

Feature-Informed Data Assimilation

Apoorv Srivastava^{a,1}, Wei Kang^{b,2}, Daniel M. Tartakovsky^{c,3}

^a*Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA*

^b*Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943, USA*

^c*Department of Energy Science and Engineering, Stanford University, Stanford, CA 94305, USA*

Abstract

We introduce a mathematical formulation of feature-informed data assimilation (FIDA). In FIDA, the information about feature events, such as shock waves, level curves, wavefronts and peak value, in dynamical systems are used for the estimation of state variables and unknown parameters. The observation operator in FIDA is a set-valued functional that involves a search process over a function of state variables, which is fundamentally different from the observation operators in conventional data assimilation. We present three numerical experiments, in which shocks and expanding waves are observed features. These examples serve to demonstrate FIDA’s ability to estimate model parameters from such noisy observations.

Keywords: observables, parameter estimation, shock

1. Introduction

Data assimilation (DA) is routinely used in many fields of science and engineering to combine a mathematical model of the system of interest with observations. Its goal is to obtain an “optimal” estimator/predictor of the system’s behavior by using observational data to improve the knowledge of the model’s structure and/or parameters. The probabilistic framework provides a natural means to quantify uncertainty in model predictions and experimental errors, both of which are ubiquitous. It describes the system parameters and model predictions in terms of their respective probability density functions (PDFs), which are informed by the model and data via the Bayes rule (see representative surveys [1, 2, 3, 4] for an overview of the field). Alternative approaches to Bayesian DA include particle filters [5, 6], various versions of (e.g., ensemble) Kalman filter [7, 8], four-dimensional variational DA [9, 10], and variational DA on statistical manifolds [11, 12].

The relative performance of these and other DA strategies depends on the model’s degree of non-linearity and the type of observational data. The model’s nonlinearity might cause its output, e.g., the state variables, to become discontinuous and/or highly non-Gaussian (possibly multi-modal), degrading

¹email: apoorv1@stanford.edu

²email: wkang@nps.edu

³email: tartakovsky@stanford.edu

15 the performance of such computationally efficient DA techniques as Kalman filters. That is one reason why successful applications of DA to problems with discontinuous solutions and shocks are relatively scarce [13, 14].

The data type enters the DA formulation through an observation operator, which accounts for noise in the data and ensures that observations can be quantitatively compared with the model’s predictions. Standard DA algorithms are designed to handle observations whose operator is a function of the state variable at a given space-time location. However, some data are equipped with a functional or a nonlocal operator that depends on the environment in a region in space. That is typical of satellite observations, whose assimilation has remained a challenge for several decades [15] because the observation operator is difficult to model or the operator’s dimension is too high for DA algorithms.

25 We focus on problems in which observational data are associated with a discontinuity or an extreme value of state variables, e.g., shock waves, level curves, wavefronts and peak value. A representative example of such problems is detonation phenomena; they exhibit shock waves whose locations are observed from images [16], without collecting any measurements of the state variables (e.g., pressure or temperature). Other applications of this kind are a combustion wavefront [17] and a peak-to-peak plot of chemical reactors [18, 19]. Typical data of this kind capture key features of the phenomena (e.g., a shock location, a spatial sub-domain in which the state variable exceeds a certain threshold) and might possess a non-differentiable observation operator that involves a search process over a function of the state variables. The need in a variety of disciplines and applications to assimilate feature based information motivates us to introduce a general mathematical formulation of feature-informed DA (FIDA) and to explore essential mathematical concepts as well as effective computational tools for FIDA.

In Section 2, a conventional formulation of DA is briefly introduced as background, followed by the formulation of FIDA problem in which the observation operator is a set-valued functional in an infinitely dimensional space. In Section 3, we report results of our numerical experiments, which suggest that the information content of feature data might be sufficient to estimate states or parameters that are not measured directly. Main conclusions drawn from this study are summarized in Section 4.

2. Problem Formulation

The formulation of a DA problem consists of a forward model, e.g., deterministic partial-differential equations or PDEs; a probabilistic representation of the model error; an observation operator that relates the observables to the model predictions; and a probabilistic model of the observation noise. The goal of DA is to combine the observations (e.g., sensor data or images) and the forward model’s predictions to obtain an optimal estimator of the state of the system as it evolves in time. It can also be applied to identify unknown parameters in the forward model.

2.1. Conventional DA

Consider a dynamical system described, at any space-time point $(\mathbf{x}, t) \in \mathbb{R}^{N_{\mathbf{x}}} \times \mathbb{R}$, by $N_{\mathbf{v}}$ state variables $\mathbf{u}(\mathbf{x}, t) = \{u_1(\mathbf{x}, t), \dots, u_{N_{\mathbf{v}}}(\mathbf{x}, t)\}$. The spatiotemporal evolution of $\mathbf{u}(\mathbf{x}, t)$ is governed by PDEs, which contain $N_{\mathbf{p}}$ uncertain parameters (random variables) $\mathbf{P} \in \mathbb{R}^{N_{\mathbf{p}}}$. To make the dependence of the state variables $\mathbf{u}$ on a realization $\mathbf{p}$ of random parameters $\mathbf{P}$ explicit, we write $\mathbf{u}(\mathbf{x}, t; \mathbf{p})$. A suitable discretization method transforms these PDEs into ordinary differential equations (ODEs). As an illustrative example, let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ be a set of M grid points in space. Let

$$\mathbf{U}(t; \mathbf{p}) = \begin{bmatrix} \mathbf{u}(\mathbf{x}_1, t; \mathbf{p}) \\ \mathbf{u}(\mathbf{x}_2, t; \mathbf{p}) \\ \dots \\ \mathbf{u}(\mathbf{x}_M, t; \mathbf{p}) \end{bmatrix} \in \mathbb{R}^N, \quad N = N_{\mathbf{v}} \cdot M \quad (2.1)$$

be a discretized trajectory at the grid points. Then, the governing PDEs are approximated with coupled ODEs

$$\frac{d\mathbf{U}}{dt} = \mathcal{M}(t, \mathbf{U}, \mathbf{p}) + \mathbf{w}(t), \quad (2.2)$$

where $\mathcal{M} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is the discretized forward model, and $\mathbf{w}(t) \in \mathbb{R}^N$ is a random model error. The solution $\mathbf{U}(t, \mathbf{p})$ represents the state of the system (e.g., temperature, pressure, and flow speed) at t at the grid points.

Let $\mathbf{y} \in \mathbb{R}^{N_{\mathbf{y}}}$ represent a collection of $N_{\mathbf{y}}$ observations at time t ,

$$\mathbf{y}(t) = H(\mathbf{U}(t; \mathbf{p})) + \boldsymbol{\nu}(t), \quad (2.3)$$

where the function $H : \mathbb{R}^N \rightarrow \mathbb{R}^{N_{\mathbf{y}}}$ is the observation operator, and the random process $\boldsymbol{\nu}(t) \in \mathbb{R}^{N_{\mathbf{y}}}$ represents sensor noise. The observation operator $H(\cdot)$ encapsulates the relationship between the observation $\mathbf{y}$ and the system state $\mathbf{U}$. For example, if a sensor capable of recording all $N_{\mathbf{v}}$ state variables were placed at the grid point $\mathbf{x}_1$, then H is a linear function

$$H(\mathbf{U}) = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \mathbf{U}, \quad (2.4)$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is square matrix of zeros, both of dimension $N_{\mathbf{v}} \times N_{\mathbf{v}}$. The measurement can take place at multiple points in space and a measurement location may not be at the grid points. In any case, H is a function defined on $\mathbb{R}^N$.

The goal of DA problem (2.2) is to numerically estimate the values of $\mathbf{U}(t; \mathbf{p})$ and $\mathbf{p}$ by combining the observations $\mathbf{y}$ with the ODE model for $\mathbf{U}(t; \mathbf{p})$.

2.2. Feature-informed DA

For systems operating in harsh environments, such as inside a rotating detonation engine, collecting data on the system state is challenging, if not impossible. On the other hand, the location and propagation of some “eye catching” characteristics, or feature events, such as shock waves, wavefronts and peaking, can be observed. FIDA problem replaces the observation operator H in (2.3) with

$$\mathcal{H} : \mathbf{u}(\cdot, t; \mathbf{p}) \rightarrow \{\mathbf{x}^* : \text{“feature event takes place at } (\mathbf{x}^*, t)\text{”}\} \quad (2.5)$$

and observations $\mathbf{y}(t)$ with $\mathcal{Y}(t)$

$$\mathcal{Y}(t) = \{\mathbf{x}^* + \mathbf{v}(t) : \text{“feature event takes place at } (\mathbf{x}^*, t)\text{”}\}. \quad (2.6a)$$

Here, $\mathbf{v}(t) \in \mathbb{R}^{N_{\mathbf{x}}}$ is the random error associated with the identification of the feature locations, whose values differ for different $\mathbf{x}^*$ at the same t . One of the goals of this study is to demonstrate that the observations $\mathcal{Y}(t)$ may contain sufficient information to estimate the system’s state and parameters. If a feature event represents a shock or discontinuity taking place at $(\mathbf{x}^*, t)$, then the data are

$$\mathcal{Y}(t) = \{\mathbf{x}^* + \mathbf{v}(t) : \|\nabla \mathbf{u}(\mathbf{x}^*, t; \mathbf{p})\| > V\}, \quad (2.6b)$$

where $V > 0$ is a large number. The observation operator $\mathcal{H}$ in (2.6b) is described by the norm of the gradient. Such data can also represent a wavefront where a state variable changes significantly in space, e.g., a combustion wavefront. Another example is data in the form of images, from which one captures the location of maximum points. In this case,

$$\mathcal{Y}(t) = \{\mathbf{x}^* + \mathbf{v}(t) : \mathbf{u}(\mathbf{x}, t; \mathbf{p}) \leq \mathbf{u}(\mathbf{x}^*, t; \mathbf{p}) \text{ for all } \mathbf{x}\}. \quad (2.6c)$$

In yet another example, one might supplement the data in (2.6a) with additional information, e.g., the maximum value of the state variables, $\mathbf{u}^* = \max_{\mathbf{x}} \mathbf{u}(\mathbf{x}, t; \mathbf{p})$,

$$\mathcal{Y}(t) = \{(\mathbf{x}^* + \mathbf{v}_1(t), \mathbf{u}^* + \mathbf{v}_2(t)) : \mathbf{u}(\mathbf{x}, t; \mathbf{p}) \leq \mathbf{u}(\mathbf{x}^*, t; \mathbf{p}) \text{ for all } \mathbf{x}\}, \quad (2.6d)$$

where $\mathbf{v}_1(t) \in \mathbb{R}^{N_{\mathbf{x}}}$ and $\mathbf{v}_2(t) \in \mathbb{R}^{N_{\mathbf{v}}}$ represent sensor noise associated with the observation of $\mathbf{x}^*$ and $\mathbf{u}^*$, respectively. A final example is the data in the form of a level surface, the set of points in space in which the value of the state variable $\mathbf{u}$ equals a constant $\mathbf{C} \in \mathbb{R}^{N_{\mathbf{v}}}$,

$$\mathcal{Y}(t) = \{\mathbf{x}^* + \mathbf{v}(t) : \mathbf{u}(\mathbf{x}^*, t; \mathbf{p}) = \mathbf{C}\}. \quad (2.6e)$$

Level surfaces are often used to identify boundaries such as the invariant sets of dynamical trajectories or boundary layers in fluid mechanics. In general, wavefronts, maximum value and level surfaces may be characterized by functions other than $\mathbf{u}(\mathbf{x}, t; \mathbf{p})$. The formula defining an observation operator should

be customized for each application to follow the physics and mathematical rules. Several examples of feature informed observations in engineering systems can be found in [20].

While the observation operator H in the standard DA, e.g., in (2.3), is a vector-valued function from $\mathbb{R}^N$ to $\mathbb{R}^{N_y}$, the FIDA observation operator $\mathcal{H}$ in (2.5) is not a function defined on a finite-dimensional space. Instead, it is a functional defined on the space of integrable functions that contains weak PDE solutions, $\mathbf{u}(\mathbf{x}, t; \mathbf{p})$. The functional formulation admits a variety of features including shocks and discontinuities. In other words, $\mathcal{H}$ involves a search process over a function of the state variables, whereas H is a function of the state variables. These search processes are difficult to linearize and, hence, pose challenges for the DA techniques, such as extended Kalman Filter (EKF) [7], 3DVAR and 4DVAR [9], that require access to the derivatives of the observation operator. Moreover, rather than being a vector of a fixed dimension, $\mathcal{H}$ is a set consisting of all points at which a feature event takes place, i.e., $\mathcal{H}$ is a set-valued functional. the number of points in $\mathcal{Y}(t)$ can be finite and change with time t , as in (2.6c), or it can be infinite, as in (2.6e). The unordered observation data in the set $\mathcal{Y}(t)$, accompanied by its variable cardinality, restrict the ensemble-based estimation of covariance matrices essential to DA methods like ensemble Kalman filter (EnKF) [1] and unscented Kalman filter (UKF) [21]. A notable exception are cases in which the observations $\mathcal{Y}$ are known to be singleton sets and, hence, can be treated as a scalar observation under the UKF framework, which allows for DA in black-box settings.

This complexity of the observation operator $\mathcal{H}$ brings both opportunities and challenges to research. The data expressed in terms of differential inequalities, as $\mathcal{Y}$ in (2.6b)–(2.6d), should be expected to have lower information content than the equality-based data $\mathbf{y}$ in (2.2). This poses the question of observability/identifiability of the system from $\mathcal{Y}$. An effective FIDA can provide estimation for systems in harsh environments where the only data available are observations of features such as shock waves. The infinite-dimensional nature of the space of measurable functions makes it imperative to ensure computational efficiency and algorithmic scalability. Learning-based approaches have emerged as promising methodologies to address the curse of dimensionality in the context of solving high-dimensional differential equations [22, 23]. For DA, a neural network surrogate for systems with continuous solutions is introduced in [24]. The use of machine learning techniques for tackling high-dimensional FIDA problems provides a compelling avenue for future research. Since a shock wave occurs in nonlinear systems only, the performance of the Kalman filter, which is proved to be optimal for linear dynamical systems, is not guaranteed and the optimality of estimation is difficult to achieve. FIDA calls for introduction of additional mathematical tools, such as set-valued analysis. Overcoming the challenges and developing FIDA algorithms are topics for long-term research that cannot be fully covered in one article. The goal of this paper is to introduce the mathematical problem formulation and to demonstrate by examples that feature-informed observations do provide valuable information for the purpose of estimating unknown state variables and parameters.

3. Numerical Experiments

We present one- and two-dimensional examples of FIDA, in which noisy observations of the shock dynamics, $\mathcal{Y}(t)$, are assimilated to reduce uncertainty in the model and initial condition parameters. The first example is the inviscid Burgers equation with a single shock; this problem admits an analytical solution. The second example deals with an initial condition for which the inviscid Burgers equation develops multiple shocks; no analytical solution is available for this case. The third example is the two-dimensional shallow-water equation, which is subject to an initial condition that results in a discontinuous solution.

In each example, the ground truth is the solution of the corresponding equations with the given parameter values, $\mathbf{p}^{\text{true}}$, from which relevant features are extracted. Observations $\mathcal{Y}$ are then generated by corrupting their ground-truth counterparts with random noise.

To simplify the presentation, the governing PDEs are solved numerically with a constant time step size Δt , and feature information is assimilated after every time step $t_k = k\Delta t$ ($k = 1, 2, \dots$). Let $\mathcal{Y}_k$ denote the feature observations at the k th time step and $\mathcal{Y}_{0:k} = \{\mathcal{Y}_i : i = 0, \dots, k\}$ denote the corresponding sequences up to time t_k .

We use the Bayesian framework to sequentially assimilate observations. At the k th step, given posterior PDF $f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}$ at $(k-1)$ th step and model dynamics, the prior beliefs $f_{\mathbf{P}_k|\mathcal{Y}_{0:k-1}}$ on the uncertain parameters are updated in light of the current observation $\mathcal{Y}_k$. The posterior PDF $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}(\mathbf{p})$ at the k th assimilation step is linked to the prior PDF $f_{\mathbf{P}_k|\mathcal{Y}_{0:k-1}}$ and, thus, to its counterpart at the previous step, $f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p})$, via the Bayesian update,

$$\begin{aligned} f_{\mathbf{P}_k|\mathcal{Y}_{0:k}} &= \frac{f_{\mathcal{Y}_k|\mathbf{P}_k, \mathcal{Y}_{0:k-1}} f_{\mathbf{P}_k|\mathcal{Y}_{0:k-1}}}{f_{\mathcal{Y}_k|\mathcal{Y}_{0:k-1}}} \\ &= \frac{f_{\mathcal{Y}_k|\mathbf{P}_k} f_{\mathbf{P}_k|\mathbf{P}_{k-1}} f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}}{f_{\mathcal{Y}_k|\mathcal{Y}_{0:k-1}}}, \end{aligned}$$

where $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}$, $f_{\mathcal{Y}_k|\mathbf{P}_k}$, and $f_{\mathbf{P}_k|\mathbf{P}_{k-1}}$ are referred to as filter density, likelihood function, and transitional PDF, respectively. This gives

$$f_{\mathbf{P}_k|\mathcal{Y}_{0:k}} \propto f_{\mathcal{Y}_k|\mathbf{P}_k} f_{\mathbf{P}_k|\mathbf{P}_{k-1}} f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}. \quad (3.1)$$

Various DA techniques estimate the posterior PDF $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}$ using (3.1) under different assumptions. For example, the Kalman filter [8] assumes linear dynamics and Gaussian distribution for all PDFs involved. The resulting posterior PDF is also Gaussian and characterized by the updated mean and covariance obtained from the Kalman filter. EKF [7] and UKF [25, 26] integrate the Kalman filter with nonlinear models using a first- and second-order approximation, respectively, of the nonlinear dynamics, while maintaining the assumption that all PDFs involved are Gaussian. In comparison, a particle filter

(PF) [27] is based on a discrete approximation of the PDFs, but makes no assumption about the model dynamics or PDFs involved.

As discussed in section 2.2, observation operators characterized by a search process and set-valued observations inhibit the direct application of most of the established DA methods. An exception, systems with a singleton set as feature observation, allows the use of UKF, as we do in the first example. The latter deals with the inviscid Burgers equation subject to an initial condition that gives rise to a single shock, whose position is used as the observation. The observation operators in the second and third examples result in non-singleton sets with varying cardinality and, hence, cannot be treated with UKF. We use a PF, with a likelihood function customized to account for set-valued observations, to assimilate the feature data in these examples. Our PF implementation follows [28, 27] and is described below for completeness.

3.1. Particle Filter

The PF relies on a weighted discrete approximation of the PDF. At the k th time step, the joint PDF $f_{\mathbf{P}_k}(\mathbf{p})$ of the uncertain (random) parameters $\mathbf{P}_k \in \mathbb{R}^{N_{\mathbf{P}}}$ is approximated by

$$f_{\mathbf{P}_k}(\mathbf{p}) \approx \sum_{j=1}^{N_{\text{par}}} w_k^j \delta(\mathbf{p} - \mathbf{p}_k^j), \quad (3.2)$$

where $\{\mathbf{p}_k^1, \dots, \mathbf{p}_k^{N_{\text{par}}}\}$ is a set of N_{par} points (aka “particles”) in the sample space of $\mathbf{P}_k$, $\delta(\cdot)$ is the $N_{\mathbf{P}}$ -dimensional Dirac delta function, and the weights $\{w_k^1, \dots, w_k^{N_{\text{par}}}\}$ satisfy $\sum_{j=1}^{N_{\text{par}}} w_k^j = 1$. PFs use importance sampling to determine $\{\mathbf{p}_k^j, w_k^j\}_{j=1}^{N_{\text{par}}}$. Specifically, the particles $\mathbf{p}_k^j$ are drawn from a so-called importance density $q_{\mathbf{P}_k}(\mathbf{p})$ that has the same support as $f_{\mathbf{P}_k}(\mathbf{p})$, and the weights w_k^j are computed as

$$w_k^j = \frac{f_{\mathbf{P}_k}(\mathbf{p}_k^j)}{q_{\mathbf{P}_k}(\mathbf{p}_k^j)}, \quad j = 1, \dots, N_{\text{par}}. \quad (3.3)$$

Assimilation of feature data $\mathcal{Y}_{0:k}$ to improve the knowledge of the parameters $\mathbf{P}_k$ is tantamount to replacing the PDF $f_{\mathbf{P}_k}(\mathbf{p})$ with the conditional PDF $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}(\mathbf{p})$, which are obtained using (3.1).

In analogy to (3.2) and (3.3), the conditional PDF $f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}$ is approximated as

$$f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p}) \approx \sum_{j=1}^{N_{\text{par}}} w_{k-1}^j \delta(\mathbf{p} - \mathbf{p}_{k-1}^j), \quad w_{k-1}^j = \frac{f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p}_{k-1}^j)}{q_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p}_{k-1}^j)}, \quad (3.4)$$

with importance density $q_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p})$. At k th step, (3.4) is used to decompose $q_{\mathbf{P}_k|\mathcal{Y}_{0:k}}(\mathbf{p})$ as $q_{\mathbf{P}_k|\mathcal{Y}_{0:k}} = q_{\mathbf{P}_k|\mathbf{P}_{k-1}, \mathcal{Y}_{0:k}} q_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}$, which gives a recursive relation for the weights using (3.1),

$$\begin{aligned} w_k^j &= \frac{f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}(\mathbf{p}_k^j)}{q_{\mathbf{P}_k|\mathcal{Y}_{0:k}}(\mathbf{p}_k^j)} \\ &\propto \frac{f_{\mathcal{Y}_k|\mathbf{P}_k}(\mathbf{p}_k^j) f_{\mathbf{P}_k|\mathbf{P}_{k-1}}(\mathbf{p}_k^j) f_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p}_{k-1}^j)}{q_{\mathbf{P}_k|\mathbf{P}_{k-1}, \mathcal{Y}_{0:k}}(\mathbf{p}_k^j) q_{\mathbf{P}_{k-1}|\mathcal{Y}_{0:k-1}}(\mathbf{p}_{k-1}^j)} \\ &\propto \frac{f_{\mathcal{Y}_k|\mathbf{P}_k}(\mathbf{p}_k^j) f_{\mathbf{P}_k|\mathbf{P}_{k-1}}(\mathbf{p}_k^j)}{q_{\mathbf{P}_k|\mathbf{P}_{k-1}, \mathcal{Y}_{0:k}}(\mathbf{p}_k^j)} w_{k-1}^j. \end{aligned} \quad (3.5)$$

It is common, e.g., [27], to deploy the transitional density $f_{\mathbf{P}_k|\mathbf{P}_{k-1}}$ as the importance density factor, $q_{\mathbf{P}_k|\mathbf{P}_{k-1}, \mathcal{Y}_{0:k}}$, such that $q_{\mathbf{P}_k|\mathbf{P}_{k-1}, \mathcal{Y}_{0:k}} = f_{\mathbf{P}_k|\mathbf{P}_{k-1}}$. This yields the weight-update rule,

$$w_k^j \propto f_{\mathcal{Y}_k|\mathbf{P}_k}(\mathbf{P}_k^j) w_{k-1}^j, \quad (3.6)$$

which we use in the ensuing examples. The likelihood density, $f_{\mathcal{Y}_k|\mathbf{P}_k}$, is central to the design of PFs and is tailored to the problem under consideration.

Algorithm 1: Particle filter implementation of recursive parameter estimation

Data: $N_{\text{par}}, \Omega, N_{\text{steps}}$

Result: $\{\mathbf{p}_{k=N_{\text{steps}}}^{j*}, w_{k=N_{\text{steps}}}^{j*}\}_{j=1}^{N_{\text{par}}}$

/ generating particles and weights $\{\mathbf{p}_0^{j*}, w_0^{j*}\}_{j=1}^{N_{\text{par}}}$ to approximate $f_{\mathbf{P}_0|\mathcal{Y}_0}$ */*

for $j = 1$ **to** N_{par} **do**

$\mathbf{p}_0^{j*} \sim U_{N_{\text{P}}}(\Omega);$

$w_0^{j*} = 1/N_{\text{par}};$

end

/ recursive parameter estimation with PF */*

for $k = 1$ **to** N_{steps} **do**

$\{\mathbf{p}_k^{j*}, w_k^{j*}\}_{j=1}^{N_{\text{par}}} = \text{PF}(\{\mathbf{p}_{k-1}^{j*}, w_{k-1}^{j*}\}_{j=1}^{N_{\text{par}}}, \mathcal{Y}_k)$

end

At initiation of the recursive filtering for parameter estimation, i.e., at step $k = 0$, the (empty) set $\mathcal{Y}_0$ contains no measurements and the filter density $f_{\mathbf{P}_0|\mathcal{Y}_0}$ is assumed to be multivariate uniform over a hyper-rectangle $\Omega \subset \mathbb{R}^{N_{\text{P}}}$. Since N_{par} particles $\mathbf{p}_0^{j*}$ ($j = 1, \dots, N_{\text{par}}$) drawn from this PDF are i.i.d, each particle is assigned a uniform weight $w_0^j = 1/N_{\text{par}}$ ($j = 1, \dots, N_{\text{par}}$). The PDF $f_{\mathbf{P}_0|\mathcal{Y}_0}$, approximated via (3.2) with $\{\mathbf{p}_0^{j*}, 1/N_{\text{par}}\}_{j=1}^{N_{\text{par}}}$, serves as the starting filter density for PFs.

Algorithm 2: FIDA via a particle filter with resampling

```

Data:  $\{\mathbf{p}_{k-1}^{j*}, w_{k-1}^{j*}\}_{j=1}^{N_{\text{par}}}, \mathcal{Y}_k$ 
Result:  $\{\mathbf{p}_k^{j*}, w_k^{j*}\}_{j=1}^{N_{\text{par}}}$ 

/* PF: propagation and weight update */
for  $j = 1$  to  $N_{\text{par}}$  do
     $\mathbf{p}_k^j = \mathbf{p}_{k-1}^{j*} + \delta \mathbf{p}_k$ ;
     $w_k^j = f_{\mathcal{Y}_k|\mathbf{P}_k}(\mathbf{p}_k^j) w_{k-1}^{j*}$ ;
end
/* Normalizing the weights */
 $t = \sum_{j=1}^{N_{\text{par}}} w_k^j$ ;
for  $j = 1$  to  $N_{\text{par}}$  do
     $w_k^j = w_k^j / t$ 
end
/* Resampling */
/* Initialize cumulative weight sum:  $c_j, j = 1, \dots, N_{\text{par}}$  */
 $c_1 = w_k^1$ ;
for  $j = 2$  to  $N_{\text{par}}$  do
     $c_j = c_{j-1} + w_k^j$ ;
end
 $i = 1$ ;
 $y_1 \sim U[0, 1/N_{\text{par}}]$ ;
for  $j = 1$  to  $N_{\text{par}}$  do
     $y_j = y_1 + (j-1)/N_{\text{par}}$ ;
    while  $y_j > c_i$  do
         $i = i + 1$ ;
    end
     $\mathbf{p}_k^{j*} = \mathbf{p}_k^i$ ;
     $w_k^{j*} = 1/N_{\text{par}}$ ;
end

```

140 To avoid the degeneracy phenomenon, wherein most of the particles are assigned a negligible weight
 after a few steps, the particles are resampled after each assimilation step: the particles with lower
 weights are eliminated with high probability, and those with higher weights are duplicated. Resampled
 particles are identified by the superscript *. At k th step, the process involves drawing N_{par} samples
 145 $\mathbf{p}_k^{j*}$ ($j = 1, \dots, N_{\text{par}}$) from the filter density $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}$, approximated via (3.2) with propagated particles
 and their associated weights $\{\mathbf{p}_k^j, w_k^j\}_{j=1}^{N_{\text{par}}}$ at first. However, since the particles $\mathbf{p}_k^{j*}$ are i.i.d. samples
 from this PDF, they have equal weights $w_k^{j*} = 1/N_{\text{par}}$ and, hence, $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}$ is approximated by (3.2)
 with $\{\mathbf{p}_k^{j*}, 1/N_{\text{par}}\}_{j=1}^{N_{\text{par}}}$, such that $f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}(\mathbf{p}_k^{j*} = \mathbf{p}_k^l) \approx w_k^l$ ($l = 1, \dots, N_{\text{par}}$). The later approximation of

$f_{\mathbf{P}_k|\mathcal{Y}_{0:k}}$ is used to propagate the particles forward.

For a deterministic model, e.g., for (2.2) without noise ($\mathbf{w} \equiv 0$), the state vector $\mathbf{u}(t)$ at any time t is completely determined once values of the parameter vector $\mathbf{p}$ are specified. While these uncertain parameters do not change with time, the PF introduces artificial dynamics to avoid the loss of diversity among the particles [28]. At k th step, evolution of the parameter vector $\mathbf{p}_k$ is described by

$$\mathbf{p}_k = \mathbf{p}_{k-1} + \delta\mathbf{p}_k, \quad (3.7)$$

where $\delta\mathbf{p}_k \in \mathbb{R}^{N_{\mathbf{p}}}$ is zero-mean Gaussian white noise with covariance matrix $\mathbf{\Sigma}_k$. Consequently, the
150 transitional PDF $f_{\mathbf{P}_k|\mathbf{P}_{k-1}}$ is $N_{\mathbf{p}}$ -dimensional Gaussian. Since elements of the parameter vector $\mathbf{p}$ are independent, the covariance matrix $\mathbf{\Sigma}_k$ is diagonal. The entries on the diagonal of $\mathbf{\Sigma}_k$ decay with k to aid the convergence.

Algorithm 1 describes the overall FIDA setup, including the spawning of initial particles at $k = 0$ and recursive parameter estimation at $k = N_{\text{steps}}$. Output $\{\mathbf{p}_{N_{\text{steps}}}^{j*}, w_{N_{\text{steps}}}^{j*}\}_{j=1}^{N_{\text{par}}}$ approximates the posterior
155 density $f_{\mathbf{P}_k|\mathcal{Y}_{0:N_{\text{steps}}}}$ after all the observations, $\mathcal{Y}_{0:N_{\text{steps}}}$, are assimilated. Assuming the parameters are identifiable from feature data, the posterior PDF $f_{\mathbf{P}_k|\mathcal{Y}_{0:N_{\text{steps}}}}$ should center around the true parameter $\mathbf{p}^{\text{true}}$. We use the mean and the 5th and 95th percentile region around the mean to judge the accuracy and spread, respectively, of the PDF $f_{\mathbf{P}_k|\mathcal{Y}_{0:N_{\text{steps}}}}$. Algorithm 2 details the implementation of PF along with the resampling procedure.

160 3.2. Example 1: Inviscid Burgers equation with a single shock

Consider a scalar state variable $u(x, t)$ whose dynamics is governed by the hyperbolic conservation law

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0, \quad f(u) = \frac{u^2}{\lambda}. \quad (3.8)$$

The value of the parameter λ in the flux function $f(u)$ is uncertain; the true state is defined by $\lambda = 2$, i.e., by the inviscid Burgers equation. Equation (3.8) is defined on the infinite domain, $|x| < \infty$, and is subject to the initial condition $u(x, 0) = u_{\text{in}}(x)$. The initial state $u_{\text{in}}(x)$ is an inverted ramp function

$$u_{\text{in}}(x) = \begin{cases} u_l & x \leq 0 \\ u_l - \alpha x & 0 < x \leq x_r \\ u_r & x > x_r, \end{cases} \quad (3.9)$$

with the upper and lower limits $u_l = 2$ and $u_r = 1$, and with $\alpha = (u_l - u_r)/x_r$; the value of the parameter x_r is uncertain, the true value is $x_r = 1$ (Figure 1). The two uncertain parameters are concatenated into the parameter vector $\mathbf{P} = [\lambda, x_r]^{\top}$, and for any parameter realization $\mathbf{p}$ the solution's dependence on parameters is emphasized by using the notation $u(x, t; \mathbf{p})$.

This forward model admits an analytical solution (see Appendix A for details). It is divided into two parts, before and after the breaking time $t^* = \lambda/(2\alpha)$, which is defined as the time when the characteristics first intersect and the shock develops. The solution before the break time, $t < t^*$, is given by

$$u(x, t; \mathbf{p}) = \begin{cases} u_l & x \leq 2u_l t/\lambda \\ \frac{u_l - \alpha x}{1 - 2\alpha t/\lambda} & 2u_l t/\lambda \leq x \leq x_r + 2u_r t/\lambda, \\ u_r & x_r + 2u_r t/\lambda \leq x \end{cases} \quad t \leq t^*. \quad (3.10a)$$

For $t \geq t^*$, it is given by

$$u(x, t; \mathbf{p}) = \begin{cases} u_l & x < x_{\text{sh}} \\ (u_l + u_r)/2 & x = x_{\text{sh}}, \\ u_r & x > x_{\text{sh}} \end{cases} \quad t > t^*, \quad (3.10b)$$

with

$$x_{\text{sh}}(t) = \frac{u_l + u_r}{\lambda} t + \frac{x_r}{2}, \quad t \geq t^*. \quad (3.10c)$$

165 This solution is observed over domain $x \in [-1, 9]$, which is discretized with $\Delta x = 0.01$ into $N = 1000$ nodes $x_m = -1 + m\Delta x$ ($m = 1, \dots, N$), resulting in the solution vector $\mathbf{U}(t; \mathbf{p}) = [u(x_1, t; \mathbf{p}), \dots, u(x_N, t; \mathbf{p})]^\top$ of dimension N . The system is advanced in time with time step $\Delta t = 0.05$, i.e., $\mathbf{U}(t; \mathbf{p})$ is evaluated at times $t_k = k\Delta t$. The base-state solution, $\mathbf{U}(t; \mathbf{p}^{\text{true}})$, is shown in Figure 1.

A feature in this example is the shock position, $x_{\text{sh}}(t)$, characterized by the observation operator

$$\mathcal{H} : \mathbf{U}(t; \mathbf{p}) \rightarrow \left\{ x^* : x^* = \frac{u_l + u_r}{\lambda} t + \frac{x_r}{2} \right\}. \quad (3.11)$$

Noisy feature observations,

$$\mathcal{Y}_k = \left\{ Y_k \equiv x_k^* + v(t_k) : x_k^* = \frac{u_l + u_r}{\lambda^{\text{true}}} t_k + \frac{x_r^{\text{true}}}{2}, \quad t_k = k\Delta t \right\}, \quad (3.12)$$

are generated by corrupting the ground-truth prediction of $x_{\text{sh}}(t_k)$ in (3.10c) (the solution corresponding to the parameter values $\mathbf{p}^{\text{true}} = [\lambda^{\text{true}} = 2, x_r^{\text{true}} = 1]^\top$), with Gaussian zero-mean white noise $v(t)$ whose standard deviation is $\sigma_{\text{obs}} = 0.1$. Note that x^* represents the shock location, x_{sh} , only after the shock initiation, $t_k \geq t^*$; at earlier times ($t_k < t^*$), it represents the center of the transition zone between u_l and u_r .

175 We use Algorithms 1 and 2 to estimate the parameter values $\mathbf{p}^{\text{true}}$ from the set of noisy observations of the shock position, $\mathcal{Y}_k$. The PDF of the initial parameter distribution, $f_{\mathbf{p}_0}$, is approximated via (3.2) with N_{par} particles $\mathbf{p}_0^j = [\lambda_0^j, x_{r0}^j]^\top$ ($j = 1, \dots, N_{\text{par}}$). These are generated by independently drawing N_{par} samples, λ^j and x_r^j for $j = 1, \dots, N_{\text{par}}$, from uniform PDFs, $\lambda^j \sim U[1.9, 2.4]$ and $x_r^j \sim U[0.6, 1.1]$.

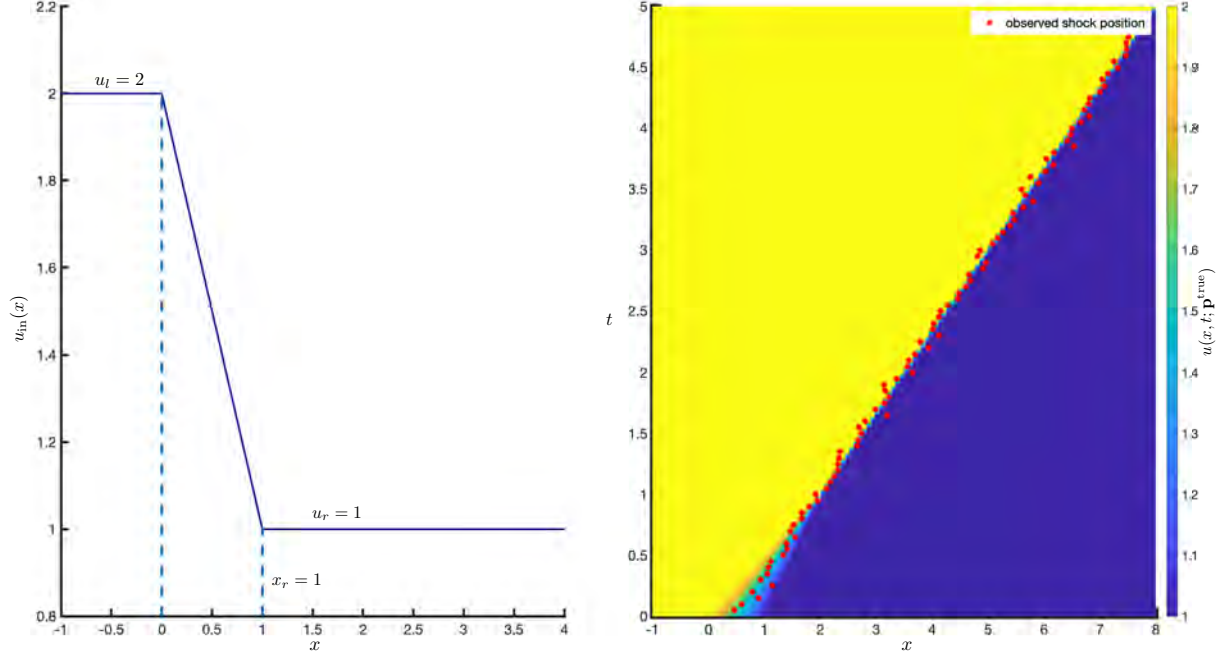


Figure 1: Initial condition $u(x, 0) = u_{\text{in}}$ (Left) and the corresponding solution $u(x, t)$ (Right) for the true state ($\lambda = 2$ and $x_r = 1$) in Example 1. Noisy observations of the shock location, $x_{\text{sh}}(t_k)$, at discrete times t_k are shown by the red dots in the right image.

The choice of a uniform importance density $q\mathbf{P}_0$ assigns equal weights $w_0^j = 1/N_{\text{par}}$ to each particle $\mathbf{p}_0^j$. Since $\mathcal{Y}_0 = \emptyset$ (no observations), $f_{\mathbf{P}_0|\mathcal{Y}_0} = f_{\mathbf{P}_0}$.

For the subsequent DA steps ($k \geq 1$), we choose the Gaussian likelihood function,

$$f_{\mathcal{Y}_k|\mathbf{P}_k} = \frac{1}{\sqrt{2\pi}\sigma_{\text{obs}}} \exp\left[-\frac{(y_k - \mu_k)^2}{2\sigma_{\text{obs}}^2}\right], \quad (3.13)$$

with the mean μ_k that coincides with the observed position of the shock, $\mu_k = Y_k$, and y_k denotes the shock position at time t_k evaluated using (3.11) with parameters $\mathbf{P}_k$. For the j th particle, the model (3.10) yields predictions of the system state, $\mathbf{U}(t_k; \mathbf{p}_k^j)$, and the shock location, $x_{\text{sh}}(t_k; \mathbf{p}_k^j)$. Hence,

$$f_{\mathcal{Y}_k|\mathbf{P}_k=\mathbf{p}_k^j} = \frac{1}{\sqrt{2\pi}\sigma_{\text{obs}}} \exp\left[-\frac{(x_{\text{sh}}(t_k; \mathbf{p}_k^j) - Y_k)^2}{2\sigma_{\text{obs}}^2}\right], \quad j = 1, \dots, N_{\text{par}}. \quad (3.14)$$

180 The 2×2 diagonal covariance matrix Σ_k , used for advancing parameters forward via (3.7), is defined by $\Sigma_k(1, 1) = \Sigma_k(2, 2) = (2.5 \times 10^{-3})/k$, and $\Sigma_k(1, 2) = \Sigma_k(2, 1) = 0$.

FIDA is performed using PFs with $N_{\text{par}} = 200, 500$, and 1000 particles. (The PFs with larger N_{par} reuse the particles from the PFs with lower N_{par} to facilitate the convergence study.) The PF with $N_{\text{par}} = 200$ particles is terminated when the sum of the standard deviations of the parameter estimators, $\sigma_\lambda + \sigma_{x_r}$, falls below the prescribed tolerance, $\varepsilon_{\text{off}} = 0.1$. Figure 2 demonstrates that this occurs at time 185 $t_{\text{off}} = 4.75$, which is used as the final time for the PFs with $N_{\text{par}} = 500$ and 1000 particles.

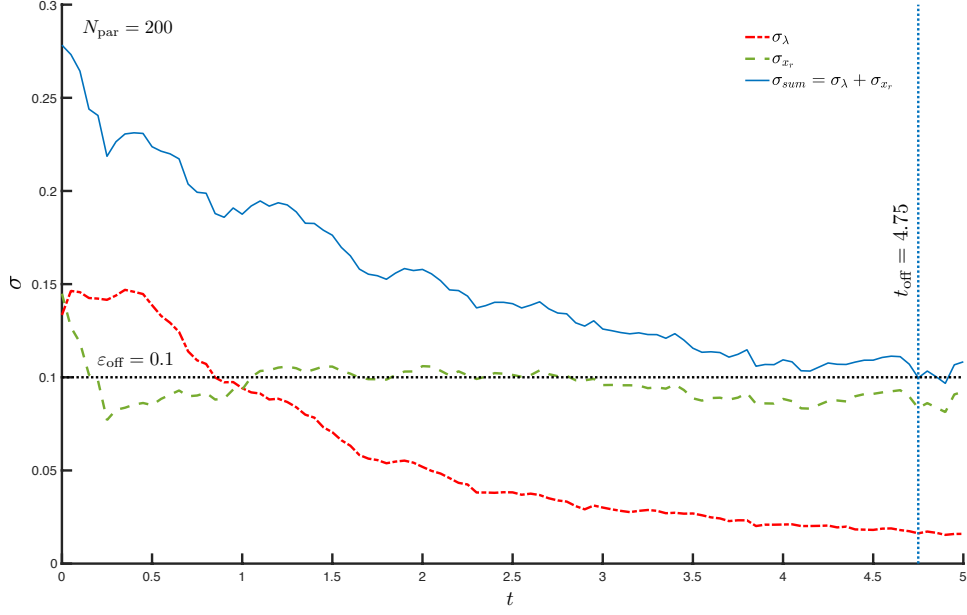


Figure 2: Cut-off time for Example 1.

Since the observation operator $\mathcal{H}$ in (3.11) results in a singleton set at any given time, it is treatable with UKF [26, 21]. For comparison with the PF estimators, the UKF-based parameter estimation is curtailed at $t_{\text{off}} = 4.75$. Figure 3 shows the convergence of the PF and UKF estimators of λ and x_r (decay of the σ_λ and σ_{x_r} estimates) with the amount of assimilated data, i.e., with the number of the DA time steps. The estimates of λ and x_r converge to their true value, $\lambda^{\text{true}} = 2$ and $x_r = 1$. However, the convergence rates for the λ and x_r are different: while the spread in the estimates of λ drops steadily as more data are assimilated, the estimates of x_r improve only in the first few DA steps and show no improvement in the subsequent steps. The analytical expression for the feature observations, Eq. (3.12), reveals that the feature varies linearly with x_r and is inversely proportional to λ . As a result, they are more sensitive to the value λ than to that of x_r . Moreover, the effect of deteriorating observation quality is reflected in the volatile estimators with large variance (enhanced noise levels). Conversely, lower noise levels result in correct and stable estimators.

Overall, these figures demonstrate that the estimates of λ and x_r converge to their true values, $\lambda = 2$ and $x_r = 1$ for sufficiently large value of N_{par} . Hence, the information content of the feature (shock location) is sufficient to identify the unknown model parameters.

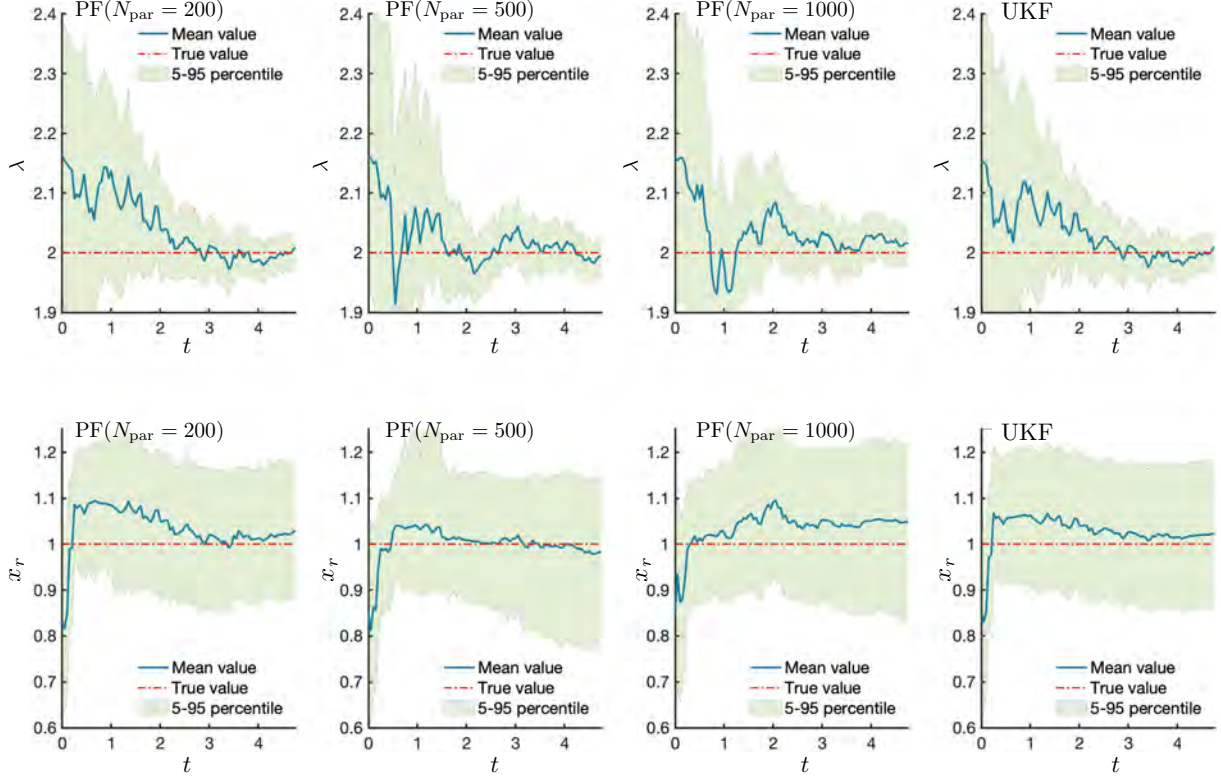


Figure 3: Convergence of the PF estimators of the uncertain parameters λ (top row) and x_r (bottom row) with the amount of assimilated data or the number of DA time steps in Example 1. The results are shown for PFs with $N_{\text{par}} = 500$ (first column), 1000 (second column), 2000 (third column) particles, and with UKF (fourth column).

3.3. Example 2: Inviscid Burgers equation with multiple shocks

To investigate FIDA's ability to handle multiple shocks, we consider the Burgers equation (3.8) subject to the initial condition

$$u(x, 0) = \begin{cases} \sin(\alpha x), & -4 \leq x \leq -1 \\ \beta & -1 < x < 1 \\ \sin(\gamma x), & 1 \leq x \leq 4 \\ 0, & |x| > 4. \end{cases} \quad (3.15)$$

In addition to the uncertain model parameter λ (3.8), the parameters α , β and γ are uncertain as well; they are arranged in the parameter vector $\mathbf{P} = [\lambda, \alpha, \beta, \gamma]^\top$. The true parameter values used in data generation are $\mathbf{p}^{\text{true}} = [2, \pi, 2, \pi]^\top$. An analytical solution for this initial-value problem is not available; instead, we solve it numerically using the causality-free algorithm [29] with the help of the `chebfun` library [30] in MATLAB.

The space is discretized with $\Delta x = 0.01$ and the state variable $u(x, t; \mathbf{p})$, where $\mathbf{p}$ is a realization of the random variable $\mathbf{P}$, is monitored over the domain $x \in [-4, 4]$ at $N = 800$ nodes $x_m = -4 + m\Delta x$

($m = 1, \dots, N$), yielding the state vector $\mathbf{U}(t; \mathbf{p}) = [u(x_1, t; \mathbf{p}), \dots, u(x_N, t; \mathbf{p})]^\top$. The system is advanced with time step $\Delta t = 0.1$, yielding the solution at times $t_k = k\Delta t$ ($k \geq 0$). Figure 4 exhibits the initial condition under true parameters $\mathbf{p}^{\text{true}}$ and the corresponding solution $\mathbf{U}(t; \mathbf{p}^{\text{true}})$.

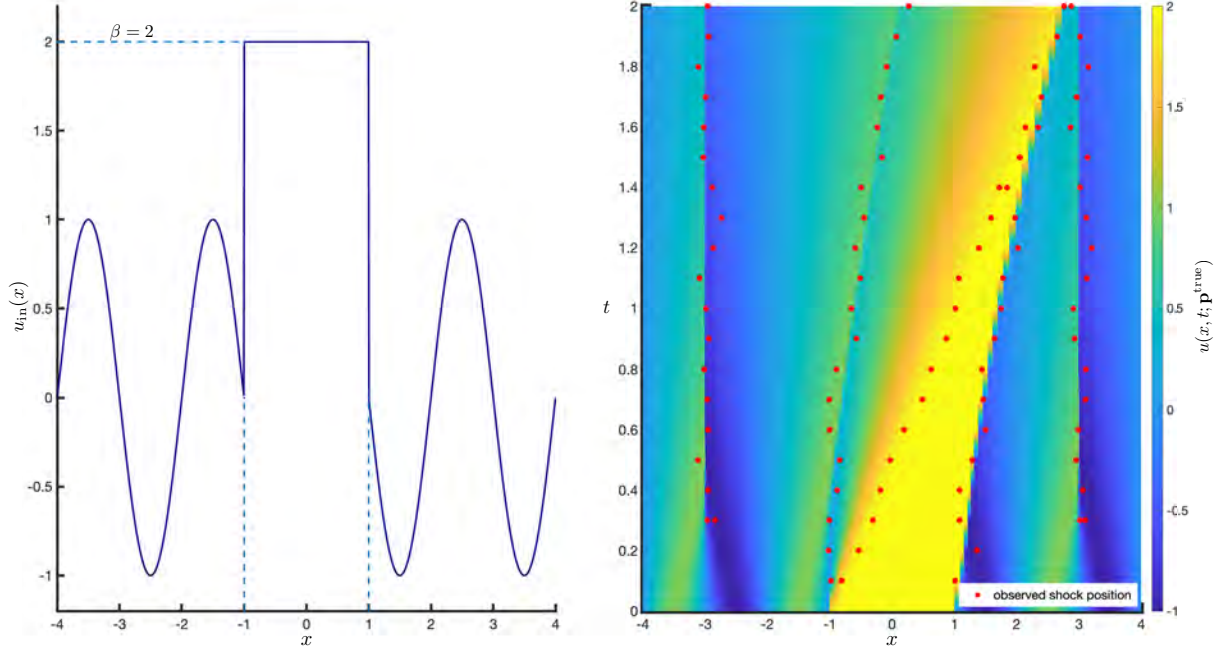


Figure 4: Initial condition $u(x, 0) = u_{\text{in}}$ (Left) and the corresponding solution $u(x, t)$ (Right) for the true state ($\lambda = 2$, $\alpha = \gamma = \pi$, and $\beta = 2$) in Example 2. The red dots in the right image show noisy observations of the shock location, $x_{\text{sh}}(t)$, at discrete times t .

Also shown in Figure 4 are noisy feature observations, i.e., “observed” shock positions. In lieu of an analytical expression for the shock locations, we compute them by evaluating the first and second derivative of the numerical solution $u(x, t; \mathbf{p})$ using `chebfun` [30]. The latter records all the points in the computational domain at which the first and/or second derivatives blow up (exceed some arbitrarily large number M); these points form the exact set of shock positions described by the observation operator

$$\mathcal{H} : \mathbf{U}(t; \mathbf{p}) \rightarrow \left\{ x^* : \frac{\partial u}{\partial x}(x^*, t; \mathbf{p}) \geq M \text{ or } \frac{\partial^2 u}{\partial x^2}(x^*, t; \mathbf{p}) \geq M \right\}. \quad (3.16)$$

The observation set,

$$\mathcal{Y}_k = \left\{ x^* + v(t_k) : \frac{\partial u}{\partial x}(x^*, t_k; \mathbf{p}^{\text{true}}) \geq M \text{ or } \frac{\partial^2 u}{\partial x^2}(x^*, t_k; \mathbf{p}^{\text{true}}) \geq M \right\}, \quad (3.17)$$

is obtained by corrupting the elements of the feature set corresponding to the true parameters $\mathbf{p}^{\text{true}}$ with zero-mean Gaussian white noise $v(t)$ with standard deviation $\sigma_{\text{obs}} = 0.1$. An arbitrary large number M in (3.16) is a parameter for the observation model. The `chebfun` library, used in this example, approximates one-dimensional functions to machine precision, and the built-in function for exact calculation

of derivatives identifies the points of discontinuities; hence, an explicit value of the threshold M is not needed. The resulting set of feature points is a finite non-singleton set, and the number of elements in the set varies with time as new shocks develop and old ones coalesce. For example, Figure 4 shows that the number of observed shocks at time t_k , N_k^{sh} , grows from $N_1^{\text{sh}} = 3$ (at $t = 0.1$) to $N_3^{\text{sh}} = 7$ (at $t = 0.3$) and then falls to $N_{10}^{\text{sh}} = 5$ (at $t = 1$). The cardinality of the observation set also varies with the uncertain parameters at any given time. This prohibits ensemble-based estimation of covariance matrices, and UKF cannot be used.

Feature observations in the set $\mathcal{Y}_k$ are assimilated using Algorithms 1 and 2 to improve the estimate of uncertain parameters $\mathbf{p}$. The PDF of uncertain parameters at $t = 0$, conditioned on $\mathcal{Y}_0 = \emptyset$, $f_{\mathbf{p}_0|\mathcal{Y}_0} = f_{\mathbf{p}_0}$, is approximated using (3.2) with $\{\mathbf{p}_0^{j*}, w_0^{j*}\}_{j=1}^{N_{\text{par}}}$. The particles $\mathbf{p}_0^{j*}$ ($j = 1, \dots, N_{\text{par}}$) are drawn from multivariate uniform distribution $U_{N^P}(\Omega)$ with $N^P = 4$ and $\Omega = [1.6, 2.1] \times [3, 3.5] \times [1.9, 2.4] \times [3, 3.5]$; these samples are assigned equal weights, $w_0^{j*} = 1/N_{\text{par}}$ ($j = 1, \dots, N_{\text{par}}$), in accordance with Algorithm 1. These particles are advanced in time ($k \geq 1$) using (3.7) with $\delta\mathbf{p}_k$, whose diagonal covariance matrix Σ_k has entries $\Sigma_k(i, i) = (2.5 \cdot 10^{-3})/k$ ($i = 1, \dots, 4$). The likelihood function $f_{\mathcal{Y}_k|\mathbf{p}_k=\mathbf{p}_k^j}$ is defined as the product

$$f_{\mathcal{Y}_k|\mathbf{p}_k=\mathbf{p}_k^j} = \prod_{l=1}^{n(\mathcal{L}_k^j)} \left[\max_{x^* \in \mathcal{Y}_k} \frac{1}{\sqrt{2\pi} \sigma_{\text{obs}}} \exp \left(-\frac{(\mathcal{L}_k^j(l) - x^*)^2}{2\sigma_{\text{obs}}^2} \right) \right], \quad (3.18)$$

where $n(\mathcal{L}_k^j)$ and $\mathcal{L}_k^j(l)$ denote, respectively, the cardinality and l th element (in any fixed order) of the set $\mathcal{L}_k^j = \mathcal{H} : \mathbf{U}(t_k; \mathbf{p}_k^j)$.

PFs with $N_{\text{par}} = 200, 500$, and 1000 particles are employed to carry out FIDA; the particles approximating $f_{\mathbf{p}_0|\mathcal{Y}_0}$ for PFs with smaller N_{par} are recycled in PFs with larger N_{par} to enable convergence study. The PF with the smallest number of particles, $N_{\text{par}} = 200$, assimilates feature observation until the sum of the standard deviations of the estimators falls below the stipulated cutoff, $\varepsilon_{\text{off}} = 0.2$, as shown in Figure 5; the corresponding time $t_{\text{off}} = 1.4$ is used as the stopping time for the PFs with $N_{\text{par}} = 500$ and 1000 . The standard deviations of the estimators based on the PF with $N_{\text{par}} = 200$ decay initially and then plateau around time $t = t_{\text{off}}$ (Figure 5).

Figure 6 demonstrates the performance of FIDA-based estimators of the uncertain parameters $\mathbf{p}$ as more feature observations are assimilated via PFs with different numbers of particles. The parameter estimators initially move towards the corresponding true values and variance decays as the observations are assimilated. However, after a few steps (generally around $k = 5$), there are only marginal variations in the estimated value and its variance, and sometimes estimators even deviate from the true value (e.g., λ for the PFs with $N_{\text{par}} = 200, 1000$). This suggests that after a few assimilation steps, additional observations neither refine the parameter estimates nor result in variance decay.

Improvement in the PF's estimator accuracy with N_{par} is negligible and signifies that the number of

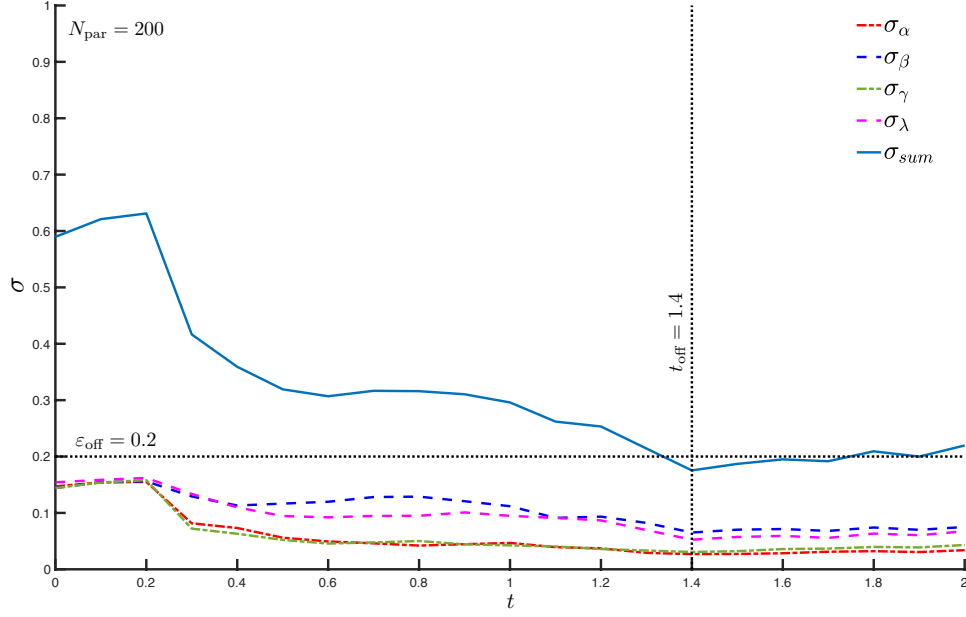


Figure 5: Cut-off time for Example 2.

particles is not a bottleneck for the current example. Different decay rates of parameters along with their converged standard deviation values, also evident from Figure 5, indicate the differences in the degree of observability of those parameters.

As in the previous example, feature observations provide useful information that improves the estimate of uncertain parameters in the model. Not all parameters benefit equally from those observations and, after a few assimilation steps, additional observations only marginally improve the estimators.

3.4. Example 3: Shallow-water equations with discontinuous initial condition

Two-dimensional shallow-water equations,

$$\frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = 0 \quad (3.19a)$$

$$\frac{\partial hu}{\partial t} + \frac{\partial}{\partial x} \left(hu^2 + \frac{1}{2}gh^2 \right) + \frac{\partial huv}{\partial y} = 0 \quad (3.19b)$$

$$\frac{\partial hv}{\partial t} + \frac{\partial huv}{\partial x} + \frac{\partial}{\partial y} \left(hv^2 + \frac{1}{2}gh^2 \right) = 0, \quad (3.19c)$$

describe the spatiotemporal evolution of the three state variables, $\mathbf{u}(\mathbf{x}, t) = \{h, u, v\}$: height of the fluid column, $h(\mathbf{x}, t)$, and two components of the vertically-averaged flow velocity, $u(\mathbf{x}, t)$ and $v(\mathbf{x}, t)$. The model parameter g is uncertain, with its true value set to $g^{\text{true}} = 10$. Equations (3.19) are defined for $t \in \mathbb{R}_+$ and $\mathbf{x} = [x, y]^\top \in [-2.5, 2.5] \times [-2.5, 2.5]$ and are subject to non-reflective boundary conditions,

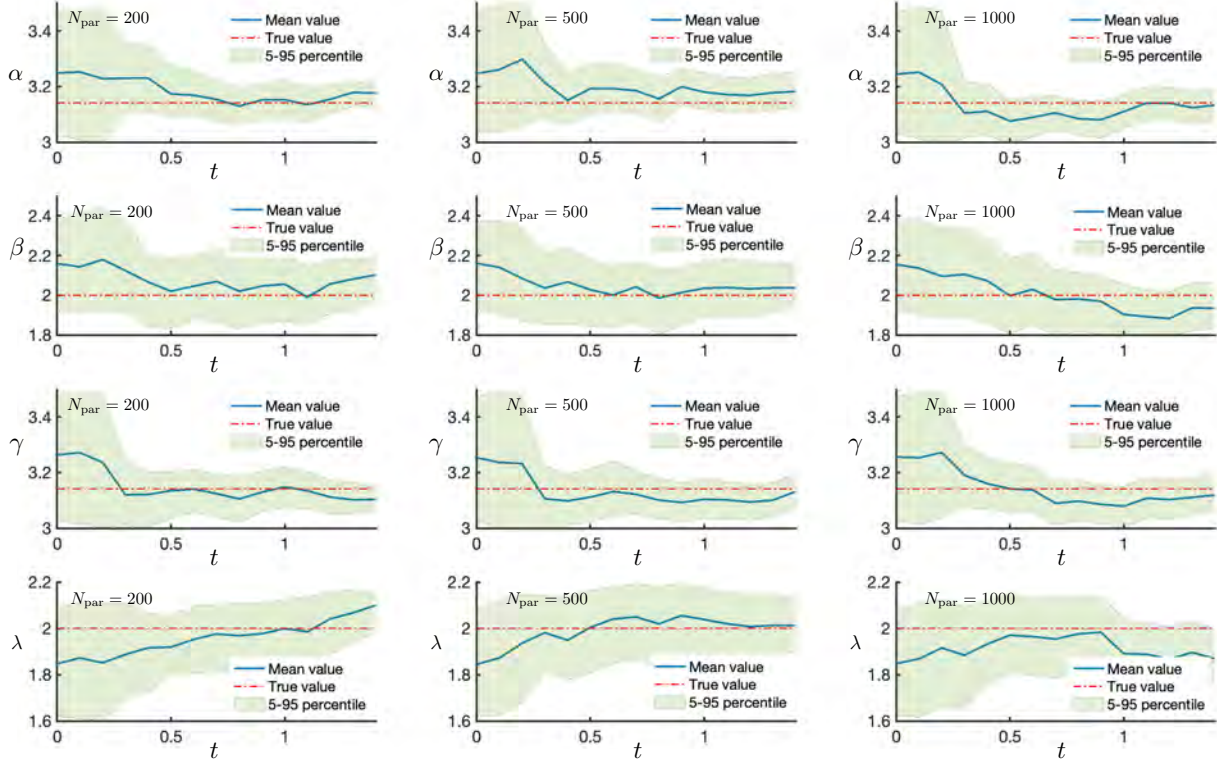


Figure 6: Convergence of the estimators of the uncertain parameters λ , α , β , and γ in Example 2.

which allow outflow from the simulation domain. The initial state of the system is

$$h(\mathbf{x}, 0) = \begin{cases} h_{\text{in}} & \sqrt{x^2 + y^2} \leq r_0 \\ h_{\text{out}} & \sqrt{x^2 + y^2} > r_0, \end{cases} \quad (3.20a)$$

$$u(\mathbf{x}, 0) = 0, \quad v(\mathbf{x}, 0) = 0, \quad (3.20b)$$

where $h_{\text{out}} = 1$ and $r_0 = 0.5$. The remaining parameter in the initial condition, h_{in} , is uncertain and its true value is $h_{\text{in}}^{\text{true}} = 25$. With this specification, the uncertain parameter vector is $\mathbf{P} = [h_{\text{in}}, g]^\top$ and the corresponding true parameter vector is $\mathbf{p}^{\text{true}} = [25, 10]^\top$. The dependence of the state variables on any parameters realization $\mathbf{p}$ is explicated by using the notation $\mathbf{u}(\mathbf{x}, t; \mathbf{p})$, $h(\mathbf{x}, t; \mathbf{p})$, $u(\mathbf{x}, t; \mathbf{p})$, and $v(\mathbf{x}, t; \mathbf{p})$. The simulation domain $[-2.5, 2.5] \times [-2.5, 2.5]$ is discretized with a square 250×250 mesh, yielding $M = 250 \cdot 250$ grid points $\mathbf{x}_m$ ($m = 1, \dots, M$). The discretized state-variables vector

$$\mathbf{U}(t; \mathbf{p}) = [h(\mathbf{x}_1, t; \mathbf{p}), \dots, h(\mathbf{x}_M, t; \mathbf{p}), u(\mathbf{x}_1, t; \mathbf{p}), \dots, u(\mathbf{x}_M, t; \mathbf{p}), v(\mathbf{x}_1, t; \mathbf{p}), \dots, v(\mathbf{x}_M, t; \mathbf{p})]^\top, \quad (3.21)$$

of dimension $N = 3M$, is advanced forward in time, with time step $\Delta t = 0.05$, using PyClaw library [31].

Features in this example are an expansion wave (discontinuities in the fluid depth h), identified using

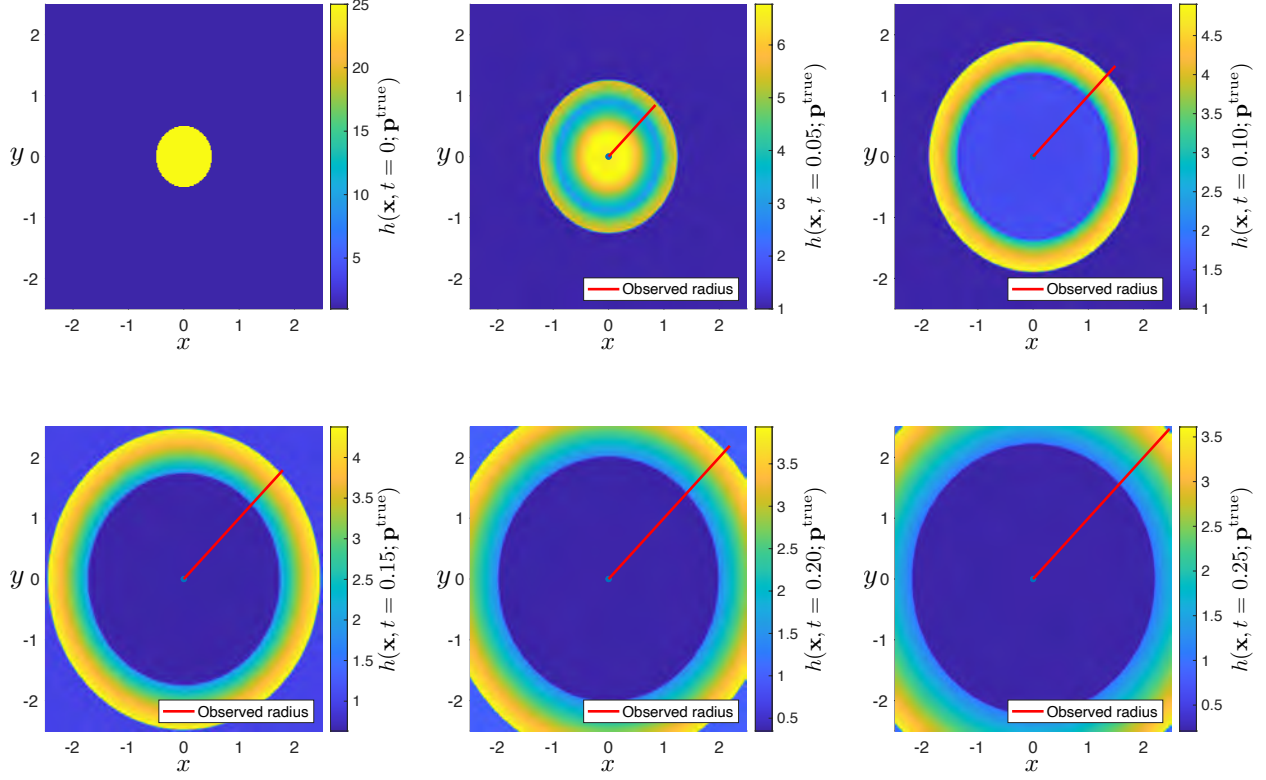


Figure 7: Temporal snapshots of water height, $h(\mathbf{x}, t)$, and observed shock radius, $r_{\text{sh}}(t)$, for the shallow-water equations in Example 3.

the gradient $\nabla h = [\partial_x h, \partial_y h]^\top$ via the observation operator

$$\mathcal{H} : \mathbf{U}(t, \mathbf{p}) \rightarrow \{\mathbf{x}^* = (x^*, y^*) : \|\nabla h(\mathbf{x}^*, t; \mathbf{p})\|_2 \geq \zeta\}, \quad (3.22)$$

for some finite ζ . The value of ζ depends on the grid size and the range of magnitudes of $h(\mathbf{x}, t)$ over the field of interest. It can be treated as a calibration parameter for the observation model and governs the accuracy of the feature data estimated through computations, in lieu of experimentally observed feature information. In the results reported below, we set $\zeta = 0.01$. For values far away from $\zeta = 0.01$, the observation operator either misses the points on the discontinuity contour ($\zeta > 0.01$) or captures spurious points ($\zeta < 0.01$). The observation set in (3.22) comprises all points lying on the (discretized) contour of discontinuity. Given the problem's radial symmetry, the discontinuity in $h(\mathbf{x}, t)$ occurs along circular contours expanding outwards with time (Figure 7). A noisy observation of the expansion wave, mimicking those observed from images, is described (at time t_k) by

$$\mathcal{Y}_k = \{\mathbf{x}^* + \mathbf{v}(t_k) : \|\nabla h(\mathbf{x}^*, t; \mathbf{p}^{\text{true}})\|_2 \geq \zeta\}, \quad (3.23)$$

where $\mathbf{v}(t_k) \in \mathbb{R}^2$ is zero-mean Gaussian white noise vector, whose diagonal covariance matrix has entries

250 $\sigma_{\text{obs}} = 0.1$ on the diagonal.

Feature observations $\mathcal{Y}_k$ are assimilated using Algorithms 1 and 2 to estimate $\mathbf{p}^{\text{true}}$. The covariance matrix Σ_k of the random perturbation $\delta\mathbf{p}_k$ in (3.7) is a 2×2 diagonal matrix, whose non-zero entries are proportional to the square of the range of the corresponding parameter in Ω : as $\Sigma_k(1,1) = (27 - 18)^2/(100k) = 0.81/k$ and $\Sigma_k(2,2) = (11 - 6)^2/(100k) = 0.25/k$. The initial parameter distribution $f_{\mathbf{P}_0}$, equivalent to filter density $f_{\mathbf{P}_0|\mathcal{Y}_0}$, is approximated via (3.2) using $\{\mathbf{p}_0^{j*}, w_0^{j*} = 1/N_{\text{par}}\}_{j=1}^{N_{\text{par}}}$. Particles $\mathbf{p}_0^{j*}$ ($j = 1, \dots, N_{\text{par}}$) are sampled from a bivariate uniform distribution $U_2(\Omega)$ with $\Omega = [18, 27] \times [6, 11]$. For ensuing steps ($k \geq 1$) with noisy feature observation $\mathcal{Y}_k$, we choose the Gaussian likelihood function,

$$f_{\mathcal{Y}_k|\mathbf{P}_k=\mathbf{p}_k^j} = \frac{1}{\sqrt{2\pi}\sigma_{\text{obs}}} \exp\left(-\frac{(r(\mathcal{L}_k^j) - r(\mathcal{Y}_k))^2}{2\sigma_{\text{obs}}^2}\right), \quad (3.24)$$

for j th particle $\mathbf{p}_k^j$ and corresponding observation set $\mathcal{L}_k^j = \mathcal{H} : \mathbf{U}(t_k; \mathbf{p}_k^j)$ in (3.22). Here, for any set S of points $\mathbf{x}^*$, $r(S) \equiv \max\{\|\mathbf{x}^*\|_2 : \mathbf{x}^* \in S\}$. Thus defined, $r(S)$ represents the distance of the farthest point from the origin and parameterizes the circular expanding wave by the distance (radius) of the point farthest from the origin.

255 We use PFs with $N_{\text{par}} = 200, 500$, and 1000 particles to assimilate feature data. Initial particles $\mathbf{p}_0^j$ ($j = 1, \dots, N_{\text{par}}$) from the PFs with smaller N_{par} are recycled in the PFs with higher N_{par} . For each PF, feature data are assimilated after every time step Δt , starting at $k = 1$ and concluding at $k = 5$ corresponding to the stopping time $t_{\text{off}} = 0.25$.

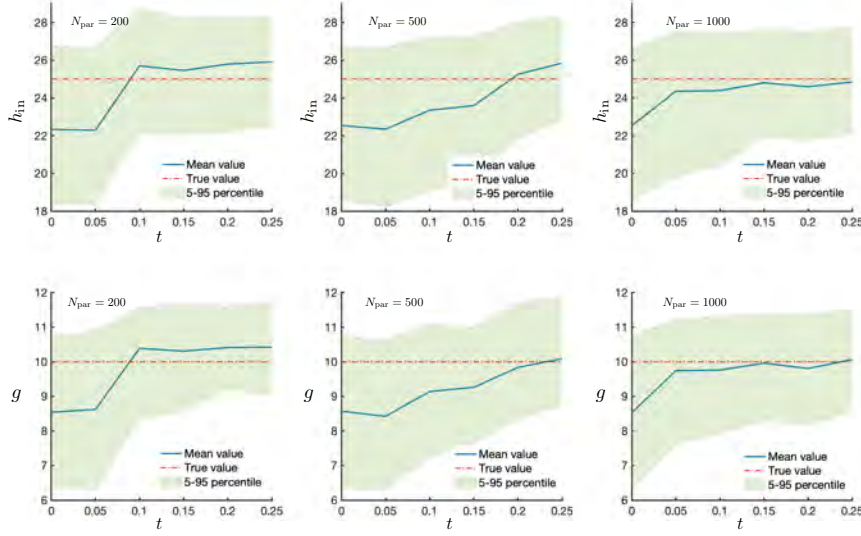


Figure 8: Convergence of the estimators of the uncertain parameters h_{in} and g in Example 3.

260 The results of the assimilation process are presented in Figure 8. Compared to other examples, The number of observations and, consequently, assimilation steps needed to obtain accurate estimates of the parameters $\mathbf{p}$ ($k = 1, \dots, 5$) is lower than in the previous two examples. As the number of

particles in the PF, N_{par} , increases, the estimates of $\mathbf{p}$ improve; yet, N_{par} has negligible effect on the prediction uncertainty of these estimators, as quantified by the 95% confidence intervals. Overall, feature observations do carry sufficient information to estimate the uncertain parameters.

4. Summary

Our study is intended to highlight the importance of a distinct class of DA problems in which observations come in the form of features, such as dynamics of shocks and wavefronts, maximum values, and level surfaces. The main difference between the conventional DA and FIDA lies in the observation operator. Much of current research on DA focuses on the developments of computationally tractable algorithms for handling the nonlinearity of a forward model [32], treating the observation operator as an afterthought. In FIDA, the observation operator is a highly nonlinear functional involving a search process over a function of the state variables that results in set-valued outputs. Nonlinearity of the observation operators and set-valued outputs present hitherto unaddressed challenges, including the question of whether the information content of feature observations is sufficient for parameter estimation and error-propagation analysis.

We present three numerical experiments in which shocks and expanding waves are observed features. These examples serve to demonstrate FIDA's ability to estimate model parameters from such noisy observations and to highlight the need for future research. As expected, our results reveal that some model parameters are more readily estimated by FIDA than others. This finding suggests the need for future research on identifiability and observability in the FIDA context, especially when the observation operator involves inequalities.

Particle filters, primarily used in our examples as the DA engine, are not scalable, and their applicability to computationally expensive forward problems is limited; the development of efficient algorithms tailored for FIDA is a priority. The ability to assimilate feature data from multiple sources and in multiple formats, or using different kinds of features simultaneously, is also of interest in many applications.

5. Acknowledgments

The research was partially supported in part by the Air Force Office of Scientific Research under grant FA9550-21-1-0381; by the National Science Foundation under award 2100927; and by Naval Research Laboratory, Monterey, California.

Appendix A. Analytical Solution for Example 1

The solution to (3.8)–(3.9) with arbitrary λ follows the analytical strategy available for the inviscid Burgers equation ($\lambda = 2$) [33]. It is presented here for the sake of completeness. Characteristic curves

$x(t)$ for (3.8) satisfy

$$\frac{dx}{dt} = L(u), \quad L(u) \equiv \frac{2u}{\lambda}; \quad x(0) = \xi, \quad (\text{A.1})$$

where the number $\xi \in \mathbb{R}$ provides a label for individual characteristics. Along any characteristics, the state variable $u(x, t)$ is constant,

$$\frac{du(x(t), t)}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} \frac{dx}{dt} = \frac{\partial u}{\partial t} + L(u) \frac{\partial u}{\partial x} = 0. \quad (\text{A.2})$$

Accounting for initial condition (3.9),

$$u(x(t), t) = u(\xi, 0) = u_{\text{in}}(\xi). \quad (\text{A.3})$$

Hence, it follows from (A.2) that

$$\frac{dx}{dt} = \frac{2u_{\text{in}}(\xi)}{\lambda} \quad \text{or} \quad x = \xi + \frac{2u_{\text{in}}(\xi)}{\lambda} t. \quad (\text{A.4})$$

Differentiating (A.3) and (A.4) with respect to x and t , and eliminating the partial derivatives of ξ ,

$$\frac{\partial u}{\partial x} = \frac{\lambda u'_{\text{in}}(\xi)}{\lambda + 2tu'_{\text{in}}(\xi)}, \quad \frac{\partial u}{\partial t} = -\frac{2u'_{\text{in}}(\xi)u_{\text{in}}(\xi)}{\lambda + 2tu'_{\text{in}}(\xi)}. \quad (\text{A.5})$$

These expressions satisfy (3.8) and, hence, provide a valid solution. However, this solution breaks down when $\lambda + 2tu'_{\text{in}}(\xi) \rightarrow 0$; the earliest time at which this occurs, t^* , is called a break time:

$$t^* = \min_x \left| \frac{\lambda}{2u'_{\text{in}}(x)} \right| \quad \& \quad u'_{\text{in}}(x) < 0 \quad \implies \quad t^* = \frac{\lambda}{2\alpha}. \quad (\text{A.6})$$

It marks the instance when the characteristics first intersect.

For $t < t^*$, combining (3.9) with (A.4) yields

$$x = \begin{cases} \frac{2u_l}{\lambda} t + \xi & \xi \leq 0 \\ \frac{2(u_l - \alpha\xi)}{\lambda} t + \xi & 0 < \xi \leq x_r \\ \frac{2u_r}{\lambda} t + \xi & \xi > x_r. \end{cases} \quad (\text{A.7})$$

Using (A.7) to eliminate ξ in (A.3) gives the solution before break time t^* ,

$$u(x, t < t^*) = u_{\text{in}} \left(x - \frac{2u_{\text{in}}(\xi)}{\lambda} t \right) = \begin{cases} u_l & x \leq 2u_l t / \lambda \\ \frac{u_l - \alpha x}{1 - 2\alpha t / \lambda} & 2u_l t / \lambda \leq x \leq x_r + 2u_r t / \lambda \\ u_r & x_r + 2u_r t / \lambda \leq x. \end{cases} \quad (\text{A.8})$$

For $t \geq t^*$, intersection of characteristics causes shocks. The characteristics first intersect, and shocks originate, at point (x^*, t^*) , where

$$x^* = u_l t^* = u_r t^* + x_r. \quad (\text{A.9})$$

If $x_{\text{sh}}(t)$ denotes shock position at time t , then at break time $x_{\text{sh}}(t^*) = x^*$. The shock speed S is given by the Rankine-Hugoniot condition,

$$S \equiv \frac{dx_{\text{sh}}}{dt} = \frac{f(u_2) - f(u_1)}{u_2 - u_1} = \frac{1}{\lambda}(u_l + u_r), \quad (\text{A.10})$$

where u_1 and u_2 are values of the variable on either side of the shock. Integrating (A.10) and using the condition $x_{\text{sh}}(t^*) = x^*$, yields an equation for the shock evolution,

$$x_{\text{sh}}(t) = \frac{u_l + u_r}{\lambda}t + \frac{u_l - u_r}{2\alpha} = \frac{u_l + u_r}{\lambda}t + \frac{x_r}{2}. \quad (\text{A.11})$$

Consequently,

$$u(x, t \geq t^*) = \begin{cases} u_l & x < x_{\text{sh}} \\ \frac{u_l + u_r}{2} & x = x_{\text{sh}} \\ u_r & x > x_{\text{sh}}. \end{cases} \quad (\text{A.12})$$

References

- [1] G. Evensen, Data Assimilation: The Ensemble Kalman Filter, Springer, New York, 2006.
- [2] K. Law, A. Stuart, K. Zygalakis, Data Assimilation: A Mathematical Introduction, Springer, New York, 2015.
- [3] P. J. V. Leeuwen, Y. Cheng, S. Reich, Nonlinear Data Assimilation, Springer International, Switzerland, 2015.
- [4] M. Asch, M. Bocquet, M. Nodet, Data Assimilation: Methods, Algorithms, and Applications, SIAM, Philadelphia, 2016.
- [5] P. Fearnhead, H. R. Künsch, Particle filters and data assimilation, Annual Review of Statistics and Its Application 5 (2018) 421–449.
- [6] A. Albarakati, M. Budišić, R. Crocker, J. Glass-Klaiber, S. Iams, J. Maclean, N. Marshall, C. Roberts, E. S. Van Vleck, Model and data reduction for data assimilation: Particle filters employing projected forecasts and data with application to a shallow water model, Comput. Math. Appl. 116 (2022) 194–211.
- [7] S. Vetra-Carvalho, P. J. van Leeuwen, L. Nerger, A. Barth, M. U. Altaf, P. Brasseur, P. Kirchgessner, J.-M. Beckers, State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems, Tellus A: Dynamic Meteorology and Oceanography 70 (1) (2018) 1–43.

- [8] W. Kang, L. Xu, Some quantitative characteristics of error covariance for Kalman filters, *Tellus A: Dynamic Meteorology and Oceanography* 73 (1) (2021) 1–19.
- [9] L. Xu, R. Daley, Towards a true 4-dimensional data assimilation algorithm: application of a cycling representer algorithm to a simple transport problem, *Tellus A: Dynamic Meteorology and Oceanography* 52 (2) (2000) 109–128.
- [10] D. Fairbairn, S. R. Pring, A. C. Lorenc, I. Roulstone, A comparison of 4DVar with ensemble data assimilation methods, *Q. J. R. Meteorol. Soc.* 140 (2014) 281–294.
- [11] F. Boso, D. M. Tartakovsky, Learning on dynamic statistical manifolds, *Proc. Roy. Soc. A* 476 (2239) (2020) 20200213. doi:10.1098/rspa.2020.0213.
- [12] F. Boso, D. M. Tartakovsky, Information geometry of physics-informed statistical manifolds and its use in data assimilation, *J. Comput. Phys.* 467 (2022) 111438. doi:10.1016/j.jcp.2022.111438.
- [13] F. E. Uilhoorn, An algorithm for finding optimal sensor placement and model uncertainty in data assimilation applied to Riemann problems, *Applied Mathematical Modelling* 103 (2022) 649–673.
- [14] C. Krause, W. Huang, D. B. Mechem, E. S. Van Vleck, M. Zhang, A metric tensor approach to data assimilation with adaptive moving meshes, *J. Comput. Phys.* 466 (2022) 111407.
- [15] J. R. Eyre, S. J. English, M. Forsythe, Assimilation of satellite data in numerical weather prediction. Part I: The early years, *Q. J. R. Meteorol. Soc.* 146 (2020) 49–68.
- [16] J. Crane, X. Shi, J. T. Lipkowitz, A. M. Kempf, H. Wang, Geometric modeling and analysis of detonation cellular stability, *Proceedings of the Combustion Institute* 38 (3) (2021) 3585–3593.
- [17] S. Balasuriya, G. Gottwald, J. Hornibrook, S. Lafortune, High Lewis number combustion wavefronts: A perturbative Melnikov analysis, *SIAM Journal on Applied Mathematics* 67 (2) (2007).
URL <https://doi.org/10.1137/050640849>
- [18] C. Piccardi, Parameter estimation for systems with low-dimensional chaos, *IFAC Proceedings Volumes* 39 (8) (2006) 291–296.
- [19] B. Peng, S. K. Scott, K. Showalter, Period doubling and chaos in a three-variable autocatalator, *J. Phys. Chem.* (1990) 5243–5246.
- [20] W. Kang, D. M. Tartakovsky, A. Srivastava, Feature-informed data assimilation - definitions and illustrative examples, arXiv: 2211.00256v1 (2022).

- [21] S. Julier, J. Uhlmann, H. F. Durrant-Whyte, A new method for the nonlinear transformation of means and covariances in filters and estimators, *IEEE Trans. Automatic Control* 45 (3) (2000) 477–482.
- 340 [22] W. E. C. Ma, S. Wojtowysch, L. Wu, Towards a mathematical understanding of neural network-based machine learning: what we know and what we don’t, *arXiv:2009.10713v2* (2020).
- [23] W. Kang, Q. Gong, Feedforward neural networks and computational functions with applications to dynamical systems, *SIAM Journal on Control and Optimization* 60 (2) (2022) 786–813.
- [24] W. Kang, L. Xu, H. Zhou, A surrogate data assimilation model for the estimation of dynamical
345 systems in a limited area, *arXiv:2307.07178v1* (2023).
- [25] S. Julier, J. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE* 92 (3) (2004) 401–422.
- [26] E. Wan, R. Van Der Merwe, The unscented kalman filter for nonlinear estimation, in: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium* (Cat. No.00EX373), 2000, pp. 153–158. doi:10.1109/ASSPCC.2000.882463.
350
- [27] B. Ristic, S. Arulampalam, N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House radar library, Artech House, 2004.
URL <https://books.google.com/books?id=cjFDngEACAAJ>
- [28] G. Kitagawa, A self-organizing state-space model, *Journal of the American Statistical Association*
355 (1998) 1203–1215.
- [29] W. Kang, L. C. Wilcox, Solving 1D conservation laws using Pontryagin’s minimum principle, *Journal of Scientific Computing* 71 (1) (2017) 144–165.
- [30] T. A. Driscoll, N. Hale, L. N. Trefethen, *Chebfun Guide*, Pafnuty Publications, 2014.
URL <http://www.chebfun.org/docs/guide/>
- 360 [31] D. I. Ketcheson, K. T. Mandli, A. J. Ahmadi, A. Alghamdi, M. Quezada de Luna, M. Parsani, M. G. Knepley, M. Emmett, PyClaw: Accessible, Extensible, Scalable Tools for Wave Propagation Problems, *SIAM Journal on Scientific Computing* 34 (4) (2012) C210–C231.
- [32] A. Spantini, R. Baptista, Y. Marzouk, Coupling techniques for nonlinear ensemble filtering, *SIAM Review* 64 (4) (2022) 921–953. doi:10.1137/20M1312204.
- 365 [33] A. Jeffrey, *Applied Partial Differential Equations: An Introduction*, Elsevier Science, 2003.
URL <https://books.google.com/books?id=xHfgL0xF-p8C>