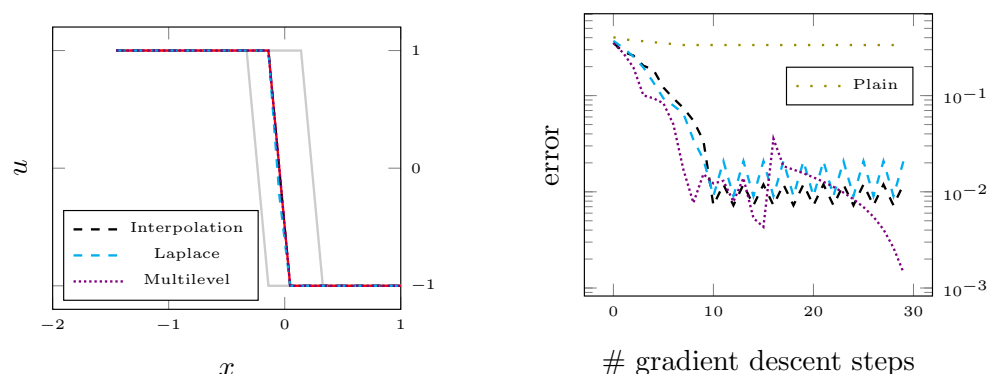


FIG. 7.1. Results for the one-dimensional example in section 7.1. Left: Snapshots (gray), exact solution (solid red), and reconstructions at  $\mu = 0$ : TSI with Laplace smoothing (cyan, dashed), multilevel smoothing (violet, dotted), and a TSI with low resolution transform (2.2) (black, dashed). Right: Corresponding gradient descent errors with the same color coding. In addition the figure shows the errors for high resolution transforms without smoothing (olive dotted).

Since the objective function typically has a kink at the minimum, its value from gradient descent steps typically oscillates near the minimum. For all optimization methods in this section, we select a learning rate that shows slight oscillations at the end to make them comparable. This is not very accurate but is sufficient to demonstrate the main point of this experiment: despite the much higher number of degrees of freedom, the Laplace and multilevel smoothing result in a convergence behavior that is comparable to the original gradient descent method applied to low resolution transforms in [49].

The concrete results are reported in Figure 7.1. First observe that a simple gradient descent method without smoothing for high resolution transforms stalls way before it reaches an acceptable minimum. Second, the value of the objective function never converges to zero. Instead, the TSI training error typically saturates at a level comparable to the discretization error of the snapshots. This effect will be observed for all numerical experiments in this paper and is discussed more thoroughly in [49].

**7.2. Testing transform resolutions.** Next, we test high versus low resolution transforms near a singularity. We choose the example

$$u(x, \mu) = \begin{cases} \frac{x-l}{-(1-\mu)-l}, & x \leq -(1-\mu), \\ \frac{x-r}{(1-\mu)-r}, & x \geq 1-\mu, \\ 0 & \text{else,} \end{cases} \quad \begin{aligned} \mathcal{P}_m &= \{0.6\}, \\ \mathcal{P}_n &= \{0.6\}, \\ \mathcal{P}_T &= \{0.9\}, \end{aligned} \quad \begin{aligned} \Omega &= [-1.5, 1.5] \\ &= [l, r], \end{aligned}$$

with snapshots and high resolution transforms discretized by piecewise linear functions on 128 cells. For the low resolution transforms we choose  $x \rightarrow X(\eta; \mu, x)$  as third order polynomials. They are defined by four values: two are used to ensure that the boundary positions do not move and two are optimized to track the two jumps. We only compare the low and high resolution transforms at the training parameter  $\mathcal{P}_T$  because by (2.2) the low order transform is independent of the target  $\mu$  for reconstructions from one single snapshot. For the high resolution transforms, we use three fixed point iterations in (4.11) and scaling  $s \in \{0.7^i : i = 0 \dots 3\}$  in increasing size.

In Figure 7.2, left, we see the TSI reconstructions at  $\mathcal{P}_T$ . We see that the high resolution transforms provide a good reconstruction, whereas the low resolution transforms do not. We have chosen the degree of the low resolution transforms sufficiently

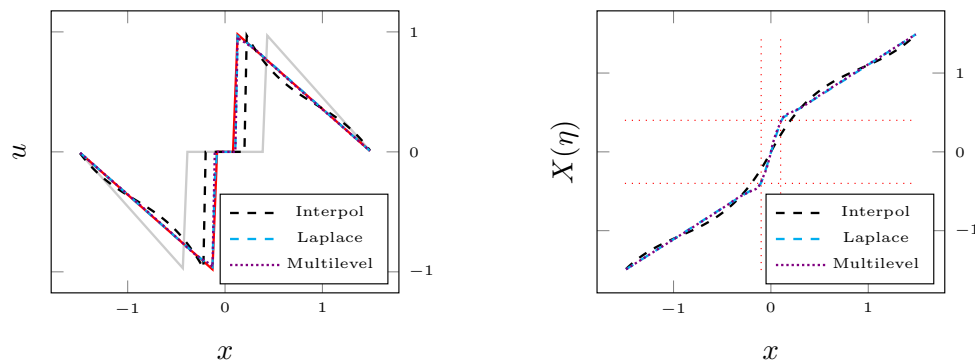


FIG. 7.2. Results for the Example in section 7.2. Left: Snapshots (gray), exact function at  $\mathcal{P}_T$  (red), and TSI with low resolution (black dashed) and high resolution with Laplace smoothing (cyan, dashed) and multilevel smoothing (violet, dotted). Right: The corresponding transform with the same color coding. The red lines are the locations of the jumps of the training snapshot (vertically) and the target snapshots (horizontally).

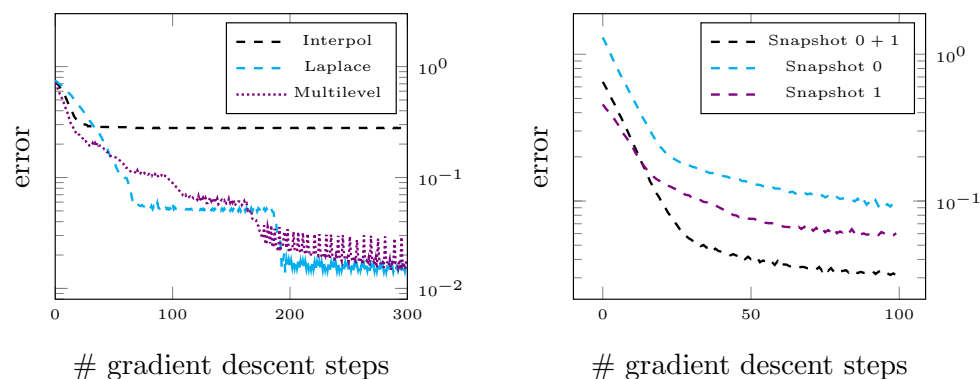


FIG. 7.3. Right: Gradient descent errors for the examples in section 7.2 with the color coding of Figure 7.2. Right: Gradient descent errors for the examples in section 7.4 for reconstruction from both snapshots or just one snapshot.

high so that in principle they can align the jumps. However, this requires a rather sharp gradient, which would create a big oscillation in the polynomial. As a result, the composition  $x \rightarrow u(X(\eta; \mu, x), \eta)$  cannot be linear in regions where the exact solution  $x \rightarrow u(\mu, x)$  is. This can be seen by carefully looking at the plot in the linear regions. Thus, the optimizer must balance the fit in the linear regions with the alignment, which is not possible to do accurately. In contrast, by Figure 7.2, right, the high resolution transforms are almost piecewise linear with kinks at the jump locations. Therefore they can simultaneously align the jumps and provide an accurate fit in the linear regions. The effect can also be clearly seen in the training errors in Figure 7.3, left, where the low resolution transforms saturate much earlier than the high resolution transforms.

**7.3. Testing transform orders.** In this section, we compare different polynomial orders in  $\eta$  of the transforms. To this end, we choose the example

$$u(x, \mu) = \begin{cases} \frac{x-l}{-p(\mu)-l}, & x \leq -p(\mu), \\ \frac{x-r}{p(\mu)-r}, & x \geq p(\mu), \\ 0 & \text{else,} \end{cases} \quad \begin{aligned} \mathcal{P}_m &= \{0.6\}, \\ \mathcal{P}_n &= \{0.6, 0.7\}, \\ \mathcal{P}_T &= \{0.8, 0.96\}, \end{aligned} \quad \begin{aligned} \Omega &= [-1.5, 1.5] \\ &= [l, r], \end{aligned}$$

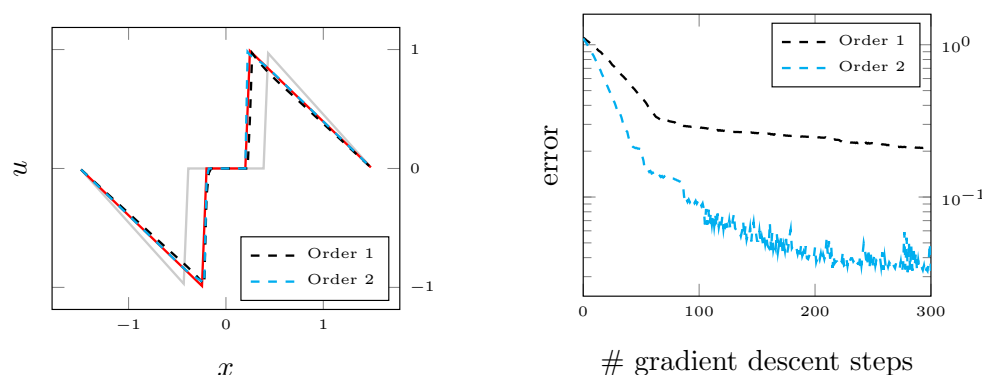


FIG. 7.4. Results for the example in section 7.3. Left: Snapshots (gray), exact function at  $\mathcal{P}_T$  (red), and TSI reconstruction with linear in  $\eta$  (black dashed) and quadratic (cyan dashed) transforms. Right: The corresponding gradient descent errors.

where  $p(\mu)$  is a second order polynomial given by  $p(0.6) = 0.4$ ,  $p(0.8) = 0.3$ , and  $p(1) = 0$ . All the other parameters are set as in the example of section 7.2. Note in particular that the jump-set is now a parabola in parameter so that a first order reconstruction as in the last section can no longer accurately recover  $u$ . For a comparison, we choose Laplace smoothing once with first order transforms (in  $\eta$ ) with  $\mathcal{P}_n = \{0.6\}$  and once with second order with  $\mathcal{P}_n = \{0.6, 0.7\}$ . To be fair, both are trained on the same training set  $\mathcal{P}_T$  as given above.

The results for  $\mu = 0.86$  are shown in Figure 7.4. In the function plot the second order reconstruction looks marginally better, but a look at the training errors reveals that they are better by a factor of about 6. The final training error for the second order case is about twice as much as for the example of section 7.2. This is encouraging because the profiles are identical, only the dynamic in parameter is changed, and in the second order example we add the training error at two training parameters, whereas we use only one in the previous example.

In principle, at this point we should present some plots of convergence rates. However, these are difficult to achieve because of the error saturation. Even if we achieve high order reconstructions, the errors need to be matched by the discretizations of the snapshots in physical space. Since this is quite difficult because of the jumps, we omit a more thorough convergence analysis.

**7.4. Two-dimensional example.** Finally, we consider one example in two spatial dimensions. We use the function

$$u(x, y, \mu) = \begin{cases} 3.0 - 10(1 - \mu)(x^2 + y^2), & \left(\frac{r}{2\mu}\right)^2 + \left(\frac{s}{\mu}\right)^2 \leq 1, \\ 0.3[(1 + \mu)\sin(x) - (1 - \mu)\sin(y)] + 0.2xy & \text{else,} \end{cases}$$

$$\begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} \cos(1.5\mu) & -\sin(1.5\mu) \\ \sin(1.5\mu) & \cos(1.5\mu) \end{pmatrix},$$

which are two parameter dependent smooth functions inside and outside of a rotated and scaled ellipse, with parameters

$$\begin{aligned} \mathcal{P}_m &= \{c(1), c(0.3)\}, \\ \mathcal{P}_n &= \{c(1), c(0.3)\}, & c(\lambda) &= 0.2\lambda + 0.05(1 - \lambda), & \Omega &= [-1, 1]^2. \\ \mathcal{P}_T &= \{c(0.6), c(0)\}, \end{aligned}$$

The snapshots and high resolution transforms are discretized by piecewise bilinear functions on a rectangular grid with  $64 \times 64$  cells. We use three fixed point iterations in (4.11) and scaling  $s \in \{0.7^i : i = 0 \dots 3\}$  in increasing size.

Figure 7.5 shows a solution plot for  $\mu = c(0.05)$  together with some of the curves  $\eta \rightarrow X(\eta; \mu, x)$  originating on the boundary of the correct ellipse. The figure also contains reconstructions from each snapshot in  $\mathcal{P}_m$ , individually. Figure 7.3, right, contains the corresponding training errors. Clearly, the reconstruction from Snapshot 1 is better than the one from Snapshot 0 because the snapshot's parameter is closer to the target  $\mu$ . However, judged by the final training error, the best reconstruction comes from the use of both snapshots, although the order of the inner transform, or equivalently  $\mathcal{P}_n$ , is unchanged. The reason is that the function  $u(x, \mu)$  is smooth and parameter dependent away from the jumps. Therefore, the training error not only depends on the alignment of the jumps but also on the approximation of the smooth part, which is clearly better for two snapshots. This lack of approximation quality in the smooth region is the likely reason why the reconstruction from Snapshot 0 alone does not fully rotate the ellipse, whereas the reconstruction from both snapshots does.

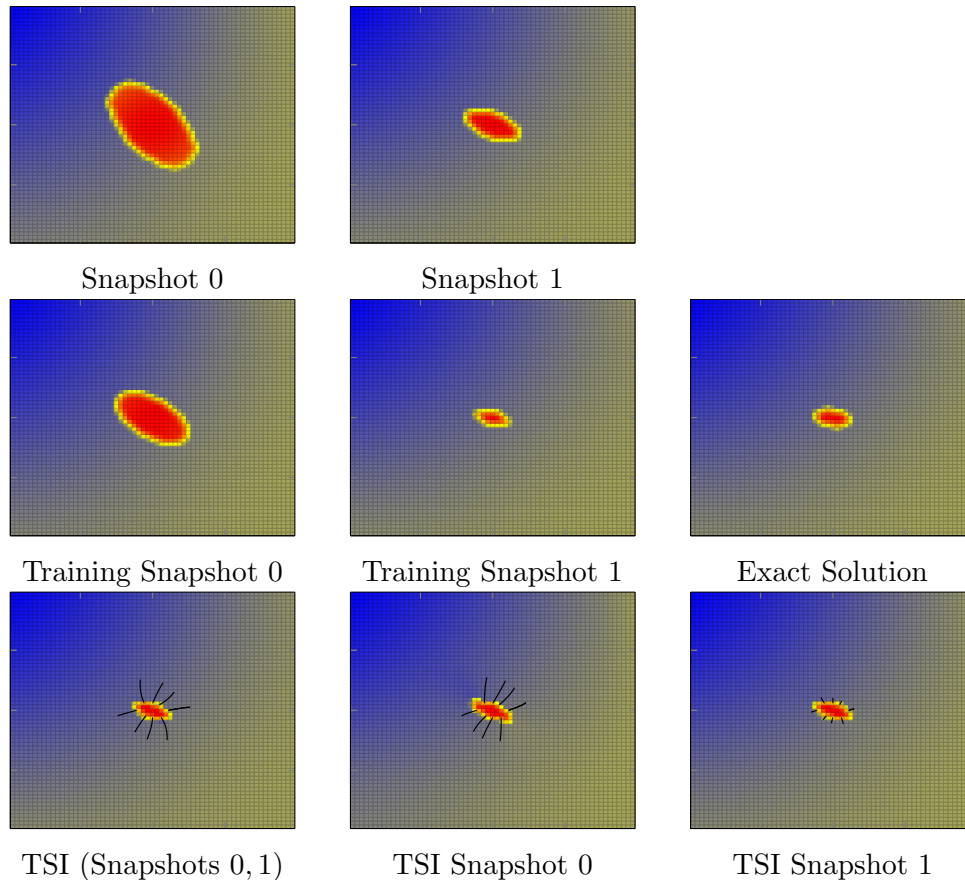


FIG. 7.5. Results for numerical tests in section 7.4. The bottom row shows the TSI reconstruction from two snapshots, only the first, and only the second snapshot. The black lines in the last row indicate several transforms  $\eta \rightarrow X(\eta; \mu, x)$  with  $\eta$  ranging from the first used snapshot to the target parameter  $\mu$  and  $x$  starting on the correct target ellipse.

We see that also in this two-dimensional example, the method can achieve a good alignment with a moderate number of gradient descent steps.

As a final remark, note that for all experiments we use a fixed step size of a conservative magnitude. This is necessary because of the nonsmooth nature of the objective function. Normally, near a minimum, the gradient becomes small so that the update between two gradient descent steps converges to zero. In our case, however, we expect a kink at the minimum, similar to minimizing an absolute value  $|x|$ . Therefore, the gradient does not become small, which is compensated by a small step size. The effect can be seen in the convergence plots, where in the end the error oscillates around an equilibrium with amplitude depending on the step size. On the other hand, far away from the minimum, the conservative choice of the step size hinders a faster convergence.

**Appendix A. Proof of Proposition 5.1.** The following proof of Proposition 5.1 is almost verbatim from [49, section 3]. Let us first estimate  $\|v \circ X_0 - v \circ X_1\|_{L_1(\Omega)}$  for  $X_s(x) := X_s(\eta, \mu, x)$  for some fixed  $\mu$  and  $\eta$  and some function  $v \in C^1(\Omega)$ . Applying the fundamental theorem for line integrals, we obtain

$$(v \circ X_0)(x) - (v \circ X_1)(x) = \int_0^1 v'(X_s(x)) \dot{X}_s(x) \, ds.$$

Note that we can regard  $|\dot{X}(s)|$  as a density and thus define a transition kernel  $\tau(s, A) = \int_A |\dot{X}_s(x)| \, dx$ . Then, with the pushforward  $\kappa(s, A) = (X_s)_* \tau(s, A) = \tau(s, X_s^{-1}(A))$  and the Lebesgue measure  $\lambda_1$  on the real line, the definition of  $S(\mu, \lambda)$  implies

$$(\lambda_1 \otimes \kappa)([0, 1] \times A) = \int_0^1 \kappa(s, A) \, ds = \int_0^1 \int_{X_s^{-1}(A)} |\dot{X}_s(x)| \, dx \, ds \leq S(\mu, \eta) |A|.$$

Thus, using the Fubini theorem for transition kernels we obtain

$$\begin{aligned} (A.1) \quad & \|v \circ X_0 - v \circ X_1\|_{L_1(\Omega)} \\ & \leq \int_{\Omega} \int_0^1 |v'(X_s(x))| |\dot{X}_s(x)| \, ds \, dx = \int_0^1 \int_{\Omega} |v'(X_s(x))| \tau(s, dx) \, ds \\ & = \int_0^1 \int_{\Omega} |v'(y)| \kappa(s, dy) \, ds = \int_{[0, 1] \times \Omega} |v'(y)| \, d(\lambda_1 \otimes \kappa)(s, y) \\ & \leq S(\mu, \eta) \int_{\Omega} |v'(y)| \, dy = S(\mu, \eta) \|v\|_{BV(\Omega)}. \end{aligned}$$

Next, we extend this estimate from functions  $v \in C^1(\Omega)$  to functions  $v \in BV(\Omega)$  by a density argument. To this end, note that for all  $\epsilon > 0$  there is a  $v_{\epsilon} \in C^1(\Omega)$  such that

$$\|v - v_{\epsilon}\|_{L_1(\Omega)} \leq \epsilon, \quad \|v'_{\epsilon}\|_{L_1(\Omega)} \leq \|v\|_{BV(\Omega)} + \epsilon.$$

Thus, to apply a density argument, it suffices to bound  $\|v \circ X_0 - v_{\epsilon} \circ X_0\|_{L_1(\Omega)}$  and  $\|v \circ X_1 - v_{\epsilon} \circ X_1\|_{L_1(\Omega)}$ . Since by assumption  $(X_s)_* \lambda(A) \leq B \lambda(A)$  for all measurable sets  $A$ , we conclude that

$$\begin{aligned} \|v \circ X_0 - v_{\epsilon} \circ X_0\|_{L_1(\Omega)} &= \int_{\Omega} |v(y) - v_{\epsilon}(y)| \, d(X_0)_* \lambda(y) \\ &\leq B \|v - v_{\epsilon}\|_{L_1(\Omega)}. \end{aligned}$$

The bound for  $\|v \circ X_1 - v_{\epsilon} \circ X_1\|_{L_1(\Omega)}$  follows analogously.

Finally, recalling that  $X_s(x) = X_s(\eta; \mu, x)$ , we apply the estimate (A.1) to the full TSI and obtain

$$\begin{aligned} & \|u_m(\cdot, \mu; X_0) - u_m(\cdot, \mu, X_1)\|_{L_1(\Omega)} \\ & \leq \sum_{\eta \in \mathcal{P}_m} |\ell_\eta(\mu)| \|u(X_0(\eta; \mu, \cdot), \eta) - u(X_1(\eta; \mu, \cdot), \eta)\|_{L_1(\Omega)} \\ & \leq \Lambda_n \max_{\eta \in \mathcal{P}_m} \|u(X_0(\eta; \mu, \cdot), \eta) - u(X_1(\eta; \mu, \cdot), \eta)\|_{L_1(\Omega)} \\ & \leq \Lambda_n \max_{\eta \in \mathcal{P}_m} [S(\mu, \eta) \|u(\cdot, \eta)\|_{BV(\Omega)}]. \end{aligned}$$

## REFERENCES

- [1] M. ABADI, A. AGARWAL, P. BARHAM, E. BREVDO, Z. CHEN, C. CITRO, G. S. CORRADO, A. DAVIS, J. DEAN, M. DEVIN, S. GHEMAWAT, I. GOODFELLOW, A. HARP, G. IRVING, M. ISARD, Y. JIA, R. JOZEFOWICZ, L. KAISER, M. KUDLUR, J. LEVENBERG, D. MANÉ, R. MONGA, S. MOORE, D. MURRAY, C. OLAH, M. SCHUSTER, J. SHLENS, B. STEINER, I. SUTSKEVER, K. TALWAR, P. TUCKER, V. VANHOUCHE, V. VASUDEVAN, F. VIÉGAS, O. VINYALS, P. WARDEN, M. WATTENBERG, M. WICKE, Y. YU, AND X. ZHENG, *TensorFlow: Large-scale Machine Learning on Heterogeneous Systems*, <https://www.tensorflow.org/>, 2015.
- [2] R. ABGRALL, D. AMSALLEM, AND R. CRISOVAN, *Robust model reduction by  $L_1$ -norm minimization and approximation via dictionaries: Application to nonlinear hyperbolic problems*, Adv. Model. Simul. Eng. Sci., 3 (2016), <https://doi.org/10.1186/s40323-015-0055-3>.
- [3] W. BANGERTH, R. HARTMANN, AND G. KANSCHAT, *deal.II—a general-purpose object-oriented finite element library*, ACM Trans. Math. Softw., 33 (2007), <https://doi.org/10.1145/1268776.1268779>.
- [4] P. BENNER, S. GUGERCIN, AND K. WILLCOX, *A survey of projection-based model reduction methods for parametric dynamical systems*, SIAM Rev., 57 (2015), pp. 483–531, <https://doi.org/10.1137/130932715>.
- [5] J. BRUNKEN, K. SMETANA, AND K. URBAN, *(parametrized) first order transport equations: Realization of optimally stable petrov–galerkin methods*, SIAM J. Sci. Comput., 41 (2019), pp. A592–A621, <https://doi.org/10.1137/18M1176269>.
- [6] J. BURKE, A. LEWIS, AND M. OVERTON, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, SIAM J. Optim., 15 (2005), pp. 751–779, <https://doi.org/10.1137/030601296>.
- [7] N. CAGNIART, R. CRISOVAN, Y. MADAY, AND R. ABGRALL, *Model Order Reduction for Hyperbolic Problems: A New Framework*, <https://hal.archives-ouvertes.fr/hal-01583224/document>, 2017.
- [8] N. CAGNIART, Y. MADAY, AND B. STAMM, *Model order reduction for problems with large convection effects*, in Contributions to Partial Differential Equations and Applications, B. N. Chetverushkin, W. Fitzgibbon, Y. Kuznetsov, P. Neittaanmäki, J. Periaux, and O. Pironneau, eds., Comput. Methods Appl. Sci. 47, Springer, New York, 2018, pp. 131–150.
- [9] Q.-Y. CHEN, D. GOTTLIEB, AND J. S. HESTHAVEN, *Uncertainty analysis for the steady-state flows in a dual throat nozzle*, J. Comput. Phys., 204 (2005), pp. 378–398, <https://doi.org/10.1016/j.jcp.2004.10.019>.
- [10] A. COHEN, *Numerical Analysis of Wavelet Methods*, JAI Press, Stamford, CT, 2003.
- [11] A. COHEN AND R. DEVORE, *Approximation of high-dimensional parametric PDEs*, Acta Numer., 24 (2015), pp. 1–159, <https://doi.org/10.1017/S0962492915000033>.
- [12] P. CONSTANTINE AND G. IACCARINO, *Reduced Order Models for Parameterized Hyperbolic Conservation Laws with Shock Reconstruction*, Tech. report, Stanford Center for Turbulence Research Annual Research Briefs, 2012, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.727.9918&rep=rep1&type=pdf>.
- [13] W. DAHMEN, *Multiscale and wavelet methods for operator equations*, in Multiscale Problems and Methods in Numerical Simulations, C. Canuto, ed., Springer, New York, 2003, [https://doi.org/10.1007/978-3-540-39810-3\\_2](https://doi.org/10.1007/978-3-540-39810-3_2).
- [14] W. DAHMEN, *How to best sample a solution manifold?*, in Sampling Theory, a Renaissance: Compressive Sensing and Other Developments, G. E. Pfander, ed., Springer, New York, 2015, pp. 403–435, [https://doi.org/10.1007/978-3-319-19749-4\\_11](https://doi.org/10.1007/978-3-319-19749-4_11).

- [15] W. DAHMEN, C. PLESKEN, AND G. WELPER, *Double greedy algorithms: Reduced basis methods for transport dominated problems*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 623–663, <https://doi.org/10.1051/m2an/2013103>.
- [16] B. DESPRÉS, G. POËTTE, AND D. LUCOR, *Robust uncertainty propagation in systems of conservation laws with the entropy closure method*, in Uncertainty Quantification in Computational Fluid Dynamics, Lect. Notes Comp. Sci. Eng. 92, Springer, New York, 2013, pp. 105–149, [https://doi.org/10.1007/978-3-319-00885-1\\_3](https://doi.org/10.1007/978-3-319-00885-1_3).
- [17] L. C. EVANS, *Partial Differential Equations*, 2nd ed., Grad. Stud. Math., AMS, Providence, RI, 2010.
- [18] J.-F. GERBEAU AND D. LOMBARDI, *Reduced-Order Modeling Based on Approximated Lax Pairs*, Tech. report, INRIA, 2012, <http://arxiv.org/pdf/1211.4153v1>.
- [19] J.-F. GERBEAU AND D. LOMBARDI, *Approximated lax pairs for the reduced order integration of nonlinear evolution equations*, J. Comput. Phys., 265 (2014), pp. 246–269, <https://doi.org/10.1016/j.jcp.2014.01.047>.
- [20] B. HAASDONK AND M. OHLBERGER, *Reduced basis method for explicit finite volume approximations of nonlinear conservation laws*, in Proceedings of the 12th International Conference on Hyperbolic Problems: Theory, Numerics, Application, College Park, MD, 2008.
- [21] B. HAASDONK AND M. OHLBERGER, *Reduced basis method for finite volume approximations of parametrized linear evolution equations*, ESAIM Math. Model. Numer. Anal., 42 (2008), pp. 277–302, <https://doi.org/10.1051/m2an:2008001>.
- [22] J. S. HESTHAVEN, G. ROZZA, AND B. STAMM, *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, Springer, New York, 2015.
- [23] A. IOLLO AND D. LOMBARDI, *Advection modes by optimal mass transfer*, Phys. Rev. E, 89 (2014), 022923, <https://doi.org/10.1103/PhysRevE.89.022923>.
- [24] S. JIN, D. XIU, AND X. ZHU, *A well-balanced stochastic galerkin method for scalar hyperbolic balance laws with random inputs*, J. Sci. Comput., 67 (2016), pp. 1198–1218, <https://doi.org/10.1007/s10915-015-0124-2>.
- [25] K. C. KIWIEL, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Math. 1133, Springer-Verlag, Berlin, 1985.
- [26] Q. LI, J.-G. LIU, AND R. SHU, *Sensitivity Analysis of Burgers' Equation with Discontinuous Shocks*, <https://arxiv.org/abs/1708.04332>, 2018.
- [27] S. MISHRA AND C. SCHWAB, *Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data*, Math. Comp., 81 (2012), pp. 1979–2018, <https://doi.org/10.1090/S0025-5718-2012-02574-9>.
- [28] R. MOJGANI AND M. BALAJEWICZ, *Lagrangian Basis Method for Dimensionality Reduction of Convection Dominated Nonlinear flows*, <https://arxiv.org/abs/1701.04343>, 2017.
- [29] N. J. NAIR AND M. BALAJEWICZ, *Transported snapshot model order reduction approach for parametric, steady-state fluid flows containing parameter-dependent shocks*, Internat. J. Numer. Methods Engrg., 117 (2019), pp. 1234–1262, <https://doi.org/10.1002/nme.5998>.
- [30] N.-C. NGUYEN, G. ROZZA, AND A. PATERA, *Reduced basis approximation and a posteriori error estimation for the time-dependent viscous burgers' equation*, Calcolo, 46 (2009), pp. 157–185, <https://doi.org/10.1007/s10092-009-0005-x>.
- [31] M. OHLBERGER AND S. RAVE, *Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing*, C. R. Math., 351 (2013), pp. 901–906, <https://doi.org/10.1016/j.crma.2013.10.028>.
- [32] M. OHLBERGER AND S. RAVE, *Reduced basis methods: Success, limitations and future challenges*, in Proceedings of the Conference Algoritmy 2016, D. Handlovičová and A. Ševčovič, eds., Publishing House of Slovak University of Technology, Bratislava, 2016, pp. 1–12.
- [33] P. PACCARINI AND G. ROZZA, *Stabilized reduced basis method for parametrized advection-diffusion PDEs*, Comput. Methods Appl. Mech. Engrg., 274 (2014), pp. 1–18, <https://doi.org/10.1016/j.cma.2014.02.005>.
- [34] B. PEHERSTORFER, *Model Reduction for Transport-Dominated Problems via Online Adaptive Bases and Adaptive Sampling*, <https://arxiv.org/abs/1812.02094>, 2018.
- [35] P. PETTERSSON, Q. ABBAS, G. IACCARINO, AND J. NORDSTRÖM, *Efficiency of shock capturing schemes for Burgers' equation with boundary uncertainty*, in Proceedings of the 7th South African Conference on Computational and Applied Mechanics, 2010.
- [36] P. PETTERSSON, G. IACCARINO, AND J. NORDSTRÖM, *A stochastic Galerkin method for the Euler equations with Roe variable transformation*, J. Comput. Phys. A, 257 (2014), pp. 481–500, <https://doi.org/10.1016/j.jcp.2013.10.011>.
- [37] R. PULCH AND D. XIU, *Generalised polynomial chaos for a class of linear conservation laws*, J. Sci. Comput., 51 (2012), pp. 293–312, <https://doi.org/10.1007/s10915-011-9511-5>.
- [38] A. QUARTERONI, A. MANZONI, AND F. NEGRI, *Reduced Basis Methods for Partial Differential Equations*, Springer, New York, 2015.



- [39] J. REISS, P. SCHULZE, J. SESTERHENN, AND V. MEHRMANN, *The shifted proper orthogonal decomposition: A mode decomposition for multiple transport phenomena*, SIAM J. Sci. Comput., 40 (2018), pp. A1322–A1344, <https://doi.org/10.1137/17M1140571>.
- [40] D. RIM AND K. MANDLI, *Displacement interpolation using monotone rearrangement*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 1503–1531, <https://doi.org/10.1137/18M1168315>.
- [41] D. RIM AND K. MANDLI, *Model Reduction of a Parametrized Scalar Hyperbolic Conservation Law Using Displacement Interpolation*, <https://arxiv.org/abs/1805.05938>, 2018.
- [42] D. RIM, S. MOE, AND R. J. LEVEQUE, *Transport reversal for model reduction of hyperbolic partial differential equations*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 118–150, <https://doi.org/10.1137/17M1113679>.
- [43] C. W. ROWLEY AND J. E. MARSDEN, *Reconstruction equations and the Karhunen–Loève expansion for systems with symmetry*, Phys. D, 142 (2000), pp. 1–19, [https://doi.org/10.1016/S0167-2789\(00\)00042-7](https://doi.org/10.1016/S0167-2789(00)00042-7).
- [44] P. SCHULZE, J. REISS, AND V. MEHRMANN, *Model reduction for a pulsed detonation combustor via shifted proper orthogonal decomposition*, in Active Flow and Combustion Control 2018, R. King, ed., Springer, New York, 2019, pp. 271–286.
- [45] T. TADDEI, S. PEROTTO, AND A. QUARTERONI, *Reduced basis techniques for nonlinear conservation laws*, ESAIM Math. Model. Numer. Anal., 49 (2015), pp. 787–814, <https://doi.org/10.1051/m2an/2014054>.
- [46] J. TRYÖEN, O. L. MAÎTRE, AND A. ERN, *Adaptive anisotropic spectral stochastic methods for uncertain scalar conservation laws*, SIAM J. Sci. Comput., 34 (2012), pp. A2459–A2481, <https://doi.org/10.1137/120863927>.
- [47] M. ULBRICH, *Optimization methods in Banach spaces*, in Optimization with PDE Constraints, M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, eds., Springer, New York, 2009, <https://www.springer.com/us/book/9781402088384#>.
- [48] G. WELPER, *h and hp-Adaptive Interpolation by Transformed Snapshots for Parametric and Stochastic Hyperbolic PDEs*, <https://arxiv.org/abs/1710.11481>, 2017.
- [49] G. WELPER, *Interpolation of functions with parameter dependent jumps by transformed snapshots*, SIAM J. Sci. Comput., 39 (2017), pp. A1225–A1250, <https://doi.org/10.1137/16M1059904>.
- [50] M. YANO, A. T. PATERA, AND K. URBAN, *A space-time hp-interpolation-based certified reduced basis method for Burgers’ equation*, Math. Models Methods Appl. Sci., 24 (2014), pp. 1903–1935, <https://doi.org/10.1142/S0218202514500110>.