DATA-DRIVEN SNAPSHOT CALIBRATION VIA MONOTONIC FEATURE MATCHING*

NEERAJ SARNA†, JAN GIESSELMANN‡, AND PETER BENNER§

Abstract. Snapshot matrices of hyperbolic equations have a slow singular value decay, resulting in inefficient reduced-order models. We develop on the idea of inducing a faster singular value decay by computing snapshots on a transformed spatial domain, or the so-called snapshot calibration/transformation. We are particularly interested in problems involving shock collision, shock rarefaction-fan collision, shock formation, etc. For such problems, we propose a realizable algorithm to compute the spatial transform using monotonic feature matching. We consider discontinuities and kinks as features, and by carefully partitioning the parameter domain, we ensure that the spatial transform has properties that are desirable both from a theoretical and an implementation standpoint. We use these properties to prove that our method results in a fast m-width decay of a so-called calibrated manifold. A crucial observation we make is that due to calibration, the m-width does not only depend on m but also on the accuracy of the full order model, which is in contrast to elliptic and parabolic problems that do not need calibration. The method we propose only requires the solution snapshots and not the underlying partial differential equation (PDE) and is therefore, data-driven. We perform several numerical experiments to demonstrate the effectiveness of our method.

1 Introduction

Several problems of practical interest are modeled using parameterized PDEs of the form

$$\mathcal{L}u(x,\mu) = 0 \quad \forall (x,\mu) \in \Omega \times D. \tag{1.1}$$

Here, $\mathcal{L}$ is some differential operator, $\mu \in D$ is some parameter which can encode, for example, different material properties, and $x \in \Omega \subset \mathbb{R}^d$ is a space point. We refer to the book [20] for an elaborate discussion on different parameterized PDEs. Note that D can contain time and in the model problem that we consider later, it is indeed the time domain. Nevertheless, the present discussion applies to general parameter domains. Often, an exact solution to the above problem is unavailable and one seeks an approximation $u(\cdot,\mu) \approx u_M(\cdot,\mu)$ in a finite-dimensional space X_M spanned by some basis $\{\phi_i\}_{i=1,\dots,M}$. The approximation $u_M(\cdot,\mu)$ is what we refer to as the full-order model (FOM). We assume that $X_M \subset L^2(\Omega)$.

In a multi-query setting, where a solution is required at several different parameter instances, computing a FOM is computationally expensive and infeasible. Reduced-order models (ROMs) aim to reduce this cost by splitting the solution algorithm into an online-offline phase. A broad description of these two phases is as follows—see [2] for further details. First, in the offline phase, one computes a snapshot matrix $\mathcal{S} \in \mathbb{R}^{M\times K}$ given as

$$\mathcal{S} := (U_M(\mu_1), \dots, U_M(\mu_K)),$$

*Submitted to the editors xxxx

Funding: N.S and P.B are supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) in the joint project "MathEnergy - Mathematical Key Technologies for Evolving Energy Grids", sub-project: Model Order Reduction (Grant number: 0324019B). J.G is supported by DFG grant SFB TRR 154, project C05.

†Corresponding author, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr 1, 39106, Magdeburg, Germany, sarna@mpi-magdeburg.mpg.de.

‡Department of Mathematics, Technical University of Darmstadt, Darmstadt, 64293, Germany, giesselmann@mathematik.tu-darmstadt.de.

§Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr 1, 39106, Magdeburg, Germany, and Faculty for Mathematics, Otto Von Guericke University Magdeburg, Gustav-Adolf-Str., 39106, Magdeburg benner@mpi-magdeburg.mpg.de.

where $U_M(\mu) \in \mathbb{R}^M$ is a vector containing all the degrees of freedom of $u_M(\cdot,\mu)$ i.e., $(U_M(\mu))_j := \langle \phi_j, u_M(\cdot,\mu)\rangle_{L^2(\Omega)}$ where $\{\phi_j\}_j$ is a set of basis functions for X_M. The parameters $\{\mu_i\}_{i=1,\dots,K}$ can be chosen uniformly, randomly, or using a greedy procedure based on an a-posteriori error indicator [6, 7, 14, 29].

In the online phase, one approximates $U_M(\mu)$ in the span of the first m left singular vectors of $\mathcal{S}$, or the so-called Proper-Orthogonal-Decomposition (POD) modes of $\mathcal{S}$. We collect these vectors in the matrix $\mathcal{U}_m(\mathcal{S})$ and with $U_m^{\text{red}}(\mu)$ we represent an approximation to $U_M(\mu)$ in range($\mathcal{U}_m(\mathcal{S})$). The online phase is efficient only if any given error tolerance of practical interest

$$\|U_m^{\text{red}}(\mu) - U_M(\mu)\|_2 \leq \texttt{TOL}, \tag{1.2}$$

can be achieved with a sufficiently small value (preferably $\ll M$) of m.

At least empirically, the singular value decay rate of the snapshot matrix is a good indicator of the decay rate of the error in (1.2); see [20, 22, 26]. Let $\sigma_i(\mathcal{S})$ denote the i-th singular value of $\mathcal{S}$. Then, for all $i \in \{1,\dots,K\}$, we find

$$\|U_M(\mu_i) - \Pi_{\text{range}(\mathcal{U}_m(\mathcal{S}))} U_M(\mu_i)\|_2 \leq \|\mathcal{S} - \mathcal{U}_m(\mathcal{S})\mathcal{U}_m(\mathcal{S})^T\mathcal{S}\|_F = \underbrace{\sqrt{\sum_{i=m+1}^{K} \sigma_i(\mathcal{S})^2}}_{=:\Xi_m(\mathcal{S})}. \tag{1.3}$$

Above, $\|\cdot\|_F$ represents the Frobenius norm, $\Pi_\square$ represents an orthogonal projection operator with $\square$ being a place holder for some finite-dimensional space, and $(\cdot)^T$ represent the transpose of a matrix. If $\{\mu_i\}_{i=1,\dots,K}$ is sufficiently dense in D then, with the above relation, we expect the error in (1.2) to decay at a similar rate as $\Xi_m(\mathcal{S})$.

For hyperbolic problems, there is ample numerical evidence (also provided by the current article) supporting that $\Xi_m(\mathcal{S})$ decays slowly resulting in an inefficient ROM [3, 17, 19, 22]. Therefore, the first step toward developing an efficient ROM is to induce a faster singular value decay in the snapshot matrix, or to so-called calibrate the snapshot matrix. Following the works in [3, 22, 32], we perform calibration by computing snapshots on a transformed domain. This results in a calibrated snapshot matrix that reads

$$\begin{gathered} \mathcal{S}_{\text{calib}} := (U_{\text{calib,M}}(\mu_1), \dots, U_{\text{calib,M}}(\mu_K)), \\ \text{where} \quad (U_{\text{calib,M}}(\mu))_j := \langle \phi_i, u_M(\varphi_M(\cdot,\mu),\mu)\rangle_{L^2(\Omega)}. \end{gathered} \tag{1.4}$$

Above, $\varphi_M(\cdot,\mu) : \Omega \to \Omega$ is a spatial transform that satisfies

$$\begin{aligned} &\text{(P1) } \varphi_M(\cdot,\mu) \text{ is a homeomorphism,} \\ &\text{(P2) } \|D_x\varphi_M(\cdot,\mu)^{-1}\|_{L^\infty(\Omega)}, \|D_x\varphi_M(\cdot,\mu)\|_{L^\infty(\Omega)} \leq \mathcal{K}_1, \end{aligned} \tag{1.5}$$

where, $\mathcal{K}_1 > 1$ is a user-defined constant and $D_\square$ denotes a weak-derivative with $\square$ being a place holder for a variable. We can think of φ_M as a way of artificially introducing the desired regularity in the snapshots along the parameter domain, which eventually results in a fast singular value decay. For further clarification, we refer to the numerous examples and arguments in [3, 24, 32] and to the later sections of our work. The properties (P1) and (P2) are desirable from both a theoretical and a numerical implementation standpoint. They will be particularly helpful in studying the m-width of a so-called calibrated manifold defined below. Later sections provide further elaboration.

Note that snapshot calibration is an offline step. In the online phase, we can use the POD modes of $\mathcal{S}_{\text{calib}}$ to approximate $U_{\text{calib,M}}(\mu)$ and then recover an approximation to $U_M(\mu)$ using $\varphi_M(\cdot,\mu)^{-1}$, or its approximation. Development of a PDE-based online algorithm that is stable, efficient and competitive with finite-element/volume/difference type approximations is another challenging task and we plan to tackle it in the future—preliminary, but noteworthy, work in this direction can be found in [3, 16, 25, 27].

We propose a data-driven and feature-matching-based algorithm to compute φ_M that satisfies (P1) and (P2). Let us elaborate on what we mean by feature matching. A feature is either a discontinuity or a kink (defined precisely later) in a snapshot $u_M(\cdot,\mu)$, and with $z_M(\mu)$ we represent its spatial location. We want the feature locations in $u_M(\varphi_M(\cdot,\mu_i),\mu_i)$ to coincide with those in some reference snapshot $u_M(\cdot,\mu_{\text{ref}})$ i.e.,

$$\varphi_M(z_M(\mu_{\text{ref}}),\mu_i) = z_M(\mu_i), \quad \forall i \in \{1,\ldots,K\}. \tag{1.6}$$

We extend $\varphi_M(\cdot,\mu_i)$ to Ω by piecewise linear interpolation. We allow for multiple-features, feature interaction and feature formation. In order to deal with these cases, we propose an adaptive selection of the reference snapshot $u_M(\cdot,\mu_{\text{ref}})$ such that (P1) and (P2) are satisfied. In Section 2 we discuss feature matching in further detail. Note that due to its data-driven nature, our algorithm treats all discontinuities the same i.e., it does not differentiate between shocks and contact discontinuities.

Most of the previous model-order reduction methods for hyperbolic equations were restricted to either periodic or extrapolated boundary conditions—for instance, see [15, 18, 22–24]. The reason being that these works relied on either a (or multiple) spatial shift, a Lie group action, or an optimal transport map, all of which have some restrictions on the boundary conditions. We show that general time-dependent boundary conditions are naturally included in the feature matching framework by defining the boundary points as additional features. The numerical experiments included in Section 5 showcase that our method works well for time-dependent boundary conditions.

In an abstract sense, an approximation of $U_{\text{calib,M}}(\mu)$ in the POD modes of $\mathcal{S}_{\text{calib}}$ is a linear approximation of the so-called calibrated snapshot manifold defined as

$$\mathcal{M}_{\text{calib},M}(D) := \{\Pi_{X_M} u_M(\varphi_M(\cdot,\mu),\mu) \;:\; \mu \in D\}. \tag{1.7}$$

A linear approximation can be accurate only if the m-width of $\mathcal{M}_{\text{calib},M}(D)$ decays fast. We prove that this is indeed the case for the calibrated manifold resulting from feature matching. We provide a bound for the m-width of $\mathcal{M}_{\text{calib},M}(D)$ in case the FOM is a finite volume (FV) scheme. Our bound depends explicitly on both m and M. To the best of our knowledge, no earlier works provide such a bound, making our work the first of its kind that provides a theoretical justification for feature matching. Note that, compared to the definition of the calibrated manifold proposed in [3], our definition is closer to what is actually used in practice—our definition uses the FOM whereas the one in [3] uses the exact solution of the evolution equation (1.1). The bounds on the m-width are discussed in detail in Section 3.

We propose to match both kinks and discontinuities. Usually, one would only match discontinuities—see for instance [3, 32]. This could be because (i) kinks get smeared out due to numerical dissipation and go undetected, or (ii) because, despite the kinks being detectable, they are not included in the set of features. For the first case, we show that, due to smearing, the FOM has sufficient regularity to ensure a fast m-width decay. For the second case, we show that matching both kinks and discontinuities provides a better calibration than only discontinuity matching. Precisely, in Section 3, we prove that both kink and discontinuity matching results in a calibrated manifold with an m-width that is $\mathcal{O}(m^{-2})$, which is $\mathcal{O}(m^{-1})$ times better than what only discontinuity matching offers. To summarize, we establish that if kinks are detectable, then it is advantageous to include them in the feature set.

In Section 5, we perform several numerical experiments showcasing the effectiveness of our method. Mindful of the above discussion, we consider highly accurate approximations in X_M where both kinks and discontinuities can be identified. For this reason, we consider the best-approximation in X_M and show that kink and discontinuity matching results in a fast singular-value decay and that both kink and discontinuity matching is better than only discontinuity matching.

Our method is explicit in the sense that we explicitly compute the feature locations and match them. In the context of model-order reduction, explicit methods have been used before (see [5, 28]), but never for problems involving multiple-features and feature interaction. Rather than using an explicit method, one can also solve an optimization problem and expect the features to be matched implicitly [16, 32]. The following reasons motivated our choice of an explicit method. Firstly, the optimization problem in implicit methods is (usually) non-convex and non-linear. If the samples $\{\mu_i\}$ are not chosen carefully, then the minimization problem can get stuck in sub-optimal local minima, resulting in a $\mathcal{S}_{\text{calib}}$ with a slow singular value decay. Secondly, explicit methods rely on shock tracking/identifying techniques that are well-studied for hyperbolic problems [4]. Thirdly, in explicit methods, it is easier to quantify (at least empirically) the error in identifying the true feature location, which is helpful in quantifying the m-width decay rate. Lastly, with an access to feature locations, it easier to satisfy (P1) and (P2), which otherwise have to be included as constraints in the optimization problem. To the best of our knowledge, none of the implicit methods can impose such constraints.

We mention that apart from snapshot calibration, in the context of hyperbolic equations, other strategies to construct an accurate approximation space include online adaptivity of basis [11, 19], embedding of the solution manifold in the Wasserstein metric space [9] and the use of auto-encoders [13]. Comparison of the approximation space resulting from snapshot calibration to these other works is an interesting question in its own right and we plan to tackle it in the future.

2 Feature Matching As a model problem, we interpret time as a parameter and consider the time-dependent hyperbolic conservation law in one space dimension given by

$$
\begin{aligned}
&\partial_t u(x,t) + \partial_x f(u(x,t)) = 0, \quad \forall (x,t) \in \Omega \times D, \quad u(x,t=0) = u_0(x) \quad \forall x \in \Omega, \\
&\qquad\qquad u(x,t) = \mathcal{G}(x,t), \quad \forall (x,t) \in \partial\Omega \times D.
\end{aligned} \tag{2.1}
$$

Above, $D := [0,T]$ is the time-domain with some final time $T > 0$, u_0 is the initial data and $\mathcal{G}$ is some (given) boundary data. We interpret the boundary conditions in a weak-sense as described in [8]. The solution vector u maps $\Omega \times D$ to $\mathbb{R}^Q$ and $f : \mathbb{R}^Q \to \mathbb{R}^Q$ is a so-called flux function, where we allow $Q \geq 1$. We restrict to a one-dimensional spatial domain with $\Omega := (x_{\min}, x_{\max}) \subset \mathbb{R}$. We consider a FV approximation space X_M where we partition Ω into M sub-intervals of the same size

$\Delta x = (x_{\max} - x_{\min})/M$ i.e.,

$$\Omega = \bigcup_{i=1}^{M} \mathcal{I}_i, \quad |\mathcal{I}_i| = \Delta x. \tag{2.2}$$

For notational simplicity, we consider a uniform spatial grid—an extension to non-uniform grids is straightforward.

For notational simplicity, we restrict our discussion to scalar problems i.e., $Q = 1$ in (2.1). An extension to systems follows by applying the proposed method to every component of the solution vector. We find φ_M such that the feature locations in $u_M(\varphi_M(\cdot, t_k), t_k)$ match to those in some reference snapshot $u_M(\cdot, t_{\text{ref}})$. The methodology used to compute φ_M drives the choice for $u_M(\cdot, t_{\text{ref}})$. For the present discussion, we choose

$$t_{\text{ref}} = 0. \tag{2.3}$$

The motivation behind our choice becomes clear as we proceed. First, we define the notion of a feature. Note that the definition implicitly assumes that the exact solution has a finite number of features, a reasonable assumption for most problems of practical interest.

Definition 2.1 (Feature). *A feature is either a discontinuity or a kink in the solution. For any $t \in D$, let there be $p(t) \in \mathbb{N}$ of such features. With $z_i(t)$ we represent the i-th feature location in $u(\cdot, t)$. Furthermore, with $z_{M,i}(t)$ we denote an approximation to $z_i(t)$ computed using $u_M(\cdot, t)$. Assuming that between the locations of discontinuities $u(\cdot, t)$ has a weak derivative, we define a kink location as a space point where this weak derivative is discontinuous. Furthermore, we define the boundary points of Ω as two additional feature locations i.e.,*

$$z_0(t) = z_{M,0}(t) = x_{\min}, \quad z_{p(t)+1}(t) = z_{M,p(t)+1}(t) = x_{\max}. \tag{2.4}$$

Without loss of generality, we assume the ordering

$$z_{M,0}(t) < z_{M,1}(t) < \cdots < z_{M,p(t)+1}(t).$$

We want to match the same type of features i.e., kinks with kinks and discontinuities with discontinuities. To distinguish between these two types of features, we associate an identifier with a feature location and define it in the following.

Definition 2.2 (Identifier). *The identifier $\Gamma : \Omega \to \{0, 1\}$ acts on a feature location and returns zero or one depending on whether there is a discontinuity or a kink at that location, respectively. For convenience, we collect all the identifiers in a vector $\gamma_M(t_k) \in \mathbb{R}^{p(t)}$ defined as $(\gamma_M(t_k))_i = \Gamma(z_{M,i}(t_k))$.*

We ask the following question. For some $t \in \{t_l\}_{l=1,\dots,K}$, given a snapshot $u_M(\cdot, t)$ and a reference snapshot $u_M(\cdot, t_{\text{ref}})$, does there exist a φ_M that satisfies (P1) and (P2) and, in the sense of (1.6), matches the features between $u_M(\varphi_M(\cdot, t), t)$ and $u_M(\cdot, t_{\text{ref}})$? We show that the answer to this question is yes if the following three conditions are satisfied

$$\begin{gathered} \text{(C1)} \quad p(t) = p(t_{\text{ref}}), \quad \text{(C2)} \quad \gamma_M(t) = \gamma_M(t_{\text{ref}}), \\ \text{(C3)} \quad \frac{1}{\mathcal{K}_1} \leq \frac{|z_{M,i+1}(t_{\text{ref}}) - z_{M,i}(t_{\text{ref}})|}{|z_{M,i+1}(t) - z_{M,i}(t)|} \leq \mathcal{K}_1 \quad \forall i \in \{0, \dots, p(t)\}. \end{gathered} \tag{2.5}$$

Above, $\mathcal{K}_1$ is the same as that defined in (1.5). The conditions (C1) and (C2) imply that the two snapshots have the same number and the same types of features. Furthermore, relative to $u_M(\cdot, t_{\text{ref}})$, (C3) prevents the features in $u_M(\cdot, t)$ from either coming too close or from moving very far away from each other. One can interpret the conditions (C1)-(C3) as a way of measuring the similarity of a snapshot to the reference snapshot, and if similar, we can find a φ_M between the two snapshots that satisfies (P1) and (P2). If (C1)-(C3) is satisfied, then we say that $u_M(\cdot, t)$ matches to $u_M(\cdot, t_{\text{ref}})$ and for convenience, represent the matching by the notation

$$\text{(C1), (C2) and (C3)} \quad \Leftrightarrow \quad u_M(\cdot, t) \leftrightarrow u_M(\cdot, t_{\text{ref}}). \tag{2.6}$$

2.1 Construction of φ_M Assume that $u_M(\cdot, t) \leftrightarrow u_M(\cdot, t_{\text{ref}})$ then feature matching provides

$$\varphi_M(z_{M,i}(t_{\text{ref}}), t) = z_{M,i}(t), \quad \forall i \in \{0, \ldots, p(t)+1\}.$$

Note that (C2) ensures that the above relation does not match discontinuities to kinks or vice-versa. Furthermore, including the endpoints of Ω as features implies that $\varphi_M(\partial\Omega, t) = \partial\Omega$. To extend $\varphi_M(\cdot, t)$ to Ω, we perform a piecewise linear interpolation, which, for $i \in \{0, \ldots, p(t)\}$ and $x \in [z_{M,i}(t_{\text{ref}}), z_{M,i+1}(t_{\text{ref}})]$, provides

$$\begin{aligned} \varphi_M(x, t) = &\left(\frac{x - z_{M,i}(t_{\text{ref}})}{z_{M,i+1}(t_{\text{ref}}) - z_{M,i}(t_{\text{ref}})}\right) z_{M,i+1}(t) \\ &+ \left(\frac{x - z_{M,i+1}(t_{\text{ref}})}{z_{M,i}(t_{\text{ref}}) - z_{M,i+1}(t_{\text{ref}})}\right) z_{M,i}(t). \end{aligned} \tag{2.7}$$

Trivially, $\varphi_M(\cdot, t)$ is continuous upto the boundary with $\varphi_M(\partial\Omega, t) = \partial\Omega$, which, due the ordering of the features in Definition 2.1, implies that $\varphi_M(\cdot, t)$ is strictly increasing. Thus, $\varphi_M(\cdot, t)$ is a homeomorphism. Furthermore, the following relation and (C3) provides (P2). For all $t \in \{t_l\}_{l=1,\ldots,K}$ and $i \in \{0, \ldots, p(t)\}$, we find

$$\begin{aligned} \frac{1}{\mathcal{K}_1} \leq D_x\varphi_M(\cdot, t)|_{(z_{M,i}(t_{\text{ref}}), z_{M,i+1}(t_{\text{ref}}))} = &\frac{|z_{M,i+1}(t_{\text{ref}}) - z_{M,i}(t_{\text{ref}})|}{|z_{M,i+1}(t) - z_{M,i}(t)|} \\ \leq &\mathcal{K}_1. \end{aligned} \tag{2.8}$$

We elaborate on why it is desirable to have (P1) and (P2).

1. *Onto property:* as mentioned in the introduction, eventually in an online phase we want to approximate the calibrated snapshot $U_{\text{calib,M}}(t)$ in span of the POD modes of $\mathcal{S}_{\text{calib}}$. We expect such an approximation to be accurate if $\varphi_M(\cdot, t)$ is an onto function. We also refer to the arguments made in [32] and our analysis in Section 3 indicating that the onto property is desirable. At least intuitively, the following example further elaborates on the desirability of the onto property. Suppose that the characteristics curves originating from $t = 0$ pass through every point in Ω for some $t^* \in D$. Then a $\varphi_M(\cdot, t^*)$ that is not onto, will discard some information in $u_M(\cdot, t^*)$, which is undesirable and inconsistent with the characteristics.
2. *Invertibility:* the analysis in Section 3 indicates that the invertibility of $\varphi_M(\cdot, t)$ is desirable.
3. *Continuity and monotonicity:* continuity and monotonicity of $\varphi_M(\cdot, t)$ ensure that, as compared to $u_M(\cdot, t)$, no new discontinuities appear in $u_M(\varphi_M(\cdot, t), t)$. For the same reason, $\varphi_M(\cdot, t)^{-1}$ should also be continuous. Points (1)-(3) imply that $\varphi_M(\cdot, t)$ should be a homeomorphism i.e., it should satisfy (P1).

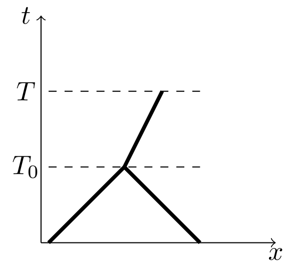


Fig. 1: Time trajectory of two discontinuities that merge to form a single discontinuity.

4. *Bounds on the derivatives:* the bound on the m-width, which we present later in Section 3, scales with $\|D_x\varphi_M(\cdot,t)\|_{L^\infty(\Omega)}$ and $\|D_x\varphi_M(\cdot,t)^{-1}\|_{L^\infty(\Omega)}$, which motivates (P2).

2.2 Open questions The above formulation leaves the following questions open. The rest of the article (tries) to answer them.

- How to handle the cases where (C1)-(C3) are not satisfied?
- How to determine the feature locations in practise?
- Why does feature matching result in a fast singular value decay?

In relation to the first question, it is easy to violate (C1). Consider Figure 1 that shows the time-trajectory of two discontinuities in an otherwise smooth function. At $t = T_0$, two discontinuities interact to form a single one, changing the value of $p(t)$ from two to one. We handle such cases by partitioning $\{t_l\}_{l=1,\ldots,K}$ into subsets and choosing (different) suitable reference snapshots such that (C1)-(C3) is locally satisfied in each of the subsets. The details are discussed in Subsection 2.3.

To answer the third question rigorously, we need decay estimates for the singular values of the calibrated snapshot matrix $\mathcal{S}_{\text{calib}}$. Such estimates are unavailable, as yet. However, later (in Section 3), we prove that feature matching results in a fast m-width decay of the calibrated manifold defined in (1.7). At least empirically, a fast m-width decay results in a fast singular value decay of the snapshot matrix. Our expectation is corroborated by the numerous numerical experiments (performed in Section 5) where we empirically establish a fast singular value decay in the calibrated snapshot matrix.

2.3 Adaptive reference snapshot selection Recall the conditions (C1)-(C3) given in (2.5). A snapshot $u_M(\cdot,t_k)$ cannot be matched to $u_M(\cdot,t_{\text{ref}})$ if either of these three conditions are violated. To handle such cases, we partition $\{t_l\}_{l=1,\ldots,K}$ into subsets containing subsequent time-instances. For each of these subsets, we find a different t_{ref} such that (C1)-(C3) is satisfied locally. The details are as follows.

We start with introducing the following notation.

Definition 2.3 (Time partitions). *We partition $\{t_l\}_{l=1,\ldots,K}$ into $N \in \mathbb{N}$ subsets (where N will be an outcome of the snapshot selection algorithm). We denote the i-th subset by $[t]_i$. With $r(i) \in \mathbb{N}$ we denote the number of elements in $[t]_i$, and with $t_{\text{ref(i)}}$ we denote the first element of $[t]_i$, where* $\text{ref}(i)$ *is an index in $\{1,\ldots,K\}$. Under this notation, $[t]_i$ reads*

$$[t]_i := \{t_{\text{ref(i)}},\ldots,t_{\text{ref(i)}+r(i)-1}\}.$$

Algorithm 2.1 presents the reference snapshot selection algorithm. The algorithm starts with the initial data as the reference snapshot, compares it to the subsequent snapshots and, in case matching is not possible, updates the reference snapshot. In addition to checking (C1)-(C3), the algorithm enforces a lower bound on the minimum distance between the features. At least empirically, one observes that the error in computing a feature location (i.e., $|z_j(t) - z_{M,j}(t)|$) is of the order of the grid-size Δx. Therefore, to have a reliable calibration we need

$$\mathcal{K}_2 \Delta x \leq \min_j |z_{M,j+1}(t) - z_{M,j}(t)| \quad \text{where} \quad \mathcal{K}_2 \geq 2. \tag{2.9}$$

The output of the algorithm are the time-indices $\{\mathrm{ref}(i)\}_{i=1,\ldots,N}$ of the reference snapshots. With these time indices, we construct $[t]_i$ as $[t]_i = \{t_{\mathrm{ref}(i)}, \ldots, t_{\mathrm{ref}(i+1)-1}\}$. Furthermore, with the help of $[t]_i$, we split the snapshot matrix as

$$\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_N), \tag{2.10}$$

where each of the sub-matrices $\mathcal{S}_i \in \mathbb{R}^{M \times r(i)}$ contain the snapshots for all $t \in [t]_i$ and can be calibrated using feature matching.

Remark 1. *We further elaborate on the importance of ensuring the lower-bound in (2.9). For proper calibration, the ordering of features observed in the numerical solution should be the same as that for the exact solution. At least empirically, we observe that the feature detection algorithm provides feature locations that are correct up to errors of size Δx. Therefore, in our numerical experiments we do not match snapshots containing features that are closer than $2\Delta x$ to any other snapshots i.e., we satisfy the lower-bound in (2.9).*

2.4 Approximation space We discuss how to use the above splitting of the snapshot matrix to construct an approximation space for the calibrated snapshot $U_{\mathrm{calib,M}}(t)$ defined in (1.4). We first consider the time interval D_i, which is a continuous analogue of $[t]_i$, and is given as

$$D_i := [t_{\mathrm{ref(i)}}, t_{\mathrm{ref}(i)+r(i)-1}]. \tag{2.11}$$

Let $\mathcal{S}_{\mathrm{calib},i}$ represent a calibration of $\mathcal{S}_i$. In the online phase, for $t \in D_i$, we approximate $U_{\mathrm{calib,M}}(t)$ in the span of the first m_i left singular-vectors of $\mathcal{S}_{\mathrm{calib},i}$ i.e., in $\mathrm{range}(\mathcal{U}_{m_i}(\mathcal{S}_{\mathrm{calib},i}))$.

We now consider the time interval d_i, which is the gap between D_i and D_{i+1}, and reads

$$d_i := (t_{\mathrm{ref}(i+1)-1}, t_{\mathrm{ref}(i+1)}). \tag{2.12}$$

Since the snapshots $u_M(\cdot, t_{\mathrm{ref}(i+1)-1})$ and $u_M(\cdot, t_{\mathrm{ref}(i+1)})$ do not match, we need information from both $\mathcal{S}_{\mathrm{calib},i}$ and $\mathcal{S}_{\mathrm{calib},i+1}$ for an accurate approximation of $U_{\mathrm{calib,M}}(t)$. Therefore, we consider the approximation space $\mathrm{range}(\mathcal{U}_{m_i}(\mathcal{S}_{\mathrm{calib},i})) + \mathrm{range}(\mathcal{U}_{m_{i+1}}(\mathcal{S}_{\mathrm{calib},i+1}))$. We summarize our above discussion.

1. For $t \in D_i$, approximate $U_{\mathrm{calib,M}}(t)$ in $\mathrm{range}(\mathcal{U}_{m_i}(\mathcal{S}_{\mathrm{calib},i}))$.
2. For $t \in d_i$, approximate $U_{\mathrm{calib,M}}(t)$ in the sum of $\mathrm{range}(\mathcal{U}_{m_i}(\mathcal{S}_{\mathrm{calib},i}))$ and $\mathrm{range}(\mathcal{U}_{m_{i+1}}(\mathcal{S}_{\mathrm{calib},i+1}))$.

2.5 Relation to the previous works To the best of our knowledge, only the works in [22, 31] propose a snapshot calibration technique for problems involving feature interaction and formation. We compare our method to both of these works. The authors in [31] propose a so-called transformed snapshot interpolation (TSI) to

Algorithm 2.1 Reference snapshot selection algorithm

Input: $\mathcal{S}$, $\mathcal{K}_1$, $\mathcal{K}_2$

Output: $\{\mathrm{ref}(i)\}_{i=1,\ldots,N}$

1: Initialize with $N \leftarrow 1$, $\mathrm{ref}(N) \leftarrow 1$ and $k \leftarrow 1$
2: $\Delta_{\min} z(t_{\mathrm{ref(N)}}) \leftarrow \min_j |z_{M,j+1}(t_{\mathrm{ref(N)}}) - z_{M,j}(t_{\mathrm{ref(N)}})|$
3: $\Delta_{\min} z(t_k) \leftarrow \min_j |z_{M,j+1}(t_k) - z_{M,j}(t_k)|$
4: Check whether the following conditions are satisfied: (C1)-(C3), $\Delta_{\min} z(t_{\mathrm{ref(N)}}) > \mathcal{K}_2 \Delta x$ and $\Delta_{\min} z(t_k) > \mathcal{K}_2 \Delta x$.
5: If the above statement returns true, increment k by on. Else, increment N by one, change $\mathrm{ref}(N)$ to k and increase k by one.
6: Till $k \leq K$, repeat from line-2.

handle shock collision problems and it differs from the current work in the following ways. Firstly, authors use an implicit method (the method requiring a solution to an optimization problem, see the introduction) to find the transform φ_M. Secondly, authors partition the time-domain using a *hp*-finite element strategy, which does not rely on a reference snapshot selection. Thirdly, it is unclear whether the transform φ_M satisfies the properties (P1) and (P2) both of which, at least according to our analysis, are crucial.

Our method differs from the shifted-POD approach (proposed in [22]) in the following sense. Firstly, shifted-POD is an iterative algorithm where each iteration calibrates a particular transport mode by shifting the spatial domain. Our spatial transform φ_M takes care of all the transport modes in one step, avoiding the need for iterations. Secondly, the shift value computation in shifted-POD requires a significant user-interference and results from either a careful observation of the snapshot matrix $\mathcal{S}$ or of its singular values. In comparison, after the snapshot matrix is computed, our method to compute φ_M is automatic. Thirdly, the shifted-POD does not cater to time-dependent boundary conditions. Note that none of the above two works study the m-width decay of the calibrated manifold.

3 Kolmogorov m-width decay In this section, we study the m-width of the calibrated manifold $\mathcal{M}_{\mathrm{calib},M}(D_i)$ defined in (1.7). Here, D_i is the continuous analogue of $[t]_i$ defined in (2.11). This section has two main highlights (i) the bound on the m-width does not only depend on the ROM dimension m but also on the FOM dimension M, and (ii) for sufficiently regular initial data u_0 and flux function f, the m-width decays fast with respect to m. Precisely, when the FOM is a FV solution, we show that

$$\delta_m(\mathcal{M}_{\mathrm{calib},M}(D_i)) = \mathcal{O}(m^{-\omega}) + \mathcal{O}(M^{-\frac{1}{2}}), \tag{3.1}$$

where the coefficient ω is related to the regularity of u_0 and $f(u_0)$ between the features. Furthermore, for any manifold $\mathcal{M} := \{h(\cdot, t) \,:\, t \in D\} \subset L^2(\Omega)$ its m-width, denoted by $\delta_m(\mathcal{M})$, is defined as

$$\delta_m(\mathcal{M}) := \inf_{\substack{\mathcal{V}_m \subset L^2(\Omega) \\ \dim(\mathcal{V}_m) = m}} \|h - \Pi_{\mathcal{V}_m} h\|_{L^2(\Omega \times D)}. \tag{3.2}$$

The M-dependency of the m-width appearing in (3.1) is introduced via the transform φ_M, which we compute using the FOM. Note that for elliptic and parabolic problems, calibration is not needed resulting in only an m-dependent m-width [1].

We now discuss the details of the result mentioned above. We restrict ourselves to a scalar conservation law i.e., $Q = 1$ in (2.1). Furthermore, we make the standard assumption that the flux function f is at least C^2 and is strictly convex. Note that for $t \in d_i$, where d_i is the gap between D_i and D_{i+1} and is as given in (2.12), calibration using feature matching is not possible and therefore, $\mathcal{M}_{\text{calib},M}(d_i)$ is irrelevant. Furthermore, since we use snapshots from both D_i and D_{i+1} to approximate the solution inside d_i, we expect this approximation to be accurate.

We start with defining a few quantities and making some assumptions. In the earlier sections, we considered a discrete space-time domain. For a large enough M, we expect the feature locations $z_i(t)$ to behave similar to the approximate feature locations $z_{M,i}(t)$. This motivates the assumption that since for all $t, t^* \in [t]_1$, we have $u_M(\cdot, t) \leftrightarrow u_M(\cdot, t^*)$, we also have

Assumption 1. $u(\cdot, t) \leftrightarrow u(\cdot, t^*), \quad \forall t, t^* \in D_i$ *for all* $i = 1, \ldots, N$.

Our results are the same for all the different D_i. Therefore, we present our results on some representative D_i that we denote by D for brevity. The above assumption allows us to define the following.

Definition 3.1 (Calibrated manifold). *Similar to* $\mathcal{M}_{\text{calib},M}(D)$*, define*

$$\mathcal{M}_{\text{calib}}(D) := \{u(\varphi(\cdot, t), t) \ : \ t \in D\}. \tag{3.3}$$

Above, φ *is the same as* φ_M *defined in* (2.7) *but with* $z_{M,j}(t)$ *replaced by the exact feature location* $z_j(t)$*. We can interpret the functions in* $\mathcal{M}_{\text{calib}}(D)$ *as a continuous-in-space analogue of those in* $\mathcal{M}_{\text{calib},M}(D)$*.*

In the next definition, we partition the space-time domain using the time-trajectory of different feature locations.

Definition 3.2 (Space-time partitioning). *Let the number of features in* $u_M(\cdot, t_{\text{ref}})$ *be* p_0 *i.e.,* $p(t_{\text{ref}} = 0) = p_0$*. For* $i \in \{0, \ldots, p_0\}$*, define*

$$\Omega_i := (z_i(0), z_{i+1}(0)), \quad \Omega_i^D := \{(x, t) \ : \ x \in (z_i(t), z_{i+1}(t)), \ t \in D\}. \tag{3.4}$$

Note that $\operatorname{clos}(\Omega) = \bigcup_{i=0}^{p_0} \operatorname{clos}(\Omega_i)$*.*

The main result of this section and its corollary are summarised below. The rest of the section proves this result.

Theorem 3.3. *The* m*-width of the calibrated manifold* $\delta_m(\mathcal{M}_{\text{calib},M}(D))$ *is bounded by*

$$\begin{aligned}
\delta_m(\mathcal{M}_{\text{calib},M}(D)) \leq & \delta_m(\mathcal{M}_{\text{calib}}(D)) \\
& + \sup_{t \in D} \|D_x \varphi_M(\cdot, t)^{-1}\|_{L^\infty(\Omega)} \|u_M - u\|_{L^2(\Omega \times D)} \\
& + \|u_M \circ \varphi_M - \Pi_{X_M} u_M \circ \varphi_M\|_{L^2(\Omega \times D)} \\
& + \sqrt{\|u\|_{L^\infty(D; BV(\Omega))} \|u\|_{L^2(D; BV(\Omega))}} \\
& \times \sqrt{\max_j \|z_{M,j} - z_j\|_{L^\infty(D)}} \times \max(1, \|D_x \varphi_M\|_{L^\infty(\Omega \times D)}).
\end{aligned} \tag{3.5}$$

Corollary 3.4. *Provided the following conditions hold*

1. *The feature identification procedure used for computing* φ_M *satisfies*

$$\max_j \|z_{M,j} - z_j\|_{L^\infty(D)} = \mathcal{O}(M^{-1}).$$

2. *There exists $\omega \geq 1$ so that for all $i \in \{0, \ldots, p_0\}$ the flux function and the initial data satisfy*

$$f \in C^{\omega+1}, \quad u_0|_{\Omega_i} \in W^{\omega,\infty}(\Omega_i). \tag{3.6}$$

Here, u_0 refers to the initial data at the beginning of the corresponding time interval D_j. Furthermore, $W^{\omega,\infty}$ represents the Sobolev-space of functions having ω weak derivatives in L^∞.

3. *For all $i \in \{0, \ldots, p_0\}$,*

$$\sup_{(x,t)\in\Omega_i^D} \frac{1}{|\beta_i(x,t)|} < \infty \quad \text{where} \quad \beta_i(x,t) := 1 + tf''(u_0(x))D_x u_0(x). \tag{3.7}$$

Then, for a convergent FV approximation scheme, using M equidistant cells, the m-width satisfies

$$\delta_m(\mathcal{M}_{\text{calib},M}(D)) = \mathcal{O}(m^{-\omega}) + \mathcal{O}(M^{-\frac{1}{2}}). \tag{3.8}$$

REMARK 2. *Note that the boundedness of β_i is equivalent to no shock forming on* $\operatorname{clos}\left(\Omega_i^D\right)$.

We make the following observations and conclusions from the above result.

1. The bound on the m-width given in (3.5) is robust under the limit $m \to \infty$ and $M \to \infty$.
2. All the terms on the right in (3.5), apart from $\delta_m(\mathcal{M}_{\text{calib}}(D))$, are M-dependent i.e., they depend on the accuracy of the full-order model.
3. For M large enough and m small enough, we expect the bound to be dominated by $\delta_m(\mathcal{M}_{\text{calib}}(D))$.
4. For a constant M, as $m \to \infty$, the bound will stagnate at a $\mathcal{O}(M^{-\frac{1}{2}})$ term. This means that as $m \to \infty$, the best approximation error of u in the ROM space is of the same order of magnitude as $\|u(\cdot,t) - u_M(\cdot,t)\|_{L^2(\Omega)}$, where $u_M(\cdot,t)$ is the FOM. Recall that the best approximation error of a (discontinuous) BV-function in a FV approximation space is $\mathcal{O}(M^{-\frac{1}{2}})$.
The practical take-away from this discussion is that it does not make sense to increase m beyond a certain limit i.e., it does not make sense to further increase m when $\|u_M(\cdot,t) - u_m^{\text{red}}(\cdot,t)\|_{L^2(\Omega)}$ and $\|u(\cdot,t) - u_M(\cdot,t)\|_{L^2(\Omega)}$ are of the same order of magnitude. Here, u_m^{red} represent a reduced-order approximation to u.
5. Note that for $u \in W^{1,\infty}(\Omega \times D)$, which allows only for kinks and no discontinuities, the best approximation error of u in the FV approximation space is $\mathcal{O}(M^{-1})$. Similarly, the last term on the right hand side of (3.5) can be improved to $\|u\|_{L^2(D;W^{1,\infty}(\Omega))}M^{-1}$.
6. The bound on the m-width in (3.5) explains that an upper-bound on $\|D_x\varphi_M(\cdot,t)^{-1}\|_{L^\infty(\Omega)}$ and $\|D_x\varphi_M(\cdot,t)\|_{L^\infty(\Omega)}$ (i.e. (P2) given in (1.5)) are desirable.
7. The bound in Theorem 3.3 and Algorithm 2.1 suggests a compromise between small and large values of $\mathcal{K}_1$—recall that $\mathcal{K}_1$ is the user-defined constant appearing in the property (P2) given in (1.5). As $\mathcal{K}_1$ increases, Algorithm 2.1 generates smaller number of reference snapshots, resulting in a calibrated snapshot matrix with a fewer number of sub-matrices. We expect that, for a given approximation accuracy, this would result in a fewer number of POD modes used to approximate the calibrated snapshot. In contrast, $\mathcal{K}_1$ scales

the $\mathcal{O}(M^{-1/2})$ part of the bound in Theorem 3.3, making it undesirable to choose a large $\mathcal{K}_1$. Numerical experiments indicate that any choice of $\mathcal{K}_1$ that is $\mathcal{O}(1)$ is acceptable.

3.1 Proof of Theorem 3.1 Triangle's inequality applied to the definition of $\delta_m(\mathcal{M}_{\text{calib},M}(D))$ provides

$$\begin{aligned}\delta_m(\mathcal{M}_{\text{calib},M}(D)) \leq & \delta_m(\mathcal{M}_{\text{calib}}(D)) \\ & + \sup_{t\in D}\|D_x\varphi_M(\cdot,t)^{-1}\|_{L^\infty(\Omega)}\underline{\|u_M - u\|_{L^2(\Omega\times D)}} \\ & + \underline{\|u_M\circ\varphi_M - \Pi_{X_M}u_M\circ\varphi_M\|_{L^2(\Omega\times D)}} \\ & + \|u\circ\varphi_M - u\circ\varphi\|_{L^2(\Omega\times D)} \\ & =: A_1 + A_2 + A_3 + A_4.\end{aligned} \tag{3.9}$$

A bound for the different A_i's is as follows.

3.1.1 Bound for A_2 and A_3 A bound for A_2 and A_3 follows from the approximation properties of a FV approximation space. The decay (in M) of A_2 is connected to the convergence of the underlying FOM, if u is in $BV \setminus W^{1,\infty}$ then A_2 will behave as $\mathcal{O}(M^{-1/2})$. Here, $BV(\Omega)$ is a space of real-valued functions with a finite total variation. Due to the approximation properties of the FV approximation space we have

$$A_3 \leq |\Omega|M^{-1/2}|u_M\circ\varphi_M|_{L^2(D,BV(\Omega))} = |\Omega|M^{-1/2}|u_M|_{L^2(D,BV(\Omega))}.$$

Note that we have used the monotonicity of φ_M in the equality above and that $|u_M|_{L^2(D,BV(\Omega))} \leq |u|_{L^2(D,BV(\Omega))}$ provided the FV scheme is total-variation-diminishing (TVD).

3.1.2 Bound for A_1 Let $g(x,t) = u(\varphi(x,t),t)$, where φ is as given in (3.3). Tracing the characteristics backwards from t to 0, we have

$$g(x,t) = u_0(\underbrace{(\mathrm{Id} + tf'(u_0))^{-1}\varphi(x,t)}_{=:X_i(\varphi(x,t),t)}) \quad \forall (x,t) \in \Omega_i \times D, \tag{3.10}$$

where u_0 is the initial data in (2.1), f is the flux-function in (2.1), and Ω_i is as defined in (3.4). Note that because the flux function is convex, while tracing the characteristics backwards in an entropy solution, they do not run into a shock. Using (3.10), the following result quantifies the regularity of g.

LEMMA 3.5. *Provided (3.6) and (3.7) hold, then $g \in L^2(\Omega; H^\omega(D))$. Here, $L^2(\Omega; H^\omega(D))$ denotes a Bochner space of L^2 functions defined over Ω with values in the Sobolev space $H^\omega(D)$.*

Proof. See Appendix A. ∎

With the regularity established in the above result, taking the linear space $\mathcal{V}_m$ (appearing in (3.2)) to be the span of first m-Fourier modes in D, we can estimate the m-width as

$$\delta_m(\mathcal{M}_{\text{calib}}(D)) \leq \|g - \Pi_{\mathcal{V}_m} g\|_{L^2(\Omega\times D)} = \mathcal{O}(m^{-\omega}). \tag{3.11}$$

Note that the (un-calibrated) solution $u(\cdot,t)$ rarely has the amount of regularity that $g(\cdot,t)$ does. In this sense, we can view calibration as a way of "artificially" introducing regularity to induce a fast m-width decay in the calibrated solution manifold.

Apart from the above result, a trivial but noteworthy case is when g is time-independent. This results in $\mathcal{M}_{\text{calib}}(D)$ consisting of a single function, which provides

$$\delta_m(\mathcal{M}_{\text{calib}}(D)) = 0 \quad \forall m \geq 1. \tag{3.12}$$

Indeed, g is time-independent provided, for all $i \in \{0, \ldots, p_0\}$, either of the following two conditions hold

$$\begin{aligned} &\text{(i) } u_0|_{\Omega_i} \equiv u_{0,i} \text{ for some constant } u_{0,i} \in \mathbb{R}, \\ &\text{(ii) } X_i(\varphi(x,t),t) \text{ is independent of } t. \end{aligned} \tag{3.13}$$

The first condition corresponds to the initial data being a constant inside Ω_i, and the second one can result inside a rarefaction fan; see Appendix B.

REMARK 3. *The result in Lemma 3.5 highlights the advantages of aligning both kinks and discontinuities. By including kinks into the set of features we can hope that u_0 is $W^{2,\infty}$ between features which makes $g \in L^2(\Omega, H^2(D))$ possible, resulting in a m-width that is $\mathcal{O}(m^{-2})$. However, if u_0 contains a kink that is not in the set of features then we expect u_0 is $W^{1,\infty} \setminus W^{2,\infty}$ between the features resulting in $g \in L^2(\Omega, H^1(D)) \setminus L^2(\Omega, H^2(D))$ and a m-width that is $\mathcal{O}(m^{-1})$.*

REMARK 4. *One can match the discontinuities in the higher-order derivatives of $u_M(\cdot,t)$ and get a faster (than presented above) m-width decay rate—precisely, matching discontinuities in the ω-order derivative results in a $\omega+1$-order decay in the m-width. However, numerically identifying the location of discontinuities in higher-order derivatives is difficult and cumbersome. As our numerical experiments indicate, for a sufficiently refined numerical approximation in X_M, kink identification is possible and for that reason, we do not consider higher-order derivatives.*

3.1.3 Bound for A_4 The estimate for $\|u \circ \varphi_M - u \circ \varphi\|_{L^2(\Omega\times D)}$ follows from the result below. The first part of the result is an extension of the result in [32] to L^2-functions and exploits the density of smooth functions in the BV-space. In the second part, we use the explicit from of the spatial transform given in (2.7) to compute $\|\varphi - \varphi_M\|_{L^\infty(\Omega\times D)}$. With the bound given in the second part, we again emphasize on the desirability of ensuring (P2).

LEMMA 3.6. *The following relations hold true.*

1. $\|u\circ\varphi - u\circ\varphi_M\|^2_{L^2(\Omega\times D)} \leq \|u\|_{L^\infty(D;BV(\Omega))}\|u\|_{L^2(D;BV(\Omega))}\|\varphi-\varphi_M\|_{L^\infty(\Omega\times D)}$.
2. *Let $\mathcal{K}_1$ be the constant given in (1.5). Then, the error $\|\varphi-\varphi_M\|_{L^\infty(\Omega\times D)}$ is bounded as*

$$\|\varphi - \varphi_M\|_{L^\infty(\Omega\times D)} \leq \max(1, \mathcal{K}_1) \max_j \|z_{M,j} - z_j\|_{L^\infty(D)}. \tag{3.14}$$

Proof. See Appendix C. ∎

4 Feature Detection It is important to note that our calibration approach can be combined with any feature detection approach and that the feature location algorithm can be used as a black-box. In order to keep this article self-contained, we explain one specific approach which was also used in our numerical experiments. This specific approach is based on the more general idea that kinks are discontinuities in the derivative i.e., discontinuities and kinks can be detected by discontinuity detection schemes using the following three steps: (i) approximate the discontinuity locations, (ii) approximate the weak derivative $D_x u(\cdot,t)$ and (iii) approximate the kink locations by applying the discontinuity detection algorithm to $D_x u(\cdot,t)$. To realize such a method, we need a discontinuity detector for which several different methods can suffice. For example, one can detect discontinuities by training a neural network [21], using the convergence properties of FOM [12], performing a multi-resolution-analysis (MRA) [30], etc.

For its ease of implementation and reasonable accuracy for the experiments considered later, we use the MRA approach and modify it slightly to suit our needs. The details of our modification are given below and for completeness, the MRA approach is discussed in Appendix D.

4.1 Discontinuity Detection Recall that our FOM corresponds to a FV approximation. With $u_i(t)$ we represent the constant value of $u_M(\cdot,t)$ inside $\mathcal{I}_i$, where $\mathcal{I}_i$ is the i-th cell defined in (2.2). The M-cells have $M+1$ faces and we collect their indices in $\mathcal{E} := \{1,\dots,M+1\}$. With x_e we represent the location of the e-th face, i.e. the face between $\mathcal{I}_e$ and $\mathcal{I}_{e+1}$. Across every face we compute the jump in $u_M(\cdot,t)$ and if the jump overshoots a given tolerance, we mark it as a potential location of discontinuity. Details are as follows.

Let $e\in\mathcal{E}$. With $J_e(t)$ we denote the absolute value of the jump in $u_M(\cdot,t)$ across the edge e i.e.,

$$J_e(t) = |u_e(t) - u_{e-1}(t)|, \quad \forall e \in \mathcal{E}. \tag{4.1}$$

Using $J_e(t)$, we define the set $\mathcal{B}(t)$ that contains the indices of faces with a potential discontinuity in the adjoining cell

$$\mathcal{B}(t) := \{e \in \mathcal{E} \ : \ J_e(t) > C \times \Delta x\}. \tag{4.2}$$

Above, C is user-defined and controls the number of faces that will be contained in $\mathcal{B}(t)$. Later, we elaborate more on the relevance of C.

To compute the discontinuity location using $\mathcal{B}(t)$, we proceed as follows. We partition $\mathcal{B}(t)$ into sub-sets $\{\mathcal{B}_i(t)\}_i$ such that each of $\mathcal{B}_i(t)$ contains indices of only the adjoining faces. For instance, if $\mathcal{B}(t) = \{1,2,4,5\}$ then $\mathcal{B}_1(t) = \{1,2\}$ and $\mathcal{B}_2(t) = \{4,5\}$. A set $\mathcal{B}_i(t)$ can have more than one element when, due to the numerical dissipation in the FV scheme, the discontinuity is spread out into a set of neighbouring cells, or when there are multiple discontinuities in succession. For both the cases, we compute the discontinuity location by taking the mean of all the face locations in $\mathcal{B}_i(t)$. Equivalently,

$$z^D_{M,i}(t) := \frac{\sum_{e\in\mathcal{B}_i(t)} x_e}{|\mathcal{B}_i(t)|} \quad \forall i \in \{1,\dots,p^D(t)\}. \tag{4.3}$$

Here $z^D_{M,i}(t)$ denotes an approximation to the the true discontinuity location z^D_i, and $p^D(t)$ denotes the total number of discontinuities.

REMARK 5. *Ideally, $\mathcal{B}(t)$ should include only those faces that have discontinuities in the adjoining cells. However, depending upon C's value and the solution's behaviour away from a discontinuity, the ideal situation might not be realized. Additional faces that do not contain discontinuities in the adjoining cells might be included in $\mathcal{B}(t)$. The inequalities given in Appendix E give some indication of how the method flags different regions. We emphasize that identifying additional feature location does not ruin the calibration procedure. It only results in additional points being matched between two snapshots. However, with any additional feature it is more likely to violate the conditions (C1)-(C3), resulting in Algorithm 2.1 generating additional reference snapshots.*

4.2 Kink detection Let $\hat{\Omega}(t) := \{z_i^D(t)\}_{i=1,\ldots,p^D(t)}$ be a set of points where $u(\cdot,t)$ is discontinuous. In Definition 2.1, we defined kink locations as points where $D_x u(t)$ has a discontinuity in $\Omega/\hat{\Omega}(t)$. Thus, to find these locations, we run the discontinuity detection algorithm on $D_x u(\cdot,t)$. To realize the algorithm we need an approximation for $D_x u(\cdot,t)$ and $\hat{\Omega}(t)$.

Let $D_x u_M(\cdot,t)$ be an approximation to $D_x u(\cdot,t)$. We find $D_x u_M(\cdot,t)$ by applying central differences to $u_M(\cdot,t)$. Let $D_x u_i(t)$ be the constant value of $D_x u_M(\cdot,t)$ in the cell $\mathcal{I}_i$. Then, $D_x u_i(t)$ is given as

$$D_x u_i(t) = \frac{u_{i+1}(t) - u_{i-1}(t)}{2\Delta x}. \tag{4.4}$$

On the continuous level, the derivative of $u(\cdot,t)$ is a Dirac-distribution at points where $u(\cdot,t)$ is discontinuous. However, on a spatially discrete level, the delta distribution is a collection of "spikes" in $D_x u_M(\cdot,t)$. To collect these spike we approximate every entry $z_i^D(t)$ by a ball of radius ϵ centered around $z_i^D(t)$. As an approximation to z_i^D we use x_e, where x_e is the location of the e-th face, $e \in \mathcal{B}(t)$, and $\mathcal{B}(t)$ is as given in (4.2). We set ϵ to $N^D \times \Delta x$ and we approximate $\hat{\Omega}(t)$ by

$$\hat{\Omega}(t) \approx \bigcup_{e\in\mathcal{B}(t)} \mathcal{B}(x_e; N^D\Delta x). \tag{4.5}$$

We choose $N^D = 3$. We use an example to motivate our choice for N^D. Let $u(\cdot,t)$ be a unit-step function with a discontinuity at $z^D = x_e + l\Delta x$, where $l \in [0,1]$. It follows that

$$D_x u_{e-1}(t) = \frac{(1-l)}{2\Delta x}, \quad D_x u_e(t) = \frac{1}{2\Delta x}, \quad D_x u_{e+1}(t) = \frac{l}{2\Delta x}.$$

For all the other intervals, $D_x u_M(\cdot,t) = 0$. Depending on the value of l, $D_x u_M(\cdot,t)$ can have a large spike in the intervals $\mathcal{I}_{e-1}$, $\mathcal{I}_e$ and $\mathcal{I}_{e+1}$. Therefore, $N^D = 3$ is a reasonable choice.

REMARK 6. *With the above method, we do not detect kinks inside the union of balls given in (4.5). However, for a small enough Δx, missing out on these kinks does not significantly increase the m-width of the calibrated manifold. This will be elucidated by numerical experiments.*

4.3 Undetected features Features can get smeared out by numerical dissipation and, depending upon the value of C given in (4.2), might go undetected. For such cases, one can show that (at least) the semi-discrete numerical solution already has sufficient regularity to ensure a fast m-width decay. Let $u_i(t)$ be as defined in Subsection 4.1 and let

$$\frac{du_i(t)}{dt} = \frac{1}{\Delta x}\left(\mathcal{F}(u_{i-1}(t), u_i(t)) - \mathcal{F}(u_i(t), u_{i+1}(t))\right)$$

be its evolution equation. Here, $\mathcal{F}$ represents a numerical flux function, which we assume is in $W^{2,\infty}$.

We first consider undetected discontinuities. Assume that $|u_{i\pm1}(t) - u_i(t)| \leq C\Delta x$, in which case we do not detect a discontinuity at the face $i-1$ and i. Then, using the regularity of $\mathcal{F}$, one can show that

$$|du_i(t)/dt| \leq 2\|\mathcal{F}\|^2_{W^{1,\infty}} C.$$

In Lemma 3.5 we proved that $\varphi(x,\cdot)\in W^{\omega}(D)$. Motivated from this, we assume that $\varphi_M(x,\cdot)\in W^{\omega}(D)$, which is equivalent to $z_{M,j}\in W^{\omega}(D)$. Then, the above bound implies that, for $x\in\mathcal{I}_i$, $u_M(\varphi_M(x,\cdot),\cdot)\in W^{\omega}(D)$. Thus, locally in $\mathcal{I}_i$, $u_M(\varphi_M(x,\cdot),\cdot)$ has the regularity needed for a fast m-width decay of the calibrated manifold.

We now consider undetected kinks. Assume that $|D_x u_i(t)-D_x u_{i-1}(t)|\leq C\Delta x$, $|D_x u_{i+1}(t)-D_x u_i(t)|\leq C\Delta x$ and $|D_x u_{i+2}(t)-u_{i+1}(t)|\leq C\Delta x$, in which case we do not detect a kink at the face $i-1$, i and $i+1$. Then, one can show that

$$|d^2u_i/dt^2|\leq 4\|\mathcal{F}\|^2_{W^{2,\infty}}(2C^2+C).$$

Following the same reasoning as above, the bound implies that, for $x\in\mathcal{I}_i$, $u_M(\varphi_M(x,\cdot),\cdot)\in W^{\omega}(D)$.

5 Numerical Experiments Let $\Xi_m(\mathcal{S})$ be as defined in (1.3). The numerical experiments show the following two things. Firstly, with kink and discontinuity matching, $\Xi_m(\mathcal{S}_{\text{calib}})$ decays much faster than $\Xi_m(\mathcal{S})$. Secondly, both kink and discontinuity matching is better than only discontinuity matching. To construct numerical approximations where both kink and discontinuity detection is possible, we consider the best-approximation in X_M. Note that in light of the discussion in Subsection 4.3, these numerical approximations are the ones were we expect the slowest m-width/singular-value decay.

Since $\Xi_m(\mathcal{S}_{\text{calib}})$ quantifies the l^2 error of approximating a calibrated snapshot in the span of the first m left singular vectors of $\mathcal{S}_{\text{calib}}$, similar to the bound in (3.5), it is possible that on increasing m, $\Xi_m(\mathcal{S}_{\text{calib}})$ stagnates at a value of $\mathcal{O}(M^{-\frac{1}{2}})$. The following experiments will provide further elaboration.

1. **Test case-1** we consider the Burgers' equation

$$\partial_t u+\partial_x\left(\frac{u^2}{2}\right)=0,\ \text{on }\Omega\times D,\quad u(\cdot,t=0)=\mathbb{1}_{[0,1]},\ \text{on }\Omega. \tag{5.1}$$

Above, $\mathbb{1}_{[0,1]}$ represents a characteristic function over $[0,1]$. We choose $\Omega=(-0.5,3.5)$ and $D=[0,4]$. On the boundary $\partial\Omega\times D$, we prescribe $u=0$.

2. **Test case-2** we consider the wave equation (rewritten as a first order system)

$$\partial_t u+A\partial_x u=0,\ \text{on }\Omega\times D, \tag{5.2}$$

where $u=(u_1,u_2)^T$ is the solution vector and the matrix A reads

$$A=\begin{pmatrix}0&1\\1&0\end{pmatrix}. \tag{5.3}$$

We choose $\Omega=(-0.5,3.5)$ and $D=[0,2]$. As the initial data, for all $x\in\Omega$, we consider

$$u_1(x,t=0)=w_1(x)+w_2(x),\quad u_2(x,t=0)=-w_1(x)+w_2(x), \tag{5.4}$$

where $w_1(x)$ and $w_2(x)$ are two sin-function bumps given as

$$\begin{aligned}w_1(x)=&\frac{1}{\sqrt{2}}(\sin(\pi x)+1)\mathbb{1}_{[0,1]}(x),\\ w_2(x)=&\frac{1}{\sqrt{2}}(\sin(\pi(x-2))+1)\mathbb{1}_{[2,3]}(x).\end{aligned} \tag{5.5}$$

As in the previous case, on $\partial\Omega\times D$, we prescribe $u=0$.

3. **Test case-3** we consider the Sod's shock tube problem that involves the Euler's equation given as

$$\partial_t \begin{pmatrix} \rho \\ \rho v \\ E \end{pmatrix} + \partial_x \begin{pmatrix} \rho v \\ \rho v^2 + P \\ Ev + Pv \end{pmatrix} = 0, \text{ on } \Omega \times D. \tag{5.6}$$

Above, ρ, v, P and E represent the density, the velocity, the pressure and the total energy, respectively. For an ideal gas, $P = (\gamma - 1)\rho e$, where γ represent the gas constant and e is the internal energy related to the total energy via $\rho e = E - \rho v^2/2$. We consider a mono-atomic ideal gas for which $\gamma = 5/3$. We choose $\Omega = (-0.5, 0.5)$ and $D = [0, 0.2]$. As the initial data, we consider a fluid at rest with the density and the pressure given as

$$\rho(x, t = 0) = \begin{cases} 1, & x \leq 0 \\ 0.125, & x > 0 \end{cases}, \quad P(x, t = 0) = \begin{cases} 1, & x \leq 0 \\ 0.1, & x > 0 \end{cases}. \tag{5.7}$$

The waves emanating from the initial discontinuity do not reach the boundary therefore, we take the boundary data from the initial values.

4. **Test case-4** we consider the linear advection equation with time-dependent boundary data

$$\begin{aligned} \partial_t u(x,t) + \beta \partial_x u(x,t) = 0, &\quad \forall (x,t) \in \Omega \times D, \\ u(x, t = 0) = (\sin(\pi x) + 1)\mathbb{1}_{[0,1]}(x) &\quad \forall x \in \Omega, \\ u(x = 0, t) = \mathbb{1}_{[0.1,0.5]}(t), &\quad \forall t \in D. \end{aligned} \tag{5.8}$$

We set $\beta = 1$, $\Omega = (-0.5, 3.5)$ and $D = [0, 1]$.

For all the test cases, we partition Ω into $M = 2 \times 10^3$ elements, and consider 10^3 uniformly placed time instances inside D. We choose $\mathcal{K}_1 = 5$, $\mathcal{K}_2 = 3$ and $C = 50$. For all the test cases, we project the exact solution onto the FV space. Details of the exact solution are given later. We compute all the $L^2(\Omega)$ inner-products with 10 Gauss-Legendre quadrature points in each cell.

5.1 Test case-1 The unique entropy solution to the problem in (5.1) reads

$$\begin{aligned} u(x,t) &:= \begin{cases} \frac{x}{t}, & x \in [0, t) \\ 1, & x \in [t, 1 + \frac{t}{2}) \\ 0, & \text{else} \end{cases}, \quad \forall t \in [0, 2), \\ u(x,t) &:= \begin{cases} \frac{x}{t}, & x \in [0, \sqrt{2t}) \\ 0, & \text{else} \end{cases}, \quad \forall t \in [2, 4]. \end{aligned} \tag{5.9}$$

The exact solution has two discontinuities at $t = 0$. One of the discontinuities gives rise to two kinks (a rarefaction fan), the other remains as a discontinuity. At $t = 2$, one of the kinks collides with a discontinuity to form a single discontinuity. Around $t = 0$, the two kinks are very close to each other and are identified as a single discontinuity in the numerical solution; see Figure 2a. As time progresses, the two kinks move away from each other and are identified correctly.

Let $\mathcal{E}(\Delta x)$ represent the maximum of the error in feature location for a grid size Δx i.e.,

$$\mathcal{E}(\Delta x) := \max_j \|z_{M,j} - z_j\|_{L^\infty(D)}. \tag{5.10}$$

Recall that $\Delta x = |x_{\max} - x_{\min}|/M$. Figure 2b shows $\mathcal{E}(\Delta x)$ for different grid sizes. We vary the number of spatial elements M from 5×10^2 to 3×10^3 in steps of 2×10^2. We choose the threshold C in the discontinuity location identification such that C/M remains constant at $2.5 \cdot 10^{-2}$. We make the following two observations. Firstly, although not monotonically, $\mathcal{E}(\Delta x)$ decreases with Δx. Secondly, $\mathcal{E}(\Delta x)$ stays close to Δx and can get smaller than Δx as Δx decreases. Thus, at least for the current feature location identification procedure and for the current test case, the assumption on the error in feature location made in Corollary 3.4 is justified.

The dashed lines in Figure 2a show the temporal locations of the reference snapshots resulting from Algorithm 2.1. The algorithm provides $N = 5$ (with N as given in Definition 2.3) different reference snapshots located at $t = 0$, $t = 0.02$, $t = 1.60$, $t = 1.92$ and $t = 1.98$, respectively. The first reference snapshot is the initial data that is matched to a few subsequent snapshots, which is a result of identifying the two close-by kinks as a single discontinuity. The second reference snapshot is at a time instance when our feature identifier can distinguish between the two kinks. The third and the fourth reference snapshot is selected because the features come too close to each other, violating either the condition (C3) given in (2.5) or the lower-bound on the minimum feature distance given in (2.9). The last reference snapshot is selected after the kink collides with the discontinuity, it matches to all the subsequent snapshots. Note that in the exact solution, the kink collides with the discontinuity at $t = 2$. However, numerically, as mentioned in Remark 6, we miss out on kinks that lie very close to a discontinuity therefore, already at $t = 1.98$ we detect only the discontinuity and not the kink that interacts with it.

Figure 2c compares $\Xi_m(\mathcal{S}_i)$ to $\Xi_m(\mathcal{S}_{\text{calib},i})$ and shows that, for all values of i and m, $\Xi_m(\mathcal{S}_{\text{calib},i})$ is smaller than $\Xi_m(\mathcal{S}_i)$. Since $\mathcal{S}_1$ contains only four snapshots, the value of $\Xi_m(\mathcal{S}_{\text{calib},1})$ does not significantly differ from $\Xi_m(\mathcal{S}_1)$. For all the other submatrices, the value of $\Xi_m(\mathcal{S}_{\text{calib},i})$, already for $m = 1$, is at least 10^{-4} times smaller than $\Xi_m(\mathcal{S}_i)$. Let us emphasize that $m = 1$ is just 0.05% of M (the dimensionality of the FOM).

For $i = 4, 5$, as m is increased, $\Xi_m(\mathcal{S}_{\text{calib},i})$ stagnates. Varying the value of M from 10^3 to 3×10^3 in steps of 2×10^2 showed that the stagnation value is $\mathcal{O}(M^{-0.8})$, which is $\mathcal{O}(M^{-0.3})$ times better than (the M-dependent part of) the bound on the m-width developed in (3.8). A possible reason for this stagnation could be the error in feature location.

For $i = 2$, the matrix $\mathcal{S}_i$ contains snapshots that are either rarefaction fans or constants between any two features, thus satisfying the condition in (3.13). This results in the calibrated manifold consisting of a single function. Ideally, the calibrated snapshot matrix should have a rank close to one and for $m = 1$, $\Xi_m(\mathcal{S}_{\text{calib},i})$ should be (very) close to zero. However, as Figure 2c depicts, because of the error in feature location, this ideal situation is not realized in practice and the value $\Xi_m(\mathcal{S}_{\text{calib},i})$ is far away from zero. Nevertheless, for $m = 13$, $\Xi_m(\mathcal{S}_{\text{calib},i})$ reaches (machine precision) zero. We attribute this convergence to the fact that the error in identifying a feature location is $\mathcal{O}(M^{-1})$ and that the calibrated manifold $\mathcal{M}_{\text{calib}}(D_i)$ consists of a single function. Observance of a similar behaviour in other experiments corroborates our claim.

5.1.1 Discontinuity matching We repeat the above experiment but with only discontinuity matching. With $\mathcal{S}^D_{\text{calib}}$ we represent the resulting calibrated snapshot matrix. Algorithm 2.1 generates two reference snapshots i.e., $N = 2$. The temporal location of these two reference snapshots are shown in Figure 3a. Both the reference snapshots are close to $t = 0$. The first reference is the initial data and is matched to a few subsequent snapshots. The second reference snapshot is at a time-instance when we can uniquely identify the two kinks, leaving us with a single discontinuity.

Figure 3b compares $\Xi_m(\mathcal{S}_{\text{calib},i})$ to $\Xi_m(\mathcal{S}^D_{\text{calib},i})$. For $i = 1$, both $\Xi_m(\mathcal{S}_{\text{calib},i})$ and $\Xi_m(\mathcal{S}^D_{\text{calib},i})$ have the same values. This is as expected, since the two close-by kinks are identified as a discontinuity. For $i > 1$ and for all $m \in [1, 20]$, $\Xi_m(\mathcal{S}_{\text{calib},i})$ is at least two orders of magnitude smaller than $\Xi_m(\mathcal{S}^D_{\text{calib},i})$. The difference is more prominent for smaller values of m. Already for $m = 1$, $\Xi_m(\mathcal{S}_{\text{calib},i})$ is four order of magnitude smaller than $\Xi_m(\mathcal{S}^D_{\text{calib},i})$. The experiment clearly establishes the benefit of including both kinks and discontinuities in the feature set.

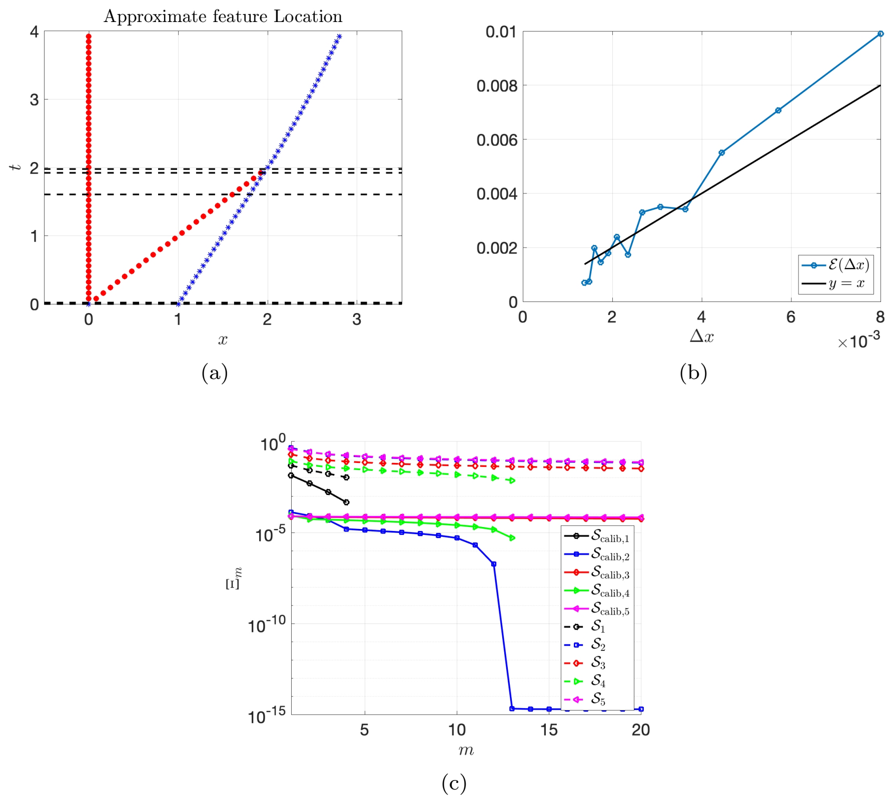
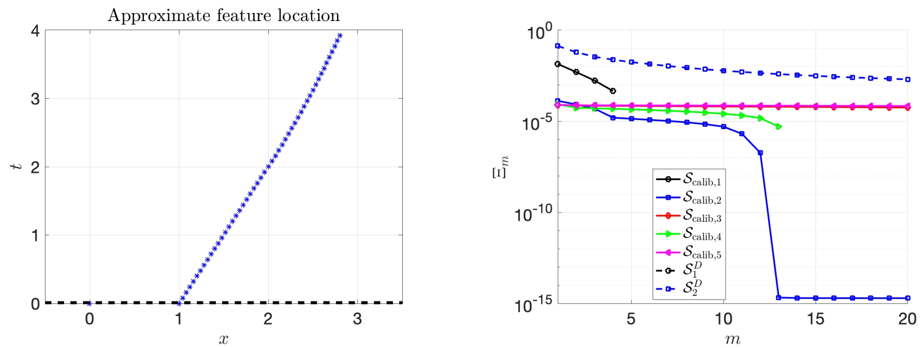


Fig. 2: *Results for test case-1. Both kinks and discontinuities included in the feature set. (a) Time-trajectory of the different features. Kink and discontinuity locations shown in red and blue, respectively. The dashed black lines show the temporal locations of the reference snapshots. (b) Error in feature location for different Δx. (c) Comparison of $\Xi_m(\mathcal{S}_i)$ to $\Xi_m(\mathcal{S}_{\text{calib},i})$. The y-axis of (c) is on a log-scale.*

5.2 Test case-2 With the help of the Riemann invariants, for all $(x, t) \in \Omega \times D$, one can conclude that the exact solution to the wave equation (5.2) is given as

$$u_1(x, t) = w_1(x - t) + w_2(x + t), \quad u_2(x, t) = -w_1(x - t) + w_2(x + t). \tag{5.11}$$



(a) (b)

Fig. 3: *Results for test case-1. Only discontinuities included in the feature set. (a) Time-trajectory of the different features. Kink and discontinuity locations shown in red and blue, respectively. The dashed black lines show the temporal locations of the reference snapshots. (c) Comparison of $\Xi_m(\mathcal{S}_{\text{calib},i})$ to $\Xi_m(\mathcal{S}^D_{\text{calib},i})$. The y-axis of (b) is on a log-scale.*

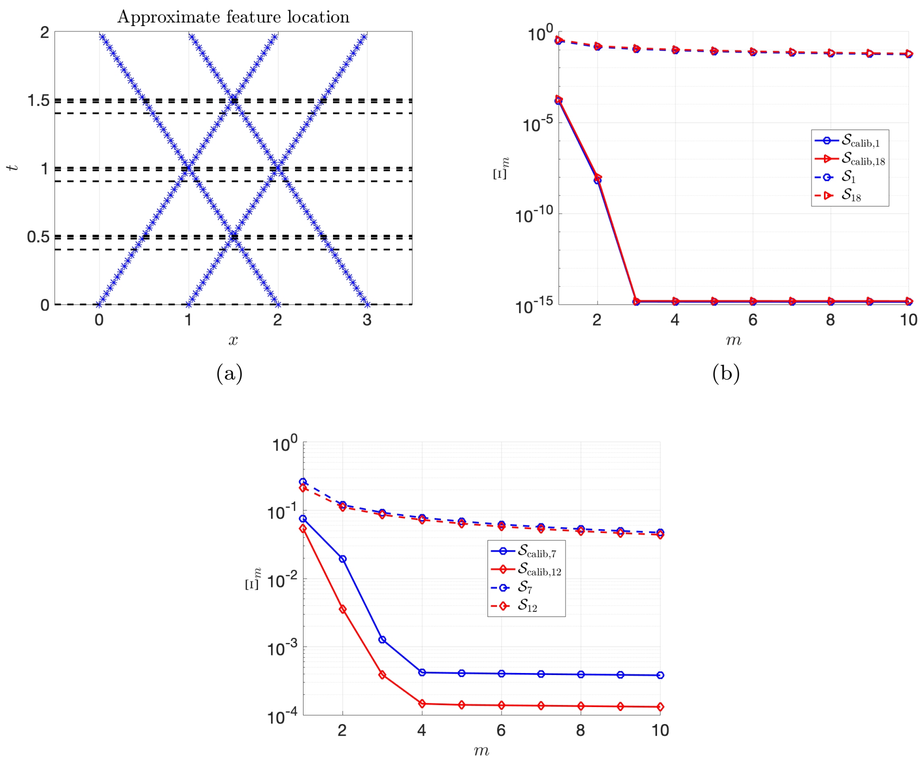
The functions w_1 and w_2 are as given in (5.5). Both u_1 and u_2 contain two discontinuities, which interact at four different time instances. For u_1, the time-trajectory of the different discontinuities is shown in Figure 4a. The algorithm accurately identifies the four discontinuities.

We discuss the results for u_1, similar results were observed for u_2. Algorithm 2.1 generates $N = 18$ different reference snapshots. The temporal locations of these snapshots are shown in Figure 4a. Similar to the previous test case, the reference snapshot changes frequently when features come close, or interact, with each other. To study Ξ_m, for the simplicity of exposition, out of the 18 different subsets $\{[t]_i\}_{i=1,\ldots,18}$, we select the first four with the largest number of snapshots. These four subsets lie inside $(0, 0.5)$, $(0.5, 1)$, $(1, 1.5)$ and $(1.5, 2)$, respectively, which are also the time-intervals with no feature interaction.

For these four subsets, Figure 4b and Figure 4c compare $\Xi_m(\mathcal{S}_i)$ to $\Xi_m(\mathcal{S}_{\text{calib},i})$. Already for $m = 1$, the value of $\Xi_m(\mathcal{S}_{\text{calib},1/18})$ is $\approx 10^{-5}$ and is machine-precision zero for $m = 3$. For the same value of m, the value of $\Xi_m(\mathcal{S}_{1/18})$ is ≈ 1. The value of $\Xi_m(\mathcal{S}_{\text{calib},7/12})$ behaves differently. For $m = 4$ and larger, it does not appear to converge to zero and stagnates at $\approx 10^{-4}$. For the same value of $m = 4$, the value of $\Xi_m(\mathcal{S}_{7/12})$ is $\approx 10^{-1}$. This is 10^3 times larger than the value of $\Xi_m(\mathcal{S}_{\text{calib},7/12})$.

Note that $\mathcal{S}_{1/18}$ contains snapshots that have two sin-bumps that do not interact with each other and have a constant speed of one. One can conclude that this results in the calibrated manifold $\mathcal{M}_{\text{calib}}(D_{1/18})$ consisting of a single function. Figure 5a shows the snapshots in $\mathcal{S}_{\text{calib},1}$. The snapshots change (very) little over time, with no change being visible. In contrast, as depicted by Figure 5b, the snapshots in $\mathcal{S}_{\text{calib},7}$ change substantially over time. This could explain the superior calibration of $\mathcal{S}_{1/18}$ as compared to $\mathcal{S}_{7/12}$.

5.3 Test case-3 An exact solution to the Sod's shock tube problem can be found in [10]. For brevity, we do not repeat the exact solution here. We present the results for velocity (v) and density (ρ). The results for pressure (P) are similar to



(a) (b) (c)

Fig. 4: *Results for test case-2. (a) Time-trajectory of the approximate feature locations. Kink and discontinuity locations shown in red and blue, respectively. The dashed black lines show the temporal locations of the reference snapshots. (b) Compares $\Xi_m(\mathcal{S}_{1/18})$ to $\Xi_m(\mathcal{S}_{\text{calib},1/18})$. (c) Compares $\Xi_m(\mathcal{S}_{7/12})$ to $\Xi_m(\mathcal{S}_{\text{calib},7/12})$. The y-axis of (b) and (c) is on a log-scale.*

that for density (ρ) and are not discussed for brevity.

5.3.1 Results for density (ρ) The initial data has a single discontinuity that splits into a rarefaction fan with two kinks and two discontinuities; see Figure 6a. The approximate feature trajectories are shown in Figure 6b. Around $t = 0$, the kinks are too close to each other and are identified as a single discontinuity. For $t \in (0.01, 0.1)$, because of a large slope inside the rarefaction fan, the algorithm is unable to distinguish between the two kinks and identifies the midpoint of the two kinks as the kink location. Only after $t = 0.1$, the spread of the rarefaction fan allows for an accurate identification of the two kinks.

Algorithm 2.1 generates $N = 11$ different reference snapshots, the location of which are shown in Figure 6b. Because the features are too close to each other, the reference snapshot changes frequently close to $t = 0$. Around $t = 0.1$, the two kinks are identified correctly and the algorithm generates an additional reference snapshot.

To study Ξ_m, out of $\{[t]_i\}_{i=1,\ldots,N}$, we select the two largest subsets. These two subsets lie inside $(0.02, 0.1)$ and $(0.1, 0.2)$, respectively. Figure 6c compares $\Xi_m(\mathcal{S}_{8/9})$

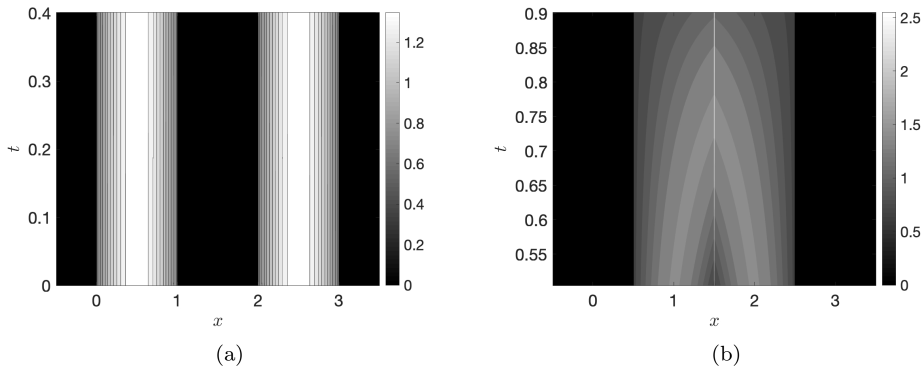


Fig. 5: *Results for test case-2. (a) and (b) show the snapshots in* $\mathcal{S}_{\text{calib},1}$ *and* $\mathcal{S}_{\text{calib},7}$*, respectively.*

to $\Xi_m(\mathcal{S}_{\text{calib},8/9})$. For both $i = 8$ and $i = 9$, $\Xi_m(\mathcal{S}_{\text{calib},i})$ decays much faster than $\Xi_m(\mathcal{S}_i)$. For $m = 10$, which is 0.5% of M, calibration provides at least one order-of-magnitude improvement, with the results for $i = 9$ being better than those for $i = 8$. Precisely,

$$\Xi_{10}(\mathcal{S}_{\text{calib},8}) \approx 5 \times 10^{-2} \times \Xi_{10}(\mathcal{S}_8), \quad \Xi_{10}(\mathcal{S}_{\text{calib},9}) \approx 1 \times 10^{-2} \times \Xi_{10}(\mathcal{S}_9). \tag{5.12}$$

As m increases, the difference between $\Xi_m(\mathcal{S}_{\text{calib},i})$ and $\Xi_m(\mathcal{S}_i)$ becomes larger. For $m = 50$, which is 2.5% of M, we find an improvement of at least two orders of magnitude

$$\Xi_{50}(\mathcal{S}_{\text{calib},8}) \approx 10^{-2} \times \Xi_{50}(\mathcal{S}_8), \quad \Xi_{50}(\mathcal{S}_{\text{calib},9}) \approx 7 \times 10^{-3} \times \Xi_{50}(\mathcal{S}_9). \tag{5.13}$$

5.3.2 Results for velocity (v) Apart from $t = 0$, $v(\cdot, t)$ has two kinks and a discontinuity. Similar to test case-1, the two kinks are identified once they have moved sufficiently far away from each other, otherwise they are identified as a single discontinuity. The discontinuity is identified accurately at all time instances; see Figure 7a.

Algorithm 2.1 generates $N = 5$ different reference snapshots. Most of these reference snapshots are close to $t = 0$. The time interval $(0.01, 0.2)$ is the largest subset of D where the reference snapshot does not change. For this time-interval, in Figure 7b we compare $\Xi_m(\mathcal{S}_i)$ to $\Xi_m(\mathcal{S}_{\text{calib},i})$. Already for $m = 1$, we find that $\Xi_m(\mathcal{S}_{\text{calib},5}) \approx 10^{-3}$, which is two orders of magnitude smaller than $\Xi_m(\mathcal{S}_5)$. For $m = 30$, which is 1.5% of M, $\Xi_{30}(\mathcal{S}_{\text{calib},5})$ is (machine precision) zero, whereas $\Xi_{30}(\mathcal{S}_5)$ is 6.4×10^{-2}.

5.4 Test case-4 An exact solution to (5.8) is given as

$$\begin{aligned} u(x,t) &= \mathbb{1}_{[0.1,0.5]}\left(t - \frac{x - x_{\min}}{\beta}\right), \quad \forall x \in (0, x_{\min} + \beta t], t \in D, \\ u(x,t) &= (\sin(\pi(x - \beta t)) + 1)\mathbb{1}_{[0,1]}(x - \beta t), \quad \forall x \in (x_{\min} + \beta t, x_{\max}), t \in D. \end{aligned} \tag{5.14}$$

For $t \in [0, 0.1)$, the solution contains two discontinuities that move to the right. At $t = 0.1$ and $t = 0.5$, two additional discontinuities enter from the left boundary.

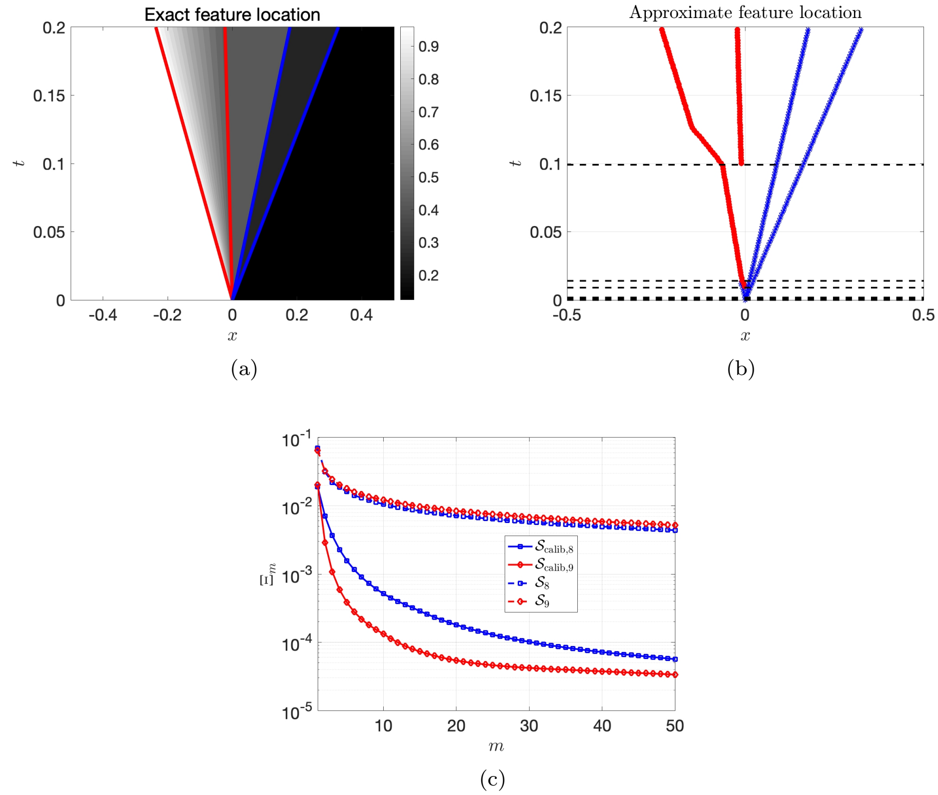
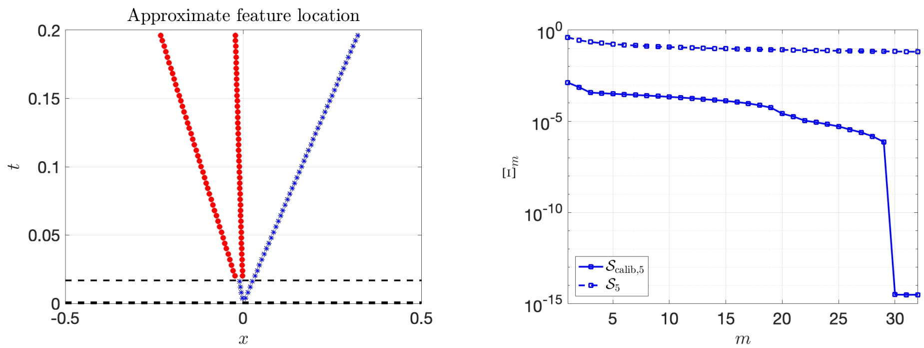


Fig. 6: *Test case-3: results for ρ. (a) and (b) show the exact and the approximate feature trajectory, respectively. Kinks are shown in red and the discontinuities in blue. Dashed lines in (b) show the temporal locations of the reference solutions. (c) Compares $\Xi_m(\mathcal{S}_{8/9})$ to $\Xi_m(\mathcal{S}_{\text{calib},8/9})$. The y-axis in (c) is on a log-scale.*

Figure 8a shows the approximate location of these discontinuities. Algorithm 2.1 generates $N = 11$ different reference snapshots. The reference snapshot changes when a new discontinuity enters from the boundary.

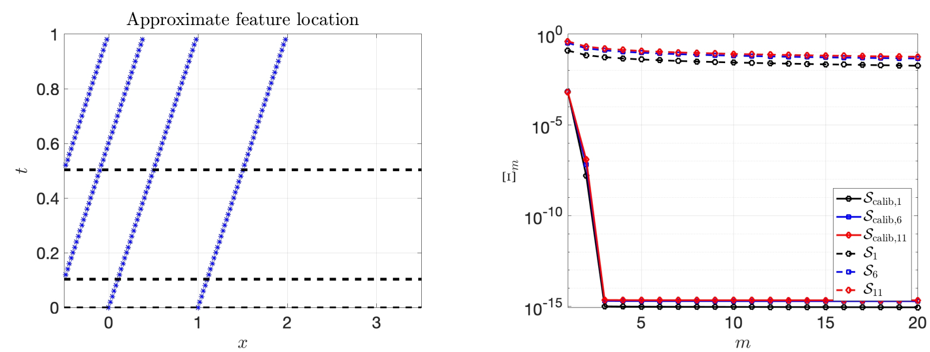
Figure 8b compares $\Xi_m(\mathcal{S}_i)$ to $\Xi_m(\mathcal{S}_{\text{calib},i})$ for the three largest subsets $[t]_i$. Clearly, $\Xi_m(\mathcal{S}_{\text{calib},i})$ decays much faster than $\Xi_m(\mathcal{S}_i)$, and is zero for $m = 3$. For the same value of m, $\Xi_m(\mathcal{S}_i)$ is $\approx 2 \times 10^{-2}$. With the above exact solution, it is easy to check that the calibrated manifold consists of a single function, which could explain the great improvement offered by calibration.

6 Conclusions We have proposed an algorithm to induce a fast singular value decay in a snapshot matrix resulting from hyperbolic equations. The algorithm relies on computing the snapshots on a transformed spatial domain with the transformation computed using feature matching between a reference and the other snapshots. The choice of the reference snapshot ensures that the transformation is a homeomorphism and has a lower and an upper bound on its weak derivative—we found these two properties desirable for both the theoretical analysis and a numerical implementation. To account for feature interaction and formation (i.e., cases where shocks collide, shocks



(a) (b)

Fig. 7: *Test case-3: results for the velocity v. (a) Approximate feature location. (b) Compares $\Xi_m(\mathcal{S}_5)$ to $\Xi_m(\mathcal{S}_{\text{calib},5})$. The y-axis in (b) is on a log-scale.*



(a) (b)

Fig. 8: *Results for test case-4. (a) Approximate feature location. (b) Compares $\Xi_m(\mathcal{S}_{1/6/11})$ to $\Xi_m(\mathcal{S}_{\text{calib},1/6/11})$. The y-axis in (b) is on a log-scale.*

form, etc.), we have proposed an adaptive reference snapshot selection technique. With this technique, we can divide the snapshot matrix into sub-matrices with each sub-matrix containing snapshots with no feature interaction/formation. In each of the sub-matrices, we perform feature matching as usual.

Under regularity assumptions on the initial data and the flux function, we have proven that feature matching results in a fast m-width decay of a so-called calibrated manifold. Our proof exploits the regularity of functions in a calibrated manifold. We have performed numerical experiments on a broad range of problems involving non-linear system of equations and time-dependent boundary conditions. Our experiments verify that feature matching is successful in inducing a fast singular value decay in a snapshot matrix. We also found that feature matching performs exceptionally well for problems where the calibrated manifold contains a single function.

We observe that although the singular values of a calibrated snapshot matrix decay fast, they can stagnate at a value that scales with the spatial grid resolution.

The stagnation is a by-product of computing the spatial transform using a numerical approximation of the exact solution and indicates that, for hyperbolic problems, not much is gained by increasing the dimension of the reduced-order model beyond a certain limit.

References

[1] M. Bachmayr and A. Cohen. Kolmogorov widths and low-rank approximations of parametric elliptic PDEs. *arXiv:1502.03117*, 2015.

[2] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

[3] N. Cagniart, Y. Maday, and B. Stamm. *Model Order Reduction for Problems with Large Convection Effects*. Springer International Publishing, Cham, 2019.

[4] B. Cockburn, G. E. Karniadakis, and C.-W. Shu. *Discontinuous Galerkin Methods: Theory, Computation and Applications*. Springer International Publishing, 2000.

[5] P. Constantine and G. Iaccarino. Reduced order models for parameterized hyperbolic conservations laws with shock reconstruction. *Annual Brief: Center for Turbulence Research*, 6(1), 2012.

[6] W. Dahmen, C. Plesken, and G. Welper. Double greedy algorithms: Reduced basis methods for transport dominated problems. *ESAIM: M2AN*, 48(3):623–663, 2014.

[7] M. Drohmann, B. Haasdonk, and M. Ohlberger. Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. *SIAM Journal on Scientific Computing*, 34(2):A937–A969, 2012.

[8] F. Dubois and P. L. Floch. Boundary conditions for nonlinear hyperbolic systems of conservation laws. *Journal of Differential Equations*, 71(1):93 – 122, 1988.

[9] V. Ehrlacher, D. Lombardi, O. Mula, and F.-X. Vialard. Nonlinear model reduction on metric spaces. application to one-dimensional conservative PDEs in Wasserstein spaces. *ESAIM: Mathematical Modelling and Numerical Analysis*, 2019.

[10] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer, Berlin, 2009.

[11] J.-F. Gerbeau and D. Lombardi. Approximated lax pairs for the reduced order integration of nonlinear evolution equations. *Journal of Computational Physics*, 265:246 – 269, 2014.

[12] L. Krivodonova, J. Xin, J.-F. Remacle, N. Chevaugeon, and J. Flaherty. Shock detection and limiting with discontinuous Galerkin methods for hyperbolic conservation laws. *Applied Numerical Mathematics*, 48(3):323 – 338, 2004.

[13] K. Lee and K. T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, 2020.

[14] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[15] S. Mowlavi and T. P. Sapsis. Model order reduction for stochastic dynamical systems with continuous symmetries. *SIAM Journal on Scientific Computing*, 40(3):A1669–A1695, 2018.

[16] N. J. Nair and M. Balajewicz. Transported snapshot model order reduction approach for parametric, steady-state fluid flows containing parameter-

dependent shocks. *International Journal for Numerical Methods in Engineering*, 117(12):1234–1262, 2019.

[17] M. Nonino, F. Ballarin, G. Rozza, and Y. Maday. Overcoming slowly decaying Kolmogorov N-width by transport maps: application to model order reduction of fluid dynamics and fluid–structure interaction problems. *arXiv:1911.06598*, 2019.

[18] M. Ohlberger and S. Rave. Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing. *Comptes Rendus Mathematique*, 351(23):901 – 906, 2013.

[19] B. Peherstorfer. Model reduction for transport-dominated problems via online adaptive bases and adaptive sampling. *arXiv:1812.02094*, 2018.

[20] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction.* Springer International Publishing, 2016.

[21] D. Ray and J. S. Hesthaven. Detecting troubled-cells on two-dimensional unstructured grids using a neural network. *Journal of Computational Physics*, 397:108845, 2019.

[22] J. Reiss, P. Schulze, J. Sesterhenn, and V. Mehrmann. The shifted proper orthogonal decomposition: A mode decomposition for multiple transport phenomena. *SIAM Journal on Scientific Computing*, 40(3):A1322–A1344, 2018.

[23] D. Rim and K. T. Mandli. Displacement interpolation using monotone rearrangement. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1503–1531, 2018.

[24] D. Rim, S. Moe, and R. J. LeVeque. Transport reversal for model reduction of hyperbolic partial differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 6(1):118–150, 2018.

[25] D. Rim, B. Peherstorfer, and K. T. Mandli. Manifold approximations via transported subspaces: Model reduction for transport-dominated problems. *arXiv:1912.13024*, 2019.

[26] C. W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15(03):997–1013, 2005.

[27] N. Sarna and S. Grundel. Model reduction of time-dependent hyperbolic equations using collocated residual minimisation and shifted snapshots. *arXiv:2003.06362*, 2020.

[28] T. Taddei, S. Perotto, and A. Quarteroni. Reduced basis techniques for nonlinear conservation laws. *ESAIM: M2AN*, 49(3):787–814, 2015.

[29] K. Veroy, C. Prud'homme, D. Rovas, and A. Patera. *A Posteriori Error Bounds for Reduced-Basis Approximation of Parametrized Noncoercive and Nonlinear Elliptic Partial Differential Equations.* 16th AIAA Computational Fluid Dynamics Conference, Orlando, Florida, 2012.

[30] M. J. Vuik and J. K. Ryan. Multiwavelet troubled-cell indicator for discontinuity detection of discontinuous Galerkin schemes. *Journal of Computational Physics*, 270:138 – 160, 2014.

[31] G. Welper. h and hp-adaptive interpolation by transformed snapshots for parametric and stochastic hyperbolic PDEs. *arXiv:1710.11481*, 2017.

[32] G. Welper. Interpolation of functions with parameter dependent jumps by transformed snapshots. *SIAM Journal on Scientific Computing*, 39(4):A1225–A1250, 2017.

Appendix A. Regularity of functions in $\mathcal{M}_{\text{calib}}(D)$. The definition of X_i provides $X_i \in C^0(\Omega_i^D)$ by the implicit function theorem and the bound on β_i.

Moreover,

$$\begin{aligned} D_t X_i(x,t) &= -\frac{f'(u_0(X_i(x,t)))}{\beta(X_i(x,t),t)} =: \hat{\mathcal{G}}(X_i(t,x),t), \\ D_x X_i(x,t) &= \frac{1}{\beta(X_i(x,t),t)} =: \tilde{\mathcal{G}}(X_i(t,x),t). \end{aligned} \tag{A.1}$$

The regularity of f and u_0 and the assumption on β imply that $\hat{\mathcal{G}}, \tilde{\mathcal{G}} \in C^{\omega-1}(\Omega_i \times D)$ which implies that $X_i \in W^{\omega,\infty}(\Omega_i^D)$ by bootstrapping.

Next, we show that $\varphi \in L^\infty(\Omega; W^{\omega,\infty}(D))$. Since $\varphi(x,t) \leq x_{\max}$, we have $\varphi \in L^\infty(\Omega \times D)$. The definition of φ in (2.7) implies that $D_t^\omega \varphi \in L^\infty(\Omega \times D)$ if $z \in W^{\omega,\infty}(D)$. When z is a kink location, following the characteristics forwards in time we find $z(t) = z(0) + f'(u_0(z_0(0))) \cdot t$, which provides the desired regularity. When z is a shock location, we proceed as follows.

For simplicity, assume that $p_0 = 1$ with a shock at $z(t)$. The argument remains the same for (non-interacting) multiple shocks. Consider the weak solution

$$u(x,t) = \begin{cases} \tilde{u}_0(x,t) & \text{in } \Omega_0^D \\ \tilde{u}_1(x,t) & \text{in } \Omega_1^D \end{cases}. \tag{A.2}$$

Above, $\Omega_{0/1}^D$ are as given in (3.4). Following the characteristics forward in time, we find

$$\tilde{u}_0(x,t) = u_0(X_0(x,t)), \quad \tilde{u}_1(x,t) = u_0(X_1(x,t)). \tag{A.3}$$

The assumption on β_i means that inside Ω_i^D characteristics of u are bounded away from intersecting each other. Thus, $\tilde{u}_0, \tilde{u}_1$ inherit their regularity from the regularity of the initial data between the features, i.e. $\tilde{u}_i \in W^{\omega,\infty}(\Omega_i^D)$ and (since intersection of characteristics is not imminent), we can find $c, \epsilon > 0$ such that $\tilde{u}_0$ has a extension $\tilde{u}_0^{\text{ex}} \in W^{\omega,\infty}(\Omega_0^{D,\text{ex}})$ (that is constant along characteristics) with

$$\Omega_0^{D,\text{ex}} := \{(x,t) \ : \ x \leq z(t) + \min(\epsilon, ct), \ t \leq T\}. \tag{A.4}$$

A similar definition holds for $\tilde{u}_1^{\text{ex}}$. By the Rankine-Hugoniot condition, z satisfies

$$d_t z(t) = \mathcal{H}(\tilde{u}_0^{\text{ex}}(z(t),t), \tilde{u}_1^{\text{ex}}(z(t),t)) \quad \text{where} \quad \mathcal{H}(a,b) := \begin{cases} \frac{f(a)-f(b)}{a-b}, & a \neq b \\ f'(a), & a = b \end{cases}. \tag{A.5}$$

Since $f \in C^{\omega+1}$ we have $\mathcal{H} \in C^\omega(\mathbb{R}^2)$ implying that z satisfies $d_t z(t) = h(z(t),t)$ with $h = \mathcal{H}(\tilde{u}_0^{\text{ex}}, \tilde{u}_1^{\text{ex}})$ and $h \in C^{\omega-1}\left(\left(\Omega_0^{D,\text{ex}} \cap \Omega_1^{D,\text{ex}}\right) \times D\right)$. Since $\Omega_0^{D,\text{ex}} \cap \Omega_1^{D,\text{ex}}$ is compact and $\tilde{u}_i^{\text{ex}}$ is Lipschitz, h is globally Lipschitz continuous providing a global solution to (A.5). Furthermore, since $h \in C^{\omega-1}$, $z \in C^\omega(D)$. Since D is closed, we have $z \in W^{\omega,\infty}(D)$ and thus $\varphi \in L^\infty(\Omega, W^{1,\infty}(D))$.

Using (3.10) the regularity of g is a direct consequence of the regularity of u_0, X_i, and φ

Appendix B. Rarefaction fan. Let $X_i(x,t)$ be as given in (3.10). We show that the second condition in (3.13) can be satisfied if Ω_i contains a rarefaction fan. Let $\Omega = (-1,2)$ and let $D = [0, 0.5]$ and consider the initial data

$$u_0(x) := \begin{cases} (f')^{-1}(0), & x \leq 0 \\ (f')^{-1}(x), & x \in (0,1) \\ (f')^{-1}(1), & x \in [1,2) \end{cases}. \tag{B.1}$$

With the above initial data, the solution reads

$$u(x,t) := \begin{cases} 0, & x \leq 0 \\ (f')^{-1}\left(\frac{x}{t+1}\right), & x \in (0, 1+t) \\ (f')^{-1}(1), & x \in [1+t, 2) \end{cases}. \tag{B.2}$$

Assume that for all $t \in D$, $u(\cdot,t)$ has a kink at both $x = 0$ and $x = 1+t$. Thus, we have two features. The kink locations are given as

$$z_1(t) = 0, \quad z_2(t) = 1+t. \tag{B.3}$$

Using the above relation, for $x \in \Omega_2 = (z_1(t), z_2(t))$, the spatial transform reads

$$\varphi(x,t) = x\,(1+t)\,. \tag{B.4}$$

For $i = 2$ and for all $x \in \Omega_2$, the definition of X_i in (3.10), the expression for u_0, and the above expression for φ provides

$$X_2(x,t) + tX_2(x,t) = x \;\Rightarrow\; X_2(x,t) = \frac{x}{1+t}. \tag{B.5}$$

Appendix C. Estimate for $\|u \circ \varphi - u \circ \varphi_M\|_{L^2(\Omega \times D)}$.

1. The following proof is an extension of the one given in [32] for L^2 functions. For some $\epsilon > 0$, define $\Omega_\epsilon : \{x \in \Omega \;:\; \mathrm{dist}(x, \partial\Omega) > \epsilon\}$. Let $u_\epsilon \in C^\infty(\Omega_\epsilon)$ be a mollification of $u(\cdot,t)$ over Ω_ϵ. Then, the following holds

$$\|u_\epsilon - u(\cdot,t)\|_{L^2(\Omega_\epsilon)} \xrightarrow{\epsilon\to 0} 0, \quad \|u_\epsilon(\cdot,t)\|_{BV(\Omega_\epsilon)} \leq \|u(\cdot,t)\|_{BV(\Omega_\epsilon)}. \tag{C.1}$$

Triangle's inequality provides

$$\begin{aligned} \|u \circ \varphi - u \circ \varphi_M\|_{L^2(\Omega_\epsilon \times D)} \leq & \|u \circ \varphi - u_\epsilon \circ \varphi\|_{L^2(\Omega \times D)} \\ & + \|u \circ \varphi_M - u_\epsilon \circ \varphi_M\|_{L^2(\Omega_\epsilon \times D)} \\ & + \|u_\epsilon \circ \varphi - u_\epsilon \circ \varphi_M\|_{L^2(\Omega_\epsilon \times D)}. \end{aligned}$$

Applying a domain transformation and using (1.5), we find

$$\|u \circ \varphi - u_\epsilon \circ \varphi\|_{L^2(\Omega \times D)} \lesssim \epsilon, \quad \|u \circ \varphi_M - u_\epsilon \circ \varphi_M\|_{L^2(\Omega \times D)} \lesssim \epsilon. \tag{C.2}$$

Because of the above two relations, it is sufficient to bound $\|u_\epsilon \circ \varphi - u_\epsilon \circ \varphi_M\|_{L^2(\Omega_\epsilon \times D)}$. For $s \in [0,1]$, define $\Phi(x,t,s) = s\varphi(x,t) + (1-s)\varphi_M(x,t)$. Using Φ, we write

$$\begin{aligned} \|u_\epsilon \circ \varphi - u_\epsilon \circ \varphi_M\|^2_{L^2(\Omega_\epsilon \times D)} = & \int_{\Omega_\epsilon \times D} \left(\int_0^1 \partial_s u_\epsilon(\Phi(x,t,s))ds\right)^2 dxdt \\ \leq & \|u_\epsilon\|_{L^\infty(D), BV(\Omega_\epsilon))} \\ & \times \int_{\Omega_\epsilon \times D} \left(\int_0^1 \partial_s |u_\epsilon(\Phi(x,t,s))| ds\right) dxdt \\ \leq & \|u\|_{L^\infty(D;BV(\Omega))} \|u\|_{L^2(D;BV(\Omega))} \\ & \times \|\varphi - \varphi_M\|_{L^\infty(\Omega \times D)}. \end{aligned}$$

Above, the last inequality follows from [32] and (C.1).

2. By definition,

$$\varphi(z_i(0),t) = z_i(t), \quad \varphi_M(z_{M,i}(0),t) = z_{M,i}(t). \tag{C.3}$$

We refer to $z_i(0)$ and $z_{M,i}(0)$ as the nodes and to $z_i(t)$ and $z_{M,i}(t)$ as the node values of a spatial transform. We introduce an intermediate (continuous and piecewise linear) spatial transform $\hat{\varphi}$ that has the same nodes as $\varphi(\cdot,t)$ and the same nodal values as $\varphi_M(\cdot,t)$ i.e., $\hat{\varphi}(z_i(0),t) = z_{M,i}(t)$. By triangle's inequality,

$$\|\varphi_M - \varphi\|_{L^\infty(\Omega\times D)} \leq \|\varphi - \hat{\varphi}\|_{L^\infty(\Omega\times D)} + \|\hat{\varphi} - \varphi_M\|_{L^\infty(\Omega\times D)}. \tag{C.4}$$

Because φ and $\hat{\varphi}$ have the same nodes, we conclude that

$$\|\varphi - \hat{\varphi}\|_{L^\infty(\Omega\times D)} = \max_j \|z_{M,j} - z_j\|_{L^\infty(D)}. \tag{C.5}$$

It is easy to check that the maximum of $|\hat{\varphi}(\cdot,t) - \varphi_M(\cdot,t)|$ occurs at either the nodes $\{z_i(0)\}_i$ or $\{z_{M,i}(0)\}_i$. Computing $|\hat{\varphi}(\cdot,t) - \varphi_M(\cdot,t)|$ at these nodes provides

$$\begin{aligned} \|\hat{\varphi}(\cdot,t) - \varphi_M(\cdot,t)\|_{L^\infty(\Omega)} &\leq \|D_x\varphi_M(\cdot,t)\|_{L^\infty(\Omega)} \max_j |z_{M,j}(t) - z_j(t)| \\ &\leq \mathcal{K}_1 \max_j |z_{M,j}(t) - z_j(t)|. \end{aligned} \tag{C.6}$$

where $\mathcal{K}_1$ is the constant in (1.5).

Appendix D. Relation to MRA. We briefly relate our feature detection method to that proposed in [30]. We specialise the formulation for a FV scheme, generalisations to arbitrary order discontinuous-Galerkin type schemes can be found in the references therein. We divide Ω into uniform $N_l = 2^l$ elements with $l \in \mathbb{N}$. Such a choice of N_l results in a hierarchy of grids parameterised by l. With $\mathcal{I}_i^l$ we represent the i-th cell at level l. With $u_i^l(t)$ we denote the FV approximation of $u(\cdot,t)$ in $\mathcal{I}_i^l$.

In the middle of every $\mathcal{I}_i^{l-1}$ lies a face that is shared between $\mathcal{I}_{2i-1}^l$ and $\mathcal{I}_{2i}^l$. Let $J_i^{l-1}(t)$ denote the jump of the FV solution across this face i.e.,

$$J_i^{l-1}(t) = |u_{2i-1}^l(t) - u_{2i}^l(t)|. \tag{D.1}$$

Thus, given u_i^l, we can compute all of J_i^{l-1}. The coefficient $J_i^{l-1}/2$ is the same as the so-called wavelet coefficient in the MRA. Define

$$D^{l-1}(t) := \max_{i\in 1,\ldots,2^{l-1}} J_i^{l-1}(t). \tag{D.2}$$

Similar to $\mathcal{B}(t)$ in (4.2), define

$$\mathcal{B}^{l-1}(t) := \{i \;:\; |J_i^{l-1}(t)| > C \times D^{l-1}(t),\ i \in \{1,\ldots 2^{l-1}\}\}. \tag{D.3}$$

At level $l-1$, cells with index in $\mathcal{B}^{l-1}$ are flagged. Due to the grid hierarchy, the cells at level l that have a discontinuity are $\{2i-1 \;:\; i \in \mathcal{B}^{l-1}\}$ and $\{2i \;:\; i \in \mathcal{B}^{l-1}\}$. Above, C is the same as that defined in (4.2).

As is clear from the definition of J_e^{l-1}, in MRA one computes the jump in the FV solution at every alternate face. Equivalently, MRA does not compute jumps at any face at level $l-1$. Therefore, a discontinuity (independent of its strength) aligned with any of these faces is not detected. Such discontinuities do not contribute to an oscillatory numerical solution. Therefore, for the purpose of flagging cells for suppressing oscillations, MRA is sufficient. However, in the present context, missing out on large shocks is undesirable. Therefore, we compute the jumps at all the faces, which allows us to detect shocks that could be aligned with cell boundaries.

Appendix E. Flagging of discontinuous regions. For simplicity, we assume that $u_M(\cdot,t)$ is a projection of $u(\cdot,t)$ onto the FV basis. At least computationally, for a small enough grid size, similar observation holds for a $u_M(\cdot,t)$ computed with a FV scheme.

1. Locally differentiable: If $u(\cdot,t)|_{\mathcal{I}_{e-1}\cup\mathcal{I}_e}$ is C^1 then Taylor expansion provides

$$J_e \leq \Delta x \|\partial_x u(\cdot,t)\|_{C^0(\mathcal{I}_{e-1}\cup\mathcal{I}_e)}. \tag{E.1}$$

2. Discontinuous: Let $u(\cdot,t)$ have a discontinuity inside $\mathcal{I}_e$. Let the point of discontinuity be $z^D = x_e + l\times\Delta x$ where $l\in(0,1)$. Furthermore, let $u(\cdot,t)$ be piecewise constant in $\mathcal{I}_{e-1}\cup\mathcal{I}_e$ with the value before and after the discontinuity being u_- and u_+, respectively. Then

$$J_e = |u_- - u_+|(1-l). \tag{E.2}$$

3. Kink: Assume that $u(\cdot,t)$ is continuous, is piecewise linear in $\mathcal{I}_{e-1}\cup\mathcal{I}_e$ and has a kink at $z^K = x_e + l\times\Delta x$. Then, assuming $u(z^K,t)=0$, $u(\cdot,t)|_{\mathcal{I}_{e-1}\cup\mathcal{I}_e}$ reads

$$u(\cdot,t)|_{\mathcal{I}_{e-1}\cup\mathcal{I}_e} = \begin{cases}(x-x_K)\partial u_- & x<z_K\\ (x-x_K)\partial u_+ & x\geq z_K\end{cases} \tag{E.3}$$

Above, ∂u_- and ∂u_+ are the left and right slopes respectively. With the above $u(\cdot,t)$, we find

$$J_e = \frac{\Delta x}{2}|(\partial u_- - \partial u_+)l^2 - 2\times\partial u_+|.$$

With the above relations and the form of $\mathcal{B}(t)$ given in (4.2), we draw the following three conclusions. First, regions where the solution is C^1 but has a large gradient might be identified as discontinuities. Second, shocks with a strength (i.e., $|u_- - u_+|$) of $\mathcal{O}(\Delta x)$ might go undetected. Third, kinks with a large left and right derivative might be identified as discontinuities. In relation to the second point, in case $J_e(t) < C\Delta x$, where C is as given in (4.2), one can show that the semi-discrete numerical solution already has the regularity necessary for a fast m-width decay.